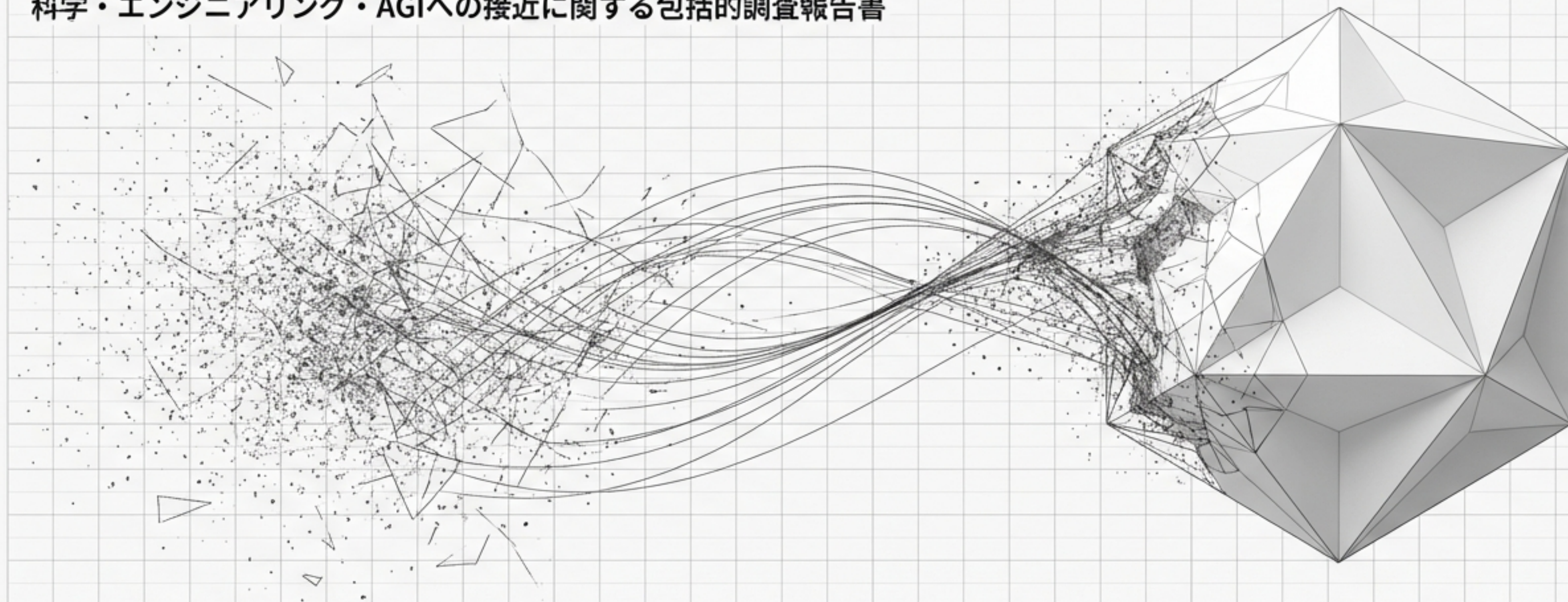


Gemini 3 Deep Think: 専門化された推論の覚醒

科学・エンジニアリング・AGIへの接近に関する包括的調査報告書



DATE: 2026年2月12日
SOURCE: Google DeepMind Report Analysis
SUBJECT: System 2 Architecture & Inference-Time Compute

エグゼクティブ・サマリー：4つの核心的変革



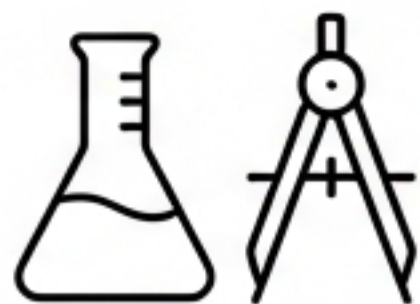
歴史的転換点

2026年2月12日、AIは「確率的な単語生成 (System 1)」から「論理的な思考プロセス (System 2)」へ移行。単なるアップグレードではなく、計算機知能のパラダイムシフトが発生。



