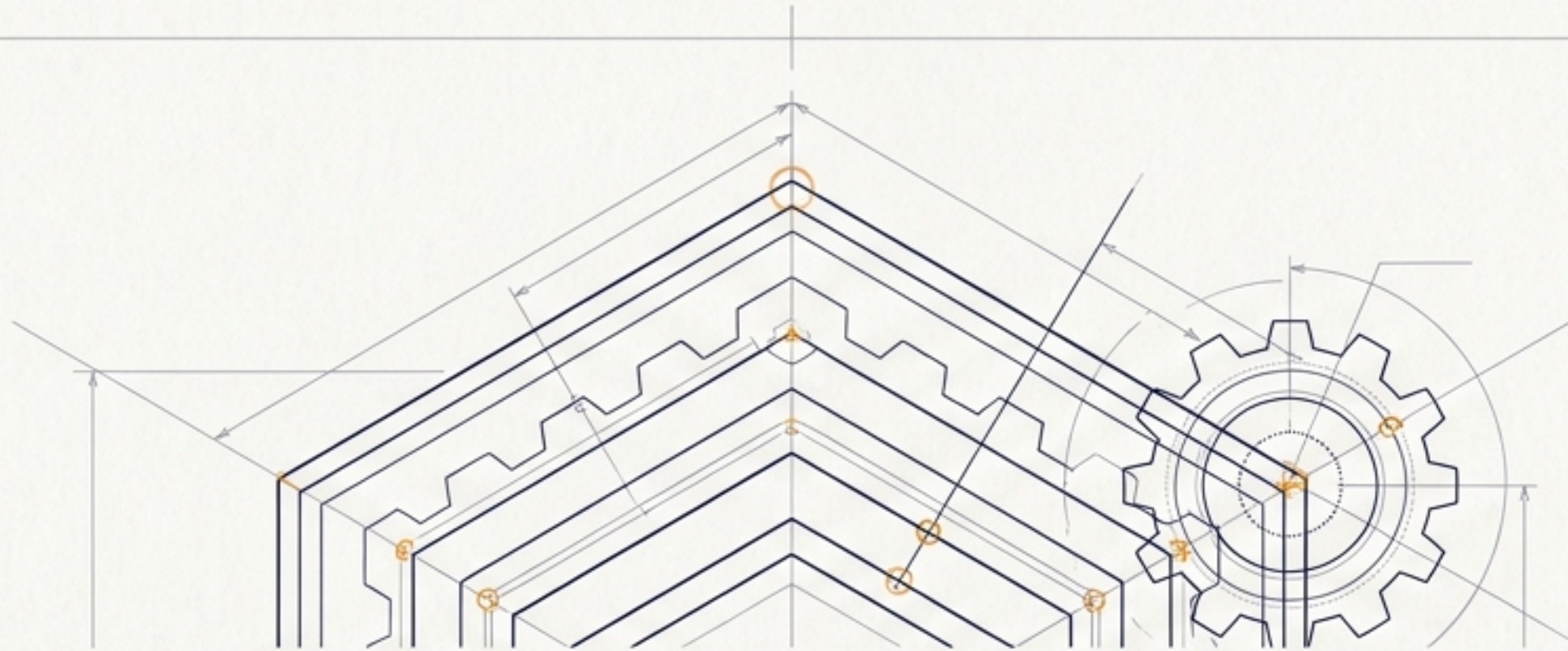
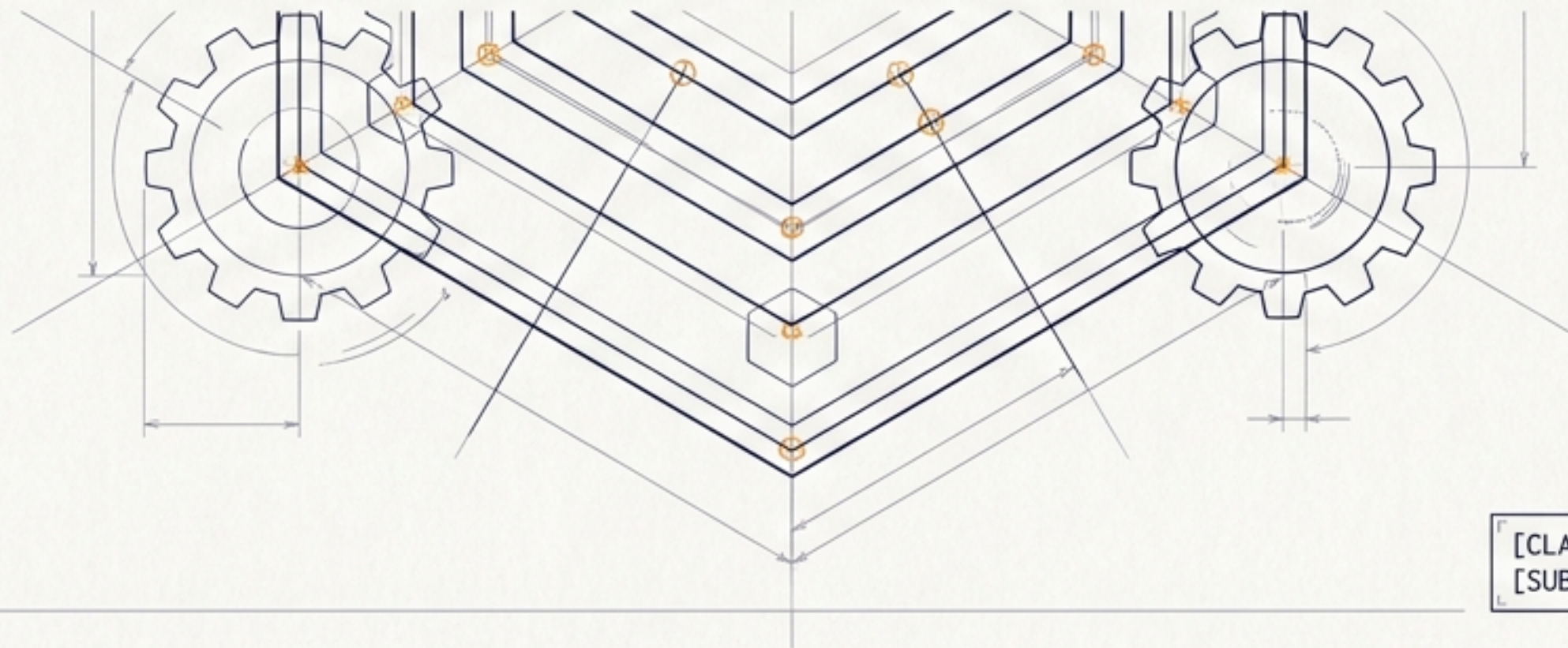




