

ARC Prize 2025

Create an AI capable of novel reasoning

header

ファイル · 110 KB



Overview

Data

Code

Models

Discussion

Leaderboard

Rules

Overview

In this competition, you'll develop AI systems to efficiently learn new skills and solve open-ended problems, rather than depend exclusively on systems trained with extensive datasets. The top submissions will show improvement toward human level reasoning.

Start

23 days ago

Close

7 months to go

Merger & Entry

Description

linkkeyboard_arrow_up

Note: This is the second ARC Prize competition on Kaggle. It builds upon the [ARC Prize 2024](#). This second competition has an updated dataset of human-calibrated problems and increased compute for participants.

Current AI systems can not generalize to new problems outside their

training data, despite extensive training on large datasets. LLMs have brought AI to the mainstream for a large selection of known tasks. However, progress towards Artificial General Intelligence (AGI) is idea constrained. Improvements in AGI could enable AI systems that think and invent alongside humans.

The Abstraction and Reasoning Corpus for Artificial General Intelligence (ARC-AGI-2) benchmark measures an AI system's ability to efficiently learn new skills. Humans have collectively scored 100% in ARC, whereas the best AI systems only score 4%. The [ARC Prize competition](#) encourages researchers to explore ideas beyond LLMs, which depend heavily on large datasets and struggle with novel problems.

This competition includes several components. The competition as described here carries a prize of \$125,000 with an additional \$600,000 available if any team can beat a score of 85% on the leaderboard. Further opportunities outside of Kaggle may also be available- to learn more visit [ARCprize.org](#).

Your work could contribute to new AI problem-solving applicable across industries. Vastly improved AGI will likely reshape human-machine interactions. Winning solutions will be open-sourced to promote transparency and collaboration in the field of AGI.

Evaluation

linkkeyboard_arrow_up

This competition evaluates submissions on the percentage of correct predictions. For each task, you should make 2 attempts to predict the exact outputs for every test input grid contained in the task. (Tasks can have more than one test input that needs a predicted output.) Each task test output has one ground truth. For a given task output, any of the 2 predicted outputs matches the ground truth exactly, you score 1 for that task test output, otherwise 0. The final score is the sum averaged of the highest score per task output divided by the total number of task test outputs.

Submission File

The submission file for this competition must be a json named submission.json.

For each task output in the evaluation set, you should make exactly 2 predictions (attempt_1, attempt_2). The structure of predictions is shown below. Many tasks have multiple outputs (a multiple dictionaries enclosed in a list), although some tasks have a single output that must be predicted. When a task has multiple test outputs that need to be predicted (e.g., task 12997ef3 below), they must be in the same order as the corresponding test inputs.

IMPORTANT: All the task_ids in the input challenges json file must also be present in the submission.json file. Both "attempt_1" and "attempt_2" must be present, even if your submission doesn't have 2 predictions.

```
{"00576224": [{"attempt_1": [[0, 0], [0, 0]], "attempt_2": [[0, 0], [0, 0]]},  
  "009d5c81": [{"attempt_1": [[0, 0], [0, 0]], "attempt_2": [[0, 0], [0, 0]]},  
  "12997ef3": [{"attempt_1": [[0, 0], [0, 0]], "attempt_2": [[0, 0], [0, 0]]},  
    {"attempt_1": [[0, 0], [0, 0]], "attempt_2": [[0, 0], [0, 0]]},  
  ...  
}
```

Timeline

linkkeyboard_arrow_up

- **March 24, 2025** - Start Date
- **October 27, 2025** - Entry deadline. You must accept the competition rules before this date in order to compete.
- **October 27, 2025** - Team Merger deadline. This is the last day participants may join or merge teams.
- **November 3, 2025** - Final submission deadline.
- **November 9, 2025** - Paper Award submission deadline.

All deadlines are at 11:59 PM UTC on the corresponding day unless otherwise noted. The competition organizers reserve the right to update the contest timeline if they deem it necessary.

Prizes

linkkeyboard_arrow_up

TOTAL PRIZES AVAILABLE: \$725,000

- 2025 Progress Prizes: \$125,000
- Grand Prize: \$600,000

In line with the spirit of the competition, participants eligible for a prize will be removed from the competition if they do not open source their solutions.

2025 Progress Prizes

- **Prizes for Top-Ranking Teams in this Competition:**
 - \$50,000
 - First Prize: \$25,000
 - Second Prize: \$10,000
 - Third Prize: \$5,000

- Fourth Prize: \$5,000
- Fifth Prize: \$5,000
- **Paper Award Prizes:** \$75,000
 - Winner: \$50,000
 - First Runner Up: \$20,000
 - Second Runner Up: \$5,000

See the **Paper Award** tab for more details on submission and evaluation.

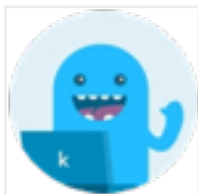
Grand Prize

A Grand Prize of an additional \$600,000 will be unlocked in the event that a team achieves a score of at least 85% accuracy on the competition leaderboard. At the end of the competition, the Grand Prize will be divided among the Top 5 teams that have achieved 85% accuracy as outlined below. In the event that fewer than 5 teams have achieved 85% accuracy, those prizes will be divided proportionately among qualifying teams.

- First Prize: \$300,000
- Second Prize: \$120,000
- Third Prize: \$60,000
- Fourth Prize: \$60,000
- Fifth Prize: \$60,000

Code Requirements

linkkeyboard_arrow_up



This is a Code Competition

Submissions to this competition must be made through Notebooks. In order for the "Submit to Competition" button to be active after a commit, the following conditions must be met:

- CPU Notebook \leq 12 hours run-time
- GPU Notebook \leq 12 hours run-time
- No internet access enabled
- External data, freely & publicly available, is allowed, including pre-trained models
- Submission file must be named submission.json

Submission runtimes have been obfuscated. If you repeat the exact same submission you will see up to 10 minutes of variance in the time before you receive your score.

