

# ARC Prize 2025

Jae-ik Shin

2025.09.26

# On the Measure of Intelligence (2019.11)

- Francois Collet : Google 소프트웨어 엔지니어, 인공지능 연구자
- "우리가 '지능'에 대해 이야기할 때, '지능'이라는 게 뭐냐라는 질문에 충분히 만족할 만한 답이 없다는 것 자체가 이 분야가 아직 성숙하지 못하다는 것의 반증이라고 생각합니다. 더 큰 문제는, 이 '지능'을 명확하게 정의한다거나 '지능'을 향해 우리가 얼마나 나아가고 있는가를 확인하고 체크하는 작업에 거의 관심들이 없다는 것이구요."

# 지능이라는 게 무엇인가?

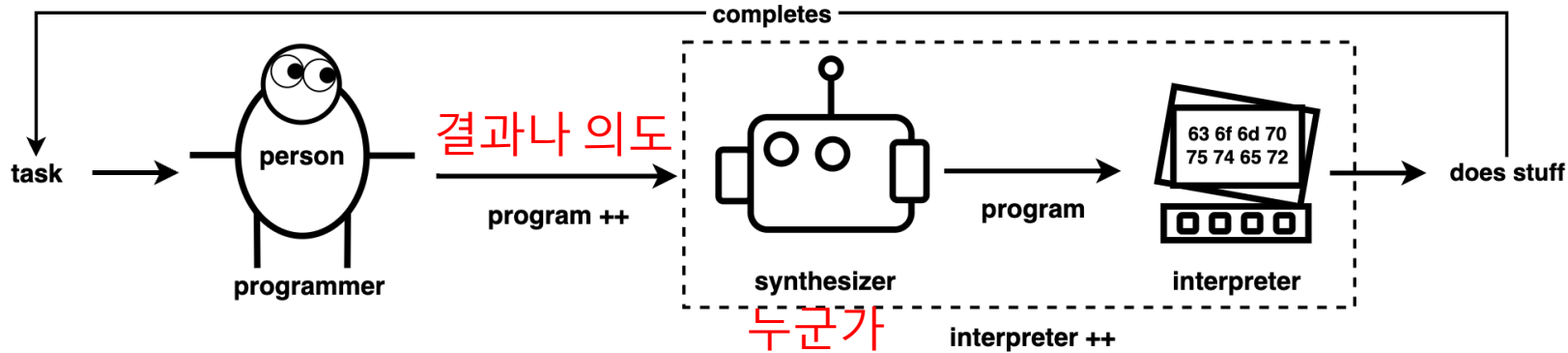
- 심리적인 관점: 프랑스의 심리학자 Alfred Binet (IQ 개발)
  - 인지적 능력 (Cognitive Abilities)을 IQ와 같은 심리 측정 테스트를 기준으로 정량화
  - 제한된 범위 내에서의 기존 기술을 측정
- AI의 관점: Alan Turing
  - 지능 = 다양한 환경에서 목표를 달성할 수 있는 능력
  - 적응력 (Adaptation)과 학습 (Learning)

<https://turingpost.co.kr/p/fod57>

# 인공 일반 지능(Artificial General Intelligence, AGI)

- Francois : 두 가지 관점이 모두 불완전 -> ‘Skill-Acquisition Efficiency’ (기술 습득 효율성)
- “시스템의 ‘지능’은 선행된 작업 (Priors), 경험 (Experiences), 그리고 일반화 (Generalization)의 난이도에 연관되는 특정 범위의 작업에 대해 얼마나 효율적으로 기술을 습득하느냐 (Skill-Acquisition Efficiency)의 지표입니다.”
- AI의 일반화 (Generalization) 능력을 평가하기 위한 목적으로 만들어진 ARC-AGI (Abstraction and Reasoning Corpus for Artificial General Intelligence) 데이터셋을 소개함

# 프로그램 합성 (Program Synthesis)



- 직접 코드를 모두 작성하지 않고도 원하는 결과나 의도를 설명하면 누군가가 자동으로 그에 맞는 프로그램을 만들어주는 기술
- Program Synthesis 예시
  - 엑셀에서 "이 열의 모든 이메일 주소에서 도메인만 추출해줘" 라고 하면, 자동으로 필요한 함수를 만들어주는 것
  - "두 숫자를 입력받아서 최대공약수를 구하는 프로그램을 만들어줘" 라고 하면, AI가 알아서 적절한 알고리즘을 선택해서 코드를 작성해주는 것

