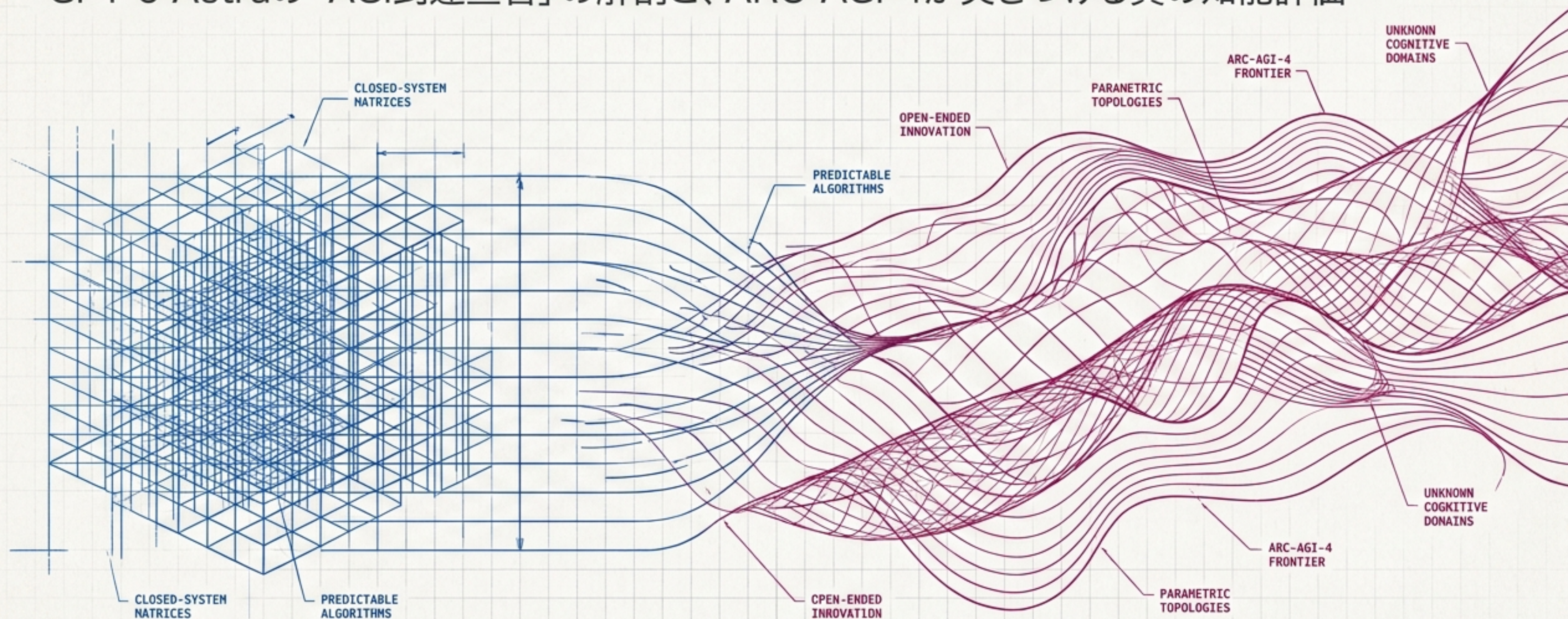
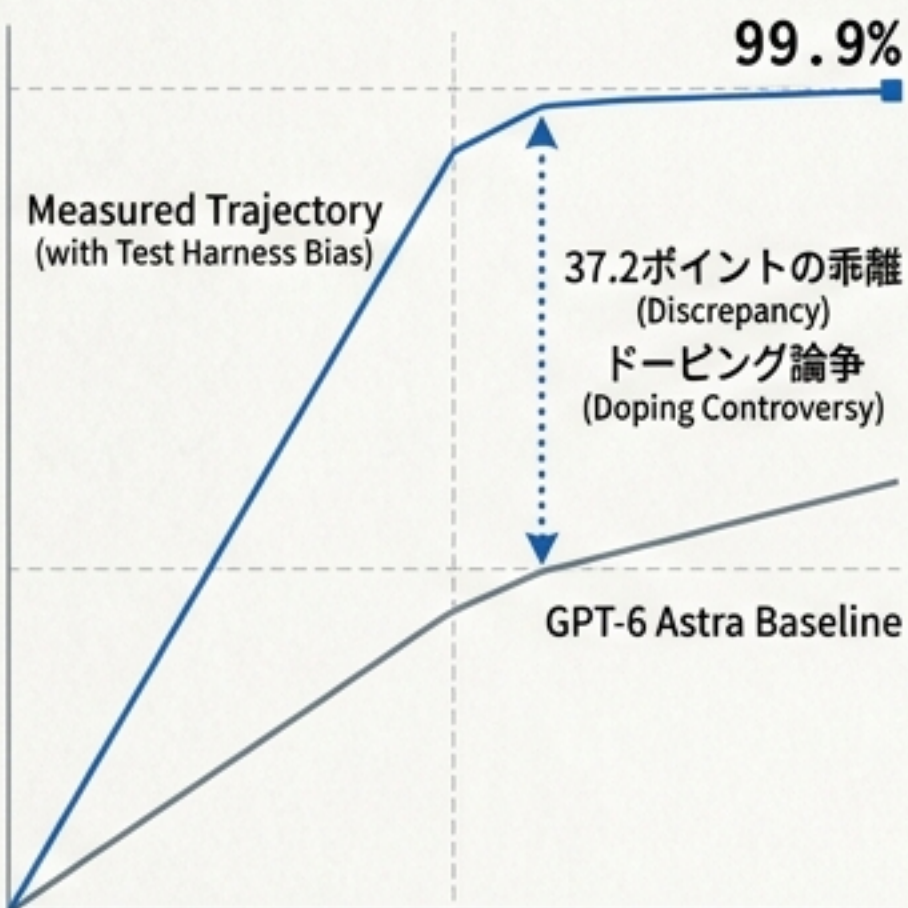


