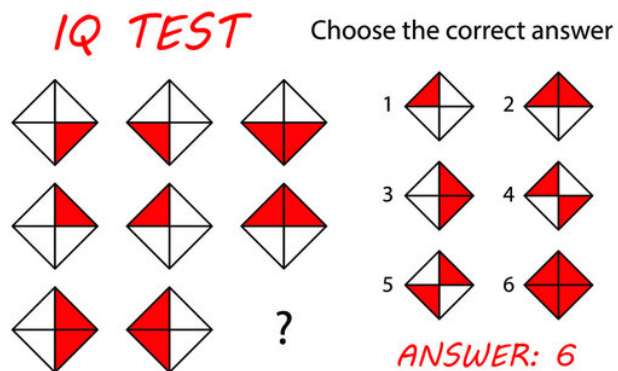Prof. R. Wattenhofer

IQ of VLMs – can they really reason?

How intelligent is ChatGPT?

Language and vision models have shown themselves to be excellent at many more general tasks over the past few years. Results have even demonstrated that LLMs can solve tasks at the levels of graduate students and reason about problems. Additionally, we see newer models being capable of handling multi-model contexts.

A natural question arises from all this: What IQ would models like ChatGPT, Claude, and Gemini score on visual reasoning IQ tests?

In this project, we will work to answer both how well many of the popular multi-modal LLMs can handle visual IQ tests and, more importantly, whether the models are actually able to reason about the answer or have memorized the solutions.

Requirements

Good programming skills in Python and knowledge of machine learning evaluation are required. Additionally, having some creative skills as you will be drawing IQ tests will be beneficial. Some light web development skills are also helpful.

We will have weekly meetings to address questions, discuss progress, and think about future ideas.

Contact

In a few short sentences, please explain why you are interested in the project and about your coding and machine learning background (i.e., your projects or relevant courses you have taken at ETH or elsewhere).

- Andreas Plesner: aplesner@ethz.ch, ETZ G95

- Frédéric Berdoz: fberdoz@ethz.ch, ETZ G60.1