

# ARC-AGI-3登場に関する検証・技術分析レポート

## Executive Summary

ARC Prize Foundation <sup>1</sup> は、ルールや目標が明示されないターン制ゲーム環境でAIエージェントの「学習効率」を測る新ベンチマークARC-AGI-3を発表・公開した。公式の公開日は2026年3月25日（公式ブログの公開日）で、日本語圏では2026年3月26日に報道されており、時差・報道タイミングの差が確認できる。<sup>2</sup> ARC-AGI-3は、64×64の抽象グリッド観測と小さな行動空間を用い、行動回数（action efficiency）を人間基準に正規化したRHAЕ（Relative Human Action Efficiency）でスコア化する設計である。<sup>3</sup> 公開時点の「半非公開（semi-private）」評価では、主要フロンティアモデルのスコアが0.00～0.37%と極めて低く、人間は設計上100%解ける（“easy for humans”基準で選別・校正）とされ、学習効率ギャップを強調する内容になっている。<sup>4</sup> 一方で、公開ゲーム（デモ）やツールキットは整備されているが、主要評価セットが「非公開」であるため、外部の独立再現は構造的に制約され、比較の厳密性は評価プロトコル遵守（特に“ハーネス”依存の回避）に強く依存する。<sup>5</sup>

## 主要発見

- 公式ブログはARC-AGI-3の公開日を「2026年3月25日」と明示し、当日サンフランシスコでローンチイベントも実施したとしている。<sup>6</sup>
- 技術報告（Technical Report）は「2026年3月24日」付で、ARC-AGI-3を「探索・目標推定・世界モデル化・計画」を要求する対話型（interactive）ベンチマークとして定義している。<sup>7</sup>
- 観測は64×64グリッド（16色）で、ターンごとにフレーム（または短いフレーム列）を受け取る。<sup>8</sup>
- スコアはRHAЕ（RHAЕ = “Ray”）で、人間（各環境10人）の「2番目に良い行動数」を基準に、 $(h/a)^2$  を上限1でクリップし、環境内はレベル重み（1～5）で加重平均、全環境平均で集計する。<sup>9</sup>
- データセット構成は（少なくとも報告書上）Public Demo 25環境、Semi-Private 55環境、Fully Private 55環境（計135環境）。<sup>10</sup>
- 公開時点のSemi-Privateスコア（公式リーダーボード）は、Google 0.37%、OpenAI 0.26%、Anthropic 0.25%、xAI 0.00%として提示されている。<sup>11</sup>
- “ハーネス”（外部の専用手続き・ツール）による人工的スコア上昇を主要なリスクとして明示し、公式リーダーボードでは「ハーネス無し」で評価する方針が明確化されている。<sup>12</sup>
- SDK/ツール類はGitHubでMITライセンスとして公開され、オフライン（ローカル）実行では約2,000 FPSをうたう。<sup>13</sup>

## 事実確認と公開状況

### 発表・公開のタイムライン

日付	出来事	一次/準一次ソース	補足
2019-11	ARC（ARC-AGI-1の源流）の枠組み（知能=スキル獲得効率）を提示	Chollet “On the Measure of Intelligence”	ARC-AGIシリーズの思想的基盤。 <sup>14</sup>

日付	出来事	一次/準一次ソース	補足
2025-07～08	ARC-AGI-3「プレビュー」競争（30日）を実施	ARC Prizeブログ（プレビュー学習）	3つの公開環境＋隠し評価でのテストが行われた。 <sup>15</sup>
2026-03-24	ARC-AGI-3 技術報告（Technical Report）日付	Technical Report PDF	“Humans 100%, frontier AI <1%”等を記載。 <sup>16</sup>
2026-03-25	ARC-AGI-3公開の公式ブログ（Published 25 Mar 2026）	ARC Prizeブログ（ローンチ）	同日、ARC Prize 2026開始日でもある。 <sup>17</sup>
2026-03-26	日本語圏での「登場」報道	GIGAZINE <sup>18</sup>	「2026年3月26日に発表」と表現。時差・報道時刻の影響が濃厚。 <sup>19</sup>