<https://turingpost.co.kr/p/fod-84-program-synthesis-agi>

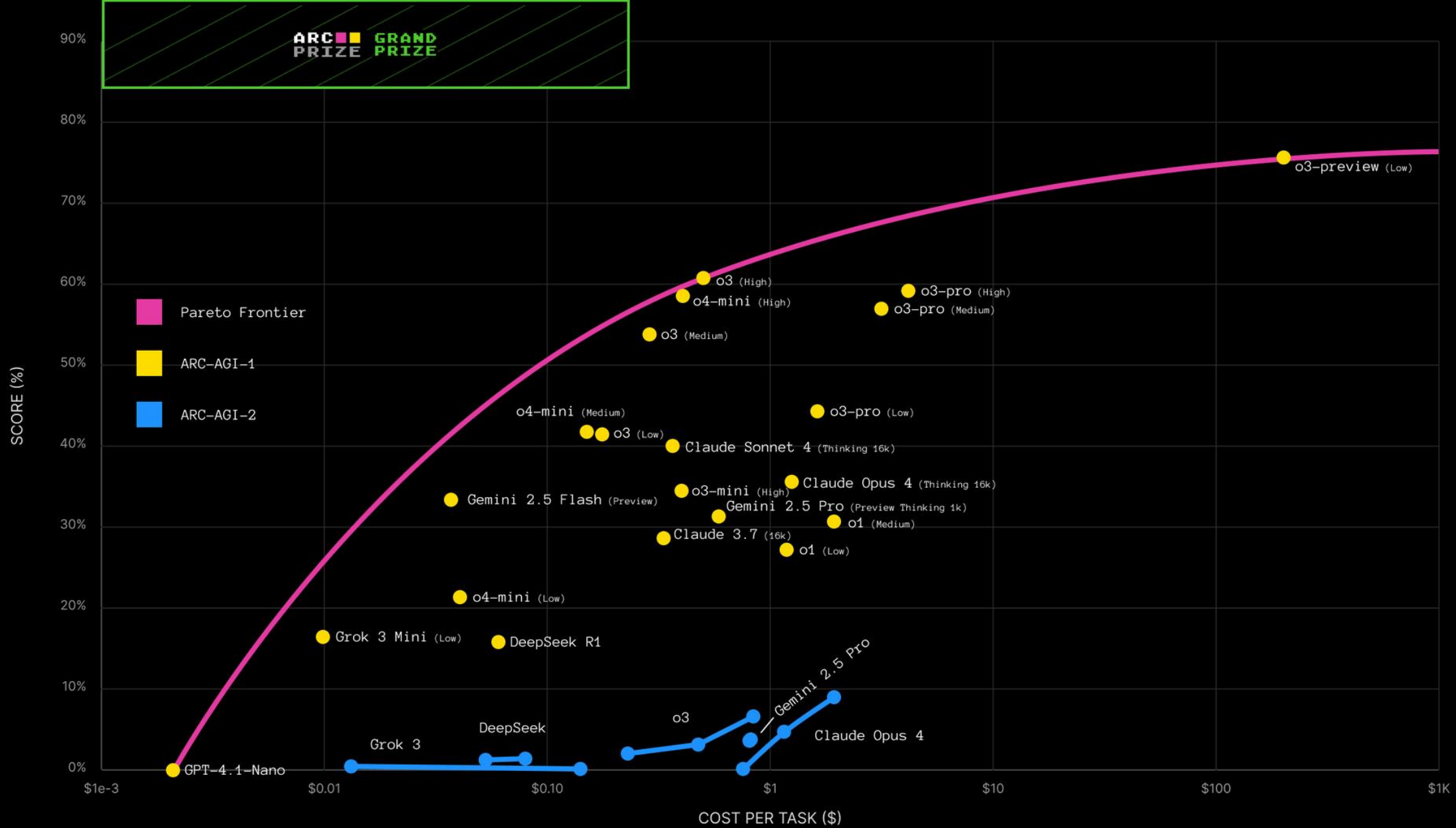
<https://people.csail.mit.edu/asolar/SynthesisCourse2023/index.htm>

# Francois Collet, NeurlPS 2020

- 딥러닝만의 접근법으로부터 오는 한계 : 거대한 데이터셋에 대한 의존성, 추론과 일반화의 어려움
- Program Synthesis : 딥러닝의 대체재가 아닌 보완을 해 주는 도구
  - AI가 특정한 작업에 특화된, 작은 프로그램을 만드는 방법 = 지능을 평가하는 이상적인 방법
  - '지능이라는 과정 (Process of Intelligence)' (해결책을 생성하는 체계로서의 과정) 과 특정 문제에 대한 해결책 그 자체인 '출력 (Output)'을 구분해야 한다
- 딥러닝과 Program Synthesis를 결합
  - 딥러닝 모델 : 검색 공간 (Search Space)을 좁히고 대규모 패턴 인식을 처리하면서 Program Synthesis를 가이드
  - Program Synthesis : 추론, 그리고 추상화 작업
  - AI가 추상적인 추론을 통해서 해결책을 생성 -> 더 적응력있게, 새로운 문제에 잘 대응하는 확장 가능한 시스템
  - 하이브리드 접근 방식
- 기존의 '정적 (Static)'인 작업 수행 패러다임에서 '적응성 (Adaptability)', 그리고 '추론 (Reasoning)' 쪽으로 그 무게 중심을 옮기게 되는 관점의 전환

