

ARC-AGI-4とGPT-6 Astraを巡る言説の検証

ベンチマークの飽和、AGIの定義、そして次世代AI評価の真実

REPORT DATE: 2026年9月13日
LEAD ANALYST: Claude Fable 5.1

不正確な言説と、検証された事実（The Corrected Reality）

流布している不正確な文章

ARC-AGI-4：Astraの急進化（ARC-3の攻略）を受けて、2026年9月に新たに発表された次期ベンチマーク。

「自律的で終わりになきイノベーションと科学的発見ができるかを測定する、さらに高度なテストになる予定です。」

検証を経て修正された正確な文脈

ARC-AGI-4は2027年Q1リリース予定の次期ベンチマーク。GPT-6 AstraのARC-AGI-3でのスコア（標準62.7%）等を踏まえ、『再帰的自己改善』と『終わりになき発明』を評価する方針が示された。



リリースは
2027年Q1予定



Astraのスコアは
標準ハーネスで
62.7%



公式な評価対象は
Open-ended
invention



ARC-AGI-4開発は
Astraリリース
前から進行

言説の解剖図 (Anatomy of the Claim)



環境がスコアを変える：ハーネス依存のメカニズム

標準ハーネス (Standard harness)

62.7%

コスト: \$26,098

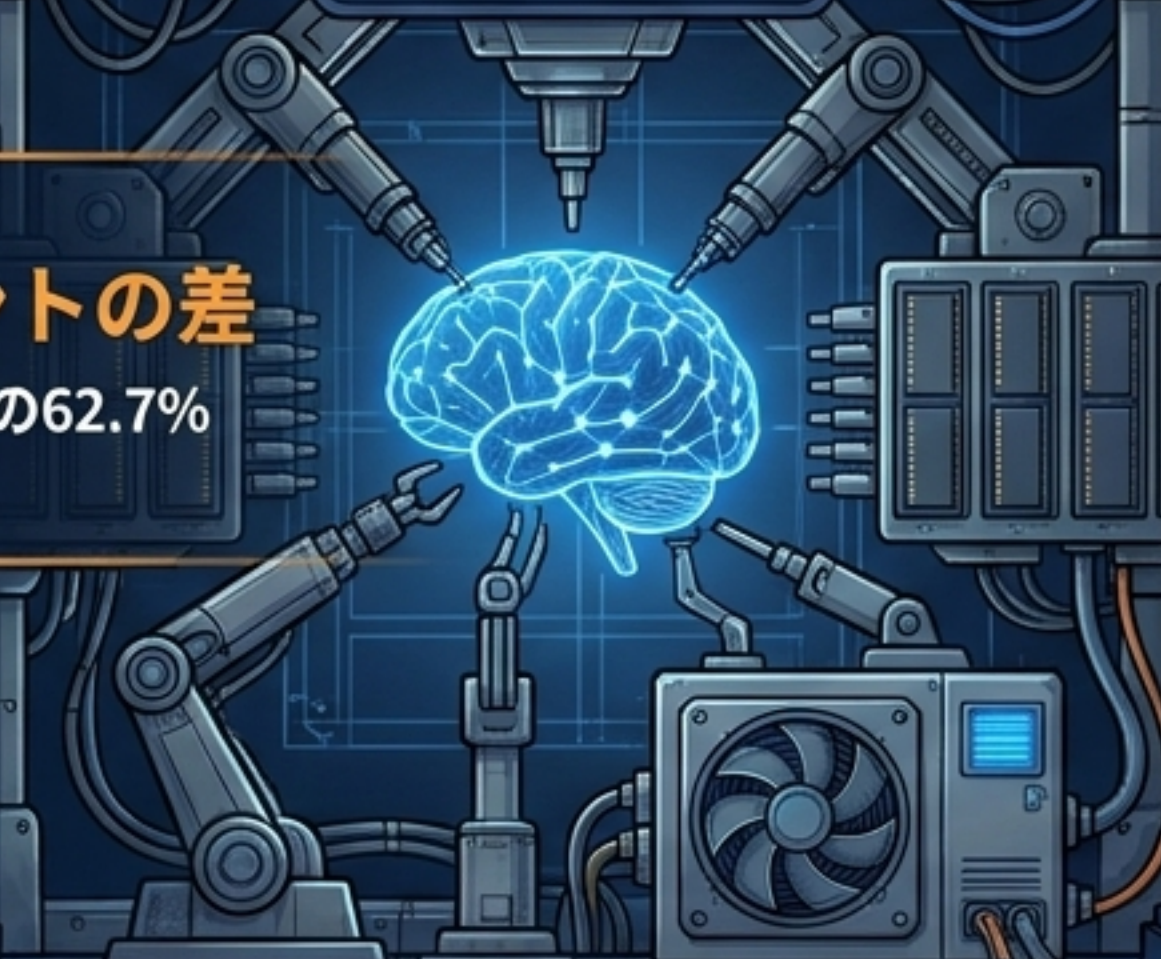


全モデルに共通適用される生の知能の測定

Provider Adapter ハーネス

99.9%

コスト: \$18,817



ベンダー独自の環境。知能の差ではなく、
メモリ処理構造の違い



同一モデルで約37ポイントの差

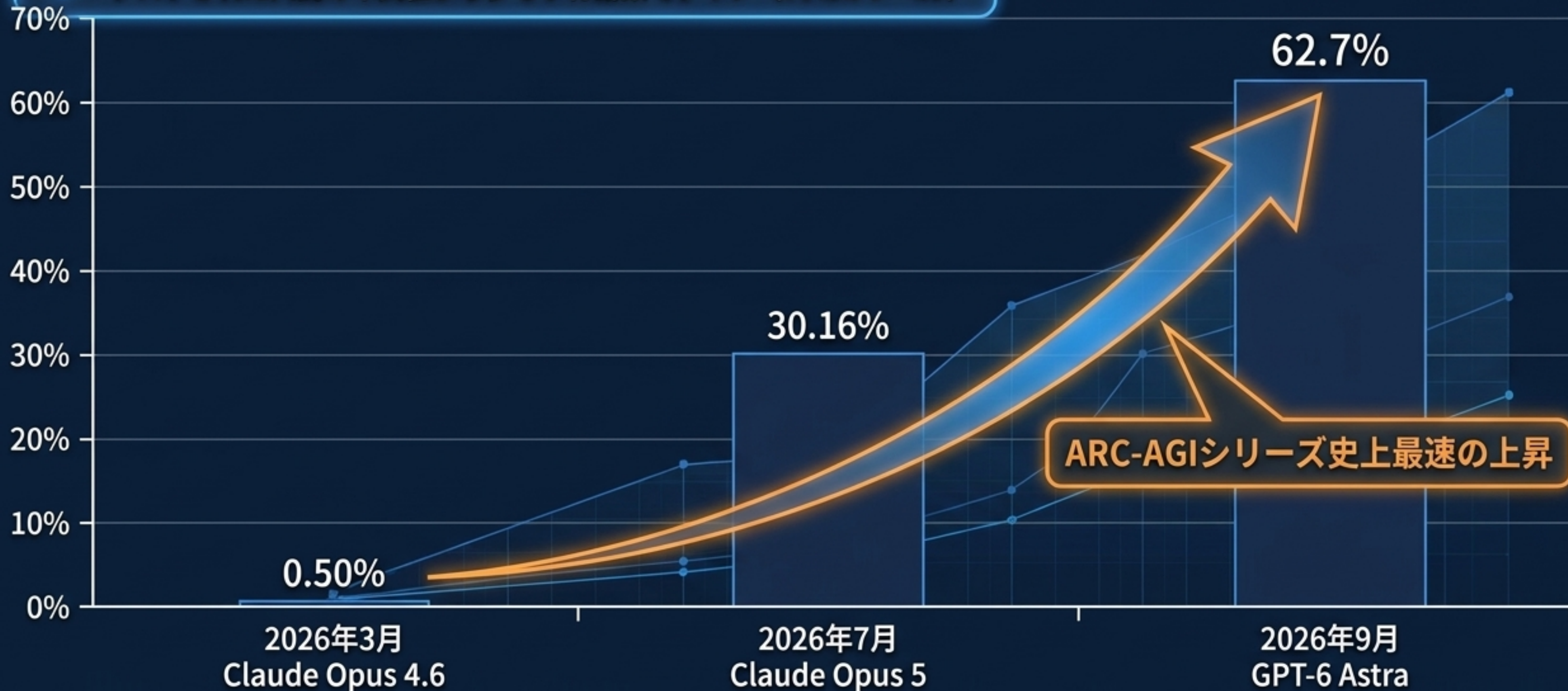
研究・比較目的では標準ハーネスの62.7%
を参照すべきである。

史上最速の進化：フロンティアモデルのスコア推移



行動効率 (Action Efficiency)

テストされた人間の中央値より少ない行動数でクリア (96%のレベル)



ベンチマークの飽和 ≠ AGI (汎用人工知能)

“ ARC 3を飽和させたなら、それはAGIなのか？
我々はそう主張していない。
— François Chollet ”