知能の壁の突破

ARC-AGI-2 スコア 84.6%。人間の平均（約 60%）を大きく上回り、未知のパターンを解明する流動性知能において人類を超越。AGI実現の最終段階へ。



科学と実装

国際物理・化学オリンピックでのゴールドメダル級成績（物理87.7%）。10年来の数学的未解決問題（オンライン劣モジュラ最適化）を自律的に解決。

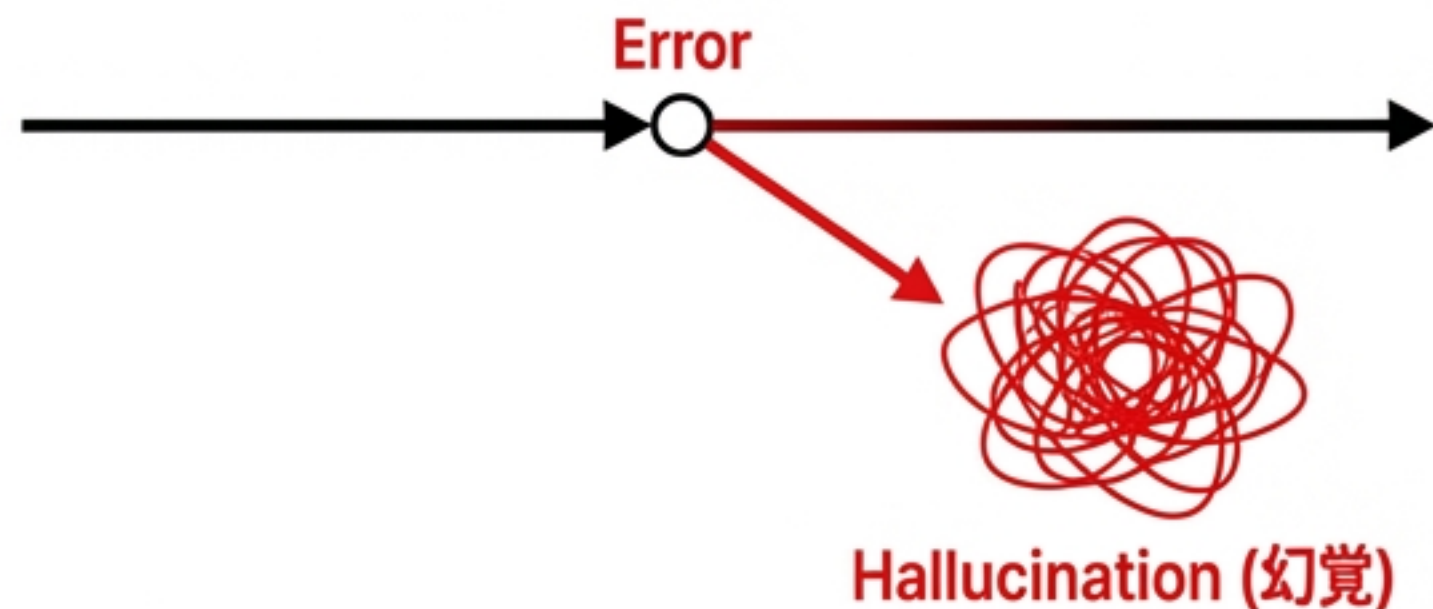


高コスト構造

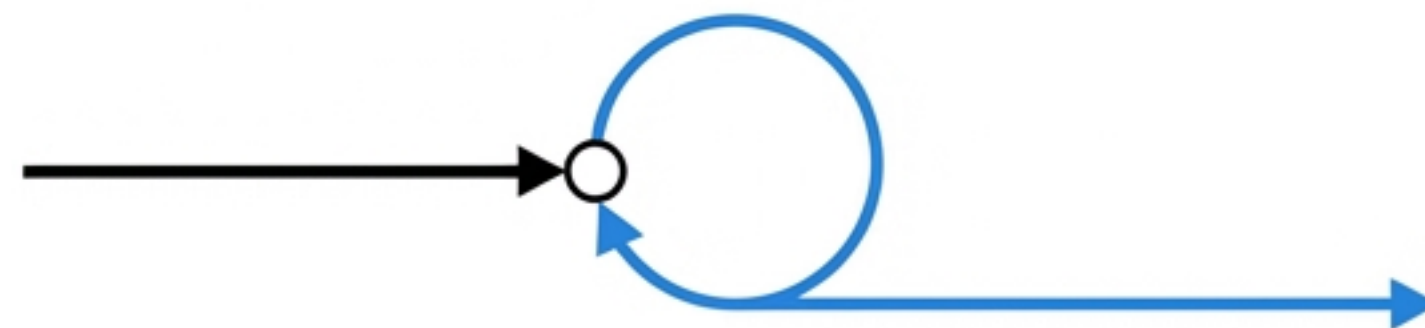
1タスクあたり \$13.62（約2,000円）の推論コスト。日常利用ではなく、科学的発見や高額なバグ修正などの「ハイステークス」な課題に特化。

System 1の限界：なぜスケーリングだけでは不十分だったのか

従来のLLM (System 1)



Deep Think (System 2)



