GeBNLP 2025

The 6th Workshop on Gender Bias in Natural Language Processing

Proceedings of the Workshop

August 1, 2025

The GeBNLP organizers gratefully acknowledge the support from the following sponsors.

BOUQuET – Benchmark and Open initiative for Universal Quality Evaluation in Translation

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-277-0

Message from the Organisation Committee

This volume contains the proceedings of the Sixth Workshop on Gender Bias in Natural Language Processing, held in conjunction with the 63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025). This year, Christine Basta, Marta R. Costa-jussà, Agnieszka Faleńska, and Debora Nozza are delighted to welcome Karolina Stańczak as a new co-organizer. Karolina brings extensive experience in the field, gained through her PhD research, and we deeply value the invaluable insights and expertise she will add to our team.

This year's workshop saw a significant increase in engagement, receiving 50 technical paper submissions and 8 ACL Rolling Review (ARR) commitment papers, totaling 58 papers. Of these, 35 archival papers were accepted, resulting in a competitive acceptance rate of 60%. The accepted papers comprise 28 long papers, 7 short papers. Additionally, we accepted 4 non-archival papers. We are particularly pleased to report a substantial increase in submissions compared to previous years. This year's 58 papers represent a notable jump from 36 papers last year, 33 papers the year before, and an average of around 19 papers in the three years prior to that. This growth underscores the increasing interest and importance of gender bias research in NLP.

The accepted papers cover a broad spectrum of natural language processing research areas, exploring key NLP tasks such as language modeling and generation, machine translation, question answering, explainable AI, classification, and gender profiling. Several papers also delve into multimodal tasks, including those incorporating vision. The research spans diverse domains, including recruitment, medical, and sports.

Furthermore, the volume introduces novel approaches to bias analysis and debiasing methods. Many papers present new monolingual and multilingual benchmarks, opening up fresh opportunities for assessment and evaluation. Beyond gender bias, numerous studies investigate other crucial social biases, including ageism, nationality, ability, and various demographic factors.

We are particularly excited by the high interest shown in low-resource and non-English languages. This year's papers feature compelling studies on languages rarely addressed in gender bias research, such as Bangla, Arabic, various African languages (Twi, Amharic), Filipino, Farsi, Maltese, Nepali, French, Japanese, German, and Italian. This multilingual focus is crucial for comprehensively addressing bias and opens the door for more inclusive research in smaller communities and low-resource linguistic contexts. A significant number of research studies in this workshop highlight important developments in gender inclusivity within NLP. Notably, this year's proceedings include studies that address both binary and non-binary gender considerations, showcasing a more comprehensive approach to understanding and mitigating gender bias.

Finally, the workshop will feature two distinguished keynote speakers: Anne Lauscher from the University of Hamburg and Maarten Sap from Carnegie Mellon University.

We are very pleased to keep the high interest that this workshop has generated over the last five editions and we look forward to an enriching discussion on how to address gender bias in NLP when we meet in a hybrid event on 1st of August 2025!

August 2025
Christine Basta, Marta R. Costa-jussà, Agnieszka Faleńska, Debora Nozza,
Karolina Stańczak. (Alphabetically ordered)

Organizing Committee

Program Chairs

Christine Basta, Alexandria University, Egypt
Marta Costa-jussà, FAIR, Meta
Agnieszka Faleńska, University of Stuttgart, Germany
Debora Nozza, Bocconi University, Italy
Karolina Stańczak, ETH Zurich, Switzerland

Program Committee

Reviewers

Ashutosh Ahuja, Starbucks
Bashar Alhafni, Mohamed bin Zayed University of Artificial Intelligence
Duygu Altinok
Andrew Aquilina, University of Pittsburgh
Matthias Aßenmacher, Ludwig-Maximilians-Universität München
Bhiman Kumar Baghel, University of Pittsburgh
Sahil Bansal, SAP SE
Marion Bartl
Akanksha Bindal
Johanna Binnewitt, Universität Köln and Federal Institute for Vocational Education and Training
Laura Cabello, Copenhagen University
Ankani Chattoraj, NVIDIA
Khaoula Chehbouni
Hongyu Chen
Hannah Devinney, Linköping University
Matthias Gallé, Cohere
Albert Gatt, Utrecht University
Vagrant Gautam
Christian Heumann, Ludwig-Maximilians-Universität München
Mahammed Kamruzzaman, University of South Florida
Gauri Kholkar, Pure Storage
Wael Khreich, American University of Beirut
Gaurav Kumar, Moveworks
Manuel Lardelli
Arun Balajiee Lekshmi Narayanan, University of Pittsburgh
Marlene Lutz, Universität Mannheim
Ananya Malik, Northeastern University
Sara Vera Marjanovic
Vera Neplenbroek
Mohan Raj
Varsha Kuppur Rajendra
Milankumar Rana
Mukund Rungta, Microsoft
Hamidreza Saffari, Polytechnic Institute of Milan
Gayathri Saranathan
Beatrice Savoldi
Mohammadamin Shafiei, University of Milan
Christina Skelton, Universität Stuttgart
Samia Touileb, University of Bergen
Stefanie Urchs, Ludwig-Maximilians-Universität München and Hochschule München
Soroush Vosoughi, Dartmouth College
Yifan Wang
Ivory Yang
Zhiwen You, University of Illinois Urbana-Champaign
Haotian Zhu

