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

# Thesis proposal

**Topic:** **Visualizing Spatial Understanding of Textual Descriptions in Language Models**

**Supervisor:** Lea Hirlimann

**Examiner:** Hinrich Schuetze

**Level:** MSc

**Summary:** This thesis investigates how language models internally represent spatial information when processing textual scene descriptions. The project proposes using frameworks such as ALFWorld, a text-based interactive environment with corresponding 3D virtual rooms from the ALFRED Benchmark, as a testbed to probe whether language models develop coherent spatial representations from language alone. This project would involve:

- assembling **training data** of pairs of textual descriptions of scenes (e.g., `"the apple is on the kitchen counter to the left of the microwave"`) and maps of the virtual rooms / coordinates of the individual objects.
- training **linear probes** on the internal activations of language models as they process textual descriptions. These probes would attempt to decode spatial coordinates or relative positions of objects, effectively projecting high-dimensional representations onto 2D or 3D spatial maps. Success would indicate that the model maintains some form of implicit spatial structure in its representations.
- (optional) An important comparison could be examining how text-only language models differ from *vision-language models (VLMs)* in their spatial representations. VLMs have direct visual grounding, so comparing their internal spatial structure to text-only models could reveal whether linguistic spatial understanding converges toward similar geometric organizations or follows fundamentally different representational strategies. This comparison could illuminate questions about the necessity of perceptual grounding for spatial cognition and whether language alone provides sufficient signal for geometric reasoning.

The project contributes to understanding whether spatial reasoning in LMs is merely linguistic pattern matching or involves some form of implicit spatial modeling, with implications for embodied AI, human-robot interaction, and theories of grounded cognition in artificial systems.

**Requirements:** interest & motivation, enthusiasm to try work with public repositories, previous experience with `torch` & `huggingface` (recommended)

**References:**


- Wes Gurnee and Max Tegmark (2024). "Language Models Represent Space and Time". In: *The Twelfth International Conference on Learning Representations*
- HV AlquBoj et al. (2025). "Number Representations in LLMs: A Computational Parallel to Human Perception". In: *CoRR*
- Mohit Shridhar et al. (2020). "Alfworld: Aligning text and embodied environments for interactive learning". In: *arXiv preprint arXiv:2010.03768*