GeBNLP 2022

The 4th Workshop on Gender Bias in Natural Language Processing

Proceedings of the Workshop

July 15, 2022

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-955917-68-1

Preface

This volume contains the proceedings of the Fourth Workshop on Gender Bias in Natural Language Processing, held in conjunction with the 2022 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT2022). This year, the organization committee changed membership: Kellie Webster made way for Christine Basta and Gabriel Stanovsky. Kellie has been one of the main reasons for the success of this workshop and we would like to thank her for her valuable and enthusiastic contribution to this workshop. We are glad to welcome our two co-organizers and look forward to sharing their insights and expertise.

This year, the workshop received 33 submissions of technical papers (12 long papers, 21 short papers), of which 28 were accepted (11 long, 17 short), for an acceptance rate of 84%. We are pleased to see an increased interest compared to our previous editions in the last three years: the submissions have increased this year to 33 papers compared to 18 papers last year and 19 papers in 2019 and 2020. Furthermore, the high quality of the submissions allowed us to have a higher acceptance rate this year of 84% compared to the previous years, where the acceptance rate was 63%, 68% and 67% respectively. Once more, we thank the Programme Committee members, who provided extremely valuable reviews in terms of technical content and bias statements, for the high-quality selection of research works.

The accepted papers cover a wide range of natural language processing research areas. From the core tasks of NLP, the papers include language modeling and generation, annotation, machine translation, word embeddings, and evaluation. New aspects regarding the analysis and the debiasing mechanisms are introduced and we are excited about the discussions these will inspire. Besides English, we have interesting studies targeting Inuktikut, Hindi and Marathi as well as Chinese, Italian, French and Spanish. All papers cover a variety of gender (and intersectional) bias studies as well as a taxonomy definition.

Finally, the workshop has two keynotes by speakers of high standing: Kellie Webster and Kevin Robinson, Google Research, and Kai-Wei Chang, University of California (UCLA-CS). We also have a panel under the theme of Evaluating gender bias in NLP and we are looking forward to the insights of this panel.

We are very pleased to keep the high interest that this workshop has generated over the last three editions and we look forward to an enriching discussion on how to address bias problems in NLP applications when we meet at a hybrid event on 15 July 2022!

July 2022
Christine Basta, Marta R. Costa-jussà, Hila Gonen, Christian Hardmeier and Gabriel Stanovsky

Christian Hardmeier was supported by the Swedish Research Council under grant 2017-930.

Organizing Committee

Organizers

Christine Basta, Polytechnic University of Catalonia and Alexandria University
Marta R. Costa-jussà, Meta AI
Hila Gonen, Meta AI and University of Washington
Christian Hardmeier, IT University of Copenhagen / Uppsala University
Gabriel Stanovsky, Hebrew University of Jerusalem

Program Committee

Chairs

Christine Basta, Universitat Politècnica de Catalunya
Marta R. Costa-jussà, Meta AI
Hila Gonen, Meta AI and University of Washington
Christian Hardmeier, IT University of Copenhagen/Uppsala University
Gabriel Stanovsky, The Hebrew University of Jerusalem

Program Committee

Gavin Abercrombie, Heriot Watt University
Jenny Björklund, Uppsala University
Su Lin Blodgett, Microsoft Research
Houda Bouamor, Carnegie Mellon University in Qatar
Ryan Cotterell, ETH Zürich
Hannah Devinney, Umeå University
Matthias Gallé, Naver Labs Europe
Mercedes García-Martínez, Pangeanic
Seraphina Goldfarb-Tarrant, University of Edinburgh
Zhengxian Gong, Computer science and technology school,soochow university
Nizar Habash, New York University Abu Dhabi
Ben Hachey, Harrison.ai
Svetlana Kiritchenko, National Research Council Canada
Shiyang Li, UC Santa Barbara
Tomasz Limisiewicz, Charles University in Prague
Gili Lior, The Hebrew University of Jerusalem
Sharid Loáiciga, University of Gothenburg
Inbal Magar, The Hebrew University of Jeruslaem
Maite Melero, BSC
Johanna Monti, L'Orientale University of Naples
Carla Perez Almendros, Cardiff University
Will Radford, Canva
Rafal Rzepka, Hokkaido University
Sonja Schmer-Galunder, Smart Information Flow Technologies
Bonnie Webber, University of Edinburgh
Lilja Øvrelid, Dept of Informatics, University of Oslo

