Enhancing AI Capabilities on the Abstraction and Reasoning Corpus: A Path Toward Broad Generalization in Intelligence

Soumya Banerjee
University of Cambridge, UK
sb2333@cam.ac.uk

Abstract

This position paper explores advancing artificial intelligence by improving its ability to generalize beyond training data—a key requirement for tasks in the Abstraction and Reasoning Corpus (ARC). Inspired by classic challenges such as the Bongard Problems, ARC tasks push AI toward more flexible, human-like intelligence. We investigate neurosymbolic systems (e.g. DreamCoder) and large language models, while emphasizing the need for diverse data sources and mathematical rigour in designing pipelines for logical reasoning.

Introduction

Modern AI systems excel at narrow, data-driven tasks but often struggle to generalize to novel situations. The ARC benchmark, comprising logic-based visual puzzles Chollet [2019], forces AI to transcend mere pattern matching by requiring abstract reasoning. Drawing inspiration from historical algorithmic challenges, we argue that bridging the gap between human and machine reasoning demands new architectures, data augmentation strategies, and refined evaluation metrics.

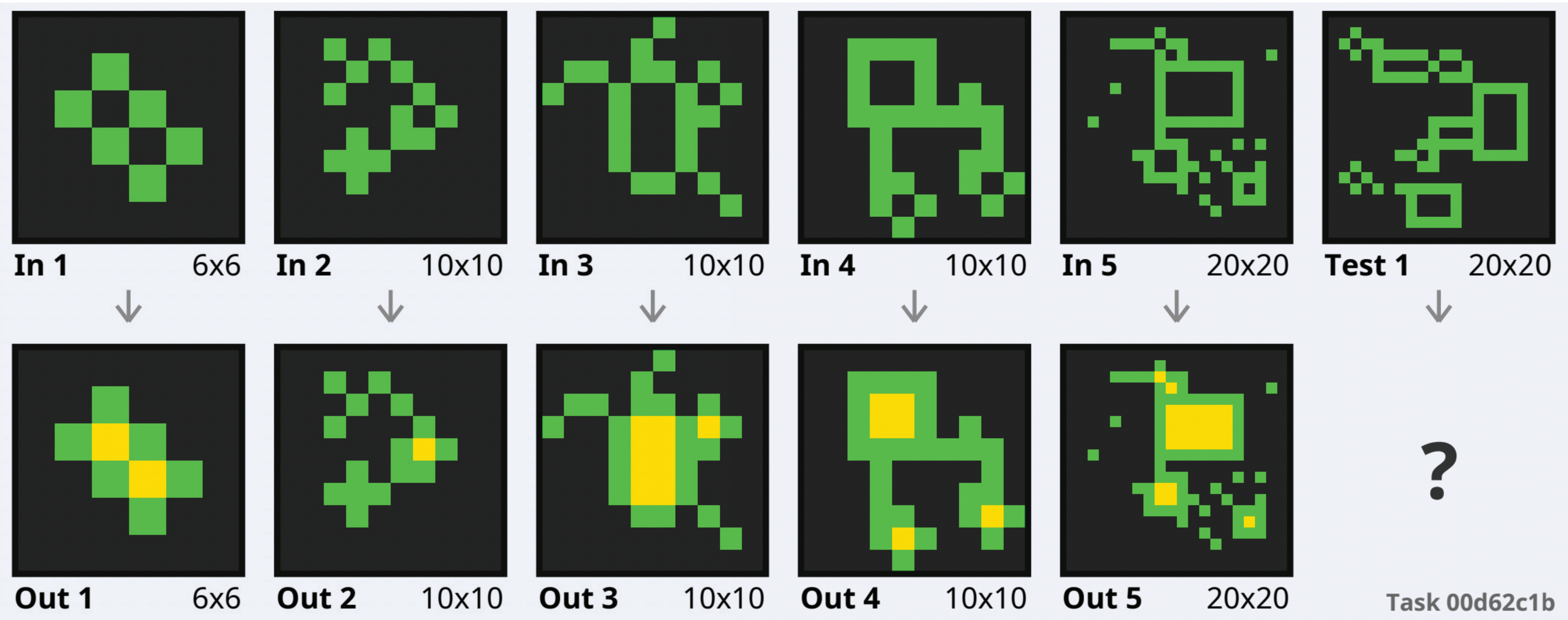

Figure: An illustration of a typical task in ARC. Each task has three training pairs shown as In (input) and Out (output) in the first three columns. The last task (denoted test) needs to be solved (shown with a question mark (?).