## 事実確認テーブル

検証項目	確認できた事実	主要ソースURL	確認状況	未確認/注記
公式発表日時	公式ブログは“Published 25 Mar 2026”	<a href="https://arcprize.org/blog/arc-agi-3-launch">https://arcprize.org/blog/arc-agi-3-launch</a> <sup>6</sup>	確認済	“3/26発表”表現は日本語報道側のタイムスタンプ差の可能性。 <sup>19</sup>
開発・運営主体	ARC Prize Foundation（非営利）と、サイト運営体としてARC Prize, Inc. 表記	<a href="https://arcprize.org/about/">https://arcprize.org/about/</a> <a href="https://arcprize.org/terms">https://arcprize.org/terms</a> <sup>20</sup>	確認済	法人格の役割分担（Foundation/Inc.の実務境界）は本件範囲では不明。
一次資料（技術報告）	“ARC-AGI-3: A New Challenge for Frontier Agentic Intelligence”（2026-03-24）	<a href="https://arcprize.org/media/ARC_AGI_3_Technical_Report.pdf">https://arcprize.org/media/ARC_AGI_3_Technical_Report.pdf</a> <sup>16</sup>	確認済	報告書ヘッダに“arXiv:submit/...”表記はあるが、arXivの公開版（absページ）は本調査では確認できず（未確認）。 <sup>16</sup>
公開デモ（人間向け）	ブラウザでプレイ可能（Play [Humans]）	<a href="https://arcprize.org/arc-agi/3/">https://arcprize.org/arc-agi/3/</a> <sup>21</sup>	確認済	タスク一覧ページ等は動的表示で、クローラ上は内容取得できない箇所あり。 <sup>22</sup>
SDK/Docs	公式ドキュメント（Quickstart、API Keys、REST API 等）	<a href="https://docs.arcprize.org/">https://docs.arcprize.org/</a> <sup>23</sup>	確認済	docs上で「AI生成の可能性があり誤り得る」注意書きあり（補助情報扱い）。 <sup>24</sup>
GitHub公開物	ツールキット/ハーネス等のレポジトリを公開	<a href="https://github.com/arcprize/ARC-AGI">https://github.com/arcprize/ARC-AGI</a> 等 <sup>25</sup>	確認済	ベンチマーク私有セット（semi/private）の中身は公開されない設計。 <sup>26</sup>

検証項目	確認できた事実	主要ソースURL	確認状況	未確認/注記
ライセンス (コード)	ARC-AGI Toolkit等はMITライセンス表記	<a href="https://github.com/arcprize/ARC-AGI/blob/main/LICENSE">https://github.com/arcprize/ARC-AGI/blob/main/LICENSE</a> <sup>27</sup>	確認済	競技提出物にはCC0やMIT-0等の“permissive public domain license”を要求。 <sup>28</sup>
入手方法	pip/uvで導入、APIキーはplatformから取得	<a href="https://docs.arcprize.org/">https://docs.arcprize.org/</a> / <a href="https://docs.arcprize.org/api-keys">https://docs.arcprize.org/api-keys</a> <sup>29</sup>	確認済	APIのSLAは無し (best-effort) 等、運用詳細は限定的。 <sup>30</sup>
ベンチマーク 結果の有無	公開時点のsemi-privateスコア表 (Table 2) を提示	Technical Report (Table 2) <sup>11</sup>	確認済	公式Webリーダーボードは動的で、本調査では同一数値の画面取得が困難 (ただし報告書に表で固定提示)。 <sup>31</sup>

## 技術仕様と評価プロトコル

ARC-AGI-3は「未知のゲーム環境で、説明なしに探索し、勝利条件を推定し、学習した機構を使って後半のレベルを解く」能力を測ることを狙う。公式ブログは「説明書もルールも目標も無い」と明示し、技術報告も「目的も指示も与えられない」ことを中核的困難として位置づける。 <sup>32</sup>

### 技術仕様の要約表

要素	仕様 (要約)	根拠	未記載/不明点
目的	エージェント型知能 (探索・モデル化・目標設定・計画実行) を「学習効率」で測る対話型ベンチマーク	<sup>33</sup>	「AGI判定」用途は公式に否定的 (測定ツールであり“litmus test”ではない立場)。 <sup>34</sup>
“ルール不明ゲーム”の定義	ルール・手順・ (少なくとも明示的な) 目標説明が与えられず、探索から勝利条件・力学を推定	<sup>32</sup>	ただし評価用のシステムプロンプトでは「goal is to win」が付与される (後述)。 <sup>11</sup>
インターフェース	ターン制。各ターンでフレーム (または短いフレーム列) を受取り、1アクションで遷移	<sup>35</sup>	マルチエージェント要素の有無は不明 (少なくとも報告書要約部では明示なし)。
観測 (入力)	64×64グリッド、各セルは16色。状態は“frame”	<sup>8</sup>	色の意味づけ (セマンティクス) は原則無し (文化的記号を避ける設計)。 <sup>36</sup>
行動 (出力)	環境ごとに異なるが、小さな行動集合 (例: 5つの基本行動+ Undo、座標指定の選択行動など)	<sup>37</sup>	具体的な全行動一覧・各環境のアクション空間は、動的ページ等で参照が必要 (本調査の静的取得では不明)。 <sup>22</sup>