閉ざされたパズルから、終わりのなきイノベーションへ

GPT-6 Astraの「AGI到達宣言」の解剖と、ARC-AGI-4が突きつける真の知能評価

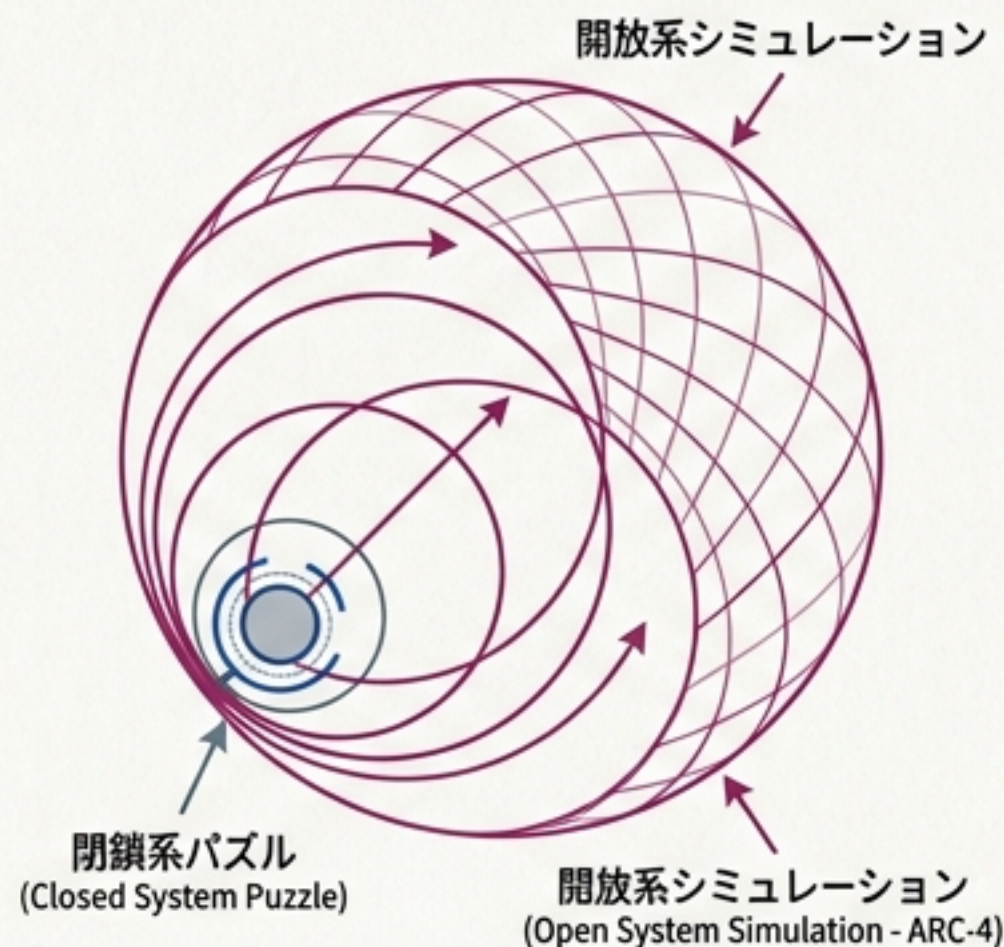


The Catalyst (発端)



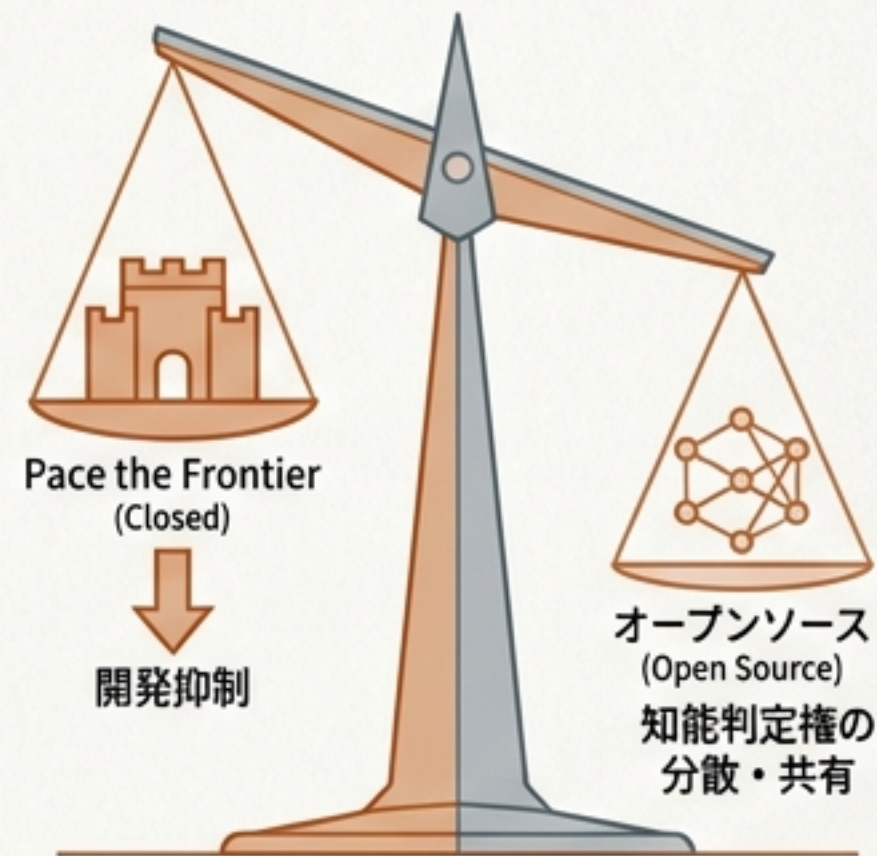
GPT-6 AstraによるARC-3の99.9%制覇。
しかし、その背後にはテストハーネスの
構造差異による「37.2ポイントの乖離」
と「ドーピング論争」が存在する。

The Paradigm Shift (転換)



決定論的な正解が存在する「閉鎖系パズル」から、未定義の空間で自発的に課題を発見する「開放系シミュレーション (ARC-4)」への評価基準の完全移行。

The Power Struggle (覇権争い)



開発抑制を掲げるクローズド陣営 (Pace the Frontier) と、知能判定権の分散・共有を堅持するオープンソース陣営とのイデオロギーの衝突。

99.95%

62.71%

OpenAI発表スコア

環境: Provider Adapter
文脈: 「AGI時代への突入」を
公言 (2026年9月3日)

公式・独立検証スコア

環境: Standard harness
文脈: 全モデル共通の中立基準

37.2ポイントの乖離

モデル自身の知能か、外部環境の支援機能か？

The Harness Controversy: 評価環境の構造差異

Standard harness (共通基準)