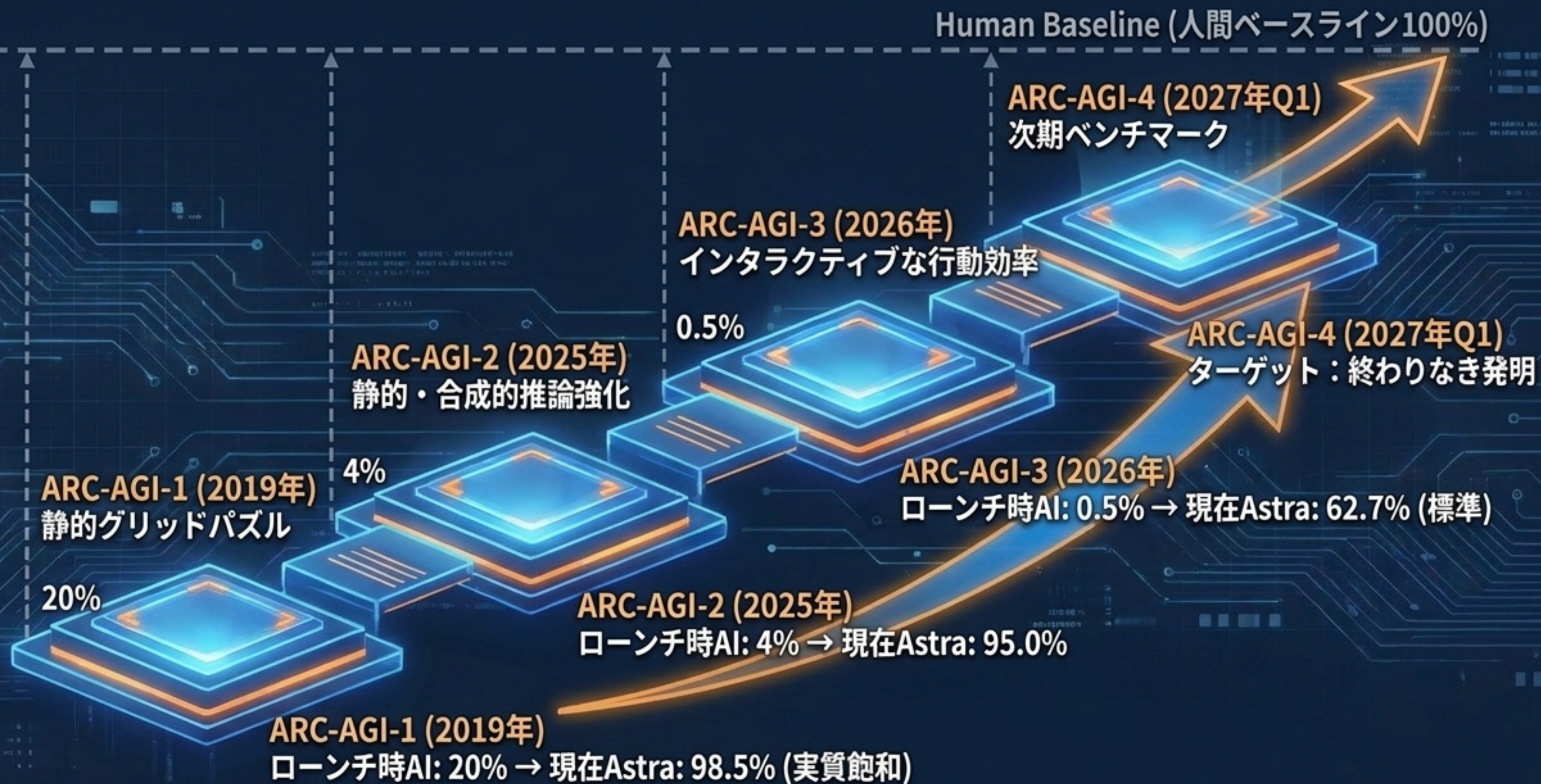
- 静的パズルの攻略 (ARC-1, 2)
- インタラクティブ推論の攻略 (ARC-3)



