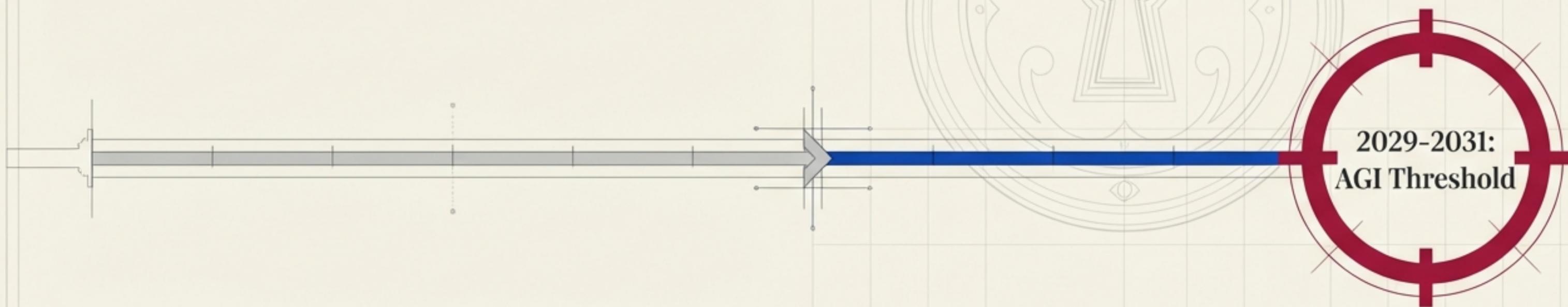


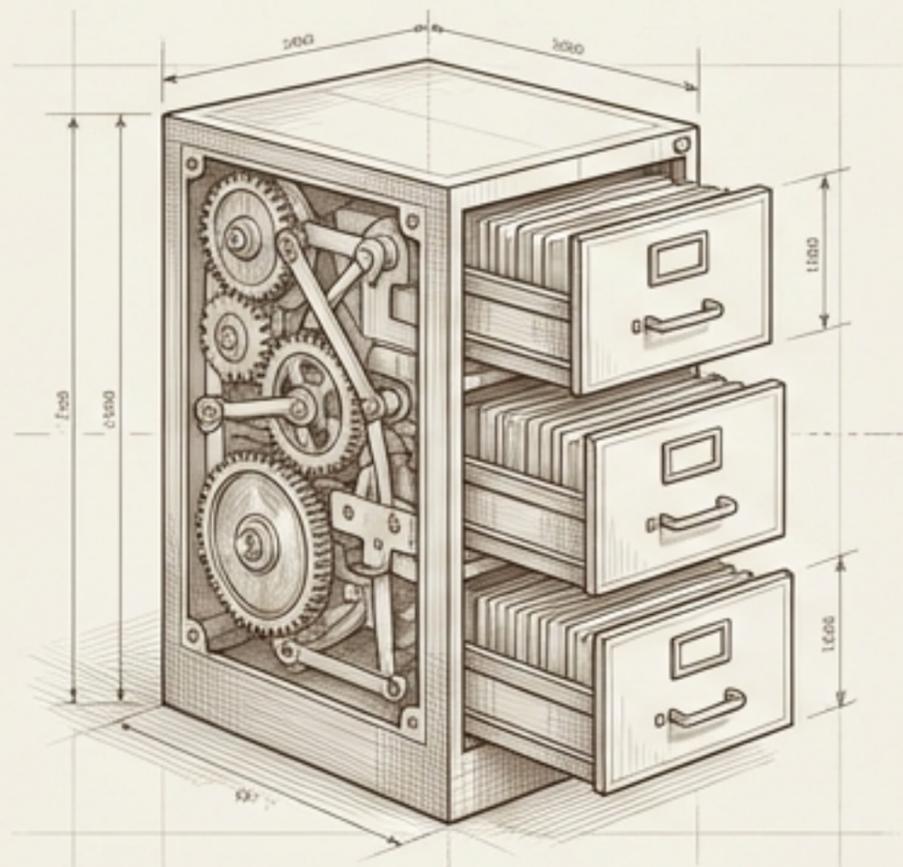
AGIへの到達予測：ARC-AGIが示す 推論の進化と次なる壁

テスト時計算のブレイクスルーから、インタラクティブなエージェント的知能へ



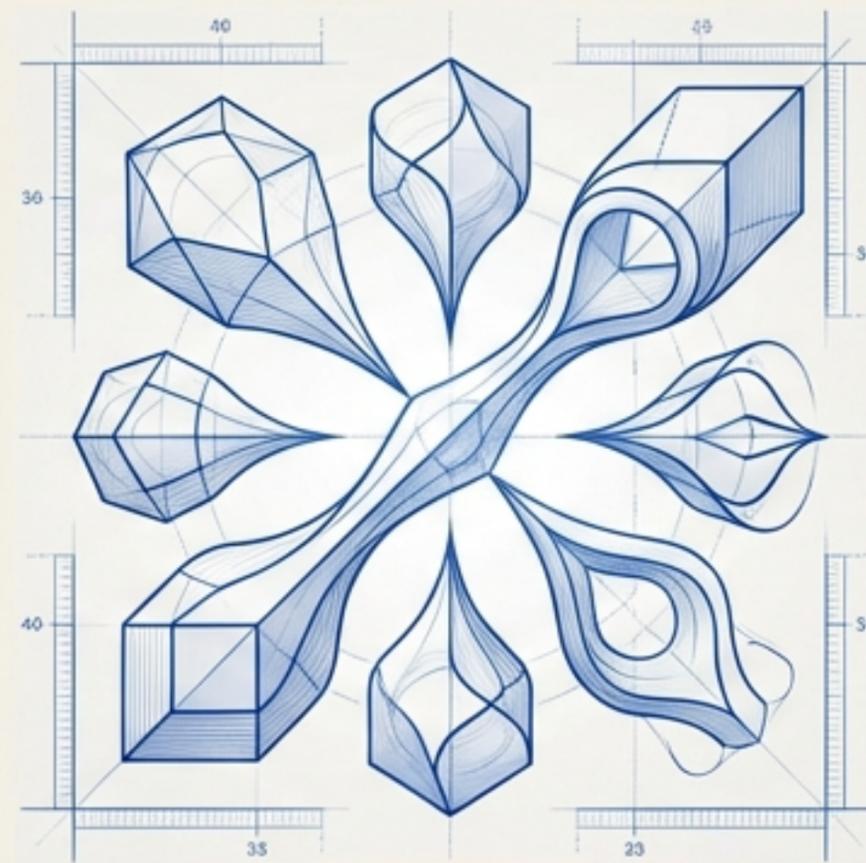