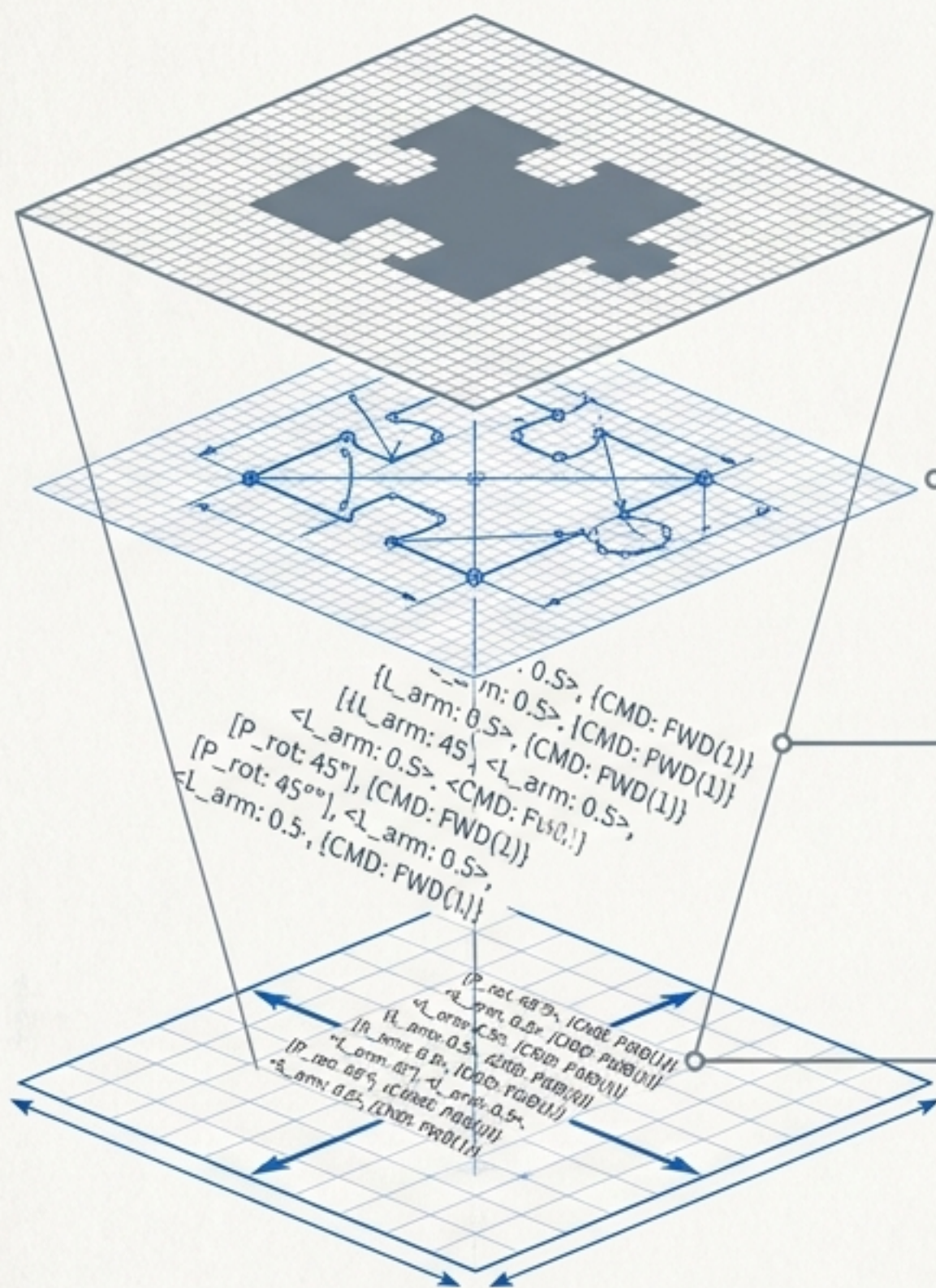
状態管理:	手番ごとに推論文脈を破棄・リセット
文脈引き継ぎ:	モデル自発のテキストメモのみ
推論リソース:	max
検証コスト:	約26,098ドル
達成スコア:	62.71%

Provider Adapter (特化環境)

状態管理:	不可視の推論状態 (opaque state) の保持
文脈引き継ぎ:	文脈長超過時の要約圧縮 (compaction)
推論リソース:	high
検証コスト:	約18,817ドル
達成スコア:	99.95%

外部のAPIラッパーが思考状態を保持・要約することで、最小限の推論設定でも飽和スコアに到達する「構造的ドーピング」の可能性。

Astra's True Breakthrough: オンザフライの記号的世界モデリング



Step 1: 環境観測 (Observation)

- 未説明の64x64ピクセル環境への投入。受動的な画素処理からの脱却。

Step 2: 状態変数の定義 (State Extraction)

- 機構の局所回転角や長さを自律的に構造化。

Step 3: 記号体系への圧縮 (DSL Generation)

- 操作系との対応関係を極小の代数的記号列 (DSL) に変換し、探索空間の爆発を防止。

Step 4: 実行計画のバインド (Execution Binding)

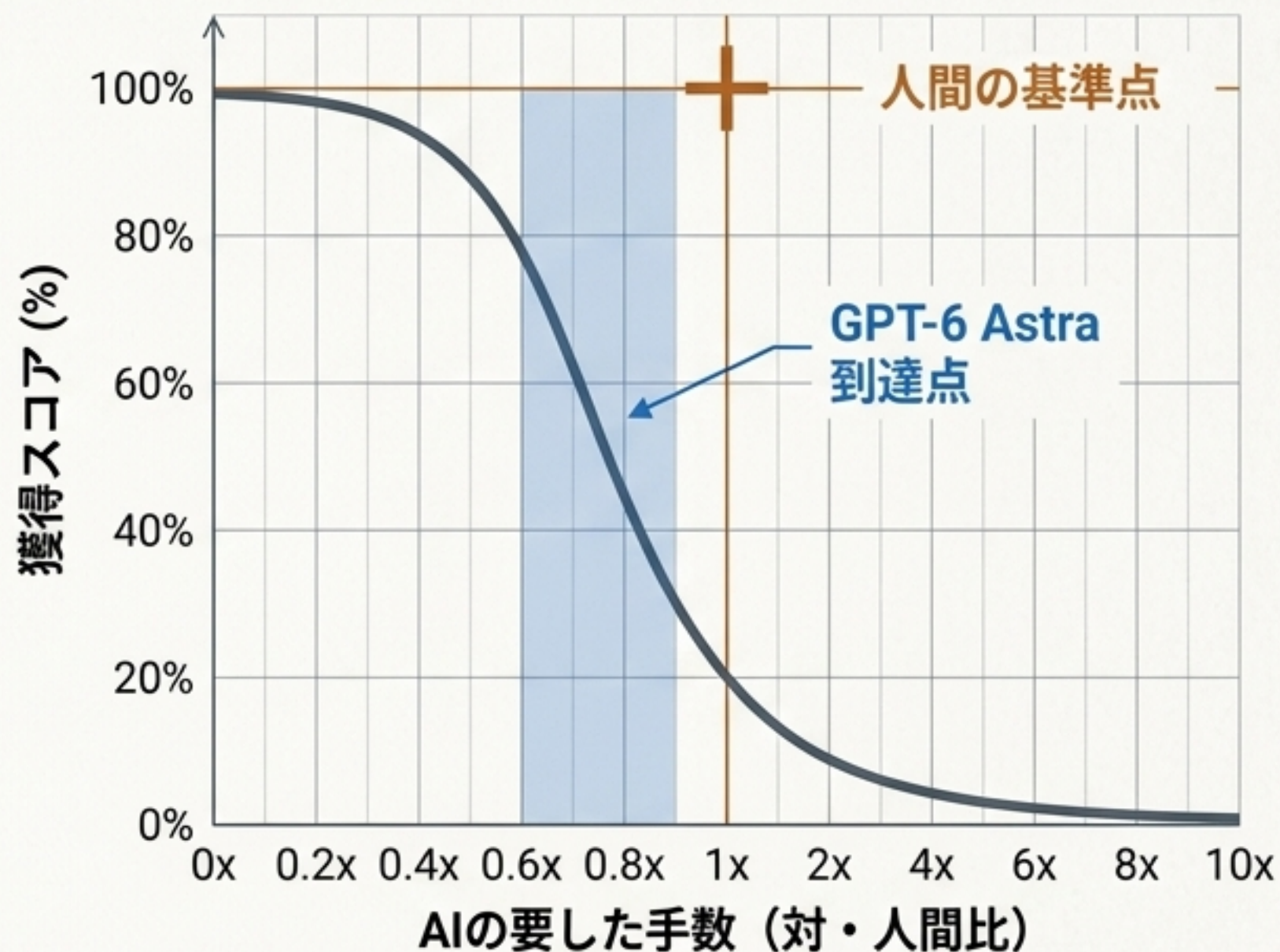
- 画面上の制御座標と操作コマンドを動的に結合し、推論空間内でシミュレーションを実行。