基盤モデルの限界を突破する「ハーネス」の設計思想

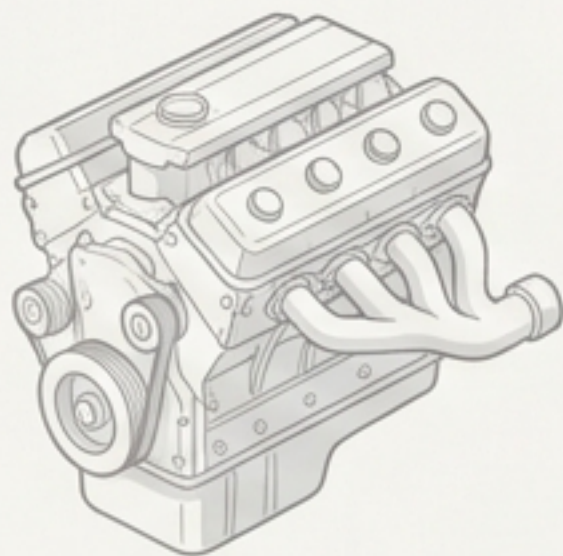
NVIDIA AVOによるARC-AGI-3完全制覇と、AIパラダイムの移行



[CLASSIFICATION: TECHNICAL ANALYSIS]
[SUBJECT: AUTONOMOUS AGENT ARCHITECTURE]

エンジン単体の限界と、実行基盤がもたらす「100.00」の衝撃

生エンジン (Bare Engine)



30%

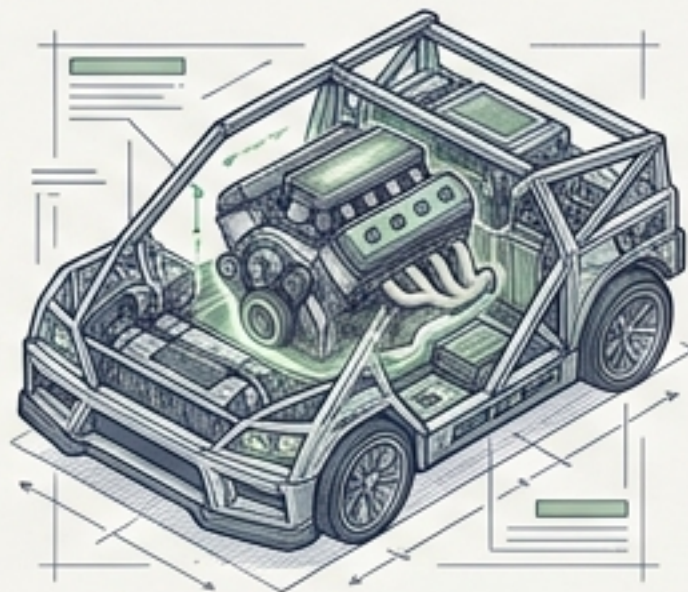


モデル単体:
Claude Opus 5 (High)

スコア: 30.16%
(ARC Prize標準設定)

結果:
記憶喪失と探索スタグネーションにより初期レベルで脱落。

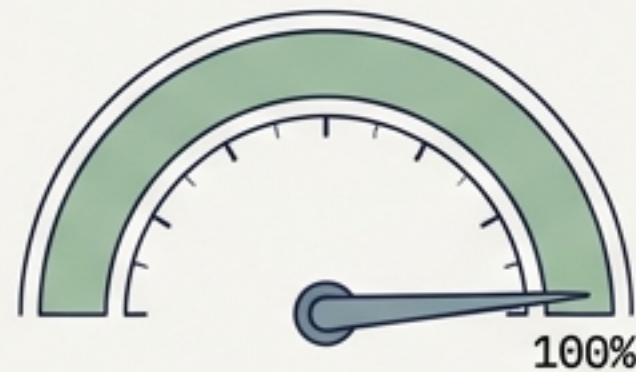
高度なシャーシ (Advanced Harness)



実行基盤適用:
NVIDIA AVO
(Claude Opus 5内包)