要素	仕様（要約）	根拠	未記載/不明点
“行動”の定義	環境状態を変える離散的相互作用。内部計算（ツール呼び出し等）は行動数に含めない	38	「内部ツール」をどこまで許すかはリーダーボード種別（公式/コミュニティ）で扱いが異なる可能性。 <sup>12</sup>
スコアリング	RHAE：レベルごとに $\min(1, (h/a)^2)$ 、環境内はレベル重み（1～5）で加重平均、全環境平均	39	“n=5”として定義される一方、環境は「少なくとも6レベルで開発」とも記載。チュートリアル等の扱い詳細は文面上やや曖昧（不明）。 <sup>40</sup>
人間ベースライン	各環境10人が試行。人間基準は「2番目に良い行動数（second-best）」	41	分散・信頼区間など統計量の提示は限定的（不明）。
データセット構成	Public Demo 25、Semi-Private 55、Fully Private 55（計135環境）	10	公式ページのメタ説明は「150+環境」等を示唆（表現差）。拡張予定/数え方差かは不明。 <sup>42</sup>
データ（環境）の生成	人間ゲームデザイナーが手作業で制作。コア知識のみ、言語・文化記号排除、既存ゲームからの新規性検査等	43	自動生成（procedural generation）比率は不明（少なくとも主要記述は手工芸的製作）。
検証（品質保証）	ランダム探索で偶然解けないこと（例：50k/1Mステップ）やリプレイ再実行性を検証	44	バグ/エクスプロイト報奨等の制度は不明。
実装・性能	エンジンはPython実装、目標1,000 FPS。ローカル開発は～2,000 FPSを推奨値として提示	45	公式推奨ハード要件（CPU/GPU/メモリ）は不明。
API	REST API。状態を持つためCookieによるセッションアフィニティ要求。レート制限の概念あり	46	APIの可用性・SLAはbest-effort（詳細は限定的）。 <sup>30</sup>
リプレイ	API実行はオンラインで録画・共有可。ローカルのみ実行は録画生成しない	47	「推論ログ」の保存は任意/環境依存の可能性（例：サンプルリプレイに“no reasoning log”表示）。 <sup>48</sup>
モデルアーキテクチャ	ベンチマーク自体はモデル非依存。例としてCNN+RL等のアプローチが言及	49	公式の推奨ベースラインモデルは不明（少なくとも報告書要約では明示なし）。

## ベンチマーク結果と再現性評価

### 提示された主要な性能指標

ARC-AGI-3は「正解率」ではなく、行動効率を人間基準で正規化したRHAEを主指標とし、公開時点のsemi-privateリーダーボード結果を技術報告内で表として提示している。<sup>50</sup>

また、技術報告の要旨は「人間は環境を100%解ける一方、2026年3月時点のフロンティアAIは1%未満」と述べ、公式ブログは「Frontier AI scores 0.26%」とより具体化している（Table 2では0.26%がOpenAIのモデルに該当）。<sup>51</sup>

## ベンチマーク結果比較表

対象	評価セット/条件	指標	結果	根拠
人間	“easy for humans”基準を満たす環境群（各環境10人試行）	完遂可能性	100% solvable（設計・選別基準）	4
フロンティアAI（公開時点）	Semi-Private（公式リーダーボード、モデルAPIで評価）	RHAE（%）	0.00～0.37%（提供者別に表提示）	11
プレビュー競争（2025）上位	3公開+3非公開環境のプレビュー設定	競争スコア（%）	12.58%（1位例）、6.71%（2位例）等	52
参考：公式ブログ記述	公開直後の概況	RHAE（%）	“Frontier AI 0.26%”	6

