Exploring Non-English Gender Bias in NLP Models

- **Supervisor**: Victor Steinborn
- **Examiner**: Prof. Hinrich Schütze
- **BSc, MSc, Open:** MSc/BSc
- **Summary:** Large pre-trained language models, such as BERT, have gained significant exposure and radically changed NLP research. Despite these successes, these models learned to reproduce social stereotypes from the training data, including gender stereotypes. Gender bias research in NLP attempts to combat these biases by identifying, measuring and mitigating them. However, the vast majority of these methods are designed for English. In this project we focus on identifying and measuring gender bias in other languages. The objective is to follow the technique of (Kaneko; 2022) and measure gender bias in Masked Language Models (MLMs), but where we use gendered attribute words in non-English languages (as opposed to English in the original approach) and compare differences in results.
- **Prerequisites:** Enthusiasm (potential for publishing results of thesis), independent work ethic, good programming background in python, knowledge of one non-English language, basic knowledge
of NLP, DL and Pytorch