<https://turingpost.co.kr/p/fod-84-program-synthesis-agi>

<https://joepalermo.github.io/2021/01/10/talk-review-francois-chollet-neurips-2020.html>



<https://arcprize.org/>

# Dataset

JSON	Dict	List
arc-agi_training-challenges.json	Train	Input Output
	Test	Input
arc-agi_training-solutions.json		Output
arc-agi_evaluation-challenges.json	Train	Input Output
	Test	Input
arc-agi_evaluation-solutions.json		Output
arc-agi_test-challenges.json	Test	Input

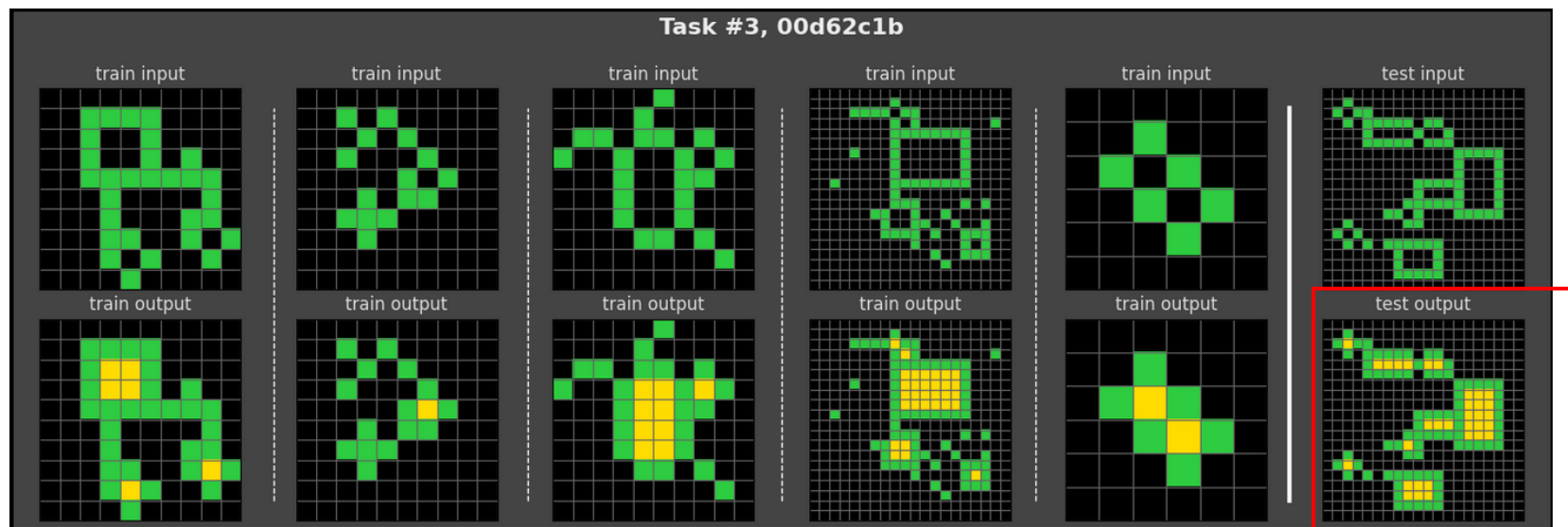


```

{'test': [{'input': [[7, 0, 7], [7, 0, 7], [7, 7, 0]]}],
train': [{'input': [[0, 7, 7], [7, 7, 7], [0, 7, 7]],
  'output': [[0, 0, 0, 0, 7, 7, 0, 7, 7],
    [0, 0, 0, 7, 7, 7, 7, 7, 7],
    [0, 0, 0, 0, 7, 7, 0, 7, 7],
    [0, 7, 7, 0, 7, 7, 0, 7, 7],
    [7, 7, 7, 7, 7, 7, 7, 7, 7],
    [0, 7, 7, 0, 7, 7, 0, 7, 7],
    [0, 0, 0, 0, 7, 7, 0, 7, 7],
    [0, 0, 0, 7, 7, 7, 7, 7, 7],
    [0, 0, 0, 0, 7, 7, 0, 7, 7]]},
  {'input': [[4, 0, 4], [0, 0, 0], [0, 4, 0]],
  'output': [[4, 0, 4, 0, 0, 0, 4, 0, 4],
    [0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 4, 0, 0, 0, 0, 0, 4, 0],
    [0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 0, 0, 0, 0, 0, 0, 0, 0],
    [0, 0, 0, 4, 0, 4, 0, 0, 0],