補足として、Semi-Privateの公開時点スコア表（Table 2）は以下のとおり提示されている：Google（Gemini 3.1 Pro Preview）0.37%、OpenAI（GPT 5.4 High）0.26%、Anthropic（Opus 4.6 Max）0.25%、xAI（Grok-4.20 Beta）0.00%。<sup>11</sup>

ここで、Google<sup>53</sup>、OpenAI<sup>54</sup>、Anthropic<sup>55</sup>、xAI<sup>56</sup>はいずれも“semi-private leaderboard”の同一表内で比較されている点が重要で、少なくとも「比較可能性（apples-to-apples）」の主張はこの枠組みに依存している。<sup>57</sup>

### 再現性・統計的妥当性の評価

外部再現性は二層に分かれる。

第一に、公開デモ環境（Public Demo）の範囲では、SDKとローカル実行が可能であり、手元での再現実験（エージェントの動作・行動数計測）は現実的である。<sup>58</sup>

第二に、Semi-Private/Privateの“本番評価”は非公開セットであるため、第三者が同一データで独立再計測することは原理的に難しい（アクセス権と機密保持の問題）。この点は、過学習・汚染対策として意図された設計でもある。<sup>59</sup>

統計的有意性の観点では、Table 2の差（例：0.37% vs 0.26%）は絶対値が極小である一方、試行の分散や推定誤差（モデル側のランダム性、環境の確率性、サーバ実行の揺らぎ）が表では提示されていないため、「この差が意味のある差か」を厳密に判断する材料は不足している。<sup>60</sup>

### 弱点・回避可能性

最大の設計上リスクとして、技術報告は「公開環境へのタスク特化（task-specific overfitting）」と「ARC-AGI-3風の合成環境などへのドメイン特化（domain-specific overfitting）」を明示し、公開セットのスコアは公式リーダーボードに載せない方針を宣言している。<sup>61</sup>

さらに、「専用ハーネスの設計がスコアを人工的に押し上げ得る」ことを具体例（同一モデルでも環境により0%と97.1%のような二峰性が出うる）で述べ、公式スコアではハーネスを使わない方針を取っている。<sup>62</sup>

スコア設計側でも、グリッチで極端に少ない手数になるケースを想定してレベルスコアの上限を人間同等（1x）にクリップするなど、回避策を組み込んでいる。<sup>63</sup>

また、偶然の総当たり（brute force）を抑制するため、ランダムプレイで非チュートリアルが解けないことの検証や、行動数が人間の5倍を超えると打ち切る運用上の上限も導入されている。<sup>64</sup>

## 再現実験の簡潔な計画

フェーズ	手順 (最小構成)	必要資源	評価指標
環境整備	SDK導入 (pip/uv) 、ローカルOFFLINEでPublic環境を実行	一般的PC (推奨要件は不明) 、Python環境	1ゲームあたりの行動数、完了レベル数、実行速度 (FPS)
ベースライン	ランダム/単純探索/状態グラフ探索など複数方針を実装し、同一ゲームで比較	実装時間、計算時間	RHAE近似 (人間hが不明なら相対比較) 、成功率、行動数分布
ログと検証	ONLINEモードでscorecard/replayを取得 (必要ならAPIキー作成)	APIキー、レート制限 (例: 600 rpm)	リプレイによるトレース一致、再実行による再現性
本番類似評価	Semi-Private相当の評価はアクセス権が必要 (外部は原則不可)	(不明: アクセス権/契約)	技術報告のTable 2 同等のRHAE (%)
堅牢性チェック	行動上限、リセット頻度、探索深さを変え、スコアの不安定性を測定	複数seed実行	平均/分散、最悪ケース、コスト (時間・API費用)

ローカル実行は「高速・無制限」を利点とし、オンライン実行は「スコアカード・共有リプレイ・リーダーボード反映」を利点とするため、研究用途ではOFFLINEで大量試行→重要実験のみONLINEで可視化、が合理的である。<sup>65</sup>

## 学術的文脈

ARC-AGI系列の中核思想は「知能=スキル獲得効率 (skill-acquisition efficiency)」という定義にあり、ARC-AGI-1は少数例から変換規則を推定する静的パズルとして提案された。<sup>66</sup>

ARC-AGI-2は同じ入出力形式を保ちつつ、より高い推論複雑性に対して“より細かい評価信号”を得る方向で拡張された、と一次論文で述べられている。<sup>67</sup>

ARC-AGI-3はここからさらに形式を変え、「静的な答え合わせ」ではなく「時間方向にわたる探索と計画」を、行動効率 (action efficiency) として測る点が差分である。<sup>68</sup>