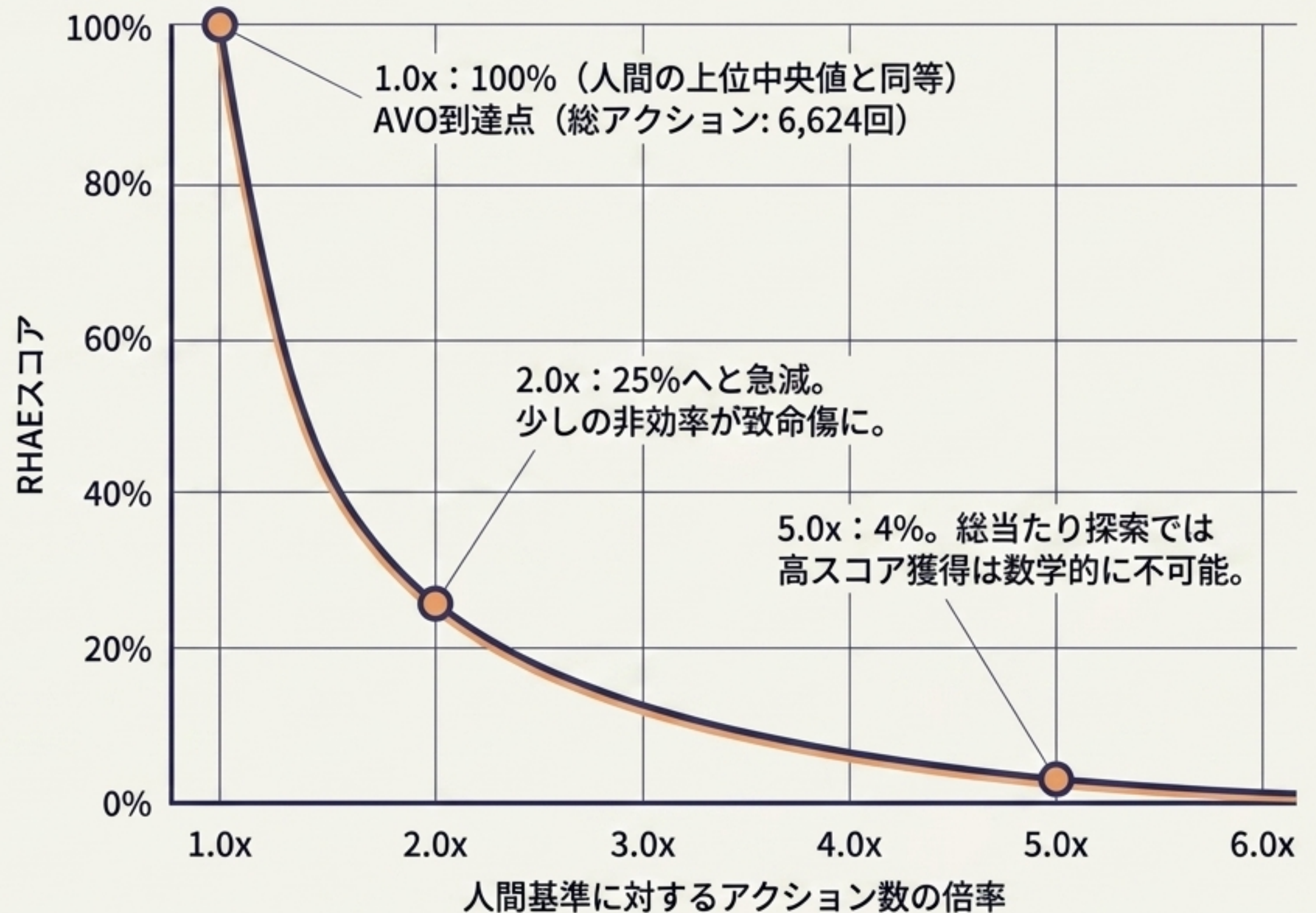
スコア: RHAE 100.00
(公開セット全183レベル完答)

インサイト:
適切な実行基盤 (Harness)
の設計が、モデルの実効性
実効性能を劇的に引き上げることを証明。

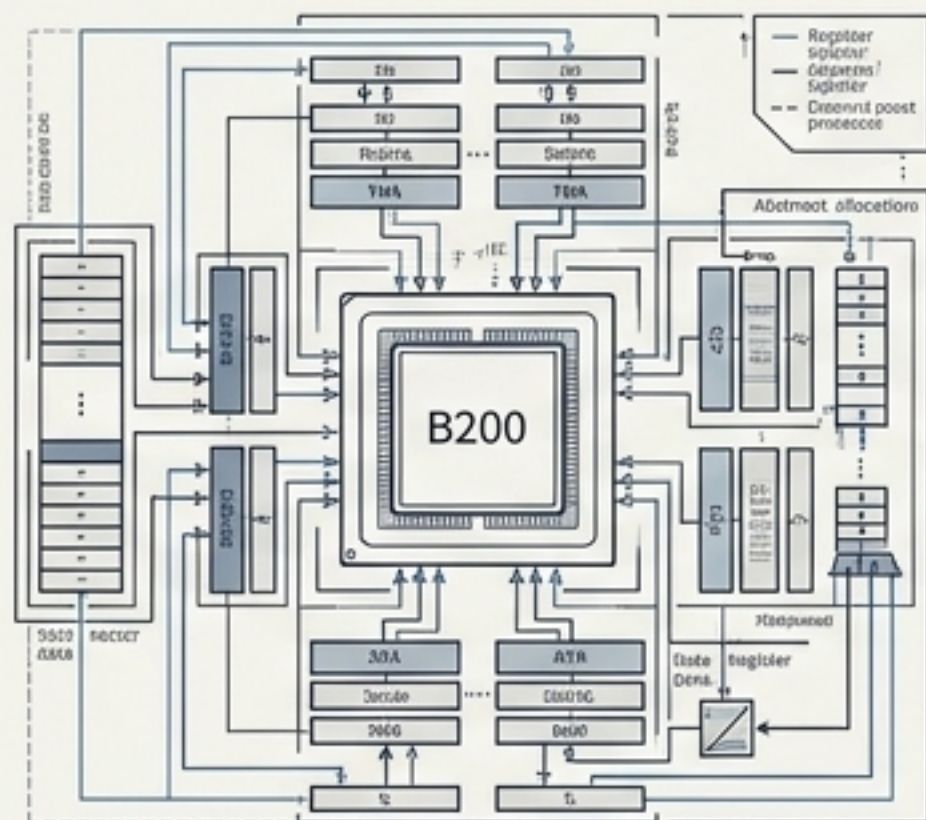


ブルートフォース（総当たり）を許さない過酷な評価指標

「後半の難関レベルへの重み付け」と「行動効率の2乗ペナルティ」。AVOは人間の初見上位陣と同等の効率（VISTA比約12%削減）で未知のロジックに到達。



低レイヤ工学と抽象パズルを繋ぐ汎用性 (Fluid Intelligence)