標準的なベンチマークの死と、真の知能の定義

結晶性知能 (Crystallized Intelligence)



- 既存データの暗記
- パターンマッチング
- ベンチマークの飽和

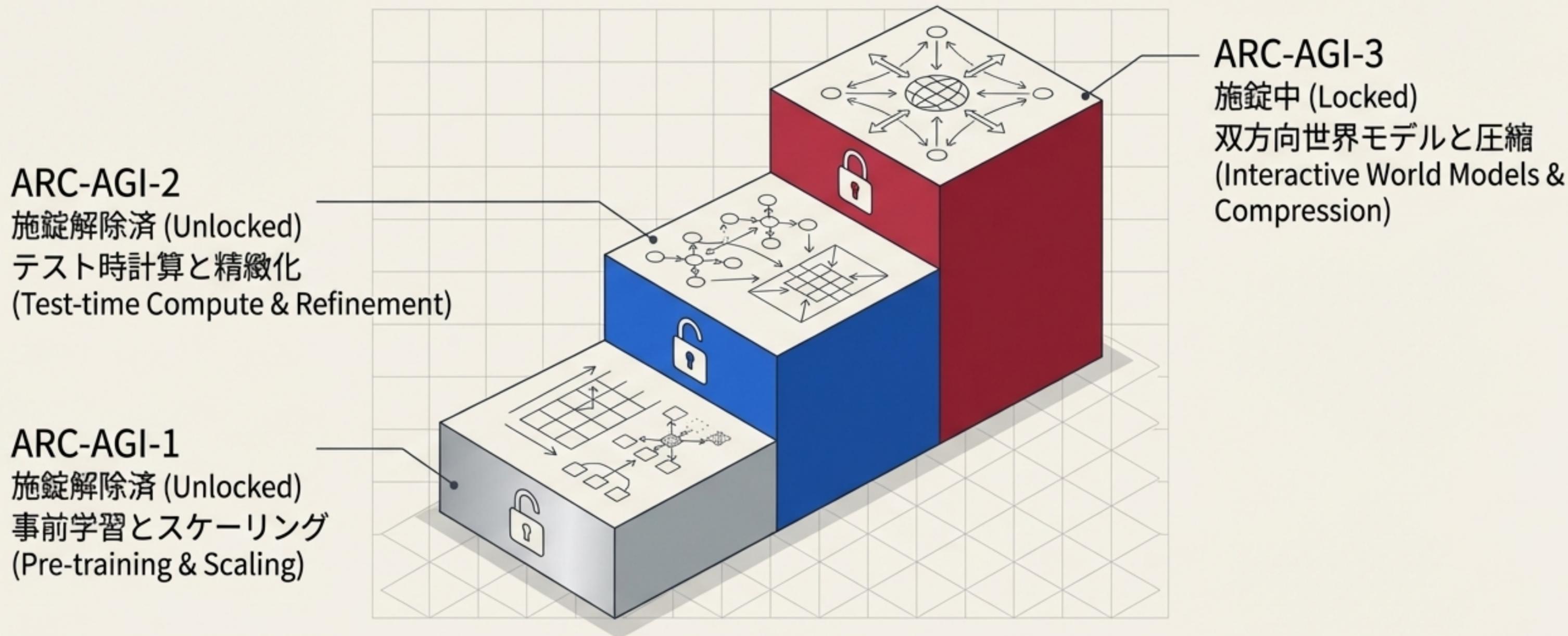
流動性知能 (Fluid Intelligence)