RHAEの設計は、ロボティクス/ナビゲーション分野で成功と経路効率を同時に測るSPL (Success weighted by Path Length) に触発されたと技術報告中で明示されている。<sup>69</sup>

この系譜は、「成功率だけでは“雑な成功 (遠回り・総当たり)”と“洗練された成功”を区別できない」という評価論の流れに整合的である。<sup>69</sup>

一方、ARC-AGI-3“プレビュー”期から、LLM単体ではなく「状態空間探索」「グラフ構築」「視覚的セグメンテーション」といった構造化探索が有力であることを示す研究も出ており、学習無し (training-free) 探索での一定の成果が報告されている。<sup>70</sup>

このことは、ARC-AGI-3が (少なくとも短期的には) 「大規模事前学習の知識」より「未知環境での探索戦略・記憶・仮説管理」をボトルネックとして浮かび上がらせる、という設計意図と整合する。<sup>71</sup>

## 倫理・社会的影響と実務的含意

ARC-AGI-3は研究コミュニティへの刺激 (賞金、オープンソース要求) を強く持つ一方で、社会的には「単一指標でAGIに近づいた/遠い」を断定する誤用が起り得る。ARC Prize側も、タスクは“AGIのリトマス試験紙ではない”旨、また自己申告スコアが比較不能になりやすい旨をTesting Policyで強調している。<sup>72</sup>

実務的には、エージェント評価を「成功/失敗」から「効率 (無駄な試行・手戻りの少なさ)」へ寄せる設計

は、現実のエージェント運用（コスト・安全）の関心と整合する可能性がある。評価の透明性を高めるためにリプレイやスコアカードを提供する点も、監査可能性に資する。<sup>73</sup>

ただし、公開と秘匿のトレードオフ（過学習対策のため非公開セットが不可欠、しかし第三者検証は難化）が残るため、政策・調達・投資判断に使う場合は「ARC-AGI-3単独」ではなく、複数評価軸との組合せが安全である。<sup>74</sup>

## 倫理・政策上の懸念と推奨アクション

論点	懸念	推奨アクション	根拠 (一次中心)
ベンチマーク“ゲーム化”	公開セットや類似合成環境への特化で、一般化能力を偽装	公式スコアは非公開セット中心+手続き固定、公開スコアはデモ用途に限定（既方針の徹底）	12
透明性 vs 秘匿性	非公開データは独立再現を阻害	少なくとも「評価手続き」「プロンプト」「サンプリング/コスト/時間」等のメタデータ公開を拡充（Verifiedの監査プロセスを活用）	75
公平性/バイアス	人間ベースラインが特定地域・募集方法に偏る可能性	人間データの統計量（分散/属性別）を追加公開し、基準の頑健性を検証	76
誤用（宣伝・政策）	“AGI達成”の誇大宣伝、逆に過小評価	「ベンチマークの測っている能力の射程」を明確化し、複数軸評価（安全・堅牢性・透明性）へ接続	77
ガバナンス整合	評価・監査の要求が国際的に増加	企業はAIリスク管理（NIST <sup>78</sup> AI RMF等）や管理規格（ISO <sup>79</sup> /IEC 42001等）と評価ログを接続	80
規制動向（EU例）	透明性・高リスク管理の要請が強化	欧州委員会のAI Act説明に沿い、透明性・ログ・文書化・人間監督を評価設計/運用に組み込む	81

なおEUではAI Actがリスクベースで義務を段階適用すると整理されており、透明性義務（生成AIの識別等）や高リスクAI義務の適用時期が明示されている（同ページ“Application timeline”など）。<sup>81</sup>

## 結論と推奨

ARC-AGI-3は、ARC系列の理念（知能=新規タスクに対する学習効率）を「対話的・時間的」な設定へ拡張し、探索・仮説更新・計画というエージェント能力を、行動効率で測る点に新規性がある。<sup>82</sup>

公開時点のsemi-privateスコアが0.00~0.37%という事実は、少なくとも「一般用途APIとして提供されるフロンティアモデルを、同一手続き（ハーネス無し）で測る」条件では、顕著なギャップが残ることを示す。<sup>57</sup>

他方、ハーネス設計・公開環境特化・非公開データによる再現性制約といった“評価研究あるある”が強く表面化するベンチマークでもあるため、社会的な意思決定に用いる際は、メタデータの充実・第三者監査・複数評価軸での triangulation が不可欠である。<sup>74</sup>