Keynote Talk

Once Upon a Bias: A Fairy Tale of Gender in Language Technology

Prof. Anne Lauscher
University of Hamburg

Abstract: This is a story of dreams, detours, and (of course) data. In this keynote, I tell the tale of how a research community—our community—set out to create gender-fair language technologies. Along the way, we met dragons like stereotypical occupations, default male pronouns, and cisnormative datasets. We tried to rescue invisible identities. We met allies, too: other communities and other research disciplines. Drawing on my own memories of our adventures I will reflect upon the challenges we tackled and the drawbacks that remain. Finally, I will open the next chapter and invite you to take a look into the future.

Bio: Anne Lauscher is a Professor of Data Science at the University of Hamburg, where her research group investigates language-based Generative AI systems with a strong focus on safety aspects and ethical concerns. Before, she was a Postdoctoral Researcher in the Natural Language Processing group at Bocconi University (Milan, Italy) where she was working on introducing demographic factors into language processing systems with the aim of improving algorithmic performance and system fairness. She obtained her Ph.D., awarded with the highest honors (summa cum laude), from the Data and Web Science group at the University of Mannheim, where her research focused on the interplay between language representations and computational argumentation. During her studies, she conducted research internships at and became an independent research contractor for Grammarly Inc. (New York City, U.S.) and for the Allen Institute for Artificial Intelligence (Seattle, U.S.). Her research gets regularly published at international top-tier Natural Language Processing (e.g., ACL, EMNLP, etc.) and Artificial Intelligence (e.g., AAAI, ICLR) venues and has been recognized with multiple awards. For instance, most recently, she received a Social Impact Award at EACL2024, and an Outstanding Paper Award at NAACL2025.

Keynote Talk

Responsible AI for Diverse Users and Cultures.

Asst. Prof. Maarten Sap
Carnegie Mellon University (CMU), Allen Institute for Ai (Ai2)

Abstract: AI systems and language technologies are increasingly developed and deployed onto users of diverse genders and cultures. Yet, they still lack contextual and cultural awareness, and are unilaterally pushed onto many users that do not necessarily want them. In this talk, I will discuss some ongoing projects towards responsible AI development for diverse users and cultures.
I will first discuss the CobraFrames formalism, a method to enhance the reasoning of models for offensive speech grounded in social contexts such as speaker and listener identities. Then, I will discuss MC-Signs, a novel benchmark to measure the cultural awareness of multimodal AI systems with respect to culturally offensive gestures. Finally, I will conclude with a study on AI acceptability, showing that lay people's opinions about when and where AI should be used varies depending on their gender, AI literacy, and more. I will conclude with some future directions towards responsible and prosocial AI.

Bio: Maarten Sap is an assistant professor in Carnegie Mellon University's Language Technologies Department (CMU LTI), and a courtesy appointment in the Human-Computer Interaction institute (HCII). He is also a part-time research scientist and AI safety lead at the Allen Institute for AI. His research focuses on (1) measuring and improving AI systems' social and interactional intelligence, (2) assessing and combating social inequality, safety risks, and socio-cultural biases in human- or AI-generated language, and (3) building narrative language technologies for prosocial outcomes. He has presented his work in top-tier NLP and AI conferences, receiving paper awards or nominations at NAACL 2025, EMNLP 2023, ACL 2023, FAccT 2023, WeCNLP 2020, and ACL 2019. His research has been covered in the press, including the New York Times, Forbes, Fortune, Vox, and more.

Table of Contents

Program

Friday, August 1, 2025

09:00 - 09:15 *Opening Remarks*

09:15 - 10:15 *Keynote Speech by Anne Lauscher*

10:15 - 11:00 *Coffee Break*

11:00 - 11:15 *Introducing MARB — A Dataset for Studying the Social Dimensions of Reporting Bias in Language Models*

Introducing MARB — A Dataset for Studying the Social Dimensions of Reporting Bias in Language Models
Tom Södahl Bladsjö and Ricardo Muñoz Sánchez

11:15 - 11:30 *GENDEROUS: Machine Translation and Cross-Linguistic Evaluation of a Gender-Ambiguous Dataset*

GENDEROUS: Machine Translation and Cross-Linguistic Evaluation of a Gender-Ambiguous Dataset
Janiça Hackenbuchner, Joke Daems and Eleni Gkovedarou

11:30 - 12:15 *Poster Session I*

12:15 - 14:00 *Lunch Break*

14:00 - 14:45 *Poster Session II*

14:45 - 15:30 *Poster Session III*

15:30 - 16:00 *Coffee Break*

16:00 - 17:00 *Keynote Speech by Marteen Sap*

17:00 - 18:00 *Lightning Talks*

18:00 - 18:15 *Closing Remarks*