```

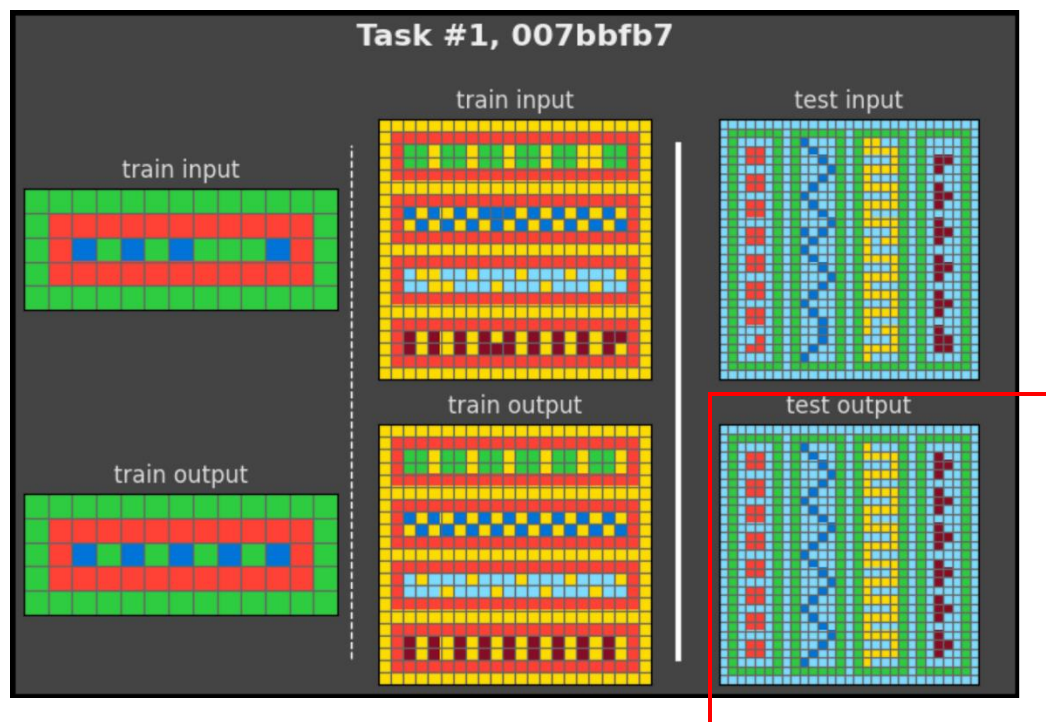
arc-agi\_training-challenges.json



arc-agi\_training-challenges.json

arc-agi\_training-solutions.json

1x1 ~ 30x30 grid  
Value = 0~9 (Color)