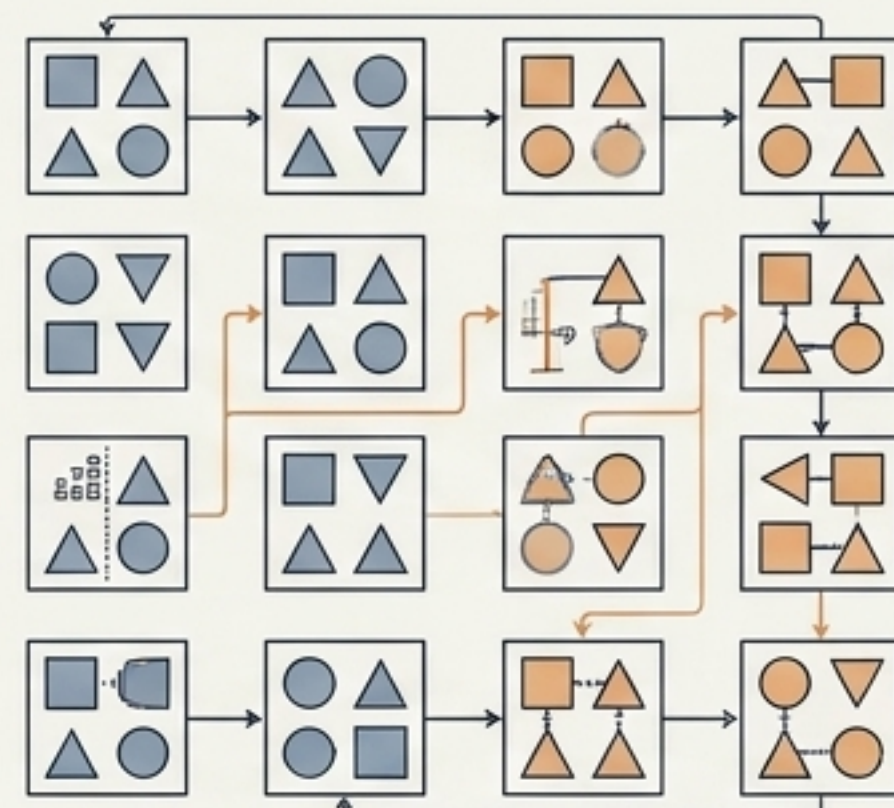


極限の低レイヤ工学

対象: Blackwell (B200) Attentionカーネル
課題: TMA非同期転送、Warp Specialization等の複雑な工学課題。

成果: 無人で7日間連続稼働。最大1,668 TFLOPS (BF16)を記録し、cuDNN比+3.5~7.0%を達成。

AVOコアエンジン

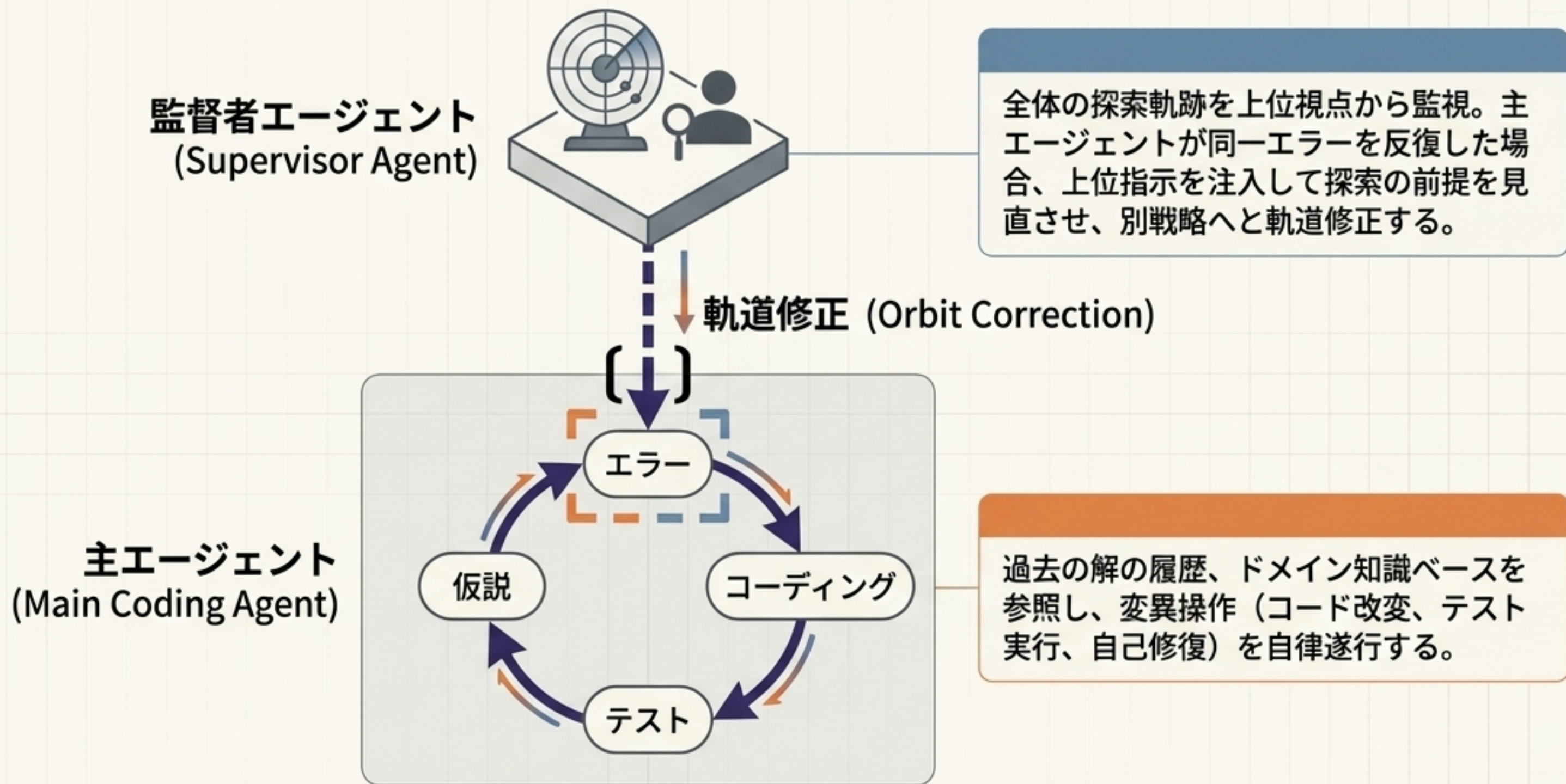


未知ルールの抽象推論

対象: ルールの事前説明がない未知の視覚的推論環境(ARC-AGI-3)。

成果: 専門的なハードウェア最適化向けに設計された認知ループを、そのままゲーム推論へと転用し完全制覇。

探索のスタグネーションを打破する「2層エージェントアーキテクチャ」