Please see the [Code Competition FAQ](#) for more information on how to submit.

Paper Award

linkkeyboard_arrow_up

You may choose to submit a paper to be eligible to win a Paper Award Prize.

To be eligible for a Paper Award, you must separately submit a paper (Kaggle Notebook, PDF, arXiv, txt, etc.) documenting and describing the conceptual approach of your eligible ARC Prize 2025 Kaggle submission. Paper submissions must be submitted within 6 days of the competition ending and must be public (thus, also open-sourced).

Paper Awards are evaluated equally on the following six components, where a score of 0 (lowest) and 5 (highest) is given in each category.

Category	Description	
Accuracy	How accurate is the submission based on its performance on the leaderboard? (Note: you will be asked to provide your submission ID to match your notebook to your submission)	
Universality	How general and universal is the Submission approach beyond the competition? Does your submission translate to other similar problems? How well does your method generalize?	

Progress	How much does the paper increase the overall chance of anyone achieving 85% on ARC Prize?	
Theory	How well does the paper describe why the Submission works (as opposed to merely describing how it works)?	
Completeness	How thoroughly and completely does the paper cover your submission to the leaderboard?	
Novelty	How novel is the Submission relative to existing public research?	

The Paper Award will be awarded to the Submission with the most points and the Paper Award Runner Up will be the second most points. Your Paper rubric evaluation will not be shared with you. In the event of a tie, the Paper that was entered first to the Competition will be the winner. In the event a potential winner is disqualified for any reason, the Paper that received the next highest score rank will be chosen as the potential winner.

Click here to go to the [Submission form](#)

Upgraded Accelerators

linkkeyboard_arrow_up

This competition has access to Kaggle's pool of powerful new L4x4 machines! These machines offer 96GB of GPU memory enabling submissions with much larger models. See [this page](#) for more on these machines.

What you need to know:

- **Quota usage** - Due to their limited availability, notebooks with L4x4s consume GPU quota at twice the rate of the older T4x2 and P100 machines. We may increase this rate as

necessary to ensure these machines are available for submission scoring.

- **Restricted Use** - L4s are only available for notebooks attached to this competition. We will build tooling to enforce this if necessary. In the meantime, attempts to circumvent this restriction may be enforced by Kaggle moderation with consequences including team bans from the competition or account bans.
- **No Internet** - All L4 sessions must have internet disabled.

Citation

linkkeyboard_arrow_up

Francois Chollet, Mike Knoop, Greg Kamradt, Walter Reade, and Addison Howard. ARC Prize 2025. <https://kaggle.com/competitions/arc-prize-2025>, 2025. Kaggle.

Dataset Description

The objective of this competition is to create an algorithm that is capable of solving abstract reasoning tasks. Critically, these are *novel* tasks: tasks that the algorithm has never seen before. Hence, simply memorizing a set of reasoning templates will not suffice.

The format is different from the previous competition, so please read this information carefully, and refer to supplementary documentation as needed.

When looking at a task, a "test-taker" has access to inputs and outputs of the demonstration pairs (train pairs), plus the input(s) of the test pair(s). The goal is to construct the output grid(s) corresponding to the test input grid(s), using **2 trials** for each test input. "Constructing the output grid" involves picking the height and width of the output grid, then filling each cell in the grid with a symbol (integer between 0 and 9, which are visualized as colors).

Only **exact** solutions (all cells match the expected answer) can be said to be correct.

Any additional information, as well as an interactive app to explore the objective of this competition is found at the [ARCPrize.org](https://arcprize.org). ***It is highly recommended that you explore the interactive app, as the best way to understand the objective of the competition.***

Task files

The information is stored in two files:

- **arc-agi_training-challenges.json**: contains input/output pairs that demonstrate reasoning pattern to be applied to the

"test" input for each task. This file and the corresponding **solutions** file can be used as training for your models.

- **arc-agi_training-solutions.json**: contains the corresponding task "test" outputs (ground truth).
- **arc-agi_evaluation-challenges.json**: contains input/output pairs that demonstrate reasoning pattern to be applied to the "test" input for each task. This file and the corresponding **solutions** file can be used as validation data for your models.
- **arc-agi_evaluation-solutions.json**: contains the corresponding task "test" outputs (ground truth).
- **arc-agi_test-challenges.json**: this file contains the tasks that will be used for the leaderboard evaluation, and contains "train" input/output pairs as well as the "test" input for each task. Your task is to predict the "test" output. **Note:** The file shown on this page is a placeholder using tasks from **arc-agi_evaluation-challenges.json**. When you submit your notebook to be rerun, this file is swapped with the actual test challenges.
- **sample_submission.json**: a submission file in the correct format

Each task contains a dictionary with two fields:

- "train": demonstration input/output pairs. It is a list of "pairs" (typically 3 pairs).
- "test": test input - your model should predict the output.

A "pair" is a dictionary with two fields:

- "input": the input "grid" for the pair.
- "output": the output "grid" for the pair.

A "grid" is a rectangular matrix (list of lists) of integers between 0 and 9 (inclusive). The smallest possible grid size is 1x1 and the largest is 30x30.

The data on this page should be used to develop and evaluate your models. When notebooks are submitted for rerun, they are scored using 100 unseen tasks found in the rerun file named **arc-agi_test-challenges.json**. The rerun tasks will contain train pairs of inputs and outputs as well as the tasks test input. Your algorithm must predict the test output. The majority of the 100 tasks used for leaderboard score only have one test input that will require a corresponding output prediction, although for a small number of tasks, you will be asked to make predictions for two test inputs.

