

6th Workshop on Gender Bias in Natural Language Processing (GeBNLP 2025)

Vienna, Austria
1 August 2025

ISBN: 979-8-3313-2408-7

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

<i>JBBQ: Japanese Bias Benchmark for Analyzing Social Biases in Large Language Models</i> Hitomi Yanaka, Namgi Han, Ryoma Kumon, Lu Jie, Masashi Takeshita, Ryo Sekizawa, Taisei Katô and Hiromi Arai	1
<i>Intersectional Bias in Japanese Large Language Models from a Contextualized Perspective</i> Hitomi Yanaka, Xinqi He, Lu Jie, Namgi Han, Sunjin Oh, Ryoma Kumon, Yuma Matsuoka, Kazuhiko Watabe and Yuko Itatsu	18
<i>Detecting Bias and Intersectional Bias in Italian Word Embeddings and Language Models</i> Alexandre Puttick and Mascha Kurpicz-Briki	33
<i>Power(ful) Associations: Rethinking “Stereotype” for NLP</i> Hannah Devinney	52
<i>Introducing MARB — A Dataset for Studying the Social Dimensions of Reporting Bias in Language Models</i> Tom Södahl Bladsjö and Ricardo Muñoz Sánchez	59
<i>Gender Bias in Nepali-English Machine Translation: A Comparison of LLMs and Existing MT Systems</i> Supriya Khadka and Bijayan Bhattarai	75
<i>Mind the Gap: Gender-based Differences in Occupational Embeddings</i> Olga Kononykhina, Anna-Carolina Haensch and Frauke Kreuter	83
<i>Assessing the Reliability of LLMs Annotations in the Context of Demographic Bias and Model Explanation</i> Hadi Mohammadi, Tina Shahedi, Pablo Mosteiro, Massimo Poesio, Ayoub Bagheri and Anastasia Giachanou	92
<i>WoNBias: A Dataset for Classifying Bias & Prejudice Against Women in Bengali Text</i> Md. Raisul Islam Aupi, Nishat Tafannum, Md. Shahidur Rahman, Kh Mahmudul Hassan and Naimur Rahman	105
<i>Strengths and Limitations of Word-Based Task Explainability in Vision Language Models: a Case Study on Biological Sex Biases in the Medical Domain</i> Lorenzo Bertolini, Valentin Comte, Victoria Ruiz-Serra, Lia Orfei and Mario Ceresa	111
<i>Wanted: Personalised Bias Warnings for Gender Bias in Language Models</i> Chiara Di Bonaventura, Michelle Nwachukwu and Maria Stoica	124
<i>GG-BBQ: German Gender Bias Benchmark for Question Answering</i> Shalaka Satheesh, Katrin Klug, Katharina Beckh, Héctor Allende-Cid, Sebastian Houben and Teena Hassan	137
<i>Tag-First: Mitigating Distributional Bias in Synthetic User Profiles through Controlled Attribute Generation</i> Ismael Garrido-Muñoz, Arturo Montejo-Ráez and Fernando Martínez Santiago	149
<i>Characterizing non-binary French: A first step towards debiasing gender inference</i> Marie Flesch and Heather Burnett	160
<i>Can Explicit Gender Information Improve Zero-Shot Machine Translation?</i> Van-Hien Tran, Huy Hien Vu, Hideki Tanaka and Masao Utiyama	171

<i>Colombian Waitresses y Jueces canadienses: Gender and Country Biases in Occupation Recommendations from LLMs</i>	
Elisa Forcada Rodríguez, Olatz Perez-de-Vinaspre, Jon Ander Campos, Dietrich Klakow and Vagrant Gautam	182
<i>Bias Attribution in Filipino Language Models: Extending a Bias Interpretability Metric for Application on Agglutinative Languages</i>	
Lance Calvin Lim Gamboa, Yue Feng and Mark G. Lee	195
<i>Surface Fairness, Deep Bias: A Comparative Study of Bias in Language Models</i>	
Aleksandra Sorokovikova, Pavel Chizhov, Iuliia Eremenko and Ivan P. Yamshchikov	206
<i>Measuring Gender Bias in Language Models in Farsi</i>	
Hamidreza Saffari, Mohammadamin Shafiei, Donya Rooein and Debora Nozza	228
<i>A Diachronic Analysis of Human and Model Predictions on Audience Gender in How-to Guides</i>	
Nicola Fanton, Sidharth Ranjan, Titus Von Der Malsburg and Michael Roth	242
<i>ArGAN: Arabic Gender, Ability, and Nationality Dataset for Evaluating Biases in Large Language Models</i>	
Ranwa Aly, Yara Allam, Rana Gaber and Christine Basta	256
<i>Assessing Gender Bias of Pretrained Bangla Language Models in STEM and SHAPE Fields</i>	
Noor Mairukh Khan Arnob, Saiyara Mahmud and Azmine Tousehik Wasi	268
<i>One Size Fits None: Rethinking Fairness in Medical AI</i>	
Roland Roller, Michael Hahn, Ajay Madhavan Ravichandran, Bilgin Osmanodja, Florian Oetke, Zeineb Sassi, Aljoscha Burchardt, Klaus Netter, Klemens Budde, Anne Herrmann, Tobias Strapatsas, Peter Dabrock and Sebastian Möller	282
<i>From Measurement to Mitigation: Exploring the Transferability of Debiasing Approaches to Gender Bias in Maltese Language Models</i>	
Melanie Galea and Claudia Borg	290
<i>GENDEROUS: Machine Translation and Cross-Linguistic Evaluation of a Gender-Ambiguous Dataset</i>	
Janiča Hackenbuchner, Joke Daems and Eleni Gkovedarou	302
<i>Fine-Tuning vs Prompting Techniques for Gender-Fair Rewriting of Machine Translations</i>	
Paolo Mainardi, Federico Garcea and Alberto Barrón-Cedeño	320
<i>Some Myths About Bias: A Queer Studies Reading Of Gender Bias In NLP</i>	
Filipa Calado	338
<i>GenWriter: Reducing Gender Cues in Biographies through Text Rewriting</i>	
Shweta Soundararajan and Sarah Jane Delany	347
<i>Examining the Cultural Encoding of Gender Bias in LLMs for Low-Resourced African Languages</i>	
Abigail Oppong, Hellina Hailu Nigatu and Chinasa T. Okolo	358
<i>Ableism, Ageism, Gender, and Nationality bias in Norwegian and Multilingual Language Models</i>	
Martin Sjøvik and Samia Touileb	379
<i>Disentangling Biased Representations: A Causal Intervention Framework for Fairer NLP Models</i>	
Yange Qian, Yilong Hu, Siqi Zhang, Xu Gu and Xiaolin Qin	393
<i>Towards Massive Multilingual Holistic Bias</i>	
Xiaoqing Tan, Prangthip Hansanti, Arina Turkatenko, Joe Chuang, Carleigh Wood, Bokai Yu, Christophe Ropers and Marta R. Costa-jussà	403

<i>Exploring Gender Bias in Large Language Models: An In-depth Dive into the German Language</i>	
Kristin Gnad, David Thulke, Simone Kopeinik and Ralf Schlüter	427
<i>Adapting Psycholinguistic Research for LLMs: Gender-inclusive Language in a Coreference Context</i>	
Marion Bartl, Thomas Brendan Murphy and Susan Leavy	451
<i>Leveraging Large Language Models to Measure Gender Representation Bias in Gendered Language Corpora</i>	
Erik Derner, Sara Sansalvador De La Fuente, Yoan Gutierrez, Paloma Moreda Pozo and Nuria M Oliver	468