



Politecnico  
di Torino

INRAE

# Towards Artificial General Intelligence Through Evolutionary Algorithms

Experimental Investigations on the ARC-AGI Benchmark

---

Master of Artificial Intelligence and Data Analytics - Politecnico di Torino

**Supervisors:**

Giovanni Squillero  
Alberto Tonda

**Candidate:**

Riccardo Daniele Turco  
student id: 328946

**The goal of this thesis is try  
to achieve AGI through the  
use of evolutionary  
algorithms.**



# Agenda

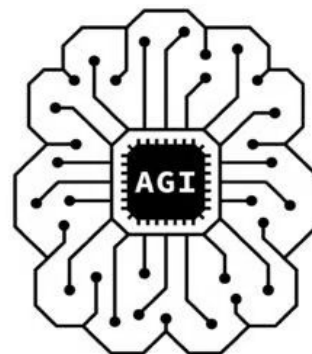
- General introduction
- Design of ARC - AGI 1 benchmark
- ARC Prize competition
- Solutions for ARC - AGI
- The ARC-AGI 2 Benchmark
- The ARC-AGI 3 Preview
- Benchmark performance
- Proposed solution
- Results
- Limitations and future work

# General introduction



## Artificial General Intelligence AGI

Aims at solving problems across diverse domains.



## ARC Benchmarks

Evaluate abstract reasoning and cognitive capabilities of machines.



## Evolutionary Algorithms EAs

Methods inspired by genetic processes and natural selection to solve search and optimization problems.

# Design of ARC - AGI 1 benchmark

- ~400 tasks divided into training, evaluation, and hidden test sets.
- Each task provides only 2-5 examples.
- **Core operations:**
  - Pattern recognition
  - Visual analogy
  - Geometric transformations
- **Goal:** assess generalization beyond memorization.

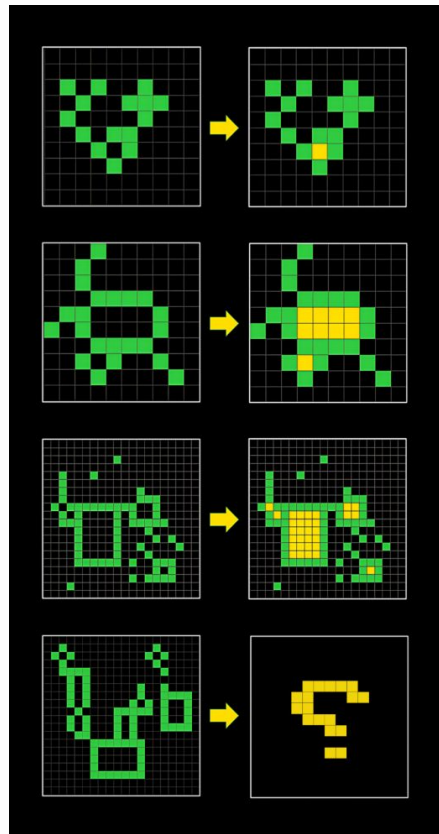


Image 1: Example of a task from the first benchmark.

# ARC Prize competitions



Exposed limitations of deep learning in general reasoning.



Showed the potential of alternative approaches:

- Program synthesis
- Neuro-symbolic models
- Evolutionary computation



Highlighted the need for better benchmarks in AI.

# Solutions for ARC - AGI

The most effective approach on the first competition are based on:

**brute-force program synthesis**  
+  
**DSL**

Place	Name	Score (Private evaluation set)
1st	the ARChitects	53.5%
2nd	Guillermo Barbadillo	40%
3rd	alijs	40%
4th	William Wu	37%
5th	PoohAI	37%

Table 1: ARC Prize 2024 Winners.

## OpenAI “o3” model

This system combined:

- LLM-driven program generation.
- Chain-of-thought reasoning.
- Advanced test-time search.

**Obtains the highest score but breach  
of the competition regulations.**

# The ARC-AGI 2 Benchmark

- Launched in March 2025.
- New set of 400 harder hidden tasks.
- Requires:
  - Symbolic Interpretation
  - Compositional Reasoning
  - Contextual Rule Application
- **No system achieved human-level performance.**



Image 2: Example of a task from the second benchmark.