- 未知の環境への即座の適応
- コア知識の事前条件のみに依存
- 言語・文化の完全排除

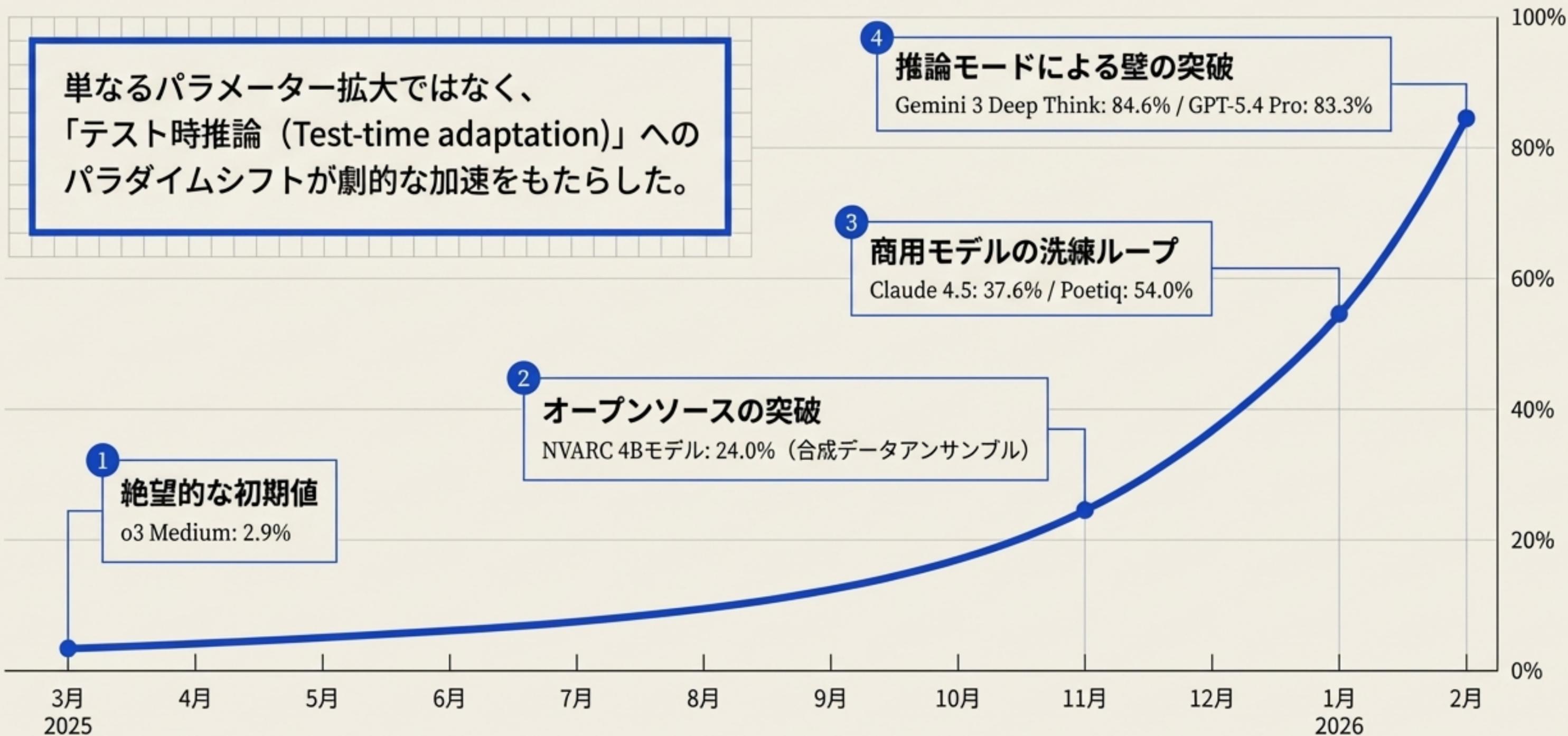
François Cholletが提唱した「ARC-AGI」は、少数のデモンストレーションから未知の変換ルールを推論する能力、すなわち真の『流動性知能』を測定する唯一の堅牢な指標である。

