5th Workshop on Gender Bias in Natural Language Processing (GeBNLP 2024)

Bangkok, Thailand
16 August 2024

ISBN: 979-8-3313-0163-7

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© (2024) 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

A Parameter-Efficient Multi-Objective Approach to Mitigate Stereotypical Bias in Language Models
Yifan Wang and Vera Demberg 1

Do PLMs and Annotators Share the Same Gender Bias? Definition, Dataset, and Framework of Contextualized Gender Bias
Shucheng Zhu, Bingjie Du, Jishun Zhao, Ying Liu and Pengyuan Liu 20

We Don't Talk About That: Case Studies on Intersectional Analysis of Social Bias in Large Language Models
Hannah Devinney, Jenny Björklund and Henrik Björklund 33

An Explainable Approach to Understanding Gender Stereotype Text
Manuela Nayantara Jeyaraj and Sarah Jane Delany 45

A Fairness Analysis of Human and AI-Generated Student Reflection Summaries
Bhiman Kumar Baghel, Arun Balajiee Lekshmi Narayanan and Michael Miller Yoder 60

On Shortcuts and Biases: How Finetuned Language Models Distinguish Audience-Specific Instructions in Italian and English
Nicola Fanton and Michael Roth 78

The power of Prompts: Evaluating and Mitigating Gender Bias in MT with LLMs
Aleix Sant, Carlos Escolano, Audrey Mash, Francesca De Luca Fornaciari and Maite Melero . 94

Detecting Gender Discrimination on Actor Level Using Linguistic Discourse Analysis
Stefanie Urchs, Veronika Thurner, Matthias Aßenmacher, Christian Heumann and Stephanie Thiemichen 140

What Can Go Wrong in Authorship Profiling: Cross-Domain Analysis of Gender and Age Prediction
Hongyu Chen, Michael Roth and Agnieszka Falenska 150

Towards Fairer NLP Models: Handling Gender Bias In Classification Tasks
Nasim Sobhani and Sarah Jane Delany 167

Investigating Gender Bias in STEM Job Advertisements
Malika Dikshit, Houda Bouamor and Nizar Habash 179

Dissecting Biases in Relation Extraction: A Cross-Dataset Analysis on People's Gender and Origin
Marco Antonio Stranisci, Pere-Lluís Huguet Cabot, Elisa Bassignana and Roberto Navigli . . . 190

Gender Bias in Turkish Word Embeddings: A Comprehensive Study of Syntax, Semantics and Morphology Across Domains
Duygu Altinok 203

Disagreeable, Slovenly, Honest and Un-named Women? Investigating Gender Bias in English Educational Resources by Extending Existing Gender Bias Taxonomies
Haotian Zhu, Kexin Gao, Fei Xia and Mari Ostendorf 219

Generating Gender Alternatives in Machine Translation
Sarthak Garg, Mozhdeh Gheini, Clara Emmanuel, Tatiana Likhomanenko, Qin Gao and Matthias Paulik 237

Beyond Binary Gender Labels: Revealing Gender Bias in LLMs through Gender-Neutral Name Predictions
Zhiwen You, HaeJin Lee, Shubhanshu Mishra, Sullam Jeoung, Apratim Mishra, Jinseok Kim and Jana Diesner 255

Is there Gender Bias in Dependency Parsing? Revisiting Women's Syntactic Resilience"
Paul Stanley Go and Agnieszka Falenska 269

From 'Showgirls' to 'Performers': Fine-tuning with Gender-inclusive Language for Bias Reduction in LLMs
Marion Bartl and Susan Leavy 280

Sociodemographic Bias in Language Models: A Survey and Forward Path
Vipul Gupta, Pranav Narayanan Venkit, Shomir Wilson and Rebecca J. Passonneau 295

Stop! In the Name of Flaws: Disentangling Personal Names and Sociodemographic Attributes in NLP
Vagrant Gautam, Arjun Subramonian, Anne Lauscher and Os Keyes 323

Evaluating Gender Bias in Multilingual Multimodal AI Models: Insights from an Indian Context
Kshitish Ghate, Arjun Choudhry and Vanya Bannihatti Kumar 338

Detecting and Mitigating LGBTQIA+ Bias in Large Norwegian Language Models
Selma Kristine Bergstrand and Björn Gambäck 351

Whose wife is it anyway? Assessing bias against same-gender relationships in machine translation
Ian Stewart and Rada Mihalcea 365

Analysis of Annotator Demographics in Sexism Detection
Narjes Tahaei and Sabine Bergler 376

An Empirical Study of Gendered Stereotypes in Emotional Attributes for Bangla in Multilingual Large Language Models
Jayanta Sadhu, Maneesha Rani Saha and Rifat Shahriyar 384

Overview of the Shared Task on Machine Translation Gender Bias Evaluation with Multilingual Holistic Bias
Marta R. Costa-jussà, Pierre Andrews, Christine Basta, Juan Ciro, Agnieszka Falenska, Seraphina Goldfarb-Tarrant, Rafael Mosquera, Debora Nozza and Eduardo Sánchez 399