# The ARC-AGI 3 Preview

- Is set to launch in 2026.
- Challenges **Interactive Reasoning**
- Based on **Game Environment**

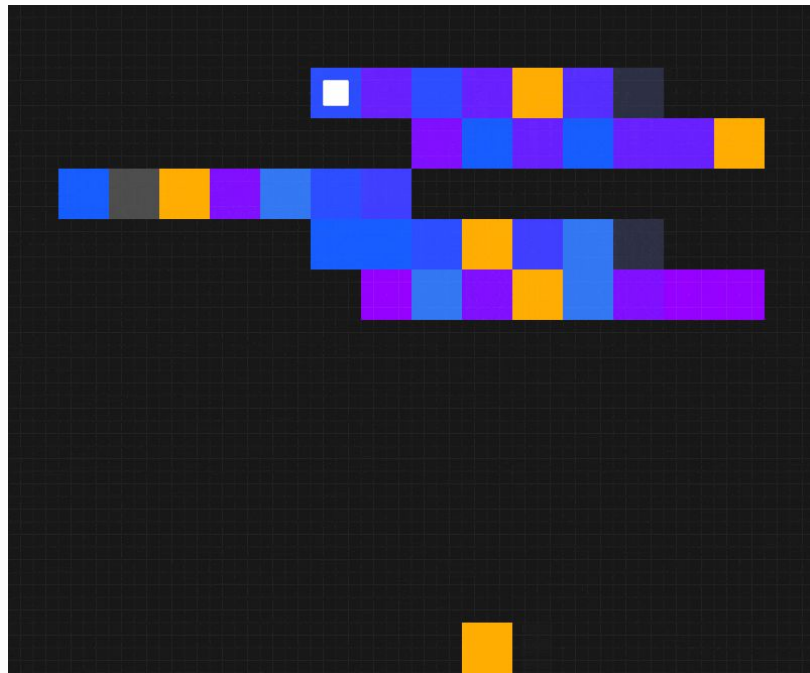


Image 3: Example of a task from the third benchmark.

# Benchmark performance

The drop in performance between the first and second competition shows us the **clear limitation** of these systems **to perform like human**.

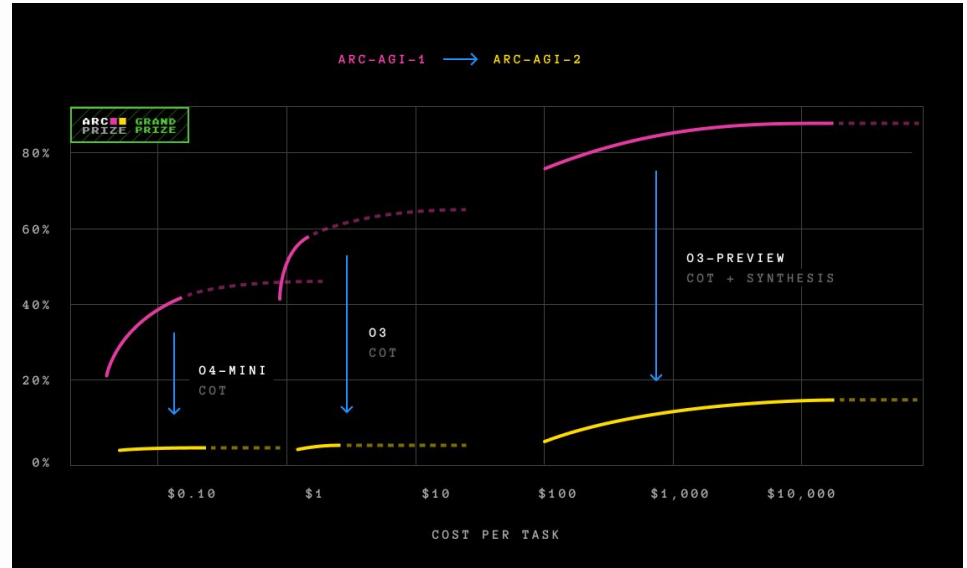


Image 4: Performance drop of openAI's o3 model when moving from the first benchmark to the second benchmark.