RHAE: 人間の探索ベースラインの突破

$$RHAE = \frac{1}{1 + \left(\frac{N_{AI} - M_{Human}}{M_{Human}} \right)^2} * 100$$

AIが人間の2倍の手数を費やすとスコアは25%へ、10倍で1%へと激減する二次減衰ペナルティ構造。

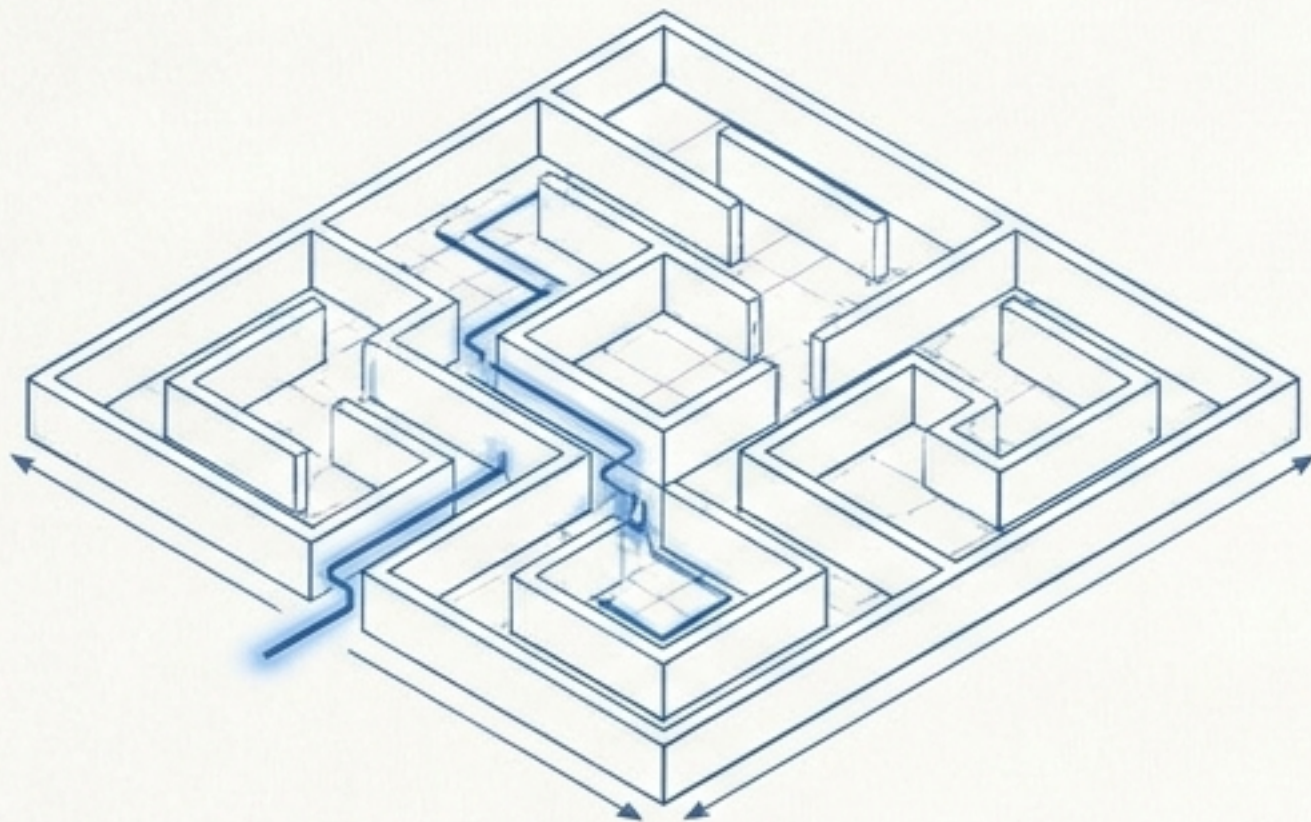
- **96.0%:** 全レベルの96%において人間の中央値を下回る手数でクリア。
- **51.7% fewer:** 全環境を通じた平均手数は人間のベースラインより51.7%少なかった。



