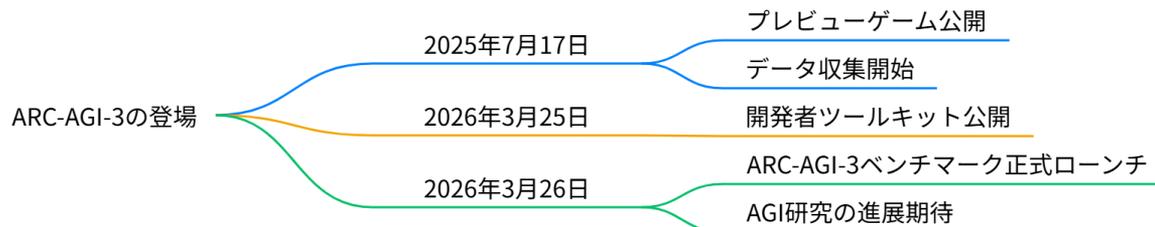


「ARC-AGI-3」で人間とAIの差が浮き彫りに

Felo AI



概要

2026年3月26日、汎用人工知能（AGI）の実現に不可欠な「知能」を測定するための新たなベンチマーク「ARC-AGI-3」が公開された [17](#)。Google の元研究員で AI 研究の第一人者であるフランソワ・ショレ氏によって開発されたこのベンチマークは、従来の静的な問題解決能力を測るテストとは一線を画し、AI エージェントが未知の環境を探索し、動的に目標を獲得し、継続的に学習する能力を評価する、世界初となる本格的なインタラクティブ推論ベンチマークである [2926](#)。

ARC-AGI-3 の最大の特徴は、人間と AI の間に存在する圧倒的な性能差を浮き彫りにした点にある [8](#)。人間が容易に 100% のスコアを達成できるのに対し、現存する最先端の AI モデルですら達成率は 1% 未満に留まっている [3732](#)。これは、現在の AI が持つ根本的な課題、すなわち未知の状況に対して人間のように柔軟に適応し、効率的にスキルを習得する能力の欠如を示唆している [1121](#)。このベンチマークは、単なる正答率ではなく「行動効率」を重視しており、AI 研究の評価軸を「暗記」から「適応と学習」へと転換させる重要なマイルストーンとなる [91135](#)。



詳細レポート

ARC-AGI-3 の登場と背景

ARC-AGI-3 は、Keras の開発者としても知られるフランソワ・ショレ氏が提唱する「Abstraction and Reasoning Corpus (ARC)」の最新版である [1 16](#)。ショレ氏は、知能を「スキル習得の効率」と定義しており、ARC シリーズはこの思想に基づき、AGI の進捗を測定するために設計された [11 19 25](#)。

これまでの「ARC-AGI-1」が既存知識を組み合わせて問題を解く能力を、「ARC-AGI-2」が流動的知能を測定する静的なテストであったのに対し、ARC-AGI-3 は AI エージェントがビデオゲームのような環境内で自ら行動を選択し、試行錯誤を通じてルールや目標を学習していく「インタラクティブ」な形式を採用している [1 7 23](#)。このアプローチは、AI が現実世界の複雑で予測不可能な状況に対応するために不可欠な、より汎用的な知能を評価することを目的としている [13](#)。

静的からインタラクティブへ：ベンチマークのパラダイムシフト

従来の AI ベンチマークの多くは、事前に定義された大規模なデータセットに対する性能を測定するものであり、モデルがテストデータに過剰に最適化（過学習）してしまう問題が指摘されてきた [39](#)。ARC-AGI-3 は、この問題を克服するために、AI がこれまで遭遇したことのない完全に新しい環境で、自然言語の指示なしにタスクを遂行する能力を問う [2](#)。このインタラクティブな形式は、AI に対して以下の 4 つの中核的な能力を要求する [11](#)。

- **即時学習 (On-the-fly Learning):** 既知の知識をその場で組み合わせて、未知の状況を理解し対応する能力。
- **探索 (Exploration):** 自らの行動選択によって、環境から必要な情報を効率的に収集する能力。
- **記憶 (Memory):** 過去の経験から得た情報を、将来の意思決定のために適切に保存・活用する能力。
- **目標獲得 (Goal Acquisition):** 最終的な目標が明示されていない状況でも、自ら中間目標を設定し、計画的に行動する能力。

これらの能力は、静的なパズルを解くだけでは測定できず、動的な環境との対話を通じて初めて評価可能となる [11 23](#)。



人間と AI の絶望的な格差

ARC-AGI-3 が AI 研究界に与えた最も大きな衝撃は、人間と AI の間に存在する「知能」の質的な違いを数値で明確に示したことである [8](#)。

対象	ARC-AGI-3 スコア
人間	100% 378
最先端 AI	< 1% 378

この圧倒的な差は、現在の AI 開発の主流である「スケーリング則（データとモデルサイズを大きくすれば性能が向上する）」だけでは、人間のような汎用的な知能には到達できないというショレ氏の持論を強力に裏付けている [21](#)。

ARC-AGI-3 では、単にタスクをクリアできたか否かだけでなく、クリアまでに要した行動のステップ数、すなわち「行動効率 (Action efficiency)」が重要な評価指標となる [11](#)。これは、環境から得た情報をいかに効率的に有効な戦略へと変換できるかを測定するものであり、「知能とは効率性である」という考えに基づいている [11](#)。人間は少ない試行回数でゲームのルールを推測し、最適な戦略を構築するが、AI はランダムサーチやブルートフォース（総当たり攻撃）に陥りがちで、効率的な学習ができていないことが示された [11](#)。