インサイト: 従来の進化的探索における単発の「変異」ステップを、自律的なコーディングエージェントのループへと完全に置換。

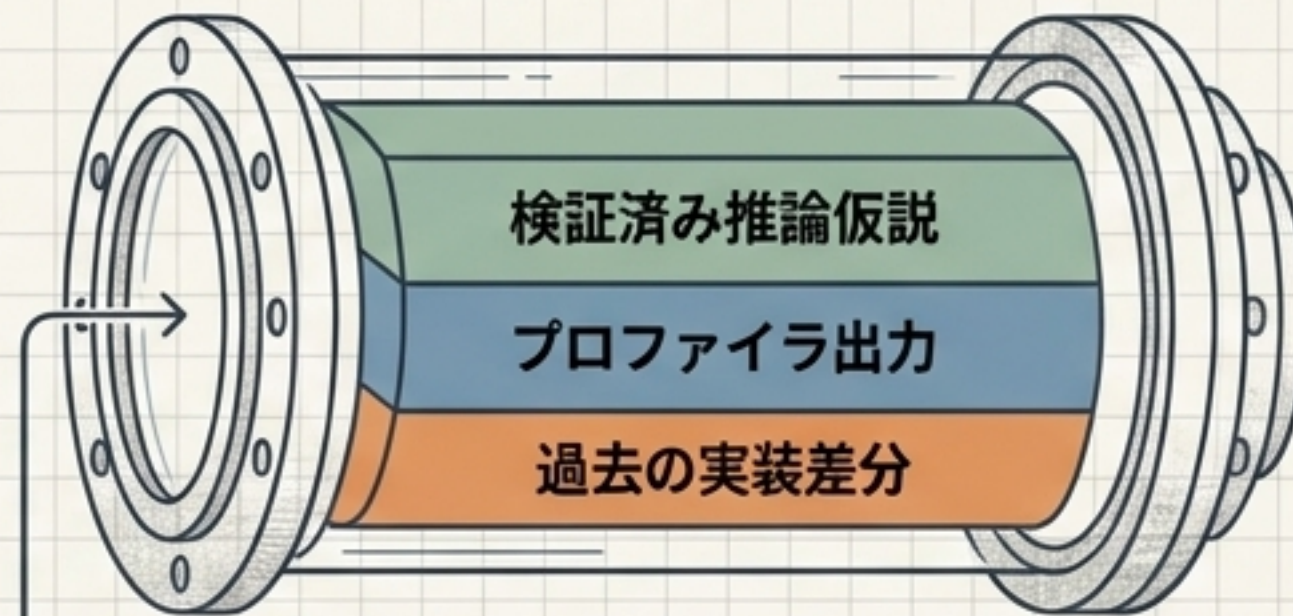
推論複利の法則 (The Compounding Reasoning Insight)

**単発のLLMは記憶喪失の天才だが、
優れたハーネスは推論を『複利』で成長させる。**



単発LLMのパイプライン

ターンごとにコンテキストがこぼれ落ち、
局所解でスタックする。



AVO / NOOAの永続メモリ

長大な試行履歴を破棄せずに引き継ぎ、新たな推論の
土台とする。NOOAではSQLiteをバックエンドに採用。

成果：7日間にわたる完全無人稼働（500以上の探索方向、40世代コミット）における
推論の連続性は、この「記憶の複利」によって担保されている。

知覚インターフェースの差異： 視覚的認識 vs 構造化データ

VISTA (MIT)



