

ARC-AGI-3: 人間とAIの「学習効率」における圧倒的な格差

HUMAN (100% RHAE)

RHAE (相対的人間行動効率) 100%



RHAEは単なる正解率ではなく、人間の行動数を基準に「どれだけ無駄なく解けたか」を測定。

評価の核心：未知への適応力



ルール説明なしの「探索と計画」

64×64のグリッド環境で、指示なしに勝利条件を推定し、行動効率を観います。



静的パズルから「対話型エージェント」へ
従来の静的な答え合わせではなく、時間軸に沿った動的な振る舞い能力が関わります。



FRONTIER AI (<1% RHAE)

RHAE (相対的人間行動効率) <1%



最新のフロンティアモデル (GPT-5.4等) でもスコアが1%に満たないという衝撃的な実態。

性能比較：人間 vs フロンティアAI

- 人間 100% に対し、AIは1%未満
2026年3月時点の最高峰AIでも、人間との学習効率には巨大な隔たりがあります。
- 厳格な「ハーネスなし」評価
外部ツールによる人工的なスコア底上げを排除し、モデル純粋の知能を測定します。

人間 100% に対し、AIは1%未満
2026年3月時点の最高峰AIでも、人間との学習効率には巨大な隔たりがあります。

主要なAIモデルと人間ベースラインのRHAEスコア比較

人間 (Easy for Humans 基準)



Google (Gemini 3.1 Pro Preview)



OpenAI (GPT 5.4 High)