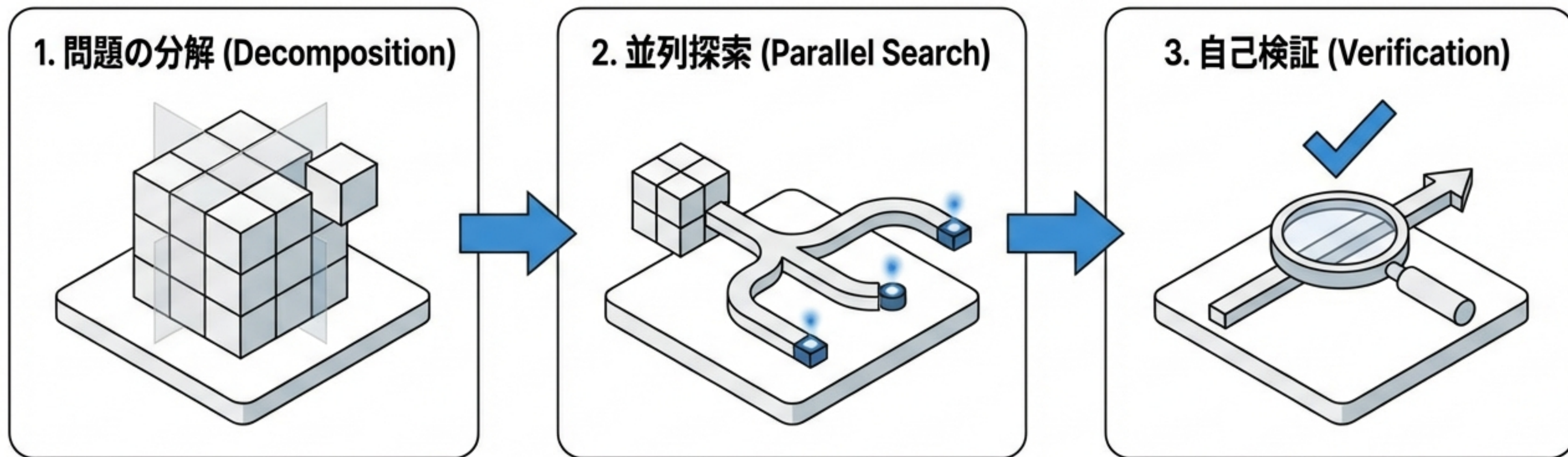
確率の罠: 従来のLLMは「次に来るもっともらしい単語」を予測するだけであり、論理の正しさを保証しない。

幻覚 (Hallucination) の連鎖: 複雑な推論において、初期の些細なミスが最終的な回答を崩壊させる。

コンテキスト: ダニエル・カーネマンの定義する「直感的生成」から脱却し、AIに「立ち止まって考える」能力を与える必要性。

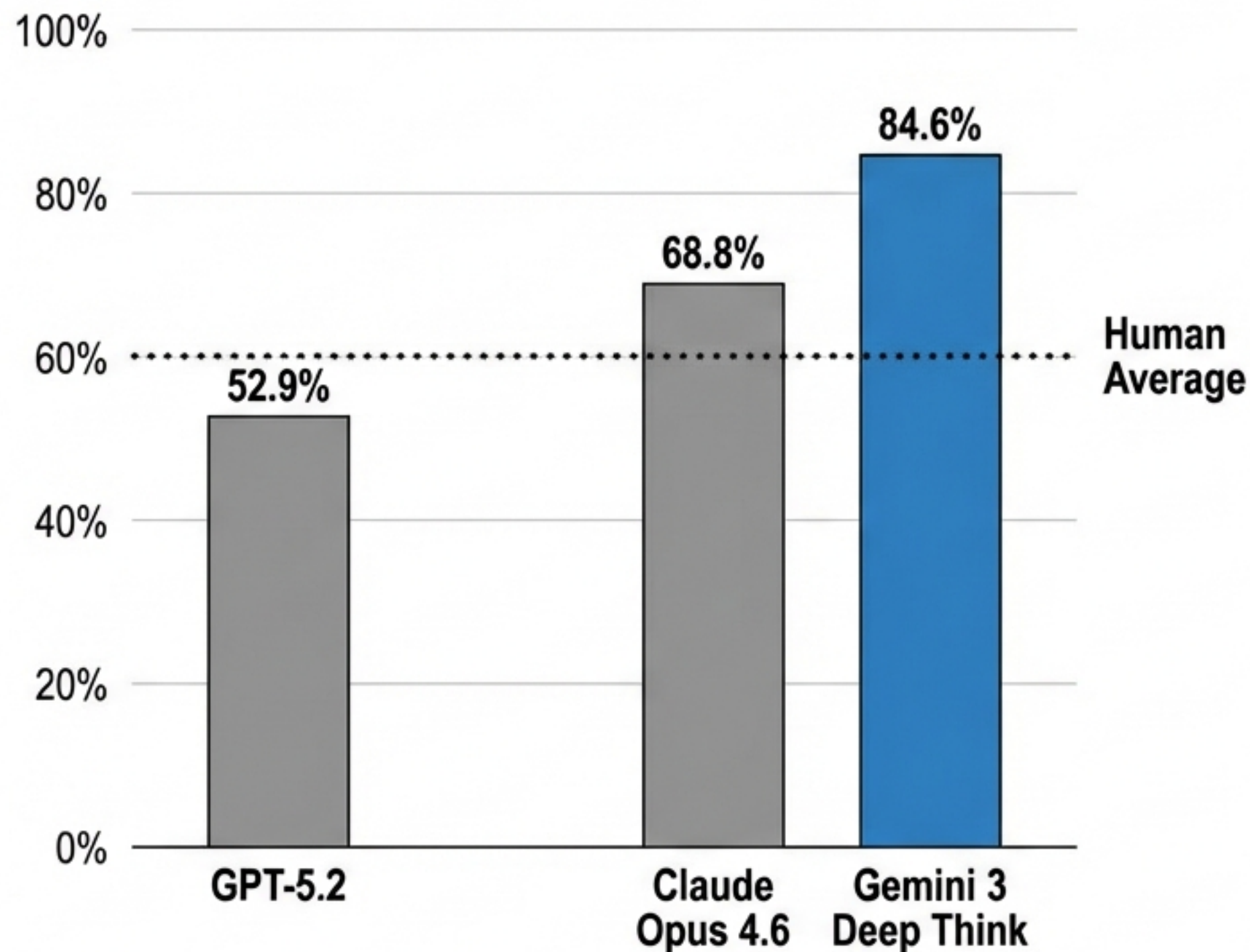
解決策：推論時間計算量 (Inference-Time Compute) の実装

