Course Name: Evaluation of LLMs

Dates: 4.-15.9.23, Mo-Fr 10-12 and 13-15

Location: Kalahari

Lecturer: Leonie Weissweiler

Course Plan

As the performance of large language models such as GPT4 improves, it becomes increasingly important to understand how they work, and to develop targeted testing methods to isolate areas in which they don't, in order to gain further understanding both of language models and language. This class will cover the major recent works in interpretability and probing, while also laying a focus on critical thinking about evaluation techniques and language.

Course Format

The course will begin with an introductory lecture about current methodology in probing and evaluation of Large Language Models. There will then be an interactive phase in which participants will decide on either separate or a joint project, which will consist of a small, but novel experiment for targeted evaluation of an LLM. The rest of the time will be spent in realising this project under supervision, and presenting and discussing the results.

If you're interested in participating in this course, please email me at weissweiler@cis.lmu.de