Metacognition and critical thinking. Instance-level demand scales and annotation by LLMs

July 2024 · Kexin Jiang Chen, José Hernández Orallo, Fernando Martínez Plumed

---

DESCRIPTION

The rapidly growing field of artificial intelligence has led to the development of advanced Large Language Models (LLMs) with impressive language skills. However, it is still unclear the extent to which these models possess metacognitive abilities, which are critical for advanced reasoning and learning. This study aims to evaluate metacognition and critical thinking abilities in LLMs, with a focus on identifying the most effective scales and dimensions for assessment. We propose a comprehensive framework, encompassing three key dimensions: the need for critical thinking processes, the difficulty of calibrating knowns and unknowns, and the difficulty in identifying relevant information. This framework is used to annotate question instances across several benchmarks from BIG-Bench and HELM, aimed at measuring advanced cognitive skills in LLMs. The annotations are generated by the state-of-the-art LLM, GPT-4. These annotations are then used as predictors to build performance models for various LLMs on these benchmarks, with the ultimate goal of determining the extent to which the benchmarks truly measure metacognitive capabilities. Our findings reveal that while many models lack metacognitive capabilities, larger models exhibit some indications of such abilities. Furthermore, the use of a multi-dimensional scale for metacognitive demands improves the predictability compared to a single integrated scale. By providing an evaluation tool for metacognition in LLMs, this study provides insights into the effectiveness of benchmarks in assessing metacognitive abilities. The finding highlights the importance of careful benchmark design and the potential of multi-dimensional scales in capturing the complex nature of metacognition.

LINKS TO MATERIALS

- Final Degree Thesis (pdf)
- Labeling with Anchors (pdf)