LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
CENTRUM FÜR INFORMATIONS- UND SPRACHVERARBEITUNG
STUDIENGANG COMPUTERLINGUISTIK

# Thesis proposal

**Topic:** **Relations in the Wild: Collecting Contextual Instances of Relation Triples for the Detection of Latent Representations**

**Supervisor:** Lea Hirlimann

**Examiner:** Hinrich Schuetze

**Level:** BSc / MSc

**Summary:** This project would address a limitation in our current knowledge representation research: the reliance on manually crafted prompt templates to probe factual knowledge in language models. This previous work uses one or two fixed templates (e.g., `"Apple's CEO is? Answer:"`, `"The CEO of Apple is? Answer:"`) (Liu et al., 2025) to elicit relations, but this approach may artificially constrain our understanding of how models internally represent relational knowledge. This project includes:

- creating a dataset mapping individual facts to their varied natural expression, by systematically collecting naturally occurring instances of relation triples from large text corpora. For a given relation (e.g., person_mother, ceo_company), the thesis would gather diverse **real-world expressions** of the same factual triple across different contexts, syntactic constructions, and linguistic registers.
- investigating whether language models maintain **unified internal representations** for the same relation regardless of surface expression and formulation. The hypothesis would be that if the neurons detected by the usage of only templates are an accurate representation of a certain relation, the ablation of those neurons should also affect the relational knowledge in natural occurring formulations.
- (optional/MSc) investigating potential **prototypical formulations** for the relation within the diverse ways a relation can be expressed.

This could reveal whether prompt template choices in prior work have systematically biased our understanding of relation encoding and whether models internally privilege certain linguistic constructions. The findings have implications for knowledge editing, fact-checking systems, and our theoretical understanding of how linguistic variability relates to semantic constancy in neural language models.

**Requirements:** motivation and curiosity, interest in interpretability, understanding of transformer/MLP-layer architecture (recommended)

**References:**

- Yihong Liu et al. "On Relation-Specific Neurons in Large Language Models". In: *The 2025 Conference on Empirical Methods in Natural Language Processing*. 2025. URL: https://openreview.net/forum?id=BPLghxmcM2