WOAH 2022

The Sixth Workshop on Online Abuse and Harms

Proceedings of the Workshop

July 14, 2022

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

Gold

Silver

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 978-1-955917-84-1

Introduction

Digital technologies have brought myriad benefits for society, transforming how people connect, communicate and interact with each other. However, they have also enabled harmful and abusive behaviors to reach large audiences and for their negative effects to be amplified, including interpersonal aggression, bullying and hate speech. Work on online abuse and harms has traditionally centred on abuse in English and other Western European languages, further widening the resource gap between Western European languages and all other languages.

As academics, civil society, policymakers and tech companies devote more resources and effort to tackling online abuse, there is a pressing need for scientific research that critically and rigorously investigates how it is defined, detected and countered. Technical disciplines such as machine learning (ML), natural language processing (NLP) and statistics have made substantial advances in this field. However, concerns have been raised about the differences in attention given to different languages and geographies. For example, English, particularly dominant forms of American English, are overrepresented in most NLP resources. Technological solutions can be developed for speakers of other dialects and languages. On the other hand, languages such as Yuruba, Urdu, Amharic, and many other languages have few to no resources available, thus providing significant challenges in developing technological systems for the detection of abuse and other harms.

For this sixth edition of the Workshop on Online Abuse and Harms (6th WOAH!) we advance research in online abuse through our theme: **On Developing Resources and Technologies for low resource Online Abuse and Harms**. We continue to emphasize the need for inter-, cross- and anti- disciplinary work on online abuse and harms. These include but are not limited to: NLP, machine learning, computational social sciences, law, politics, psychology, network analysis, sociology and cultural studies. Continuing the tradition started in WOAH 4, we invite civil society, in particular individuals and organisations working with women and marginalised communities who are often disproportionately affected by online abuse, to submit reports, case studies, findings, data, and to record their lived experiences. We hope that through these engagements WOAH can directly address the issues faced by those on the front-lines of tackling online abuse.

Speaking to the complex nature of the issue of online abuse, we are pleased to invite Mona Diab, Murali Shanmugavelan, Gebre Gebremeske, Daniel Borkan, Lucas Dos Santos, Alyssa Lees, and Rachel Rosen to deliver keynotes. In addition to our invited keynotes, we received 47 submissions out of which 24 were accepted. Of the accepted papers, 20 were long papers and 4 were short papers. These papers will be presented in our poster session. We thank the reviewers for their dedication and efforts in providing in-depth and timely reviews.

With this, we welcome you to the Sixth Workshop on Online Abuse and Harms. We look forward to a day filled with spirited discussion and thought provoking research!

Aida, Bertie, Kanika, Lambert, and Zeerak.

Organizing Committee

Organizer

Kanika Narang, Meta AI
Aida Mostafazadeh Davani, University of Southern California
Lambert Mathias, Meta AI
Bertie Vidgen, The Alan Turing Institute
Zeerak Talat, Simon Fraser University

Program Committee

Chairs

Kanika Narang, Meta AI
Aida Mostafazadeh Davani, University of Southern California
Lambert Mathias, Meta AI
Bertie Vidgen, Alan Turing Institute
Zeerak Talat, Simon Fraser University

Emergency Reviewers

Francielle Vargas, University of São Paulo
Alon Halevy, Facebook AI
Qifan Wang, Meta AI
Shaoliang Nie, Facebook Inc
Yi-ling Chung, The Alan Turing Institute
Sheikh Sarwar, Amazon.com

Program Committee

Piush Aggarwal, FernUniversität in Hagen, Computational Linguistics
Khalid Alnajjar, University of Helsinki
Nikolay Babakov, Skoltech Institute of Science and Technology
Xingjian Bai, University of Oxford
Dan Bateyko, Georgetown University Law Center
Thales Bertaglia, Maastricht University
Noah Broestl, University of Oxford
Tommaso Caselli, Rijksuniversiteit Groningen
Yung-sung Chuang, Massachusetts Institute of Technology
Leon Derczynski, IT University of Copenhagen
Nemanja Djuric, Aurora Innovation
Lucie Flek, CAISA Lab, Faculty of Mathematics and Informatics, Philipps University of Marburg
Paula Fortuna, TALN, Pompeu Fabra University
Simona Frenda, Università degli Studi di Torino and Universitat Politècnica de València
Sara E. Garza, FIME-UANL
Shlok Gilda, University of Florida
Lee Gillam, University of Surrey
Darina Gold, University of Duisburg-Essen
Udo Hahn, Friedrich-Schiller-Universität Jena
Adeep Hande, Indian Institute of Information Technology,Tiruchirappalli
Alex Hanna, Google
Christopher Homan, Rochester Institute of Technology
Ruihong Huang, Texas A&M University
Pica Johansson, Alan Turing Institute
Srecko Joksimovic, University of South Australia
David Jurgens, University of Michigan
Brendan Kennedy, University of Southern California
Ashiqur Khudabukhsh, Carnegie Mellon University