開発ロードマップとコミュニティの反応

ARC-AGI-3 は、2026 年 3 月 26 日の正式公開に先立ち、いくつかのステップを経て開発が進められてきた [7](#)。

- **2025 年 7 月 17 日:** 3 つのプレビューゲームが公開され、人間と AI のパフォーマンスに関するデータ収集が開始された [11](#)。このプレビュー期間を通じて、一部のゲームがランダムサーチに対して脆弱であるといった課題が特定され、正式版の改善に活かされた [11](#)。
- **2026 年 3 月 25 日:** AI エージェントを統合するための開発者ツールキットがオープンソースで公開された [37](#)。
- **2026 年 3 月 26 日:** ARC-AGI-3 ベンチマークが正式にローンチ [7](#)。

このベンチマークは「ARC Prize 2026」というコンペティションと連動しており、世界中の研究者や開発者がスコアを競い合うことで、AGI 研究の進展を加速させることが期待されている [222](#)。

コミュニティからは、AGI の真の能力を測る画期的な試みとして大きな期待が寄せられている一方で [9 17 35](#)、一部からは「知性ではなく、単なる粘り強さを試すテストだ」といった批判的な意見も出ている [24](#)。

AGI 研究へのインプリケーション

ARC-AGI-3 の登場は、AGI 研究の今後の方向性に大きな影響を与える可能性がある。Google の Gemini 3 Deep Think な

どが **ARC-AGI-2** で人間を超えるスコアを記録し、**AGI** 到達への期待が高まっていたが、**ARC-AGI-3** の結果は、現行モデルが持つ能力の限界を明確に示した [33 42 43](#)。

このベンチマークは、AI 研究の焦点を、特定のタスクにおける性能向上から、未知の環境への適応能力や一般化能力といった、より本質的な「知能」の探求へとシフトさせる触媒となるだろう [17 36](#)。単にモデルを巨大化させるアプローチの限界が示された今、**ARC-AGI-3** が提示する課題を克服する新しいアーキテクチャや学習方法の開発が、真の **AGI** 実現に向けた次なるブレークスルーの鍵を握っている [21 36](#)。

1. [AI の知能をルール不明のゲームで測定する「ARC-AGI-3」が登場](#)
2. [ARC-AGI-3](#)
3. [ARC-AGI-3 の発表 世界で唯一の飽和していないエージェント型 ...](#)
4. [AI の知能をルール不明のゲームで測定する「ARC-AGI-3」が登場](#)
5. [Introducing ARC-AGI-3 : r/LocalLLaMA](#)
6. [ARC Prize - Explore All Tasks](#)
7. [AI の知能をルール不明のゲームで測定する「ARC-AGI-3」が登場](#)
8. [Announcing ARC-AGI-3 The only unsaturated agentic ...](#)
9. [インタラクティブ推論ベンチマークを用いた AI 知能の再定義](#)
10. [ルールを理解してタスクをクリアする」ことを ...](#)
11. [ARC-AGI-3 Preview: 30-Day Learnings](#)
12. [ARC-AGI-3 は、人間と AI の両方で高い IQ \(145 以上\) を確実 ...](#)
13. [ARC-AGI-3 Quickstart](#)
14. [ARC-AGI 3 が紹介されました : r/agi](#)
15. [Measuring Intelligence: ARC-AGI-3 is a Test For Agents](#)
16. [2026-03 事前資料 \(1\) : 抽象推論 \(ARC-AGI を内包 未分化\)](#)
17. [Measuring AI Intelligence: ARC-AGI Redefines Progress to ...](#)
18. [ARC-AGI-3 : r/singularity](#)
19. [What is ARC-AGI?](#)
20. [【完全ガイド】ARC-AGI とは? ~概要と始め方~](#)
21. [ARC-AGI-3 の衝撃と動画・マルチモーダル競争の激化](#)
22. [ARC Prize 2026](#)
23. [「AI が賢い」をどう測るか—ARC-AGI-3 が変えた — 野崎秀吾](#)
24. [ARC AGI 3 はバカげてる : r/LocalLLaMA](#)
25. [ARC-AGI 3 は、あらかじめ決められたゴールがない ... - Reddit](#)
26. [ARC-AGI-3](#)

27. [Gemini の能力は人間に近づいたか、ARC-AGI ベンチマークが ...](#)
28. [ARC-AGI 3: March 2026](#)
29. [ARC-AGI-3 は、人間と AI の両方で高い IQ \(145 以上\) を確実 ...](#)
30. [AI リーダー・研究者・発信者・批評家たちの発信情報 \(2026 ...](#)
31. [ARC-AGI が解かれた!? #AI - Qiita](#)
32. [AGI の度合いを測るベンチマーク「ARC-AGI-3」が公開。現状 ...](#)
33. [主要な AI モデルが AGI テストで全滅：汎用人工知能の高い壁](#)
34. [ARC-AGI 3 が紹介されました - Reddit](#)
35. [インタラクティブ推論ベンチマークを用いた AI 知能の再定義](#)
36. [主要な AI モデルが AGI テストで全滅：汎用人工知能の高い壁](#)
37. [ARC prize が AI テスト用のオープンソース ARC-AGI-3 開発者ツール ...](#)
38. [未知への適応力を鍛える：ARC1 から ARC3 まで紐解く AGI ...](#)
39. [AI ベンチマークはもはや限界、新たなテスト手法の登場相次ぐ](#)
40. [Gemini 3 Deep Think のアップグレード版が登場、知性を測定する ...](#)
41. [AI が苦手なパズル：ARC-AGI ベンチマークが示唆する ...](#)
42. [Google Gemini 3 Deep Think、ARC-AGI-2 で 84.6%達成](#)
43. [バブル、そして AGI。イノーバウィークリーAI インサイト記事で ...](#)