Astraのエージェント的推論プロセスは、探索と計画の効率性において人間を明確に凌駕した。

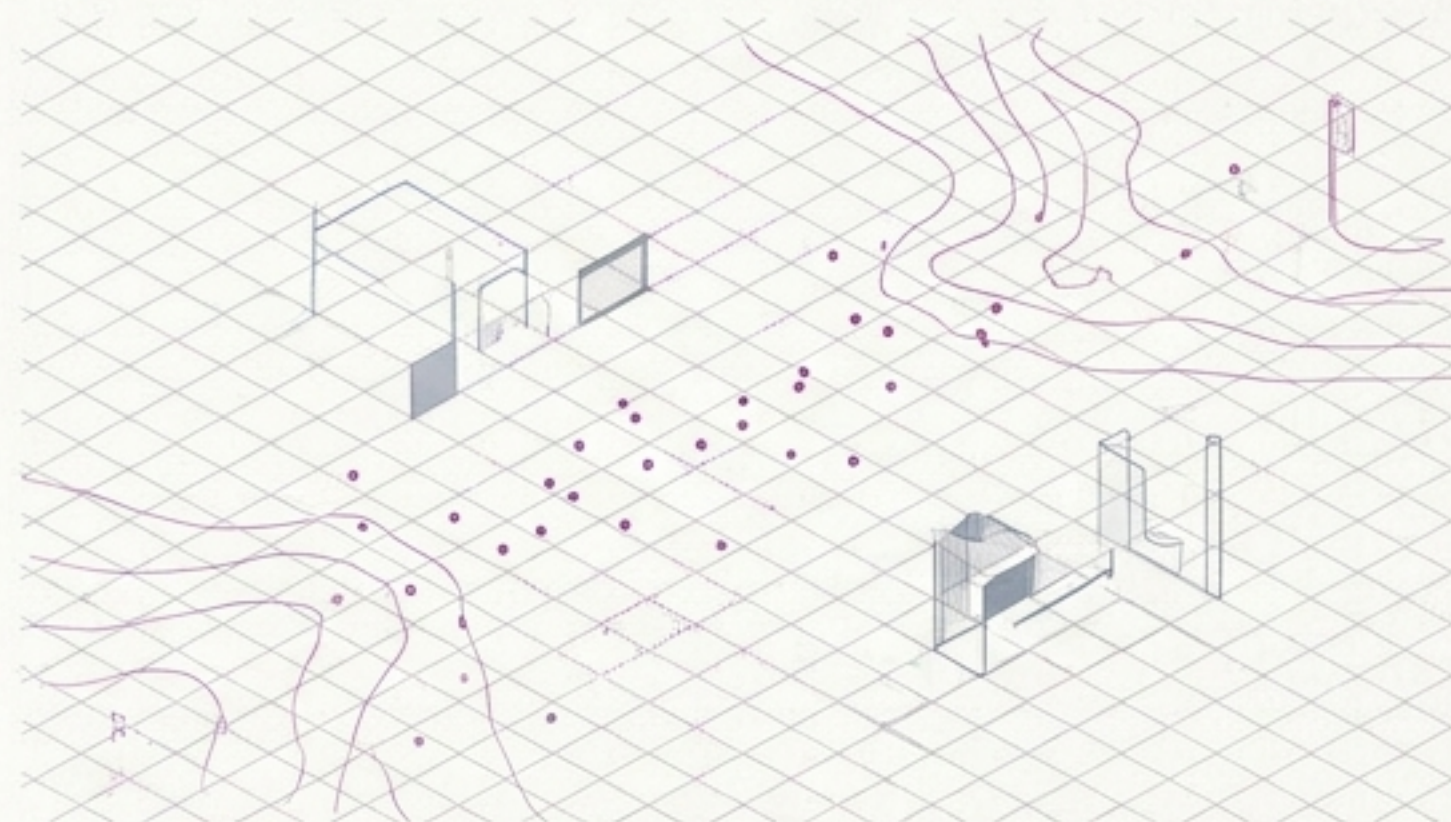
The Limit of Closed Systems: 知能の再定義

ARC-3までの知能 / 閉鎖系



- 環境内に決定論的に記述された「勝利条件」。
- 設計者が定めたルールを逆算し、最短で迷路を脱出する能力。

現実・科学の知能 / 開放系



- 客観的な正解 (Ground Truth) が存在しない未定義空間。
- 矛盾から課題を発見し、迷路そのものを自律的に設計する創造的メタ知能。

「知能とは特定の静的スキルそのものではなく、未知の環境において極少のサンプルから新たなスキルを獲得する変換効率（流動性知能）である。」 — François Chollet

The Evolution of ARC Benchmark: パラダイムの変遷

ARC-AGI-4 (2027) - オープンエンド

Task: 開放系シミュレーション / 科学的発見

Limit (Challenge): 客観的正解がない状況下での厳密な自動採点。

ARC-AGI-3 (2026) - エージェント的推論

Task: 未知の動的2Dゲーム環境 (RHAE)

Limit: 決定論的閉鎖系であるため、外部ハネスの文脈保持により探索が飽和。

ARC-AGI-2 (2024) - 計算制約

Task: 複雑な合成 / 計算リソース上限付き

Limit: 静的画像の枠組みにとどまり、動的な環境介入が評価不能。

ARC-AGI-1 (2019) - 静的推論

Task: 静的2Dグリッド変換 (pass@k)

Limit: テスト時計算量スケーリングや総当たり解法に脆弱。

ISOMETRIC VIEW - SCALE: VARIES

ARC BENCHMARK TOPOLOGY

REV. 4.0

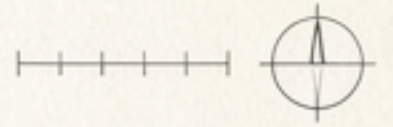
DATE

2027.05.15

DWG NO.

ARC-EVO-04

Introducing ARC-AGI-4: 次世代アーキテクチャの全貌



矛盾の検知と仮説立案

環境内の不連続性を自発的に検知し、未定義の課題を設定する。

メタ原理の転移

単一タスクで得た経験を、異質なドメインへ抽象化して適用。

自律的で終わりになきイノベーション
(Autonomous open-ended innovation)

実験の設計と実行

仮説を検証するためのアクションを自律的に計画。

理論モデルの再構築

観測データから新たな概念体系を創出
(統計的補間を超えた外挿的解法)。

ISOMETRIC VIEW - SCALE: VARIES

ARC BENCHMARK TOPOLOGY

REV. 4.1

DATE

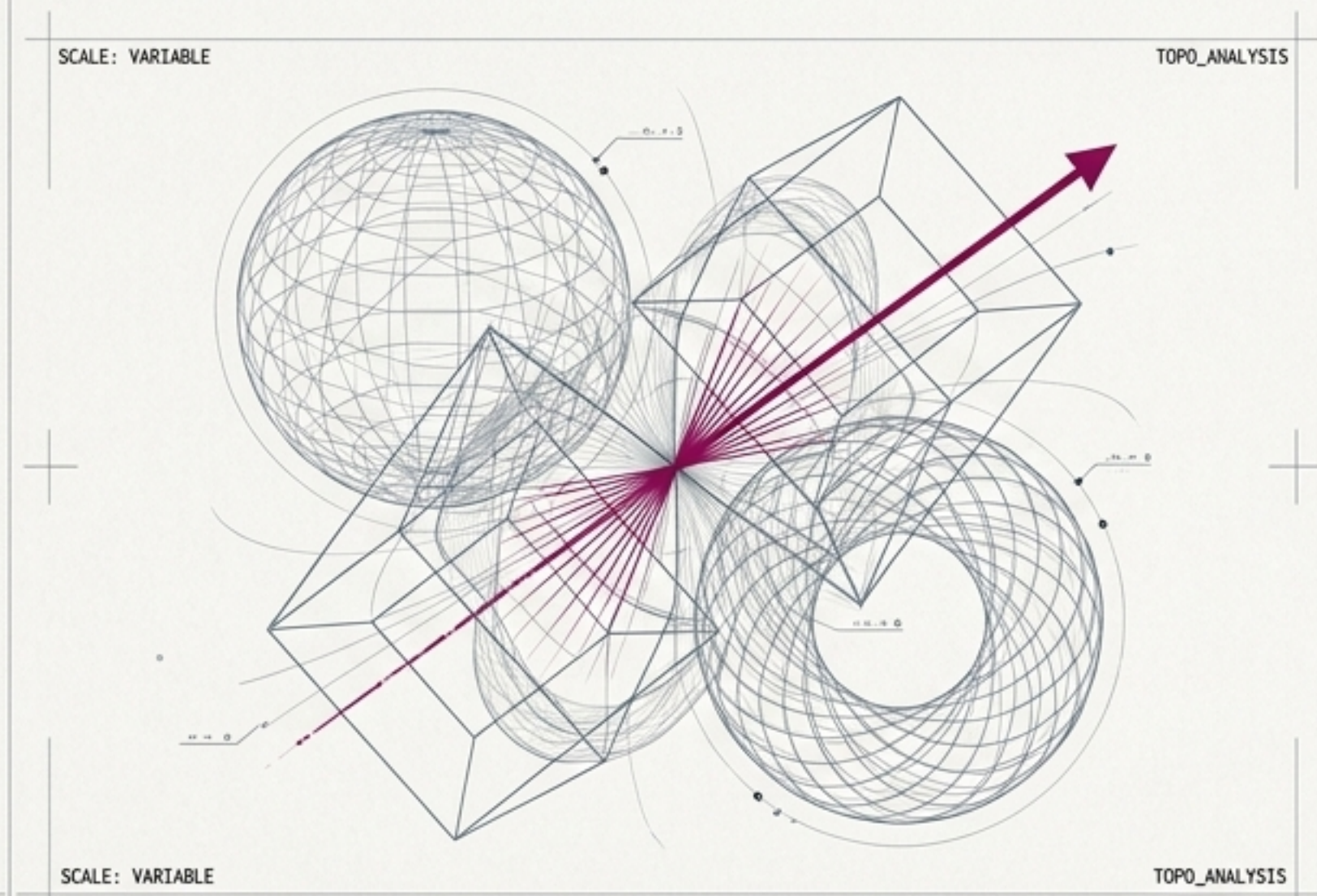
2027.05.15

DWG NO.