推奨として、研究者は「公開デモ=開発用」「非公開セット=主張の根拠」という位置づけを守りつつ、探索・記憶圧縮・仮説管理の一般原理（他ドメインへ移植可能な設計）に焦点を当てるべきである。<sup>83</sup>

政策立案者は、ARC-AGI-3のような“高信号”評価を歓迎しつつも、単一ベンチマークでAGIや安全性を断定しないガイダンス（透明性・ログ・検証可能性の最低要件）を整備し、産業界の誇大広告・過小評価の双方を

抑制するのが望ましい。 77

企業は、ARC-AGI-3のRHAЕ思想（無駄な行動を抑える効率）を自社エージェント評価に取り込みつつ、ドメイン固有制約（業務ツール、権限、失敗コスト）を加えた社内ベンチと組み合わせ、監査可能なリプレイ/ログ運用を標準化することが実務上の価値につながる。 84

---

1 3 4 5 7 8 9 10 11 12 16 26 33 35 36 37 38 39 40 41 44 45 49 50 51 52 53 54 56

57 60 61 62 63 64 68 69 71 75 76 78 83 <https://arcprize.org/media/>

**ARC\_AGI\_3\_Technical\_Report.pdf**

[https://arcprize.org/media/ARC\\_AGI\\_3\\_Technical\\_Report.pdf](https://arcprize.org/media/ARC_AGI_3_Technical_Report.pdf)

2 6 17 32 43 79 <https://arcprize.org/blog/arc-agi-3-launch>

<https://arcprize.org/blog/arc-agi-3-launch>

13 27 <https://github.com/arcprize/ARC-AGI/blob/main/LICENSE>

<https://github.com/arcprize/ARC-AGI/blob/main/LICENSE>

14 18 66 82 <https://arxiv.org/abs/1911.01547>

<https://arxiv.org/abs/1911.01547>

15 <https://arcprize.org/blog/arc-agi-3-preview-30-day-learnings>

<https://arcprize.org/blog/arc-agi-3-preview-30-day-learnings>

19 <https://gigazine.net/news/20260326-arc-agi-3/>

<https://gigazine.net/news/20260326-arc-agi-3/>

20 <https://arcprize.org/about>

<https://arcprize.org/about>

21 42 <https://arcprize.org/arc-agi/3/>

<https://arcprize.org/arc-agi/3/>

22 <https://arcprize.org/tasks?v=3>

<https://arcprize.org/tasks?v=3>

23 29 55 <https://docs.arcprize.org/>

<https://docs.arcprize.org/>

24 <https://docs.arcprize.org/toolkit/overview>

<https://docs.arcprize.org/toolkit/overview>

25 <https://github.com/arcprize/arc-agi>

<https://github.com/arcprize/arc-agi>

28 **ARC Prize 2026**

<https://arcprize.org/competitions/2026>

30 46 [https://docs.arcprize.org/rest\\_overview](https://docs.arcprize.org/rest_overview)

[https://docs.arcprize.org/rest\\_overview](https://docs.arcprize.org/rest_overview)

31 <https://arcprize.org/arc-agi/3/leaderboard>

<https://arcprize.org/arc-agi/3/leaderboard>

34 59 72 74 77 <https://arcprize.org/policy>

<https://arcprize.org/policy>

47 73 84 <https://docs.arcprize.org/recordings>

<https://docs.arcprize.org/recordings>

48 <https://arcprize.org/replay/32d2653f-4bb1-461a-b23e-1fdd5229cb71>

<https://arcprize.org/replay/32d2653f-4bb1-461a-b23e-1fdd5229cb71>

58 65 <https://docs.arcprize.org/local-vs-online>

<https://docs.arcprize.org/local-vs-online>

67 <https://arxiv.org/abs/2505.11831>

<https://arxiv.org/abs/2505.11831>

70 [2512.24156] Graph-Based Exploration for ARC-AGI-3 Interactive Reasoning Tasks

<https://arxiv.org/abs/2512.24156>

80 AI Risk Management Framework | NIST

[https://www.nist.gov/itl/ai-risk-management-framework?utm\\_source=chatgpt.com](https://www.nist.gov/itl/ai-risk-management-framework?utm_source=chatgpt.com)

81 AI Act | Shaping Europe's digital future

<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>