



ARC - Reasoning with Language and Vision

The Abstraction and Reasoning Corpus (ARC) is like an IQ test for machine intelligence, allowing us to quantify progress towards human-level AI. It was introduced in 2019 by François Chollet as a [Kaggle](#) competition. Rather than focusing on a single specific challenge, the aim of ARC is to provide an ensemble of challenges to measure the ability to handle and adapt to new unseen environments. One task of the ARC usually consists of three given samples with input and output pairs. For a fourth sample, only the input is given and the output must be produced. These challenges seem easy for humans to solve, however machines, at least for now, struggle with them a lot. The implicit assumption is that we humans can rely on existing "priors" and knowledge that help us tackle these tasks. Therefore, a central aspect of ARC is to make these priors more explicit and try to incorporate them into learning systems.

One way to include existing priors and knowledge to mimic human intuition is to leverage the expertise of pretrained models, i.e. large language models. We have already developed a promising approach which uses language as an intermediate reasoning step and can solve ARC tasks. The whole pipeline consists of multiple parts, combining the vision and language domain. The overarching goal is to create a system, which can learn and tackle the ARC challenge in an end-to-end fashion. The exact scope of the thesis is up to discussion. Amongst some possible directions, we envision to develop a trainable vision captioning module and further boost the performance of the language models, i.e. by finetuning them to ARC.

Other inputs or directions are welcomed. If you have your own ideas which are similar or just interested in ARC in general we can set up a meeting to discuss a possible thesis.

Requirements: Strong motivation, knowledge in deep learning, or a solid background in machine learning, Python and libraries such as TensorFlow or PyTorch. Expertise in Computer Vision and/or previous experiences or projects with (large) language models is an advantage. We will have weekly meetings to discuss open questions and determine the next steps.

Interested? Please contact us for more details!

Contact

- Joël Mathys: jmathys@ethz.ch, ETZ G63
- Benjamin Estermann: besterma@ethz.ch, ETZ G60.1