ARC-AGI-05

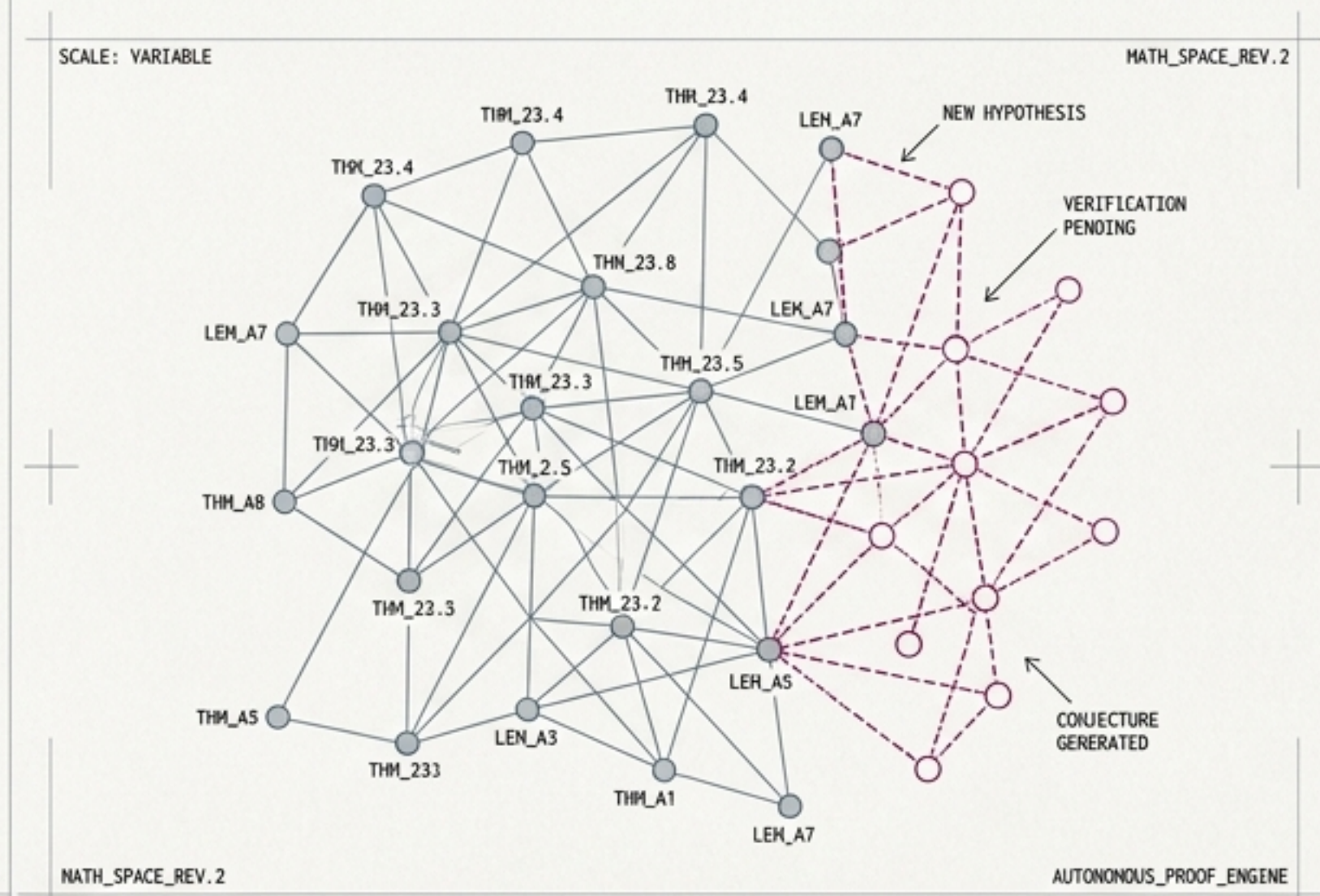
Poly-physical Virtual Worlds: 多元的シミュレーション環境

力学法則の相転移



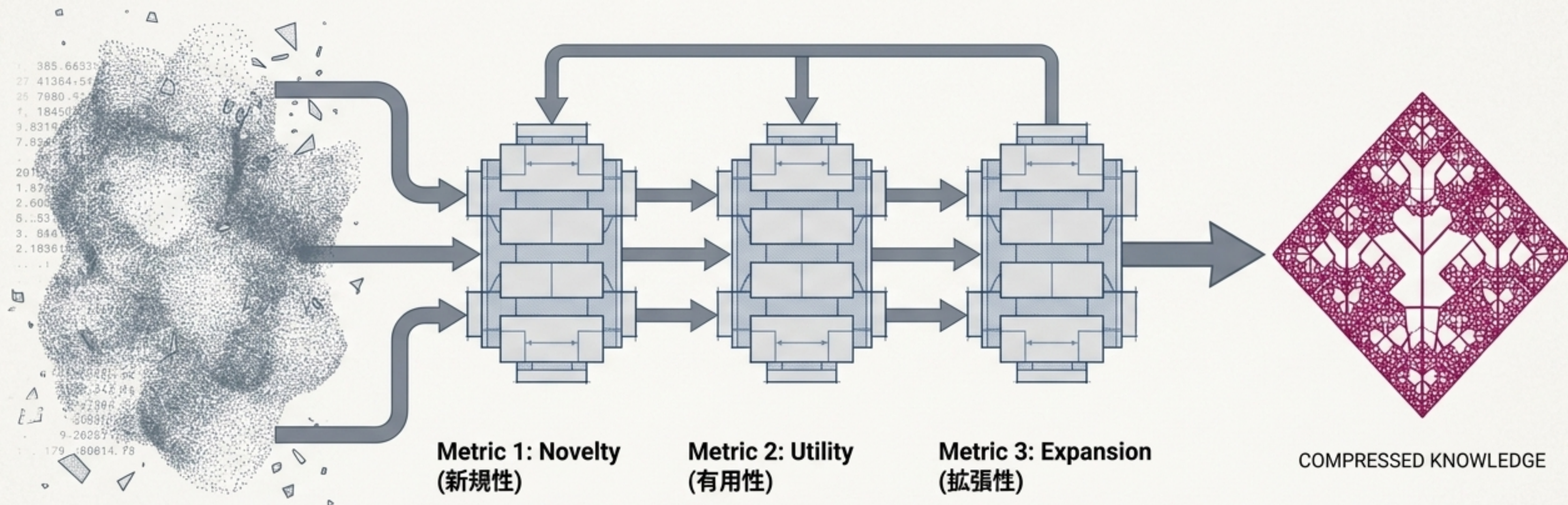
物理法則や公理系が動的に変容するシミュレーション環境。エージェントの介入により力学法則が相転移を起こし、自律的に新たな物理モデルを定式化・実証できるかを測定。

