



Thesis proposal

Topic: Developing a benchmark for multilingual gender-neutral generalization with neopronouns

Supervisor: Leonor Veloso

Examiner: Prof. Hinrich Schütze

Level: BSc

Summary: The inclusion of gender-neutral language and neopronouns in languages with marked gender in data, as well as how LLMs currently handle and predict them, is still an understudied topic. Recent years have seen the development of gender-inclusive datasets and models for languages such as German (Amrhein et al. 2023; Lardelli, Lauscher, and Attanasio 2024), Portuguese (Veloso, Coheur, and Ribeiro 2023), and Dutch (Boven, Du, and Nguyen 2024).

Regarding the evaluation of LLMs regarding neopronouns, Bunzeck and Zarriß (2024) have created the first benchmark dataset designed to evaluate language models' ability to handle English neopronouns, for the question-answering (QA) task.

This project entails a review of the currently existing literature with the purpose to curate a multilingual gender-inclusive dataset that contains language-specific neopronouns. In the subsequent step, that dataset can be used to follow the approach developed by Bunzeck and Zarriß (2024) in order to study the performance of multilingual LLMs in the gender-inclusive QA task across gendered languages.

Requirements:

- Enthusiasm!
- Basic knowledge of Python and data handling-related libraries (like pandas and numpy)

References:

- Chantal Amrhein et al. (2023). "Exploiting biased models to de-bias text: A gender-fair rewriting model". In: *arXiv preprint arXiv:2305.11140*
- Manuel Lardelli, Anne Lauscher, and Giuseppe Attanasio (2024). "GeFMT: Gender-Fair Language in German Machine Translation". In: *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2)*, pp. 37–38
- Leonor Veloso, Luísa Coheur, and Rui Ribeiro (2023). "A rewriting approach for gender inclusivity in Portuguese". In: *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 8747–8759
- Bastian Bunzeck and Sina Zarriß (2024). "The SlayQA benchmark of social reasoning: testing gender-inclusive generalization with neopronouns". In: *Proceedings of the 2nd GenBench Workshop on Generalisation (Benchmarking) in NLP*, pp. 42–53
- Goya van Boven, Yupei Du, and Dong Nguyen (2024). "Transforming Dutch: Debiasing Dutch Coreference Resolution Systems for Non-binary Pronouns". In: *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pp. 2470–2483