「回答する前に、思考する」



標準モデルに対し1.4倍～2.3倍のレイテンシ（遅延）を許容し、その時間を「思考」に充てる。
これは数学者が証明を組み立て、エンジニアがデバッグする認知的プロセスそのものである。

AGIへのマイルストーン：ARC-AGI-2 分析



流動性知能の証明:
ARC-AGIは学習データに含まれない
「未知の視覚パターン」を解くテスト。
記憶（暗記）では突破できない。

人間超え: 一般成人の平均スコアを
20ポイント以上上回る。

フランソワ・ショレの警鐘:
「未解決タスクを人間以上の効率で
学習できるようになった時点で、
AGIの重要な要件は満たされた」

学術分野での支配的成果：オリンピック級の知性

Physics (IPhO)



87.7%

Gold Medal Level

物理法則に基づいた厳密な計算と概念統合の両立。

Chemistry (IChO)



82.8%

Gold Medal Level

定量的な推論能力の証明。

Math (IMO)



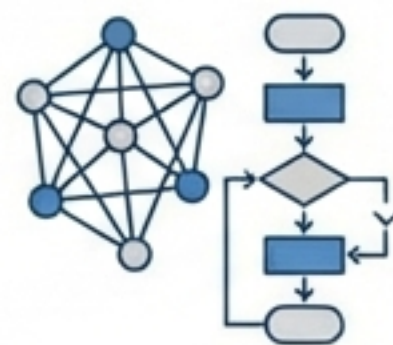
Gold Medal Level

数千の論理分岐から唯一の正解を導出。

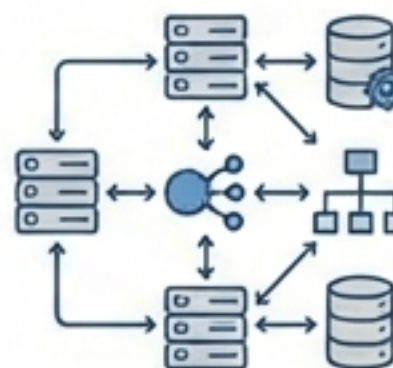
エンジニアリング：オートコンプリートを超えて

3455

Codeforces Elo Rating (World Top 0.1%)



アルゴリズム設計
グラフ理論、動的計画法を駆使し、
時間計算量・空間計算量を自ら評価
して最適解を選択。



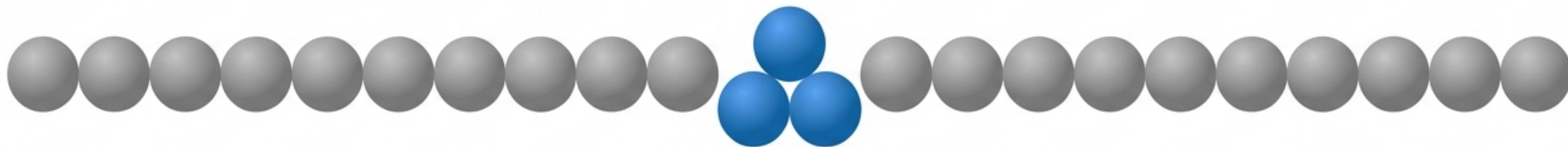
システムアーキテクチャ
単なるコード生成ではなく、分散
システムの設計やボトルネック解消
が可能。



デバッグ
人間のシニアエンジニアでも困難な
「数百万ドルの損失」に直結する
複雑なバグの特定。

“記憶に基づいたコーディングでは到達不可能な領域。”

「Aletheia」プロジェクト：自律的な発見の時代



Case Study: オンライン劣モジュラ最適化 (Online Submodular Optimization)