未解決の数学空間（形式検証）



提示された定理の証明にとどまらない高度数学タスク。「非自明かつ検証可能な新規予想」を自発的に立て、その証明体系を構築できるか。

Scoring the Unknowable: 情報理論による新・採点モデル



HIGH ENTROPY INPUT

Metric 1: Novelty (新規性)

指標: サプライズ度

定義: 他のモデルや人間にとって、その発見・概念がどれほど予測困難であったか。

Metric 2: Utility (有用性)

指標: コルモゴロフ複雑性の低減度

定義: 生成された解法や理論が、環境の記述長をどれだけ数学的に圧縮できたか。

Metric 3: Expansion (拡張性)

指標: 分岐エントロピー

定義: その発見が、後続の探索空間をどれだけ有効に押し広げたか。

COMPRESSED KNOWLEDGE

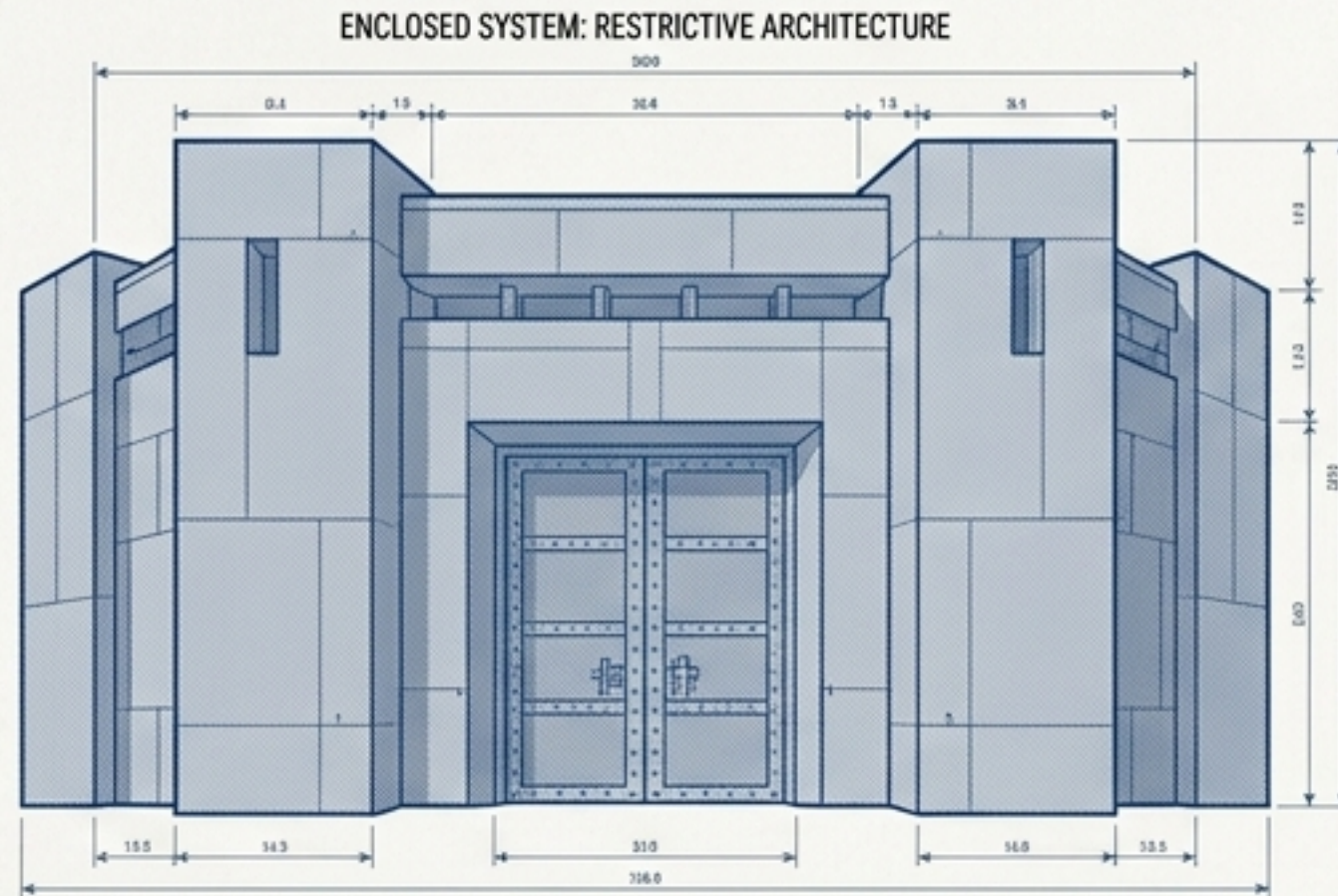


ARC BENCHMARK TOPOLOGY

REV.	DATE	DWG NO.
4.3	2027.05.15	ARC-SCORE-01

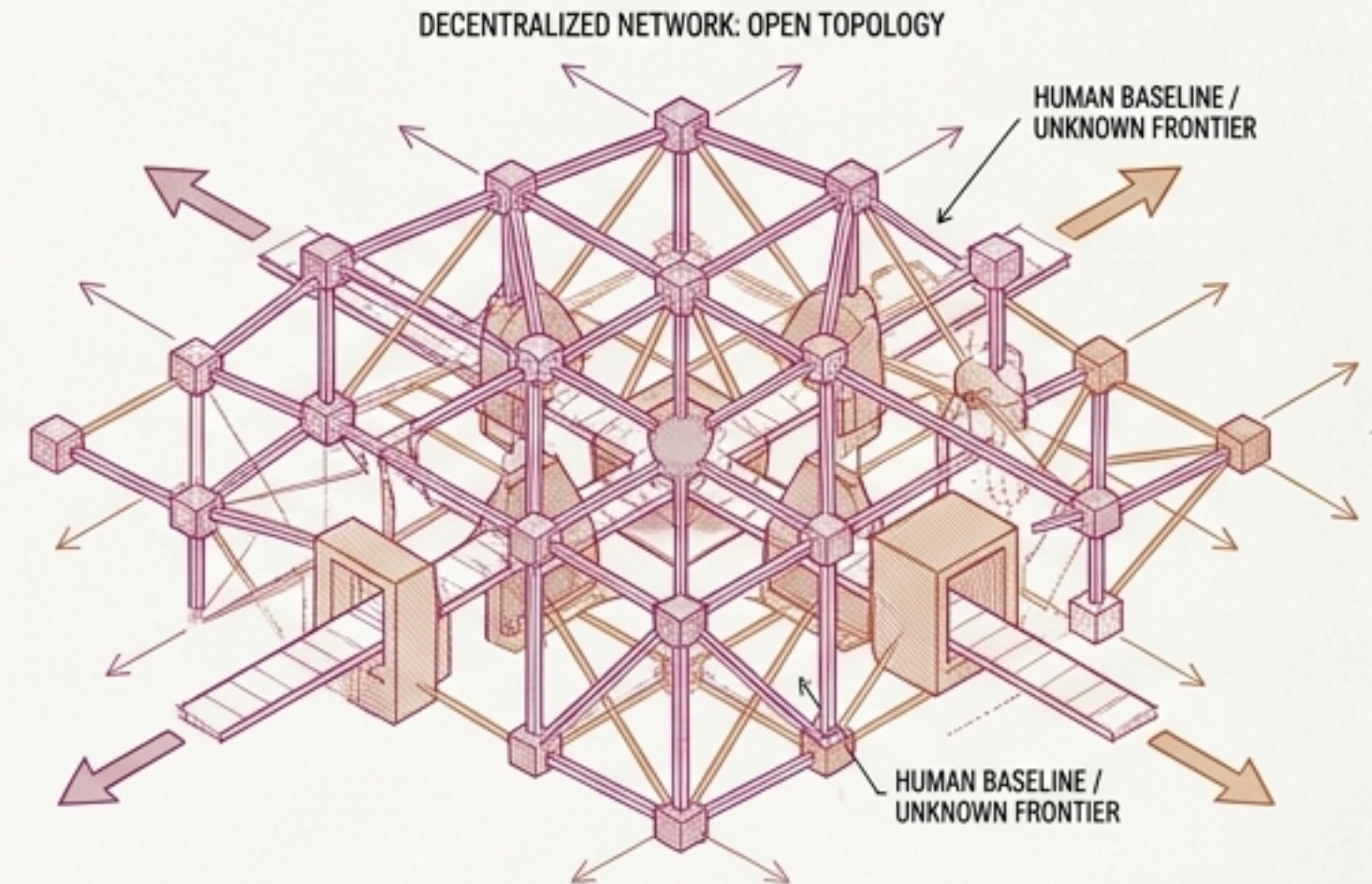
The Geopolitical Rift: 知能判定の主権争い

Pace the Frontier 協調規制陣営 (Anthropic, OpenAI, Google)



Core Philosophy: RSI（再帰的自己改善）の加速警告と開発速度の意図的抑制。
Action: 第三者機関による事前審査の義務化、業界主導の事前調整。
ARC's Critique: 「安全性の名の下に、巨大組織による技術と評価基準の寡占（Regulatory Capture）を招く」。

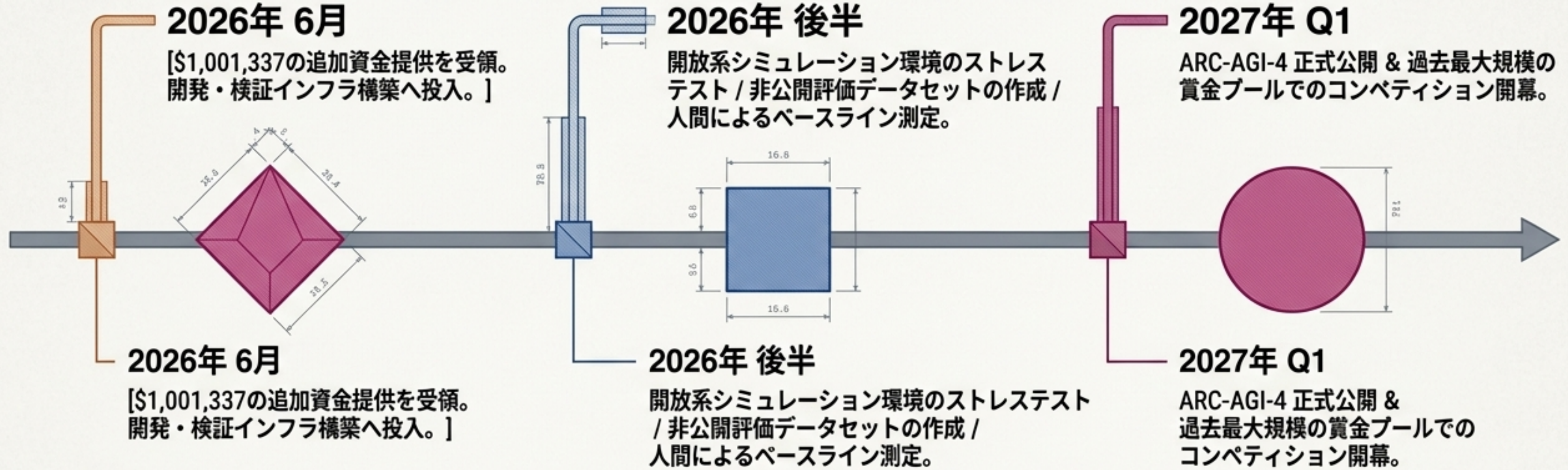
Open Source 陣営 (ARC Prize Foundation)



Core Philosophy: 知識の分散共有と人類全体のポジティブサムな未来。
