

MSc thesis or guided research project

Känguru: LLM mathematical reasoning challenge

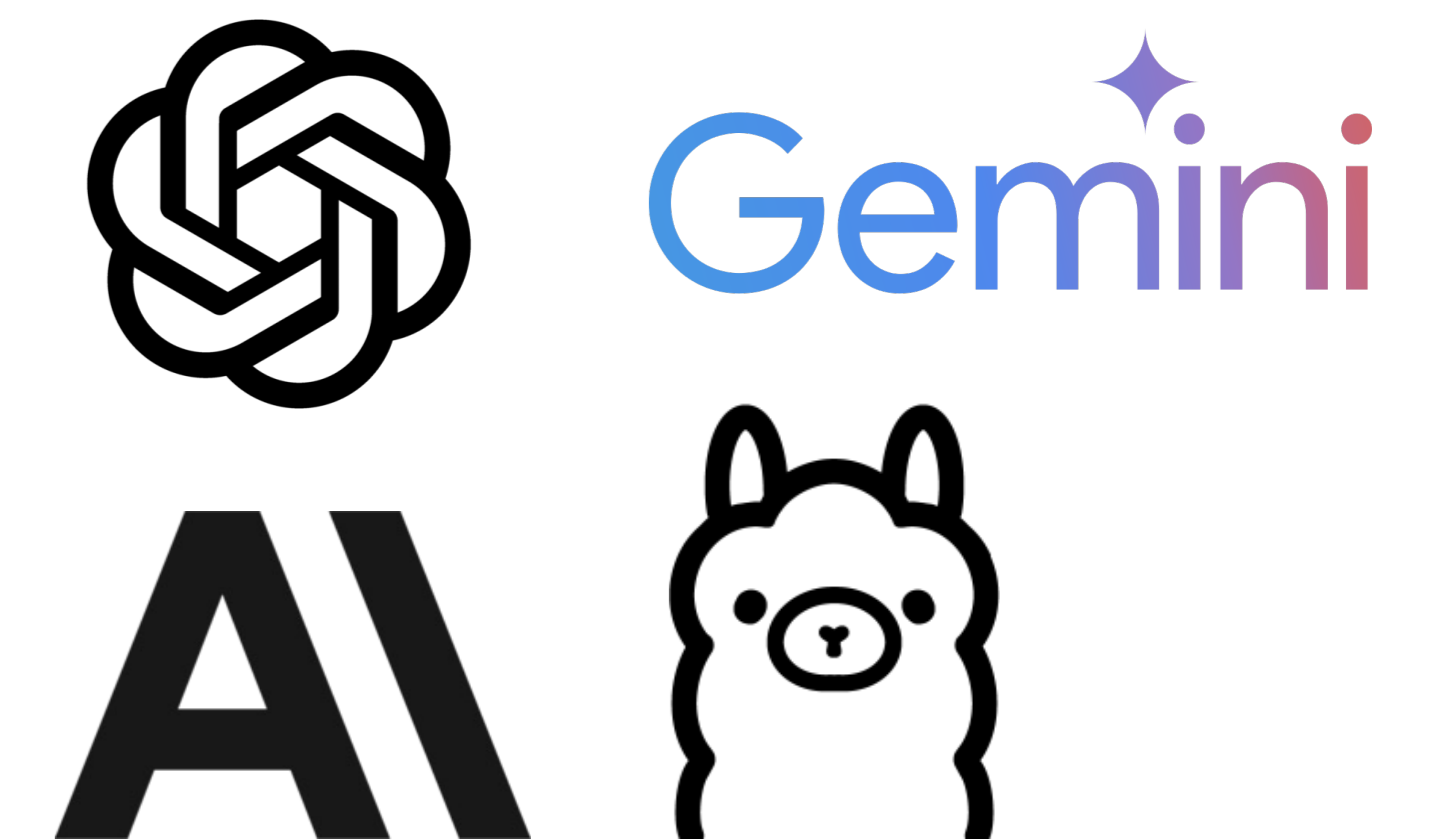
What



- ▶ “Känguru der Mathematik” is a mathematical reasoning test taken by almost a million German students each year
- ▶ We have all exercises for all age groups from 1998 until 2024 and their solutions (3,900 multiple choice exercises)
- ▶ We have the results of German students from 2001 onwards

Work packages

- ▶ Process existing pdfs to extract the exercises (as text), illustrations (as images), and solutions as well as previous results
- ▶ Release a public mathematical reasoning benchmark dataset (in German; but other languages are available)
- ▶ Evaluate commercial and open LLMs on the benchmark dataset (for different input modalities)
- ▶ Exploratory data analysis and statistical testing



Scientific questions

- ▶ Historic development of students’ mathematical reasoning skills in Germany over the years (normalized by LLM)
- ▶ Exploring differences in which types of problems are easy/hard to humans/LLMs respectively (at different ages)
- ▶ Potential insights into successful didactics of mathematics



Expectations

- ▶ Having (some) experience with LLM APIs and being keen on the required engineering involved
- ▶ Being creative and comfortable in finding the best tools for the job (especially for pre-processing)
- ▶ Being able to independently formulate and quantitatively verify scientific hypothesis
- ▶ Ideally, having had prior exposure to HPC environments (SLURM managed Linux cluster) using (multiple) GPUs

Interested?

Please send

- ▶ CV,
- ▶ current and previous transcript(s), and
- ▶ a short paragraph describing why you are interested in the project

to niki.kilbertus@tum.de



Apply here!