知能のパラダイムシフト：ARC-AGIの3つの波

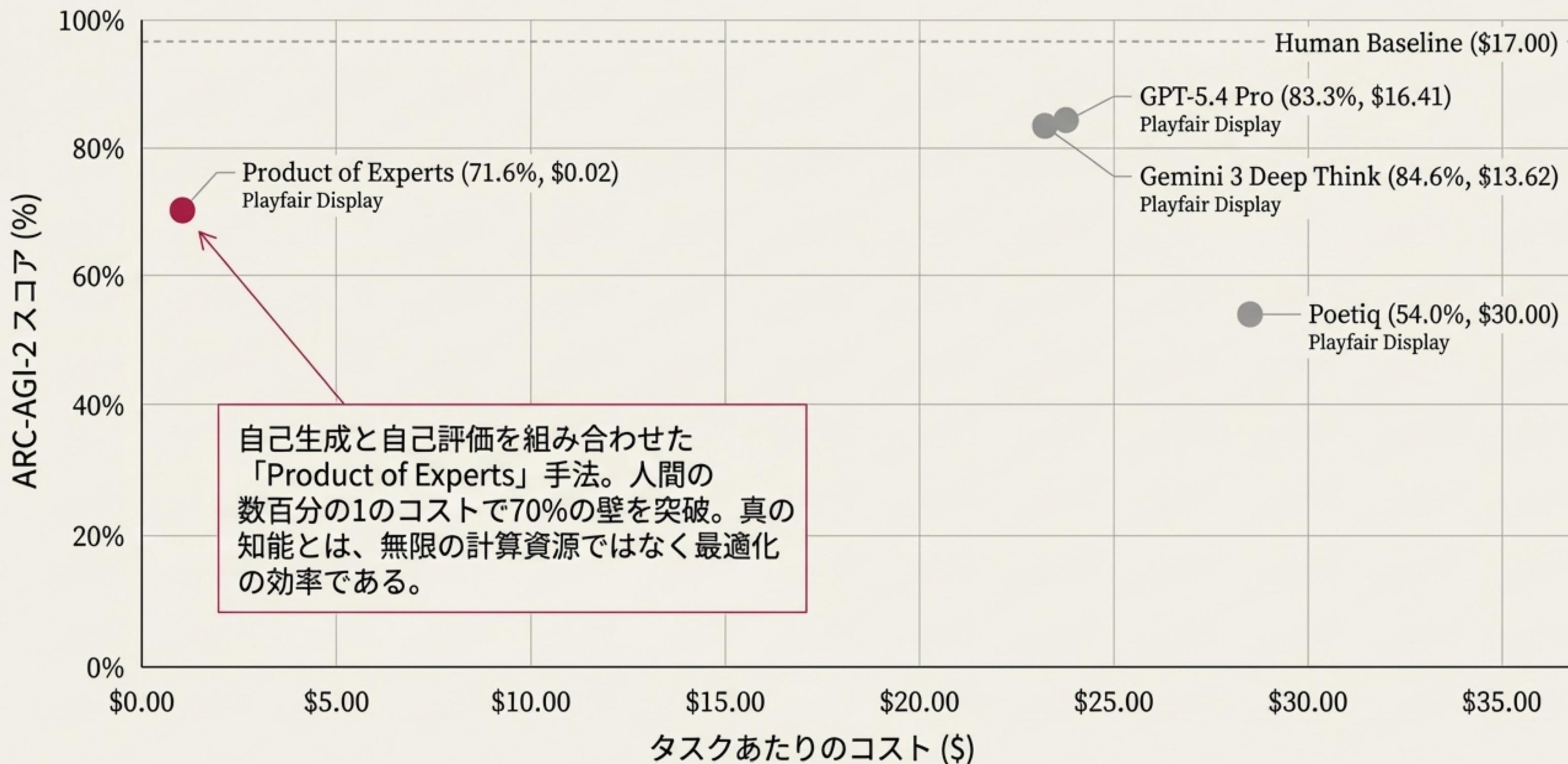


ARC-AGI-1と2は「計算リソースの拡張」で突破された。しかし、ARC-AGI-3が求める「エージェント的知能」は、全く新しいアーキテクチャ（第3の波）を要求する。