# Proposed solution

Based on: **Geometric Representations + Evolutionary Algorithm**

## 1 Geometric Representations

## 2 Initial Analysis

Representation	Description
Pixel	The grid is seen as a list of individual pixels defined by its position (x, y) and color.
Row	The grid is interpreted as a sequence of rows. Each row is treated as a unit enabling actions that operate with it.
Column	The grid is interpreted as a sequence of columns. Each column is treated as a unit enabling actions that operate with it.
Rectangle	The grid is parsed into rectangles characterized by its color, width, height, and absolute position relative to the top-left corner.
Figure	Connected groups of pixels are clustered into figures characterized by a color, a shape pattern and an absolute position.
Color Figure	Extension of Figure Representation, figures can be composed of multiple colors and are further split into inner regions and borders.
First/Second Diagonal	The grid is decomposed into diagonals that can be manipulated as a sequence.
Color Layer (discarded)	Each color in the grid is isolated into a layer, where pixels of the target color are marked and others are ignored.
Border (discarded)	The grid is analyzed to detect and represent open and closed borders formed by contiguous pixels.

Table 2: Geometric representations description

# Proposed solution

## 3 Initial Population Generation

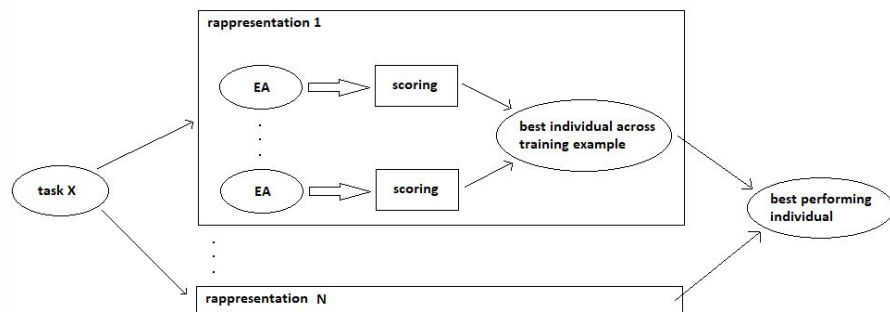


Image 5: Single-Example model

## 4 Evolutionary Algorithm Execution

- Single-Example Evolution
- Multi-Example Evolution

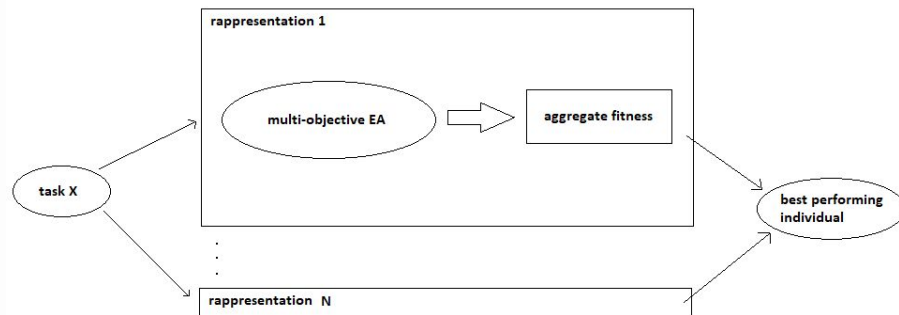


Image 6: Multi-Example model

# Proposed solution

## 5 Mutation Operators

- add,      - tweak,      - swap

## 6 Parent Selection

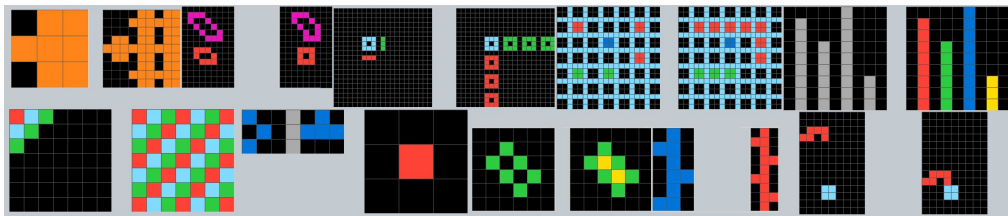
- tournament selection,      - lexicase selection

## 7 Survival Selection

## 8 Solution Generation

# Results

- **Experimental setup:** 10 ARC-AGI tasks, each configuration run 10 times.



- **Single-Example Evolution** → More stable performance
  - Avg. 2/10 tasks solved
  - Best run: 3/10 tasks solved
- **Multi-Example Evolution** → Conceptual advantages, Larger search space
  - Avg. 1/10 tasks solved
  - Best run: 3/10 tasks solved

# conclusioni

## ■ Limitations

- Dependence on Predefined Representation
- fixed parameters
- complexity of the search space

## ■ Future work

- automatic parameter tuning
- new representations
- fitness functions
- hybrid approaches

**THANK YOU**

**FOR YOUR TIME**