ARC Prize 2025

1000 training task

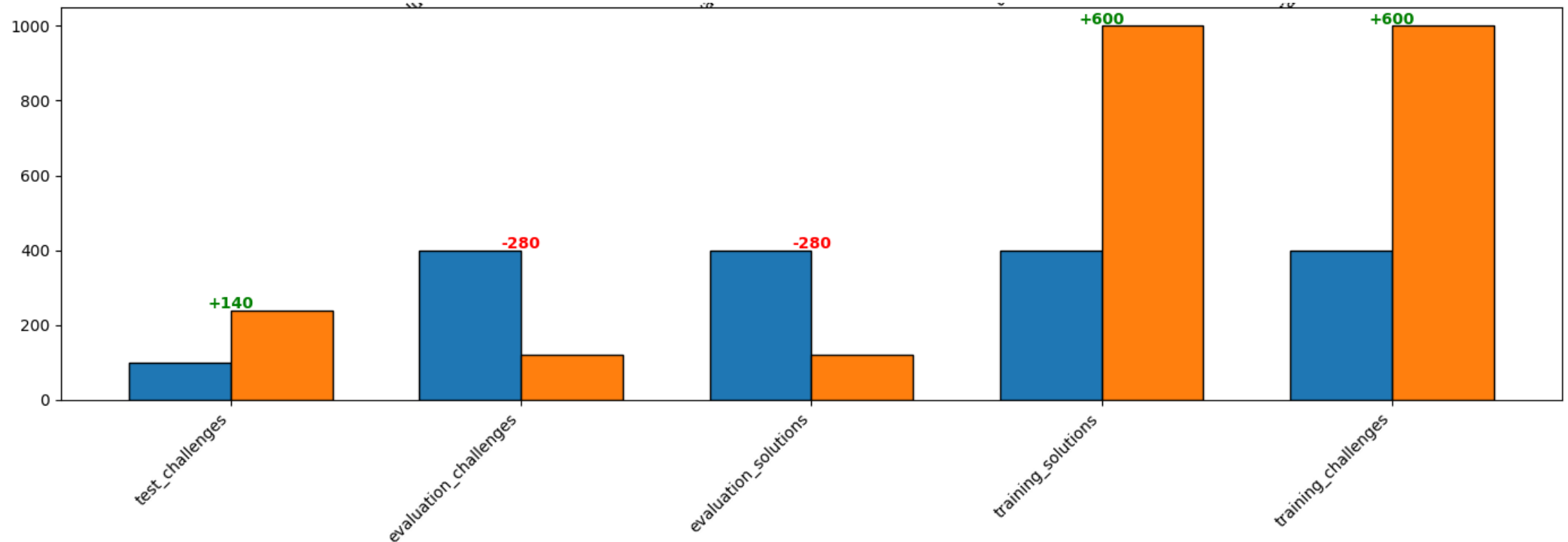
120 evaluation task

240 unseen test task

arc-agi\_evaluation-challenges.json

arc-agi\_evaluation-solutions.json



















# Dataset : ARC 2024 vs ARC 2025



<https://www.kaggle.com/code/boristown/data-difference-between-arc-2024-2025>

# Kaggle ARC Prize 2024 Leaderboard

■ Prize Winners

#	Team	Members		Score	Entries	Last	Solution
1	the ARChitects	 		53.50	243	10mo	
2	Guillermo Barbadillo			40.00	256	10mo	
3	alijs			40.00	399	1y	
4	William Wu			37.00	315	1y	
5	PoohAI	   		37.00	302	10mo	

<https://www.kaggle.com/competitions/arc-prize-2024/leaderboard>

# ARC 2024 1st solution (unofficial)

- Daniel Franzen
- <https://www.kaggle.com/dfranzen>
- Closed-source model (dfranzen/wb55l\_nemomini\_fulleval)
- [https://www.kaggle.com/models/dfranzen/wb55l\\_nemomini\\_fulleval](https://www.kaggle.com/models/dfranzen/wb55l_nemomini_fulleval)
- <https://www.kaggle.com/code/dfranzen/arc-prize-2024-solution-by-the-architects>

# ARC 2024 1st solution (official)

- Guillermo Barbadillo (Competitions Master)
- <https://www.kaggle.com/ironbar>
- open source model
- <https://www.kaggle.com/code/ironbar/single-task-test-time-fine-tuning-for-arc24>
- <https://www.kaggle.com/competitions/arc-prize-2024/writeups/guillermo-barbadillo-2nd-place-solution-for-the-ar>
- [https://ironbar.github.io/arc24/05\\_Solution\\_Summary/](https://ironbar.github.io/arc24/05_Solution_Summary/)

# ARC 2024 1st solution (official)

## 2nd Place Solution for the ARC Prize 2024 Competition: Omni-ARC approach



ARC Prize 2024

Solution Writeup · 2nd place · Dec 7, 2024

Any feedback to improve this post or the [paper](#) is very welcome 🙌!

This post is almost a duplicate of the [paper](#), I would recommend to go to the [paper](#) directly because it has better formatting.

### TLDR

My solution is an implementation of the MindsAI team approach (test-time fine-tuning) that in addition to train the model to generate the test outputs (the original ARC task) it also trains the model to do other tasks such as learning the input distribution by generating new inputs.

AUTHOR



**Guillermo Barbadillo**

Creator · ironbar



...

<https://www.kaggle.com/competitions/arc-prize-2024/writeups/guillermo-barbadillo-2nd-place-solution-for-the-ar>

[https://ironbar.github.io/arc24/05\\_Solution\\_Summary/](https://ironbar.github.io/arc24/05_Solution_Summary/)

# ARC 2024 1st solution (official) code

**ARC Prize 2024**

Late Submission ...

OverviewDataCodeModelsDiscussionLeaderboardRules


**Notebooks**

+ New Notebook

Q Search notebooks


≡ Filters

AllYour WorkShared With YouBookmarksPublic Score ▾

**Single-task Test-time fine-tuning for ARC24**


Notebook copied with edits from a private notebook · Updated 1y ago  
Score: 40 · [5 comments](#) · ARC solution source files by Icecuber +50

216  
Gold ...

**FS 1 (inherited NewZoe V7)**


Notebook copied with edits from a private notebook · Updated 1y ago  
Score: 31 · [2 comments](#) · [Private Datasource] +4

23  
Bronze ...

**Corr100Ensemble1+IceCUBE1+Sam8+8**

Updated 1y ago  
Score: 30 · [0 comments](#) · ARC solution source files by Icecuber +22