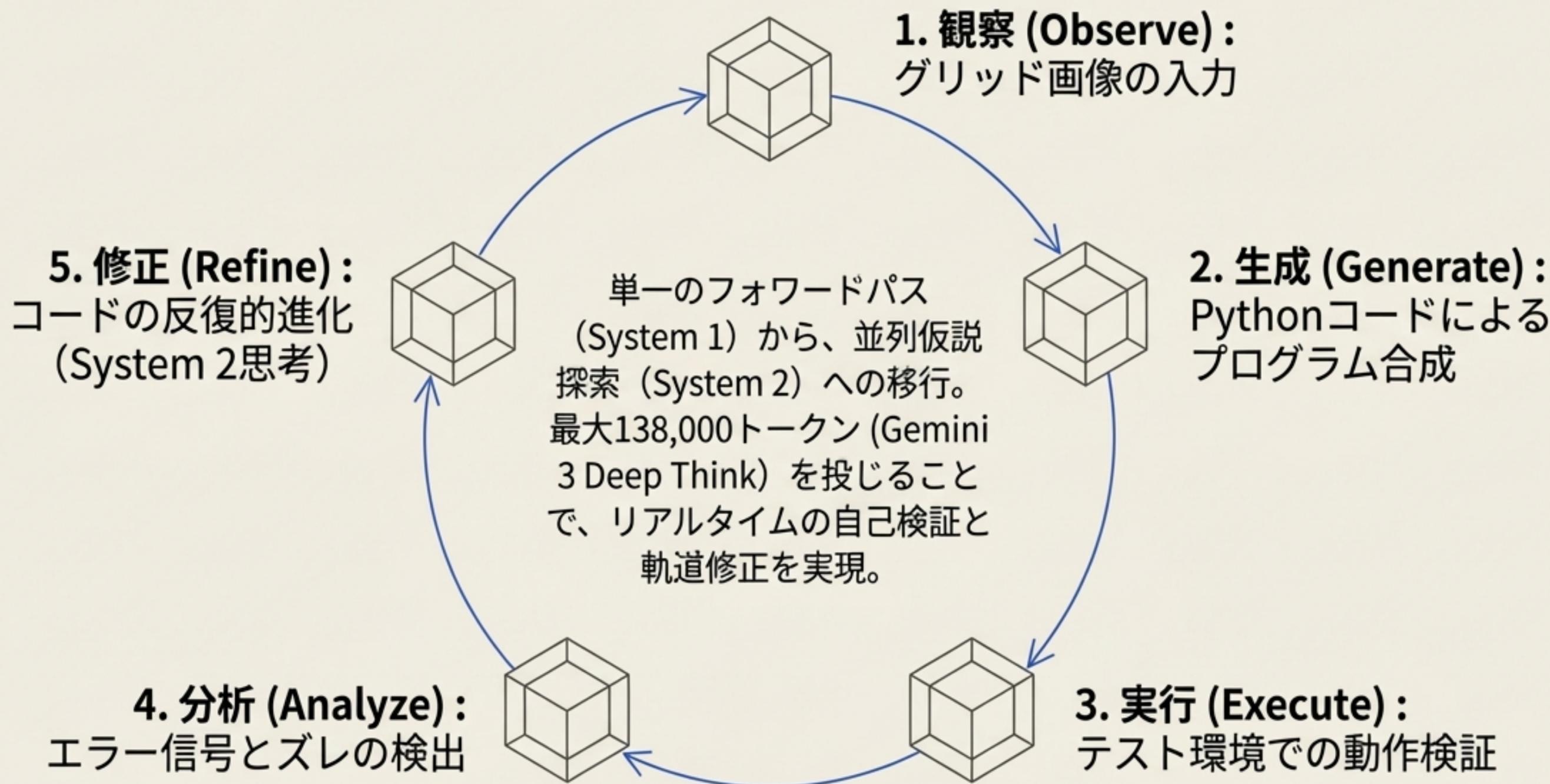
ARC-AGI-2攻略の軌跡：わずか11ヶ月での飛躍



知能の本質は「効率」にある：コスト vs パフォーマンス



ARC-2の「鍵」：進化型洗練ループ (Refinement Loops)