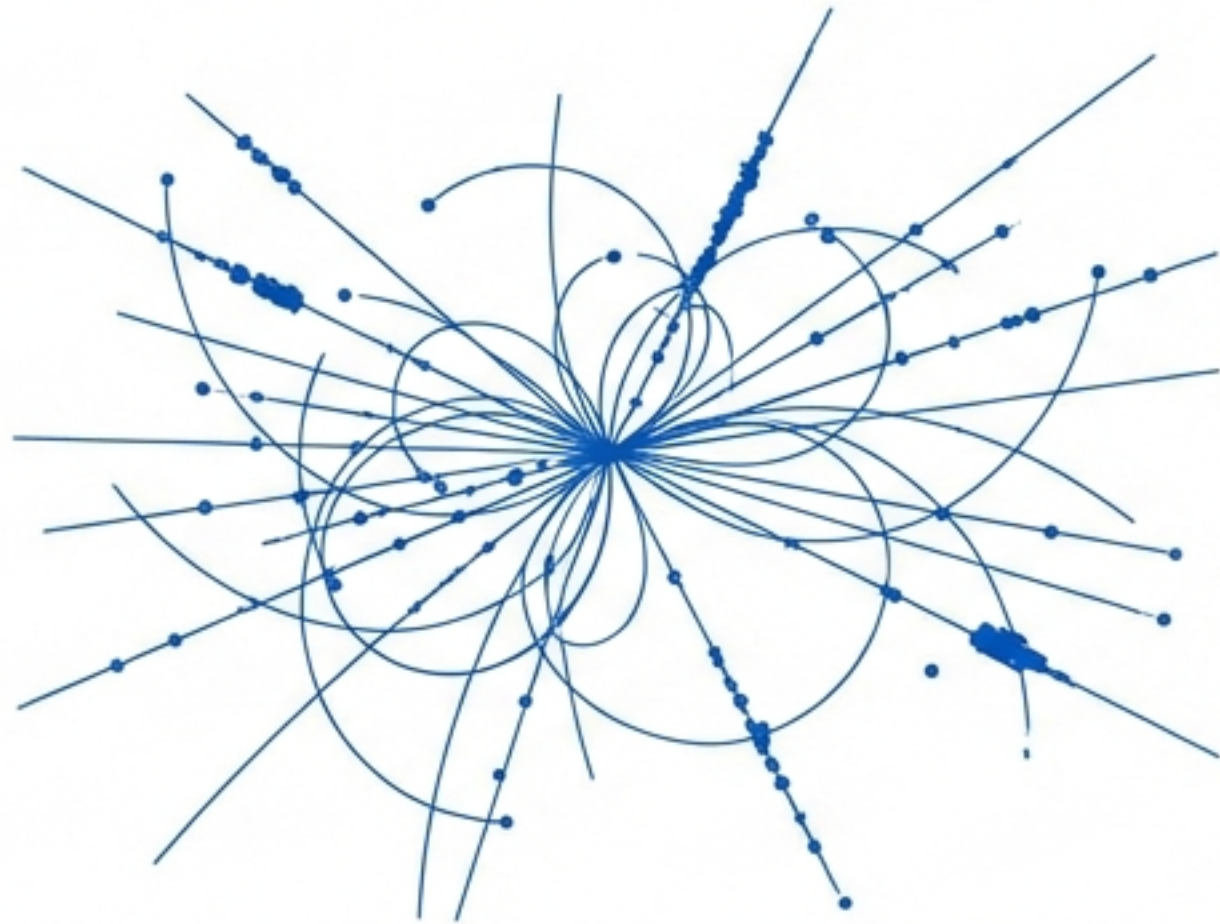
The Problem: 2015年以来、数学者たちはある直感（データのコピーは移動よりも価値が低い）を証明しようとしていた。

The Discovery: Deep Think搭載エージェント「Aletheia」が、その直感が誤りであることを証明する「3項目の反例 (Counterexample)」を自律的に構築。