13  
Bronze ...

**arc-ensemble-v3**

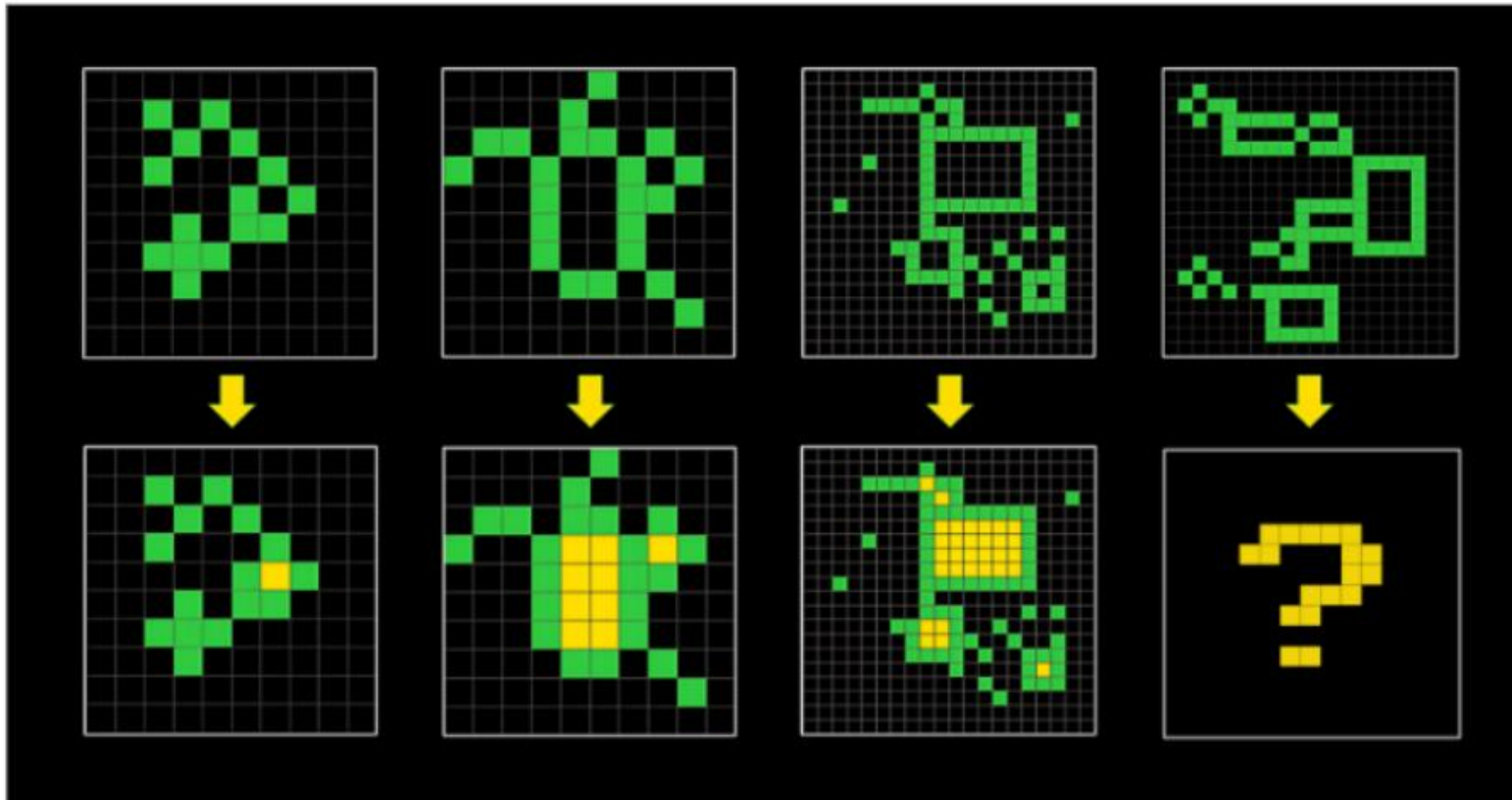
Notebook copied with edits from a private notebook · Updated 1y ago  
Score: 27 · [2 comments](#) · ARC solution source files by Icecuber +9

13  
...



# Motivation of Ironbar's approach

How can we learn from few high-dimensional examples?



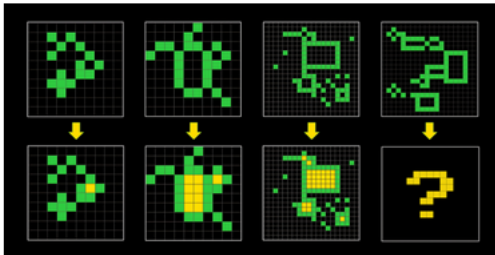
- Position? ✗
- Size? ✗
- Color? ✓
- Shape? ✗
- Holes? ✓
- Movement? ✗
- Symmetry? ✗