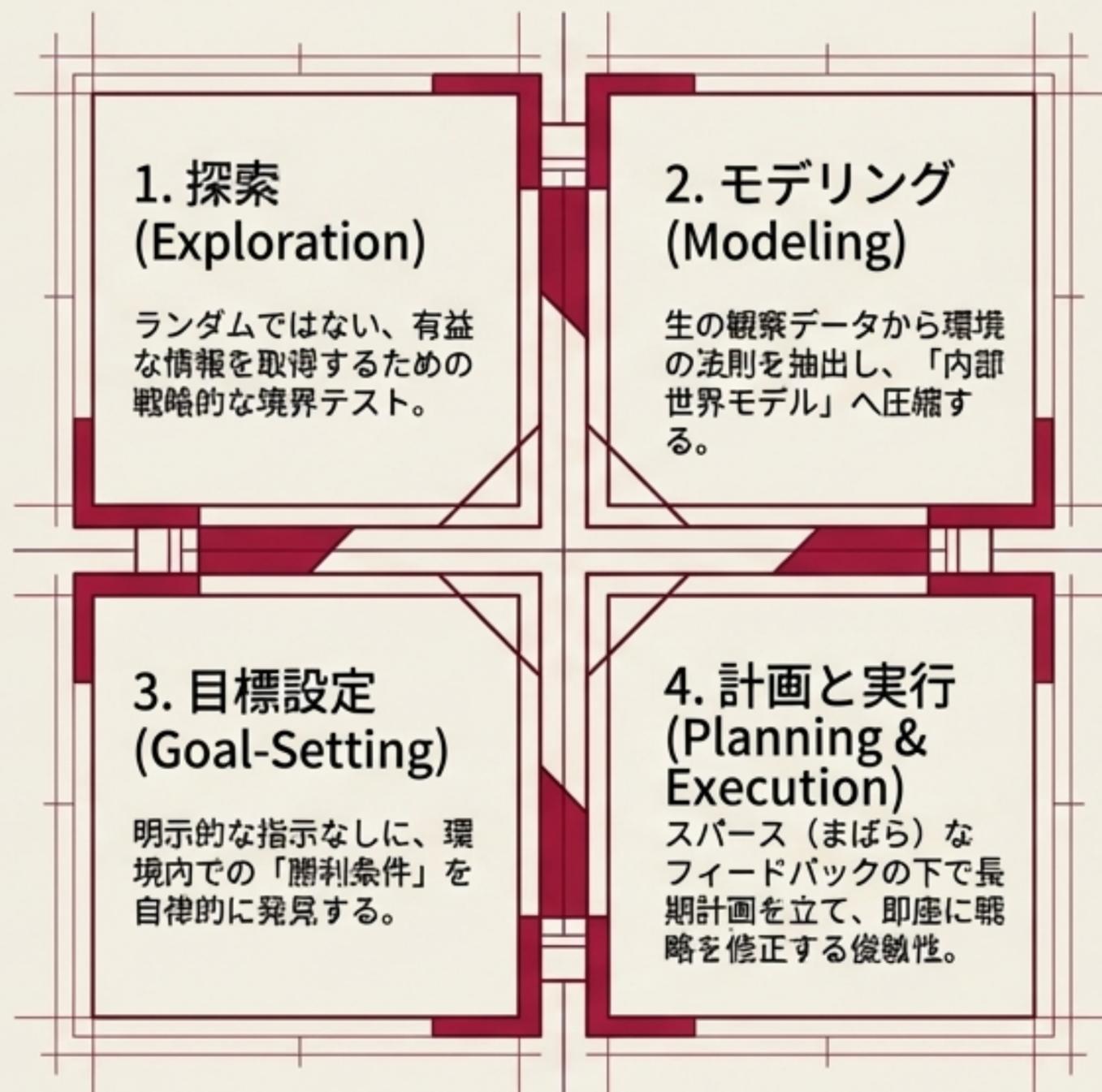


次なる絶壁：ARC-AGI-3へのパラダイムシフト

	[ARC 1 & 2]	[ARC 3]
評価フォーマット	静的な2Dグリッド画像	インタラクティブなターン制動的環境
ゴールの提示	入出力例から明示的に類推	非明示 (環境からの自律的発見が必須)
評価基準	最終的な出力の正確性 (Accuracy)	行動効率性 (RHAE: 人間との相対比較)
AIの攻略パラダイム	テスト時計算とプログラム合成	未解決 (リアルタイム世界モデル構築)