Significance: AIは既存知識の整理だけでなく、人類が到達していない「新しい数学的真理」を発見し始めた。

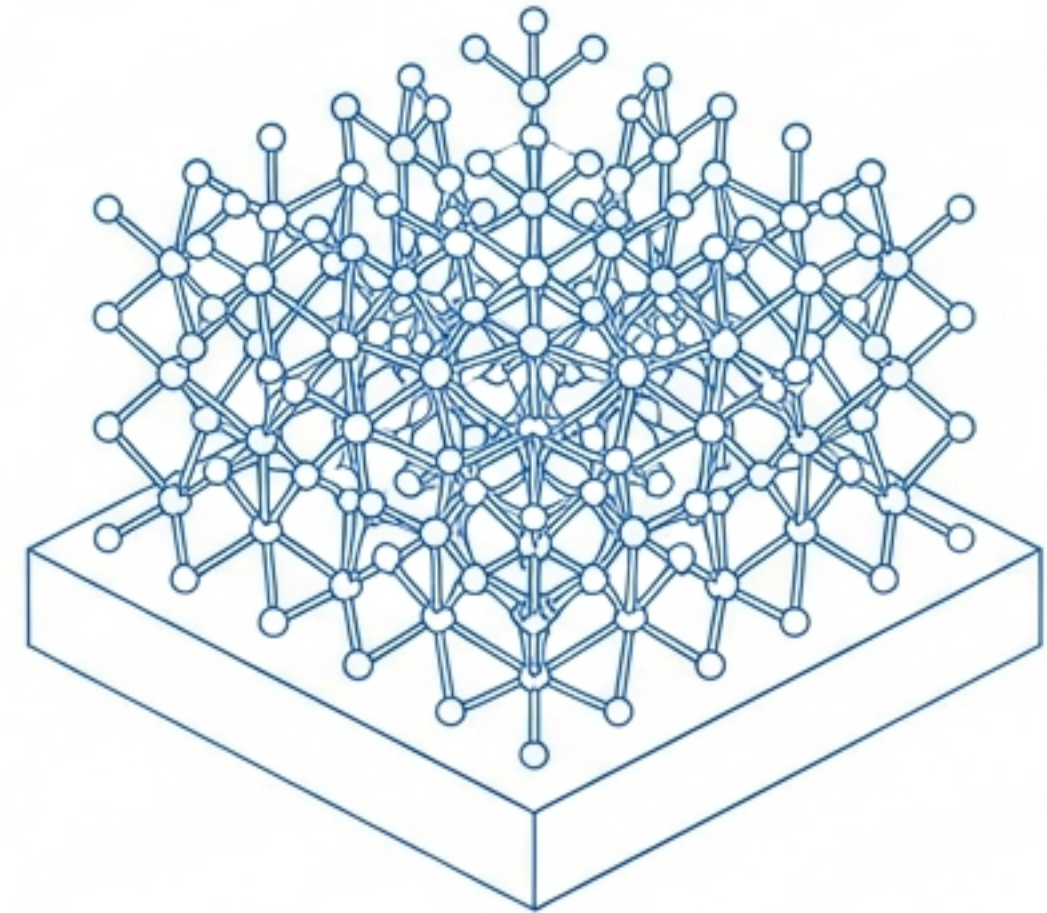
応用科学：理論から製造現場へ

Rutgers Univ (Theoretical Physics)



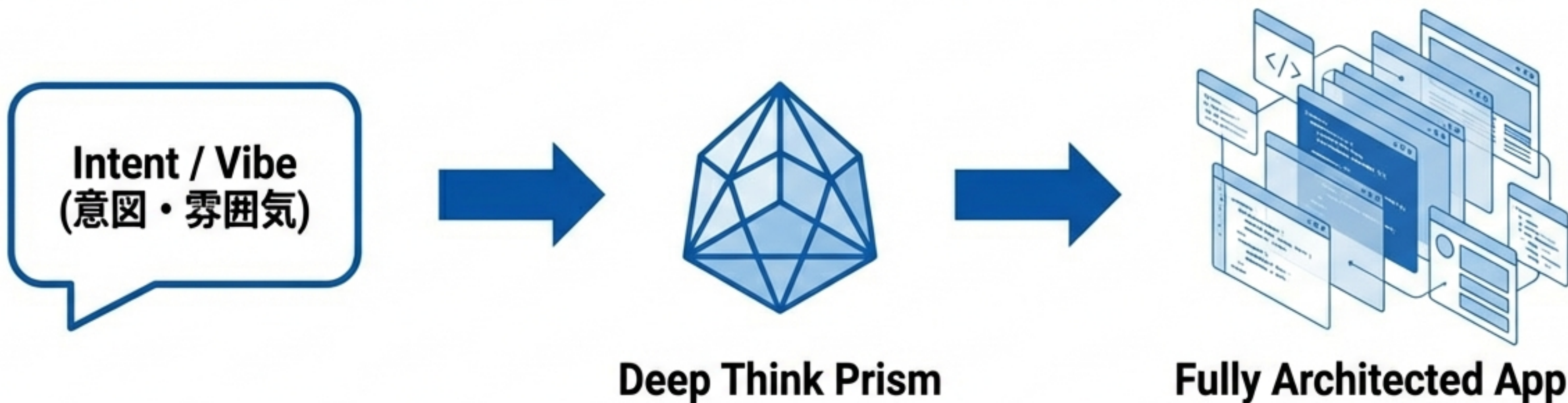
専門家の査読をすり抜けた論文の論理的誤りを Deep Think が指摘。批判的思考能力の証明。

Duke Univ (Material Engineering)



半導体材料（結晶成長）。データが不完全な「泥臭い」課題において、100マイクロメートルの薄膜を成長させる最適なレシピを設計。

「Vibe Coding」の台頭とGoogle Antigravity



Definition: Vibe Coding (バイブ・コーディング)