Table of Contents

Program

Friday, July 15, 2022

08:30 - 08:40	*Opening Remarks*
08:40 - 09:25	*Keynote 1: Kellie Webster and Kevin Robinson, Google Research*
09:30 - 10:00	*Oral papers 1*
	Challenges in Measuring Bias via Open-Ended Language Generation Afra Feyza Akyürek, Muhammed Yusuf Kocyigit, Sejin Paik and Derry Tanti Wijaya
	Looking for a Handsome Carpenter! Debiasing GPT-3 Job Advertisements Conrad Borchers, Dalia Gala, Benjamin Gilburt, Eduard Oravkin, Wilfried Bounsi, Yuki M Asano and Hannah Kirk
10:00 - 10:30	*Break*
10:30 - 11:15	*Keynote 2: Kai-Wei Chang, UCLA Computer Science*
11:15 - 12:00	*Oral papers 2*
	Fewer Errors, but More Stereotypes? The Effect of Model Size on Gender Bias Yarden Tal, Inbal Magar and Roy Schwartz
	Analysis of Gender Bias in Social Perception and Judgement Using Chinese Word Embeddings Jiali Li, Shucheng Zhu, Ying Liu and Pengyuan Liu
	On the Dynamics of Gender Learning in Speech Translation Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri and Marco Turchi
12:00 - 13:30	*Lunch break*
13:30 - 14:30	*Poster session*
	Analyzing Hate Speech Data along Racial, Gender and Intersectional Axes Antonis Maronikolakis, Philip Baader and Hinrich Schütze

Friday, July 15, 2022 (continued)

Don't Forget About Pronouns: Removing Gender Bias in Language Models Without Losing Factual Gender Information
Tomasz Limisiewicz and David Mareček

Uncertainty and Inclusivity in Gender Bias Annotation: An Annotation Taxonomy and Annotated Datasets of British English Text
Lucy Havens, Beatrice Alex, Benjamin Bach and Melissa Terras

Debiasing Neural Retrieval via In-batch Balancing Regularization
Yuantong Li, Xiaokai Wei, Zijian Wang, Shen Wang, Parminder Bhatia, Xiaofei Ma and Andrew Arnold

Gender Biases and Where to Find Them: Exploring Gender Bias in Pre-Trained Transformer-based Language Models Using Movement Pruning
Przemyslaw Joniak and Akiko Aizawa

Gendered Language in Resumes and its Implications for Algorithmic Bias in Hiring
Prasanna Parasurama and João Sedoc

The Birth of Bias: A case study on the evolution of gender bias in an English language model
Oskar Van Der Wal, Jaap Jumelet, Katrin Schulz and Willem Zuidema

Assessing Group-level Gender Bias in Professional Evaluations: The Case of Medical Student End-of-Shift Feedback
Emmy Liu, Michael Henry Tessler, Nicole Dubosh, Katherine Hiller and Roger Levy

Unsupervised Mitigating Gender Bias by Character Components: A Case Study of Chinese Word Embedding
Xiuying Chen, Mingzhe Li, Rui Yan, Xin Gao and Xiangliang Zhang

An Empirical Study on the Fairness of Pre-trained Word Embeddings
Emeralda Sesari, Max Hort and Federica Sarro

Mitigating Gender Stereotypes in Hindi and Marathi
Neeraja Kirtane and Tanvi Anand

A Taxonomy of Bias-Causing Ambiguities in Machine Translation
Michal Měchura

Friday, July 15, 2022 (continued)

On Gender Biases in Offensive Language Classification Models
Sanjana Marcé and Adam Poliak

Gender Bias in BERT - Measuring and Analysing Biases through Sentiment Rating in a Realistic Downstream Classification Task
Sophie Jentzsch and Cigdem Turan

Occupational Biases in Norwegian and Multilingual Language Models
Samia Touileb, Lilja Øvrelid and Erik Velldal

Indigenous Language Revitalization and the Dilemma of Gender Bias
Oussama Hansal, Ngoc Tan Le and Fatiha Sadat

What changed? Investigating Debiasing Methods using Causal Mediation Analysis
Sullam Jeoung and Jana Diesner

Why Knowledge Distillation Amplifies Gender Bias and How to Mitigate from the Perspective of DistilBERT
Jaimeen Ahn, Hwaran Lee, Jinhwa Kim and Alice Oh

Incorporating Subjectivity into Gendered Ambiguous Pronoun (GAP) Resolution using Style Transfer
Kartikey Pant and Tanvi Dadu

14:30 - 15:00 *Oral papers 3*

HeteroCorpus: A Corpus for Heteronormative Language Detection
Juan Vásquez, Gemma Bel-Enguix, Scott Andersen and Sergio-Luis Ojeda-Trueba

Worst of Both Worlds: Biases Compound in Pre-trained Vision-and-Language Models
Tejas Srinivasan and Yonatan Bisk

15:00 - 15:30 *Break*

15:30 - 16:15 *Panel discussion*

Friday, July 15, 2022 (continued)

16:15 - 16:45 *Oral papers 4*

Evaluating Gender Bias Transfer from Film Data
Amanda Bertsch, Ashley Oh, Sanika Natu, Swetha Gangu, Alan W. Black and Emma Strubell

Choose Your Lenses: Flaws in Gender Bias Evaluation
Hadas Orgad and Yonatan Belinkov

16:30 - 17:00 *Discussion and closing*