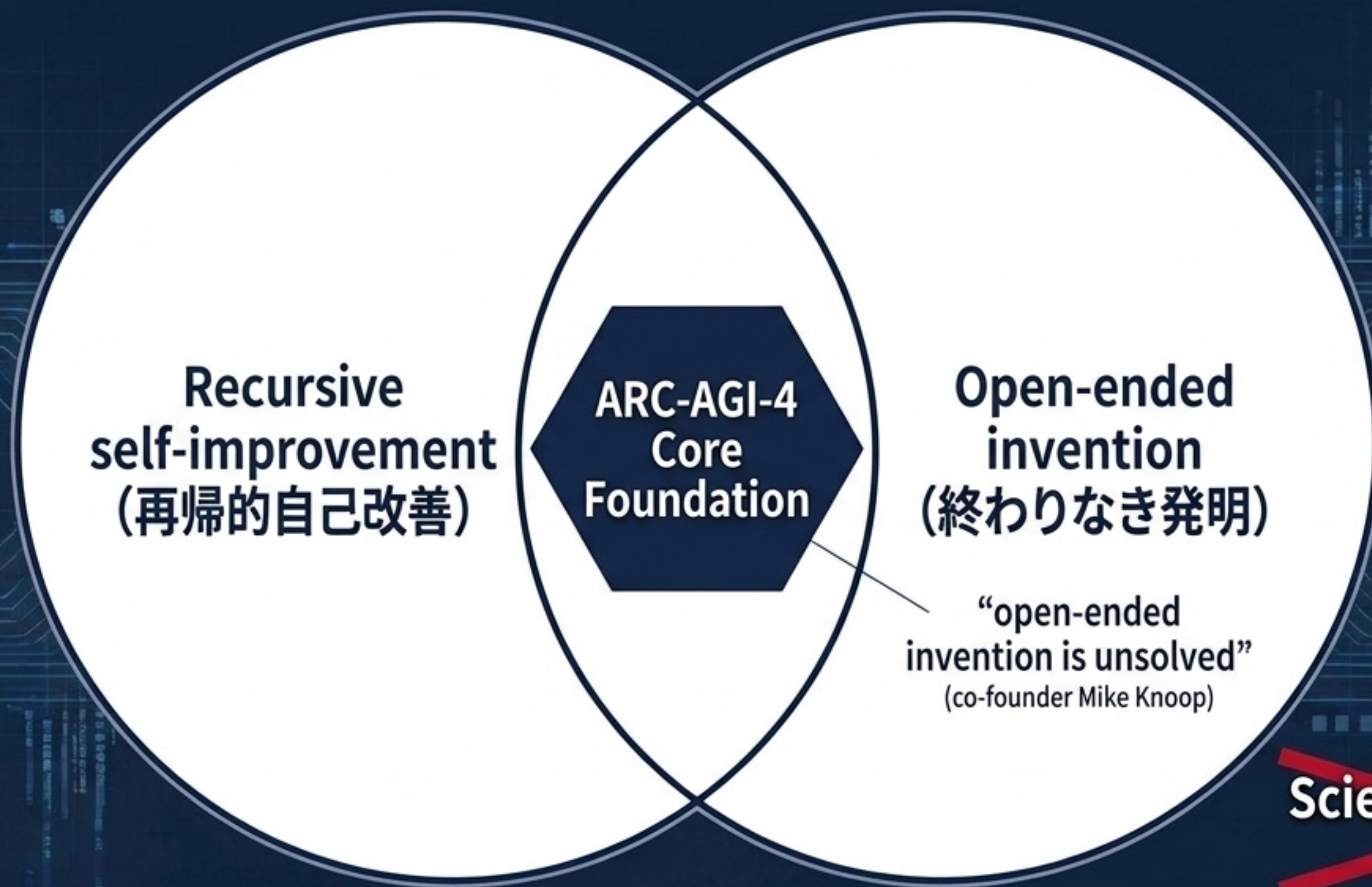
人間レベルの新しい
スキル獲得効率
(AGI)