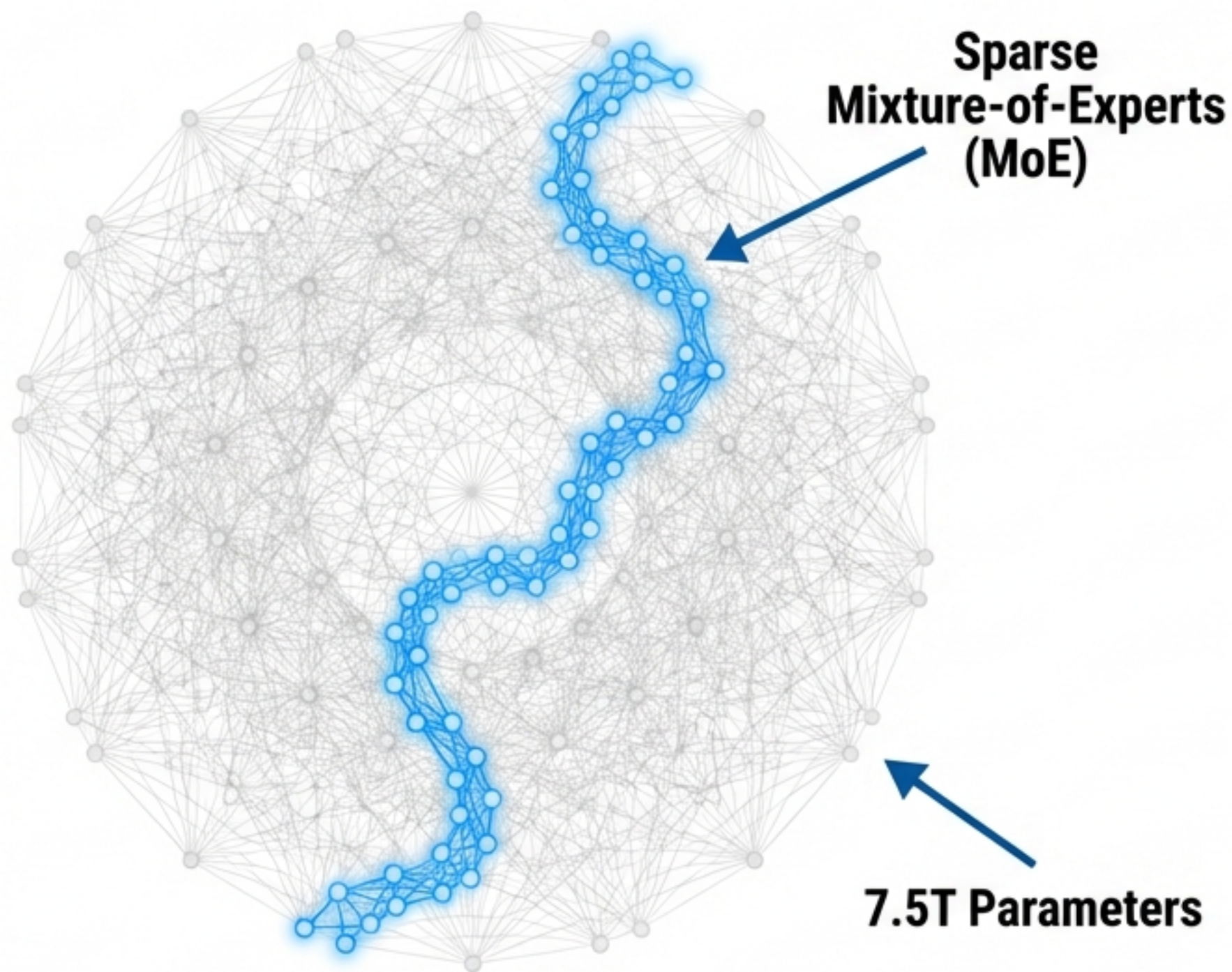
厳密な仕様書や構文ではなく、アプリケーションの「雰囲気 (Vibe)」や「意図」を伝えるだけで、AIが複雑なコード体系を構築する手法。

Platform: Google Antigravity

エディタ、ターミナル、ブラウザに直接アクセスし、実装・テスト・デバッグを自律実行。

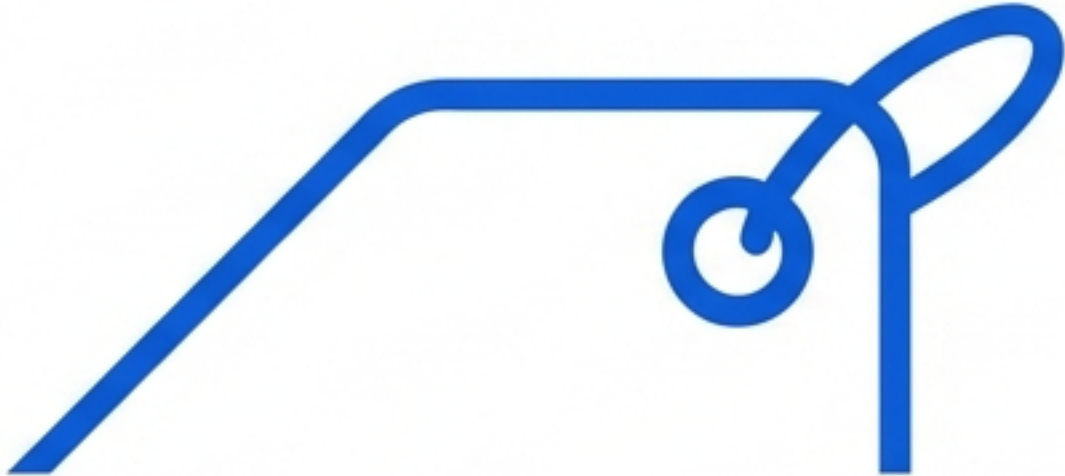
Evidence: WebDev Arena (Elo 1487)

技術的基盤：7.5兆パラメータの巨大脳



- **メカニズム**: すべてのニューロンを使わず、トークンごとに最適な「エキスパート（サブネットワーク）」へ動的にルーティング。
- **コンテキストウィンドウ**: 100万トークン（1 Million Context Window）。長大な推論過程での情報損失を防ぐ。
- **ネイティブ・マルチモーダル**: 単一スタックでテキスト、画像、コードを処理。