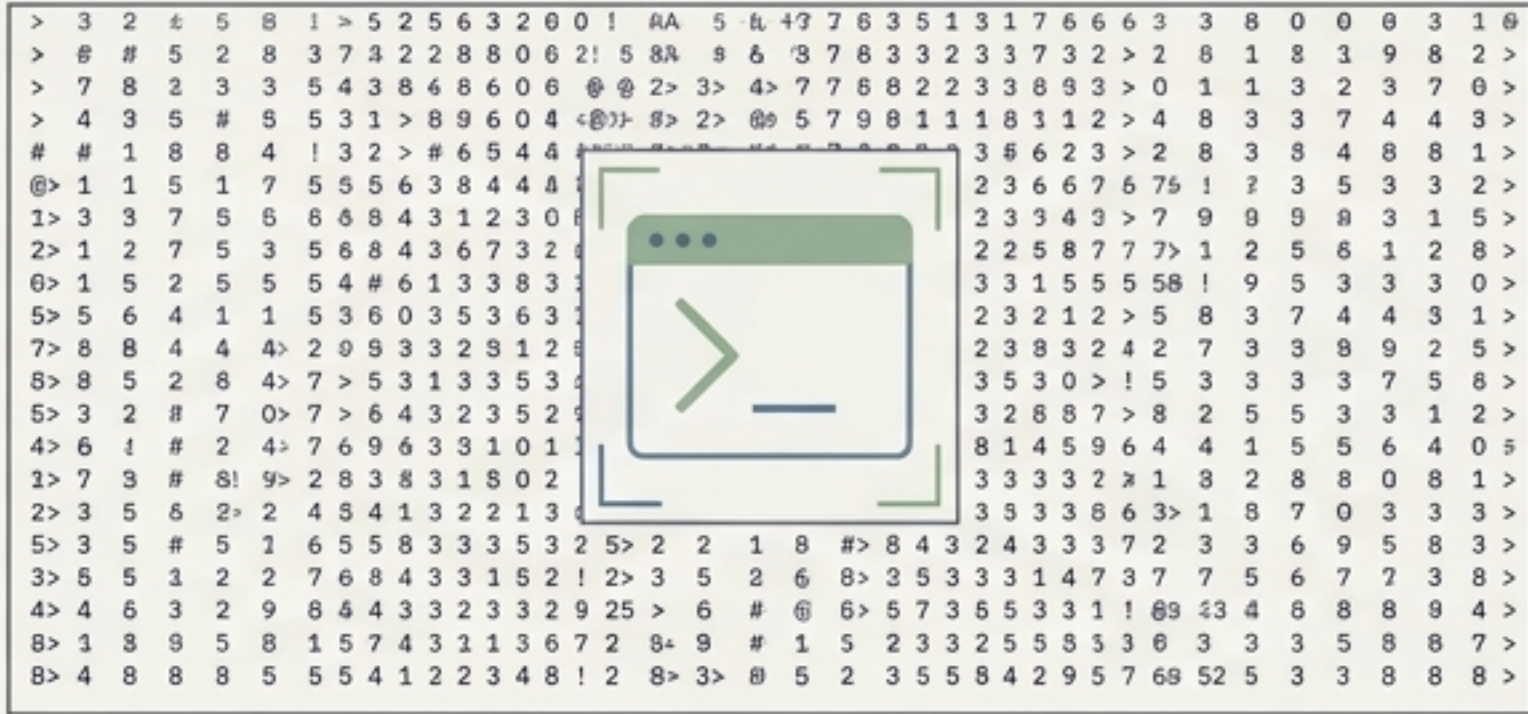
👁️ アプローチ: 直接知覚・自然言語推論

🖼️ 入力: 512x512 高解像度PNG画像

0.0005 2.204 50.29
0.0055 9.200 152.0479
0.0035 6.204
0.0012 0.003

特徴: 人間の目と同じように視覚として直接知覚し、ロスレス視覚メモリを利用。

AVO (NVIDIA)



🤖 アプローチ: 構造化データからの直接論理パース

📄 入力: 64x64 テキストグリッド

0.0030 2.204 25.83
0.0003 8.204 152.0479
0.0035 0.204
0.0022 0.006

特徴: 画像トークンを使用せず、マトリックスのような生データの羅列から背後の遷移メカニズムを推論。

インサイト: AVOは視覚的ヒューリスティクスに頼らず、数値・記号配列から直接的に背後の遷移メカニズムを推論する能力を実証。

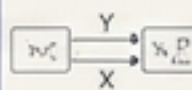
ARC-AGI-3 ハーネスアーキテクチャ比較マトリクス

システム	基盤LLM	アプローチ	RHAEスコア	総アクション	知覚入力
NVIDIA AVO	Claude Opus 5	汎用変異ループ	100.00	<u>6,624</u>	64x64 テキスト
Tycho (Actor-Req)	Claude Opus 5	プログラムの世 界モデル	100.00	6,641	64x64 構造化データ
VISTA (MIT)	Claude Opus 5	直接視覚知覚	100.00	7,542	512x512 PNG
NOOA (Research)	GPT-5.6 Sol	オブジェクト指向	85.10 (19/25完了)	-	型付き I/O
Claude Opus 5 単体	-	単体呼び出し	30.16	計測不能	ARC Prize 標準

単なる100点満点ではなく、Tychoと互角、VISTA比で約12%削減された「行動効率の極致」。

モデル性能を損失なく引き出す 「NOOA」の6つの設計原則

Typed input/output



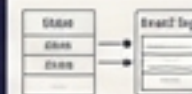
Pydantic等での厳格なバリデーション。
構文解析エラーを抑制。

Code as action

```
def run_action():  
    return 0  
    "say": "hi",  
    "order": "generate"  
    return val()
```

JSON形式のツール呼び出しを排し、直接実行可能なPythonコードを生成。

Explicit object state



エージェントの状態をクラス属性として保持し、履歴のコンパクションから分離。

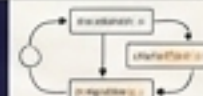
Agent Interaction Pipeline

Pass by reference



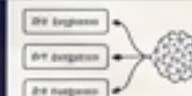
巨大データダンプを避け生オブジェクトの参照を渡す。
コンテキスト窓の浪費防止。

Programmable loop



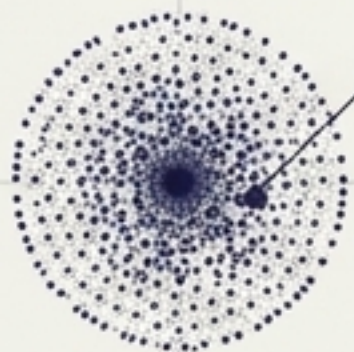
オーケストレーション自体を編集可能な通常Pythonで実装し動的再構成を支援。

Model-callable APIs



記憶の検索・整理をモデルから呼べるAPIとして公開し、能動的記憶管理を実現。

コンピュータ需要のパラダイムシフトと エコシステム戦略



現在: Single-turn Prompting

従来の対話型AI。ユーザーの単一
入力に対する単発の応答。

消費: 数百～数千トークン

戦略的インサイト

NVIDIAが自社モデルのみに依存せず、オープンなエージェント
基盤やNOOA等のハーネス技術を公開する背景には、自律探索
ループ普及に伴う「推論アクセラレータ需要の創出」という
プラットフォーム戦略が存在する。

