



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND SPRACHVERARBEITUNG
STUDIENGANG COMPUTERLINGUISTIK



Thesis proposal

Topic: Hate speech detection data generation for a low-resource language

Supervisor: Axel Wisiosek, Haotian Ye

Examiner: Hinrich Schütze

Level: BSc / MSc

Summary: Current LLMs heavily rely on data both for pre-training and fine-tuning to achieve good performance. However, data is only available to a limited extent for a large part of the world's languages. On the other hand, multilingual LMs often suffer from language imbalance, which is shown by performance gaps between high- and low-resource languages. The goal of the project is to collect low-resource language data for the task of hate speech detection and evaluate whether LLMs are capable of performing the task of similar quality as for high-resource languages, e.g. English. Possible experiments include:

- data generation (positive/neutral/hateful statements) and quality evaluation
- zero-shot/few-shot hate speech detection
- few-shot in-context learning
- compare instruction-tuned models (ChatGPT, Alpaca, etc.) regarding their restrictions in processing hateful/offensive content

Requirements: good programming/data processing skills, knowledge of the Transformers library.