

# Poetiq AIの快挙と課題：AGIベンチマーク75%スコアの深層分析

革命的な主張：AIが「知能の壁」を突破

超難関ベンチマーク「ARC-AGI-2」

「記憶」が通用せず、未知の課題を解く真の推論能力（流動性知能）を測るテスト。



54.2%

従来の最先端AI (GPT-5.2)

人間には容易、AIには絶望的な壁



75%

Poetiq AI (主張値)

GPT-5.2と独自の「再帰的自己改善」システムで人間を大幅に超える結果を達成。

| プレイヤー              | ARC-AGI-2 スコア |
|--------------------|---------------|
| 人間 (平均)            | 60%           |
| 従来の最先端AI (GPT-5.2) | 54.2%         |
| Poetiq AI (主張値)    | 75%           |

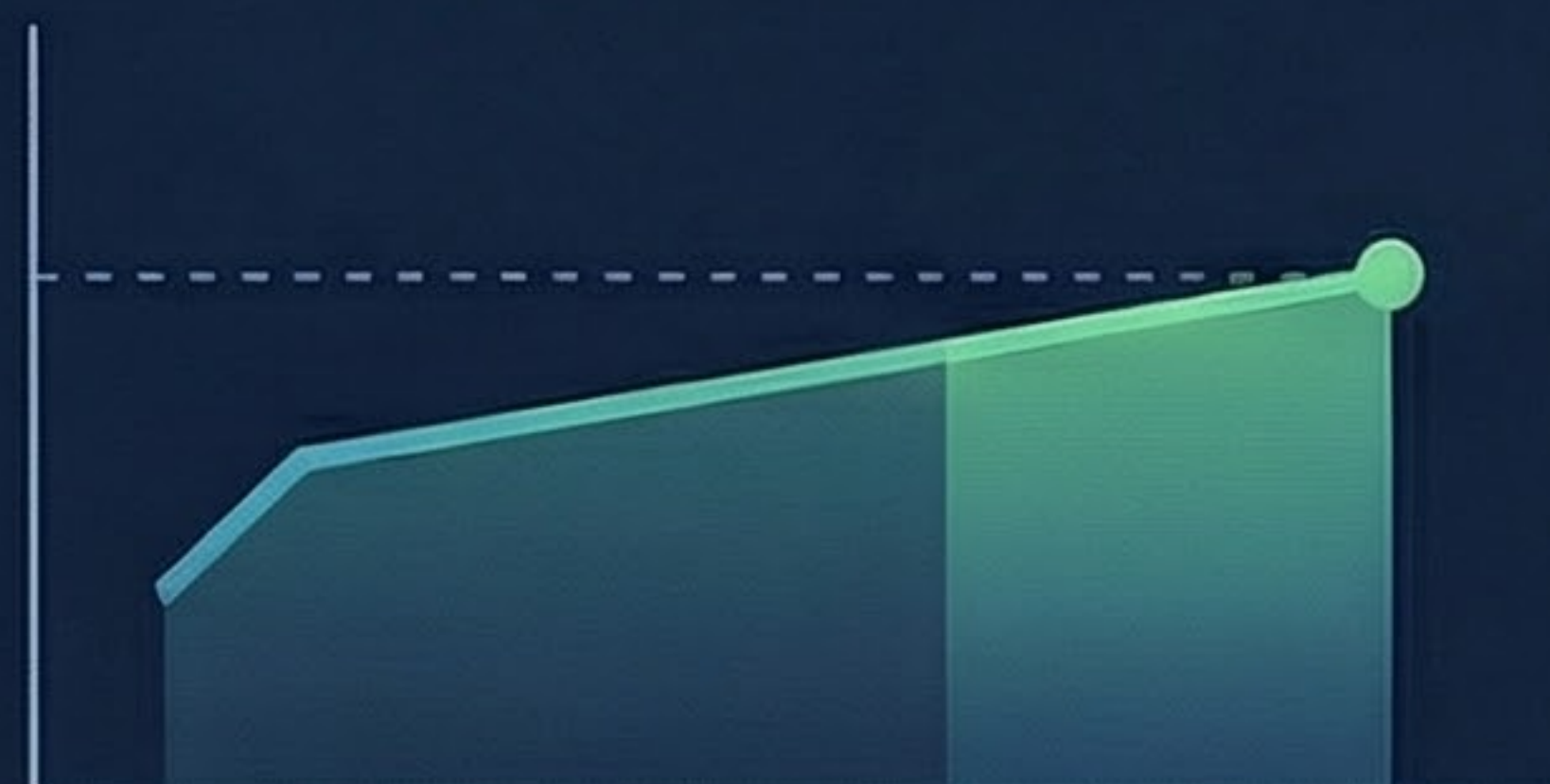
冷静な分析：スコアの信頼性とコストの壁

以前の公開スコア(65.3%)



過去の「スコア乖離」という懸念

約11ポイント下落した実績があります。



現実的な検証後スコア予測「62%」

それでも人間の平均スコア(60%)を超え、歴史的な快率となる可能性があります。



1問数ドル？「富豪的アプローチ」の課題

GPT-5.2 X-Highの利用は高コストで、現時点での実用化は限定的です。

| アプローチ                   | 精度 (%) | 1タスクあたりコスト (概定) |
|-------------------------|--------|-----------------|
| NVARC (NVIDIA)          | ~28%   | ~\$0.2          |
| Poetiq (GPT-5.2 X-High) | 75%    | ~\$2-3+         |
| 平均的な人間                  | 60%    | ~\$10           |