思考のコスト：推論の経済学



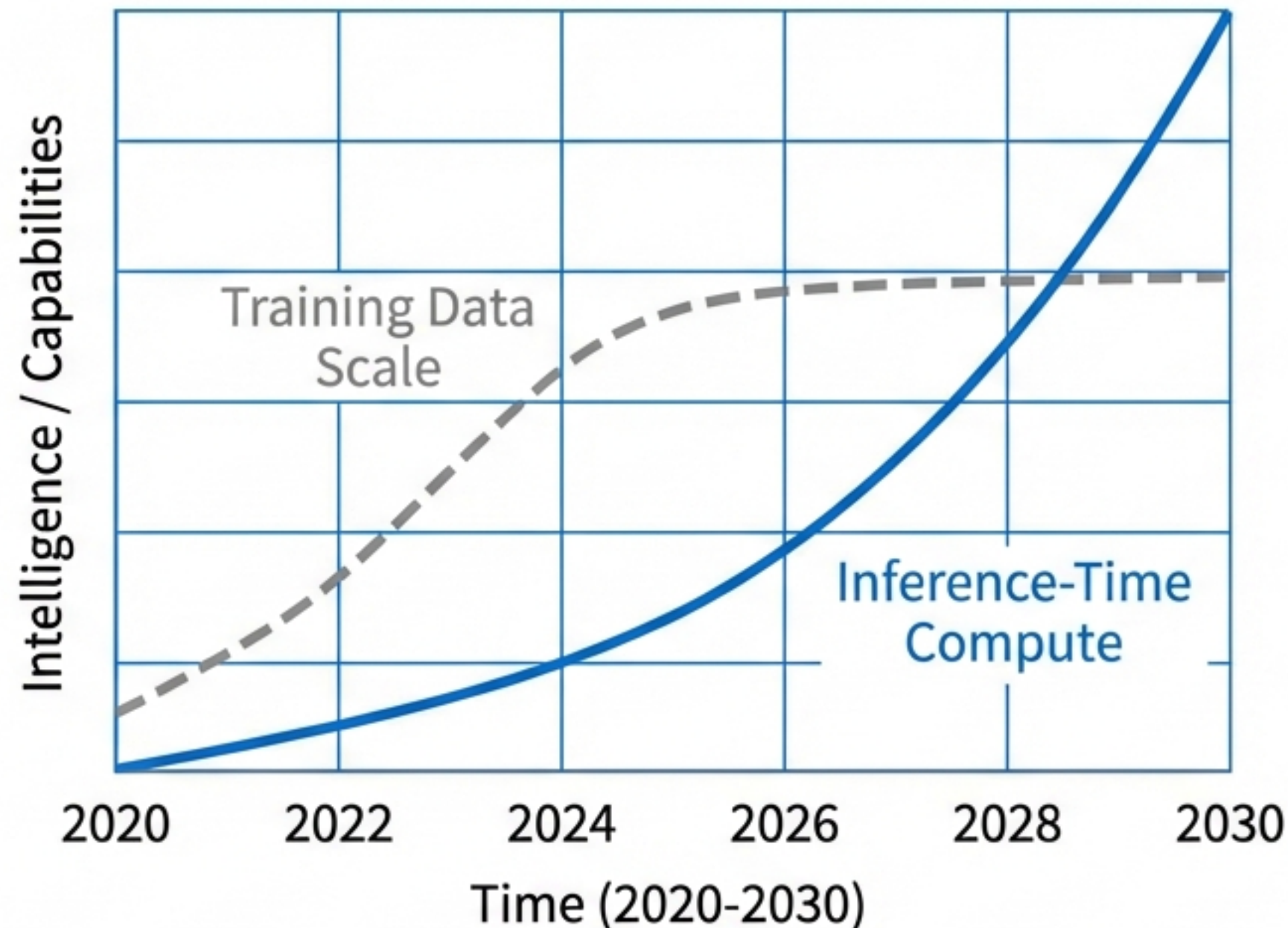
\$13.62 / Task
vs \$0.01 (Standard LLM)



- 戦略的示唆: 利用の階層化 (Tiered Usage)
 - Gemini 3 Flash: 日常会話、UI操作（安価・高速）。
- Gemini 3 Deep Think: 科学的発見、未解決問題、高額なバグ修正（高価・低速）。

Redefining Scaling Laws: The Finite Data Era

スケーリング則の再定義：有限データ時代



- **Context:** 2025-2026年、Web上の高品質テキストデータは枯渇した (The Finite Data Era)。
- **New Law:** More Inference Compute = More Intelligence.
- **Mechanism:** 学習データの量 (Scale) ではなく、推論時の計算量 (Thinking Time) と「研究の味 (Research Taste)」の向上に注力。

リスクと安全性：諸刃の剣

Slopacolypse



Andrej Karpathyの警告。AI生成の無価値なコンテンツ（Slop）によるデジタル空間の汚染。

Security Threats



自律的にゼロデイ脆弱性を発見する能力。セキュリティの常識を覆すリスク。

Access Control



GoogleによるAPIアクセスの厳格な制限。能力の高さは、そのまま脅威の高さに直結する。

Redefining Scaling Laws: The Finite Data Era

結論：ポスト・チューリングテストの世界へ



2026年、AIは「ツール」から「思考するパートナー」への不可逆的な進化を遂げた。

Final Thought:

「何ができるか」という問いは終わった。これからは「その知能をいかに管理し、共存するか」が問われる。

我々は文明の基盤（科学、工学、知識生産）が再構築される時代の入り口に立っている。