人間ができてAIができないことを提案できなくなるまで、我々はベンチマークを作り続ける。

知能測定ロードマップ：ARC-AGIの系譜

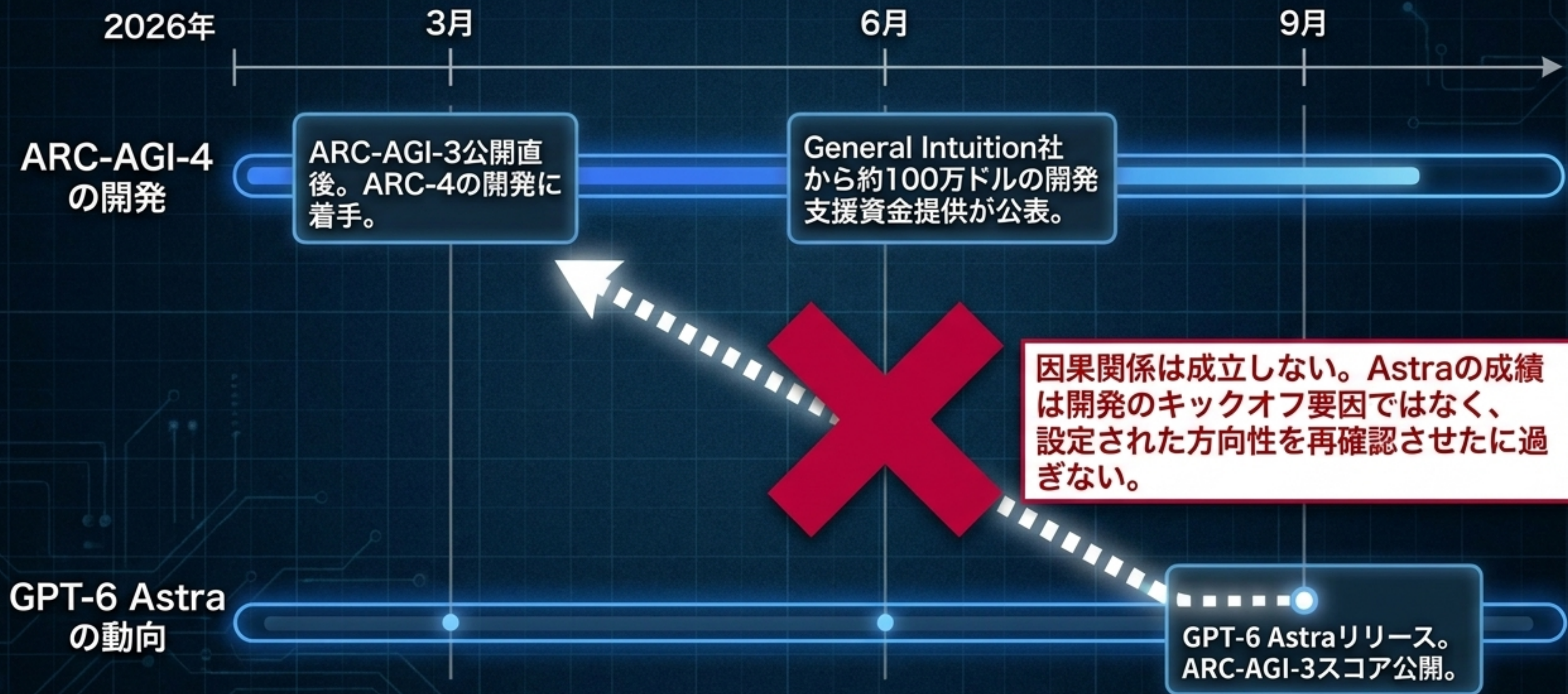


ARC-AGI-4のターゲット：次期ベンチマークが測るもの



~~Scientific Discovery
(科学的発見)~~

開発の真のタイムライン：因果関係の否定



技術と法の交差点：AIの発明能力と知財制度の衝突

The Bottleneck

Technology / ARC-AGI-4

Law / IP Framework

The Bottleneck

- Open-ended inventionの定量化
- 自律的な発明・研究能力の評価確立

構造的矛盾：AIが自律的に発明能力を証明しても、現行法では特許権の主体になれない。

- Inventorship (発明者性) の壁
- DABUS事件の決着：発明者は自然人に限る

論点の移動：特許実務の次なる焦点

発明者性 (Inventorship) の議論



自然人に限定される法理が
主要法域で確立済み



進歩性のハードル上昇

AI普及により、「当業者」の
平均水準が底上げされ、特
許取得の難易度が上昇。



出願量の爆発と 審査負荷

R&Dの自動化による大量出
願。USPTOやEPO等での
AI審査導入が加速。



人間寄与の立証

AI支援発明において、人間が
「どこにどう寄与したか」の緻
密なトラッキングが不可欠に。



Data: IP専門家の85%が既に何らかの形で
AIを利用している (2025年 Questel調査)

The Final Takeaways

01

Capability vs Scaffolding

GPT-6 Astraの急進化は事実
ただが、「99.9%」はアダプ
ター環境への依存が大きい。
生の知能（62.7%）とメモリ
処理構造を切り分けて評価
すべきである。

02

Trajectory of Evaluation

ARC-AGI-4（2027年Q1予定）
はAstraの副産物ではなく、
AI評価の必然的な次の一手で
ある。目標は「再帰的自己改
善」と「終わりのなき発明」の
測定へと移行する。

03

Implications for Society

ベンチマークがAIの発明能
力を定量化する未来におい
て、知財議論の焦点は「誰が
発明したか」から、「いかに
進歩性を証明し人間の寄与を
切り出すか」へと完全にシ
フトする。