Jonathan Karr | Discussion Leader Cover Letter

Chen, Y., Mahoney, C., Grasso, I., Wali, E., Matthews, A., Middleton, T., ... & Matthews, J. (2021, July). Gender Bias and Under-Representation in Natural Language Processing Across Human Languages. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 24-34).

This paper examines gender bias and under-representation in Natural Language Processing (NLP) systems, focusing on word embeddings trained on Wikipedia corpora across nine languages. The authors aim to extend gender bias measurement methods beyond English and identify structural gaps in NLP pipelines. Although they don't explicitly have a research question, they "extend an influential method for computing gender bias from Bolukbasi et al. to eight additional languages, modifying and translating the original defining sets and profession sets to include languages with grammatically gendered nouns".

The study faces several challenges, including grammatical differences across languages, inconsistent NLP tool support, and disparities in available corpora. The authors claim that gender bias exists in word embeddings across multiple languages, that NLP pipelines disproportionately underrepresent most human languages, and that adapting bias measurement techniques beyond English reveals systemic inequalities.

To support these claims, they analyze Wikipedia corpora using word vector analysis with Word2Vec and Principal Component Analysis (PCA) to define gender direction and weighted averaging techniques for grammatical gender differences.

While limited by its reliance on Wikipedia and its focus on nine languages, the study highlights critical gaps in NLP infrastructure. The findings emphasize the need for more inclusive NLP models to mitigate gender bias and ensure fair representation in AI-driven decision-making.