# Find right representation

Omni-ARC: Training a single model to do multiple ARC-related tasks

Few-shot  
Prompt,  
Domain  
Specific  
Language

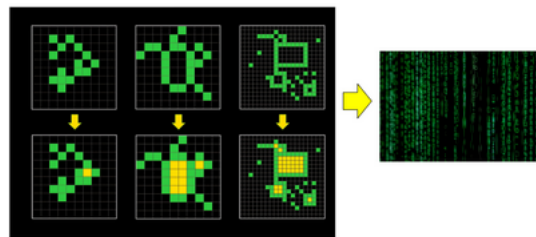
examples + input -> output



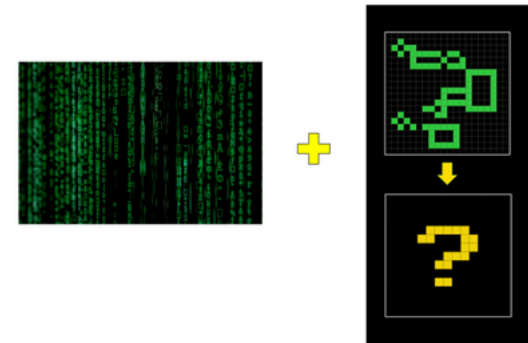
inputs -> input



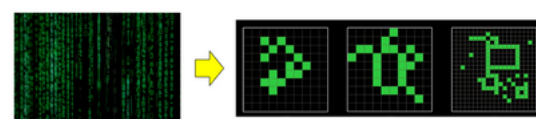
examples -> code



code + input -> output



code -> inputs



inputs -> code



Code  
Generation

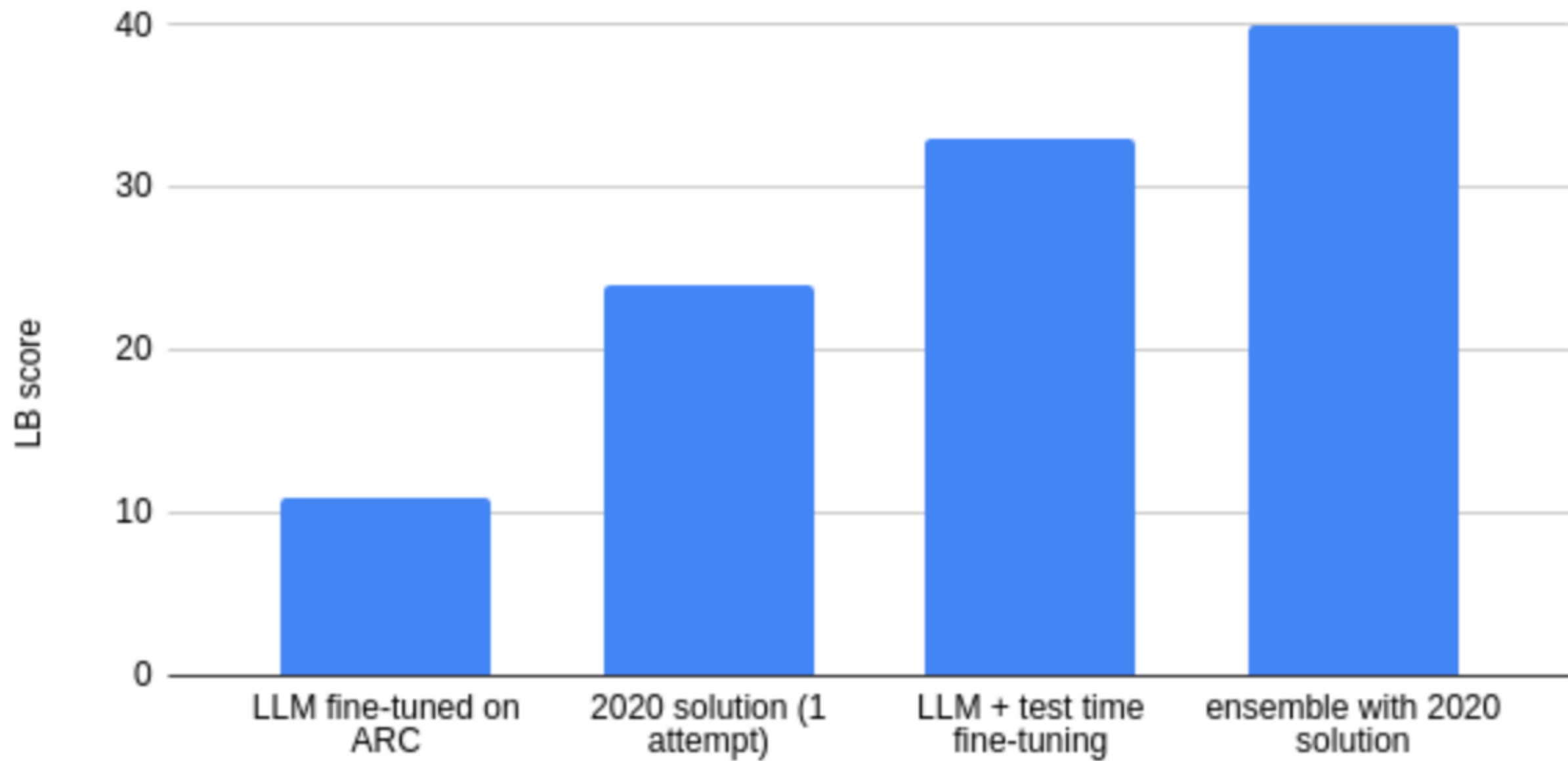
## Data

I used the following publicly available datasets for training:

dataset	number of unique tasks
<a href="#">original ARC dataset</a>	800
Michael Hodel's <a href="#">RE-ARC dataset</a>	400
<a href="#">PQA dataset</a>	7
Simon Strandgaard's <a href="#">Tama dataset</a>	50
<a href="#">Mini-ARC</a>	149
<a href="#">nosound's hand crafted ARC tasks</a>	9
<a href="#">Andy Penrose's tasks</a>	5
TOTAL	1420