未来: Long-horizon Autonomous Agents

長期自律エージェント。1つの課題解決のために
多数の推論ループを実行し連続稼働する。

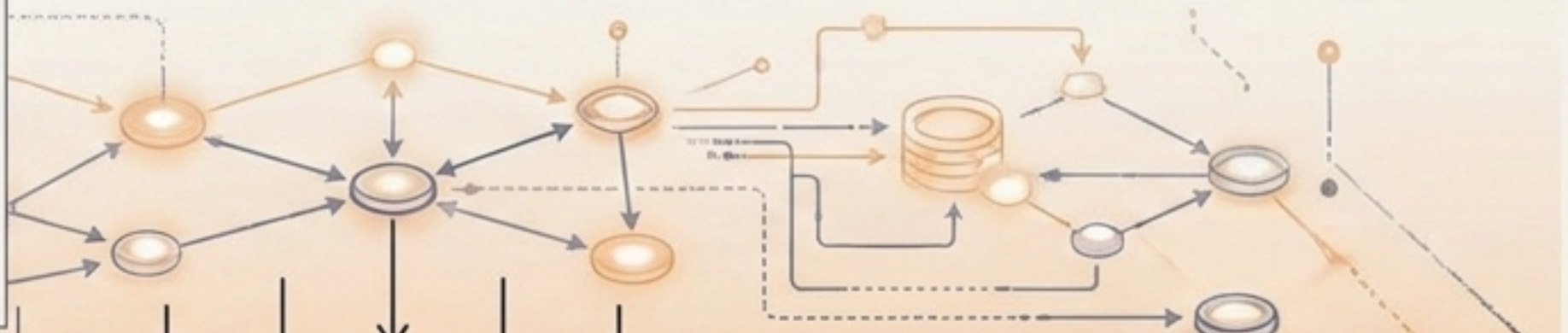
消費: 数百万～数億トークン / 7日間連続稼働

セキュリティ境界の再定義： 提案レイヤーと執行レイヤーの分離

推論・提案レイヤー (Proposer)

エージェントがプログラマブルなコードを生成し試行錯誤する動的空間。

※課題：プロンプトベースのガードレールは無力化する。



最小権限の原則による決定論的審査

Telemetry:
andeslayer: pontane-proposer
preacerv-eust: 3.0
neeeoss-tsms: 19.0
vask.seat: 6eu2h036358ez46a859
tsaer-sontemt: -208

Telemetry:
1 took: Generation: 38%
1 took: Generation: 38%
1 took: Generation: 38%

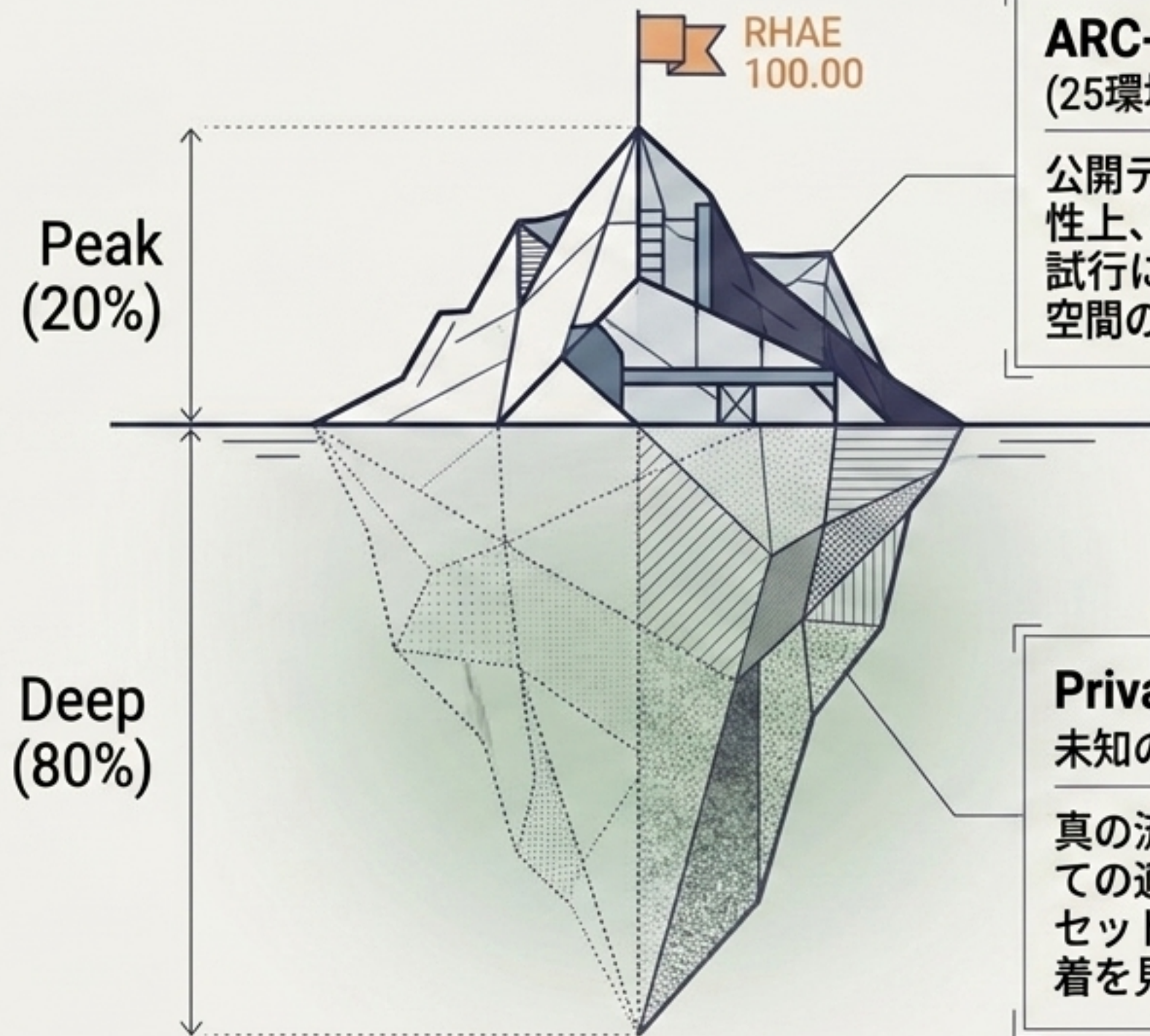


認可・執行レイヤー (Executor)