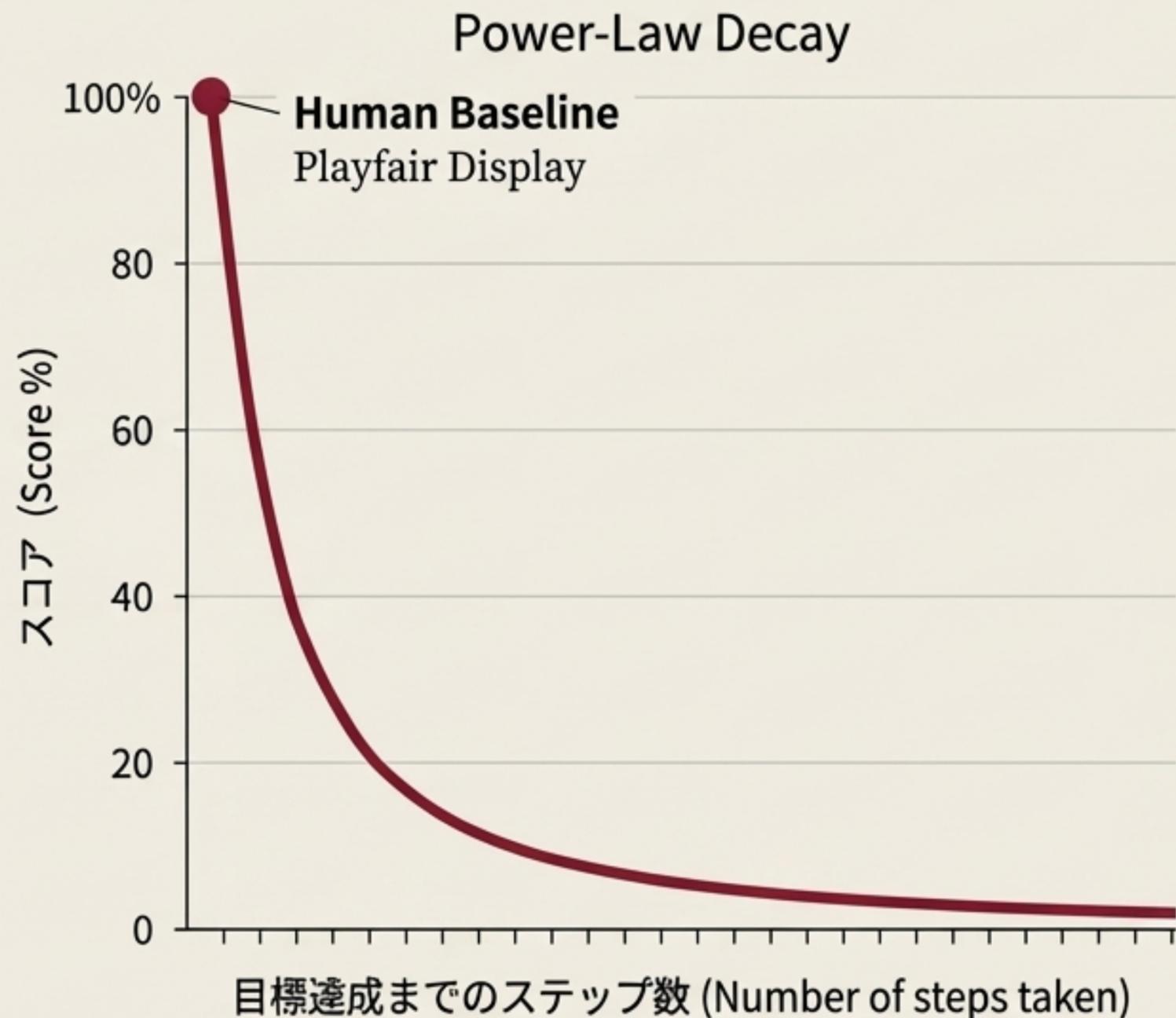
評価の軸が「受動的な静的推論」から、「未知のゲーム環境を自力で解明する動的エージェント」へと完全に移行した。