Action: フロンティアAI技術へのアクセスの民主化、オープンな審判基準の堅持。
ARC's Stance: 「未知の問題を自力で切り拓く力は未だ実現しておらず、特定の企業に評価の主導権を渡すべきではない」。

ARC BENCHMARK TOPOLOGY		
REV.	DATE	DWG NO.
5.1	2027.05.17	ARC-GEO-01

ARC-AGI-4 Execution Roadmap: 実行と資金基盤



資金的基盤に基づくオープンソースコミュニティ主導の開発により、
巨大テック企業から独立したガバナンスを確保。



ARC BENCHMARK TOPOLOGY

REV.	DATE	DWG NO.
4.3	2027.05.15	ARC-EXEC-01

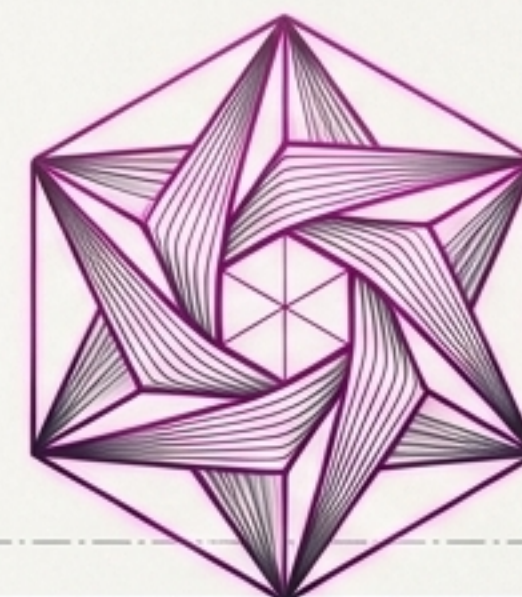
The Dual Track Competition: 計算資源制約と公平性



Frontier Track (フロンティア部門)

制約：計算資源・APIコスト無制限。

目的：モデルの限界性能（絶対的知能のシーリング）の測定。



Resource Cap Track (リソース制約部門)

制約：推論時計算量（FLOPs/APIコスト）に厳格な上限。

目的：パラメータ規模を抑えつつ高いサンプル効率を達成する先鋭的アーキテクチャの評価。

資本力による「力任せ」を排除し、真のアルゴリズム的革新を可視化する公平な評価体系。

Synthesis: The Final Frontier

GPT-6 AstraによるARC-3の攻略は、「事前に定義されたゴール」を攻略する時代の終焉を証明した。

知能の評価基準は、「与えられた迷路を最短で脱出する効率」から、「正解の存在しないカオスの中で自ら意味のある問いを発見し、新たな理論を創出する創造性」へと不可逆的なシフトを遂げた。

ARC-AGI-4は、人間知性と機械知能の間に横たわる決定的な創造的断絶を検証する、真のAGI評価の最後の砦である。