NVIDIA OpenShell等のセキュアランタイム。ファイルアクセスや通信などの実権限行使を決定論的に執行する。



公開セットの限界と 「真の流動性知能」への道程



ARC-AGI-3 Public Evaluation Set (25環境 / 183レベル)

公開テストにおける満点。ただし、タスク特性上、体系的な反復行動や特定パターンの試行によって解に到達しやすい側面（探索空間のハック）が指摘されている。

Private Evaluation Set 未知の汎化性能テスト

真の流動性知能（Fluid Intelligence）としての適応力の証明は、今後の非公開テストセットでの客観的評価によって最終的な決着を見る。

結論：モデルのスケールから「ハーネス設計」の時代へ

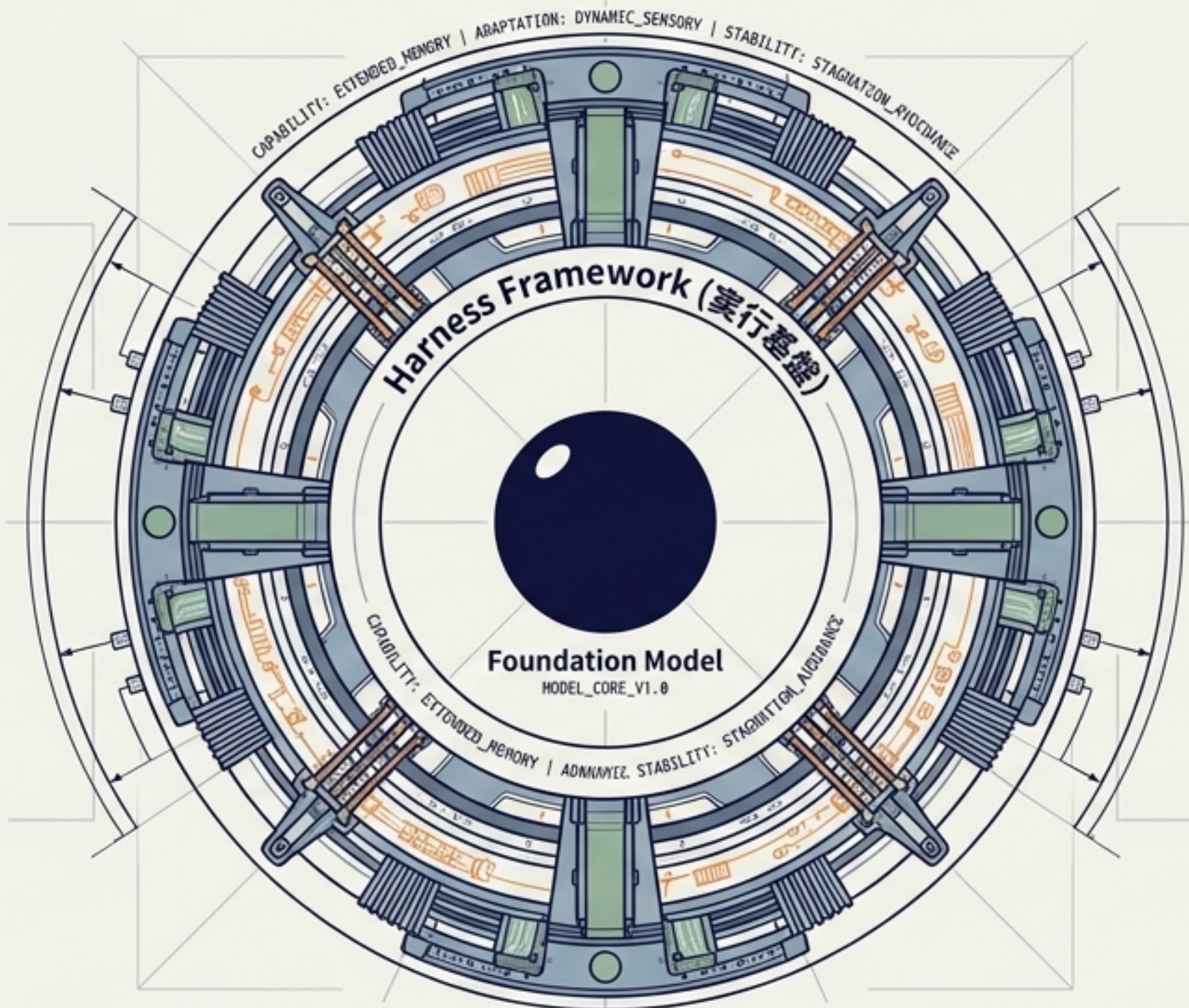
Harness is the new competitive moat.

DOMAIN: COMPETITIVE_ADVANTAGE |
FOCUS: SYSTEM_INTEGRATION

記憶の維持、知覚の適合、
スタグネーションの回避。

これらはもはやモデル単
体ではなく、周囲を囲む
「実行基盤」が担うべき
領域である。

DOMAIN: COMPETITIVE_ADVANTAGE |
FOCUS: SYSTEM_INTEGRATION



Cross-domain Architecture.

DOMAIN: AGENT_BLUEPRINT |
ARCHITECTURE: SELF-AUTONOMOUS_LOOP

GPUカーネルの極限最適化
と未知の論理パズル。

対極にあるドメインを単一
の「自律変異ループ」で
制覇した AVOの実績は、
エージェント設計の確固た
るブループリントとなる。

DOMAIN: AGENT_BLUEPRINT |
ARCHITECTURE: SELF-AUTONOMOUS_LOOP

生の推論エンジンをいかに手なずけ、複利的な思考回路へと組み替えるか。
エージェント時代の覇権は、その「シャーシ」の設計力にかかっている。