LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND SPRACHVERARBEITUNG
STUDIENGANG COMPUTERLINGUISTIK

Thesis proposal

Topic:	**Investigating Media Bias in Language Models**
Supervisor:	Molly Kennedy
Examiner:	Hinrich Schütze
Level:	BSc

Summary: As language models increasingly shape user experiences in digital communication, understanding and addressing biases in these models is essential. While much research focuses on gender or occupational biases Bolukbasi et al. 2016, this project will explore media bias, examining how political, ideological, or cultural slants in training data influence model outputs Bender et al. 2021.

Aims:

- Data Collection: Gather datasets from various sources reflecting different media biases (e.g., political leaning or regional perspectives) Kulshrestha et al. 2017.

- Bias Analysis: Examine how training data influences language model outputs, focusing on media bias.

- Linguistic Features: Identify specific words or phrases associated with biased outputs Gentzkow and Shapiro 2010.

- Evaluation and Mitigation: Propose strategies for reducing bias using data balancing or prompt engineering Sheng et al. 2021.

Requirements: Programming skills in Python, Machine Learning Basics, Natural Language Processing (NLP) knowledge, Data Handling (Pandas, Numpy), Bias Measurement Techniques, Data Science and Statistics.

References:

- Tolga Bolukbasi et al. (2016). "Man is to computer programmer as woman is to homemaker? debiasing word embeddings". In: *Advances in neural information processing systems* 29

- Emily M Bender et al. (2021). "On the dangers of stochastic parrots: Can language models be too big?" In: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 610–623

- Juhi Kulshrestha et al. (2017). "Quantifying search bias: Investigating sources of bias for political searches in social media". In: *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pp. 417–432

- Matthew Gentzkow and Jesse M Shapiro (2010). "What drives media slant? Evidence from US daily newspapers". In: *Econometrica* 78.1, pp. 35–71

- Emily Sheng et al. (2021). "Societal biases in language generation: Progress and challenges". In: *arXiv preprint arXiv:2105.04054*