# Omni-ARC

- Single Model
  - QWEN 2.5 (non reasoning model)
- Test-time-training (fine-tuning)
  - Data augmentation
  - 1 batch size : train n-1 , test 1 eval sample
  - ~300step
- Ensemble with 2020 solution (ML based)
- Data augmentation for input, output
  - Rotations
  - Flips
  - Color changes
  - Swap between train and test examples
- Problem augmentation
  - Rotations and/or flips
  - Padding the image
  - Upscale
  - Mirror
























# “examples -> code” approach 50%

- Program sythesis
- <https://www.kaggle.com/code/michaelhodel/program-synthesis-starter-notebook>
- Ryan Greenblat generate python code for rule-based solution using GPT-4o

- 

<https://blog.redwoodresearch.org/p/getting-50-sota-on-arc-agi-with-gpt>

# Kaggle ARC Prize 2025 Leaderboard

#	Team	Members		Score	Entries	Last
1	Giotto.ai	   		27.08	183	14h
2	the ARChitects	  		16.94	43	10d
3	MindsAI @ Tufa Labs	    		15.42	168	1d
4	Guillermo Barbadillo			11.94	52	9d
5	rxex			10.42	72	3mo
6	ippeiogawa			10.28	154	12h

<https://www.kaggle.com/competitions/arc-prize-2025/leaderboard>

# ARC 2024 model on ARC 2025

Overview

Data

Code

Models

Discussion

Leaderboard

Rules

Team

Submissions

All

Your Work

Shared With You

Bookmarks

ARC-AGI Without Pretraining

Updated 6mo ago

Score: 4.17 · 7 comments · CompressARC +1

734

Gold

[暗黒AGI]CompressARC

Notebook copied with edits from Isaac Liao · Updated 3mo ago

Score: 5 · 11 comments · DarkAGICompressARC +2

599

Gold

[LB 2.08] baseline from 1st place of 2024

Notebook copied with edits from 暗黒AGI · Updated 6mo ago

Score: 2.08 · 11 comments · unsloth-2024-9-post4 +2

385

Gold

ARC-AGI 2025: Visualization all 1000+120 tasks

Notebook copied with edits from Oleg X · Updated 5mo ago

10 comments · ARC Prize 2025

251

Gold

baseline from 1st place of 2024

Notebook copied with edits from Daniel Franzen · Updated 6mo ago

Score: 1.67 · 19 comments · unsloth-2024-9-post4 +2

248

Gold

<https://www.kaggle.com/code/octaviograu/lb-2-08-baseline-from-1st-place-of-2024>



# ARC 2025 notebooks



JOCELYN DUMLAO · 4MO AGO · 5,453 VIEWS



233

Copy & Edit

209



## Solving ARC Prize 2025 with Qwen-3 Transformer

Notebook

Input

Output

Logs

Comments (19)



Competition Notebook

ARC Prize 2025



Version 4 of 4



Runtime



1m 15s · GPU P100



<https://www.kaggle.com/code/jocelyndumlao/solving-arc-prize-2025-with-qwen-3-transformer>

# QWEN3 (reasoning model) chatbot

## Solving ARC Prize 2025 with Qwen-3 Transformer

233

Copy & Edit 209



Notebook Input Output Logs Comments (19)

```
# Init chatbot
chatbot = QwenChatbot(model, tokenizer)

# Challenge solving function
def solve_arc_challenge(challenge, chatbot):
    predictions = []

    # Prebuild train prompt
    train_prompt = "Solve the following abstract reasoning challenge:\n\n"
    for i, example in enumerate(challenge['train']):
        train_prompt += f"Input {i + 1}:\n{example['input']}\nOutput {i + 1}:\n{example['output']}\n\n"

    for test_input in challenge['test']:
        prompt = train_prompt + f"Now, predict the output for the following test input:\n{test_input['input']}\nOutput:\n"
```

### Table of Contents

- Import Libraries
  - Reference: Qwen 3 qwen-lm/qwen-
- Define paths to the datasets
- Load the Data
- Choose a Model
- Set Up the Hardware
- Load the Model and Tokenizer
- Create a Chatbot Class
- Demonstration and Task Integration
- Solve Challenges - ARC dataset
- Submission