エージェント的知能の解剖学：ARC-AGI-3の4つの柱



1秒間に1000フレームを処理する独自のPythonベース環境内で、これら人間の認知発達に近いプロセスが統合的に評価される。

ブルートフォースの無効化：RHAE（相対的人間行動効率）

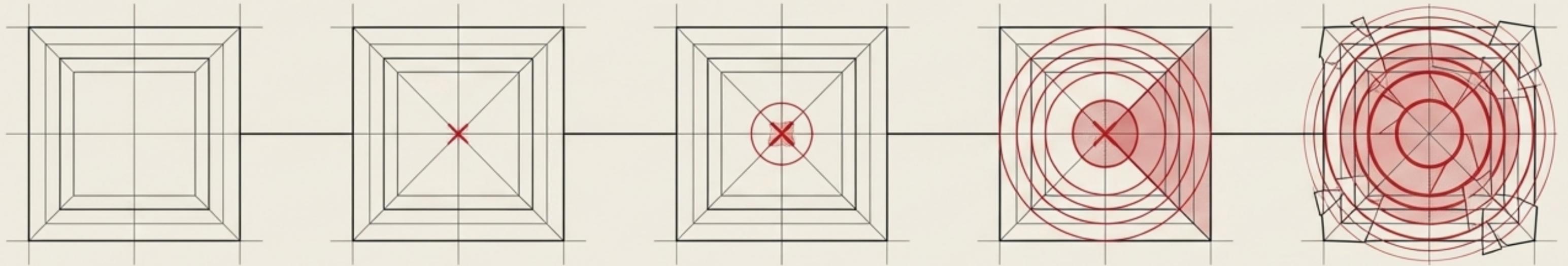


- 新評価基準「RHAE」は、最終的なクリア可否ではなく「**行動の無駄のなさ**」を測定する。
- 人間は事前のコア知識から素早く仮説を立て、最小手数でクリアする。
- AIが強化学習や試行錯誤（総当たり）で10倍の手数をかけると、スコアは「べき乗則」により劇的に低下し1%未満となる。

現在の最高峰システム（StochasticGoose等）でも、行動効率はわずか「**約12.58%～13%**」に留まる。

既存のスケーリング則の崩壊

Googleの研究「Towards a Science of Scaling Agent Systems」が示す冷酷な現実。



- **並列探索の限界:** 直列的な行動（前の結果に依存する行動）が求められる環境では、多数の仮説を並列生成するARC-2の手法は機能しない。
- **破滅的故障モード:** 不要な推論ステップや微小なエラーが連鎖的に増幅し、計算リソースを増やすほどかえって性能が低下する。

単にエージェントの数を増やすアプローチはすぐに天井に達する。自己回帰的な「次トークン予測」に最適化された現在のLLM/VLMでは、エージェントシステムの天井にすでに直面している。

ARC-AGI-3突破への鍵：第3のパラダイムシフト



Pillar 1: MDLに基づく情報圧縮 (Minimum Description Length)

- 確率的な「パターンの暗記・予測」からの脱却。
- 観察された現象を最も簡潔なルールとして「圧縮」することに報酬を与える内部世界モデルの構築（CompressARCやTRMの発展系）。

Pillar 2: 継続学習と神経記号的メモリ (Continual Learning & Neuro-symbolic Memory)

- 固定された重みへの依存からの脱却。
- 事前のプロンプトなしに、プレイを通じて学んだルールをリアルタイム（オンザフライ）で自身の内部表現に組み込む動的なエピソード記憶。

AGI到達のタイムライン：2029年～2031年の予測



ARC-AGI-3は既存技術の延長（スケーリング）では解決できず、第1波と同等の根本的なパラダイム移行期間を要する。専門家のコンセンサスは、リリースから3～5年の開発サイクルを示唆している。

真の人工汎用知能（AGI）の入り口

“

ARC-AGI-3において人間と同水準の行動効率（70%以上のスコア）が観測された時、人類は真の人工汎用知能（AGI）の入り口に足を踏み入れる。

”

-
- ベンチマークは飽和していない。ARC-AGI-3は、唯一残されたAGIへの北極星（North Star）である。
 - AI開発の最適化目標は、直ちに「予測(Prediction)」から「圧縮とエージェンシー（Compression & Agency）」へと転換されなければならない。