Motivation and Background

Human intelligence is adaptive, rooted in embodied experience and a vast reservoir of prior knowledge. In contrast, AI processes input as tokens or pixels without direct sensory experience. This fundamental difference means that intelligence in machines may be qualitatively distinct. Historical challenges such as the Bongard Problems illustrate early attempts to formalize abstract reasoning, while ARC pushes current systems to achieve a similar breadth of generalization.

Current Approaches and Challenges

Existing methods on ARC include neurosymbolic systems like DreamCoder and large language models (LLMs) such as GPT-4. While these models leverage vast data, their reliance on pattern recognition often leads to limitations in logical precision and consistency. We observe that many winning approaches use test-time augmentation and brute-force heuristics, raising concerns about whether they truly capture the essence of reasoning or simply exploit shortcuts.

The Role of Data

Enhancing ARC performance hinges on more diverse data. Controlled trials with human participants can provide insight into natural reasoning strategies, while synthetic data augmentation can expand training sets—much like AlphaGo's use of massive game datasets in its breakthrough. By integrating both real and augmented data, we aim to promote AI systems that generalize robustly and reason abstractly.

Building Math-Inspired Pipelines

To narrow the gap between machine and human reasoning, we propose pipelines that combine neural networks with mathematically inspired structures. Incorporating concepts from symbolic logic, topology, and category theory may yield architectures that support rigorous logical reasoning. Such hybrid systems can benefit from the exploratory power of LLMs while enforcing the precision required for true abstraction.

Complementing Human Problem Solving

Instead of striving for AI systems that autonomously solve all ARC tasks, we advocate a cooperative approach. AI can complement human reasoning by providing hypotheses and pattern insights while humans apply context, intuition, and abstract reasoning to refine these suggestions. Such human-AI collaboration could yield more robust and innovative solutions to complex problems.

Conclusions

ARC serves as a critical benchmark for developing AI systems capable of abstraction and logical reasoning. Our work underscores that intelligence in machines and humans is fundamentally different Holm and Banerjee [2024], and a unified theory may require redefining key concepts such as “reasoning” and “understanding”. By combining diverse data, math-inspired neural architectures, and collaborative human-AI strategies, we pave the way for broader generalization and deeper insights into machine intelligence.

Future Directions and Impact

Looking ahead, integrating symbolic reasoning with LLMs and other deep learning methods offers a promising path toward achieving AI that can reason as flexibly as humans Bober-Irizar and Banerjee [2024]. This hybrid approach not only advances ARC performance but may also inform broader AI research, with potential applications in scientific discovery, complex systems analysis, and beyond.

References

Mikel Bober-Irizar and Soumya Banerjee. Neural networks for abstraction and reasoning. *Scientific Reports*, 14:1–20, 11 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-73582-7. URL https://www.nature.com/articles/s41598-024-73582-7.

François Chollet. On the measure of intelligence. 11 2019. doi: 10.48550/arxiv.1911.01547. URL https://arxiv.org/abs/1911.01547v2.

Halfdan Holm and Soumya Banerjee. Intelligence in animals, humans and machines: a heliocentric view of intelligence? *AI and Society*, pages 1–3, 4 2024. ISSN 14355655. doi: 10.1007/S00146-024-01931-1/METRICS. URL https://link.springer.com/article/10.1007/s00146-024-01931-1.

For more details, please visit: https://osf.io/preprints/osf/3jq4t_v1

Acknowledgements: We thank the Accelerate Programme for Scientific Discovery for funding this work.