LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND SPRACHVERARBEITUNG
STUDIENGANG COMPUTERLINGUISTIK

Thesis proposal

Topic: **Translation-Induced Framing Shifts in LLM Bias Evaluations**

Supervisor: Molly Kennedy

Examiner: Hinrich Schütze

Level: MSc

Summary: The student will measure how LLM bias/framing judgments change when evaluating (i) original non-English news vs (ii) machine-translated variants, across multiple translation conditions (systems, directions, and prompt-language choices). They will analyze which linguistic phenomena correlate with judgment flips (e.g., modality/hedging, named-entity handling, sentiment-bearing terms) and run controlled ablations and statistical comparisons across several models and languages.

References:

- Kai Hartung et al. (2023). "Measuring sentiment bias in machine translation". In: *International Conference on Text, Speech, and Dialogue*. Springer, pp. 82–93
- Yafu Li et al. (2025). "Lost in Literalism: How Supervised Training Shapes Translationese in LLMs". In: *arXiv preprint arXiv:2503.04369*
- Jun Wang, Benjamin Rubinstein, and Trevor Cohn (2022). "Measuring and mitigating name biases in neural machine translation". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2576–2590