Thomas Kleinbauer, Saarland University
Vasiliki Kougia, University of Vienna
Ralf Krestel, ZBW & Kiel University
Sheng Li, University of Georgia
Zi Lin, University of California, San Diego
Jeremiah Liu, Google Research
Hongyin Luo, MIT
Diana Maynard, University of Sheffield
Mainak Mondal, Institute of Engineering and Management
Smruthi Mukund, Amazon
Isar Nejadgholi, National Research Council Canada
Debora Nozza, Bocconi University
Ali Omrani, University of Southern California
Alexander Panchenko, Skolkovo Institue of Science and Technology
Kartikey Pant, Salesforce
Viviana Patti, University of Turin, Dipartimento di Informatica
John Pavlopoulos, Athens University of Economics and Business
Matúš Pikuliak, Kempelen Institute of Intelligent Technologies
Vinodkumar Prabhakaran, Google
Michal Ptaszynski, Kitami Institute of Technology
Masoumeh Razzaghi, Texas A&M University-Commerce
Georg Rehm, DFKI
Björn Ross, University of Edinburgh
Paolo Rosso, Universitat Politècnica de València
Dana Ruiter, Saarland University
Paul Röttger, University of Oxford
Nazanin Sabri, University of Tehran
Qinlan Shen, Oracle
Karthik Shivaram, Tulane University
Marian Simko, Kempelen Institute of Intelligent Technologies
Jeffrey Sorensen, Google Jigsaw
Gerasimos Spanakis, Maastricht University
Arjun Subramonian, University of California, Los Angeles
Sajedul Talukder, Southern Illinois University
Tristan Thrush, Hugging Face
Sara Tonelli, FBK
Dimitrios Tsarapatsanis, University of York
Gareth Tyson, QMUL
Avijit Vajpayee, Amazon
Ingmar Weber, Qatar Computing Research Institute
Jing Xu, Facebook AI
Fan Yang, Nuance Communications
Seunghyun Yoon, Adobe Research
Samira Zad, Florida International University
Aleš Završnik, Institute of criminology at the Faculty of Law Ljubljana
Torsten Zesch, Computational Linguistics, FernUniversität in Hagen

Keynote Talk: Multilingual hate speech detection: From labeling to systems, challenges and opportunities

Mona Diab
George Washington University - Facebook AI

Abstract: Assessing social media content is quite challenging due to the subjective nature of the material where context plays a pivotal role. In this talk, I highlight the challenges of dealing with nuanced language due to inherent characteristics of dialects as manifested in the Arabic language as well as in English. I will talk about challenges in labeling and building systems where the amount of labeled data is on the low. However such challenges can be mitigated with smart designs while also heeding diversity and inclusion in the process.

Bio: Mona Talat Diab is a computer science professor at George Washington University and a research scientist with Facebook AI. Her research focuses on natural language processing, computational linguistics, cross lingual/multilingual processing, computational socio-pragmatics, and applied machine learning. Besides this, she also has special interests in Arabic NLP and low resource scenarios. Diab completed her Ph.D. in computational linguistics at the University of Maryland, Linguistics Department and University of Maryland Institute for Advanced Computer Studies (UMIACS) in 2003, under the supervision of Philip Resnik. She was also a postdoctoral research scientist at Stanford University (2003–2005) under the mentorship of Dan Jurafsky, where she was a part of the Stanford NLP Group. After her postdoc at Stanford, Diab took a position as principal investigator at the Center for Computational Learning Systems (CCLS) in Columbia University, where she was also adjunct professor in the computer science department. In 2013 she joined the George Washington University as an associate professor, where she was promoted to full professor in 2017. Diab is the founder and director of the GW NLP lab CARE4Lang.

Keynote Talk: Keynote by Murali Shanmugavelan

Murali Shanmugavelan
Oxford Internet Institute - Data and Society, NYC

Bio: Murali Shanmugavelan researches caste in media and communication studies and digital cultures. His PhD from the School of Oriental and African Studies (SOAS) University of London was focused on everyday communicative practices of caste. He has over 15 years of experience developing, managing and implementing projects focused on developing media and ICT policies and practice; outreach and strategic communications; and innovations in mobile applications in multi-disciplinary and cross-cultural settings.

Keynote Talk: Social media and hate speech in time of war: The case of Tigray

Gebre Gebremeskel
The Centre for Mathematics and Computer Science (CWI), Netherlands

Abstract: Hate Speech has been around in Ethiopia before social media, but with very limited reach. With the coming of social media companies that have no or little business interest to lose in low-resourced languages such as those in Ethiopia, diaspora activists that have nothing or little to lose from engaging in online hate speech, and several technical and institutional challenges, hate speech on social media slowly became mainstream in Ethiopia, tearing societies apart and eventually serving as an animating force for a genocidal war on Tigrayans. In this speech, I will briefly assess the normalization of hate speech in Ethiopia, the factors that led to this, and the role hate speech and social media played during the Tigray war, social media hate speech detection and monitoring, and what should be done going forward.

Bio: Gebrekirstos G. Gebremeskel is the founder and chief editor of Tghat.com, founder of mermru.com, and a PhD candidate at Radboud University Nijmegen, Netherlands. He has a double masters degree: MSc in Human Language Science and Technology from the University of Malta and MA/MSc in Linguistics (research) from the University of Groningen. Tghat was founded in November 2020 following the start of the war on Tigray in response to the Ethiopian government's imposition of media and telecommunications blackout as part of the war on Tigray. Tghat has been engaged in documenting, researching and writing about the Tigray war. Merrmru.com, is a website dedicated to collecting and developing Natural Language Processing tools and resources for the learning and the computational processing of Geez-based languages such as Tigrinya, Geez and Amharic. The website has an extensive capability to take any Tigrinya verb and provide tens of thousands of inflections. His PhD research focuses on the intersection of Information Retrieval, Recommender Systems, NLP and their impacts on society. Some of his academic publications can be found in Google Scholar.
He has previously worked as a researcher at the CWI Amsterdam, interned at Yahoo! And worked for other companies. Gebrekirstos also writes for other outlets, speaks in different platforms and events, and appears on local and international media including Al Jazeera and the BBC to offer analysis and views on the Tigray war, Ethiopia and the Horn of Africa. He tweets at @gebrekirstosG. His more extended bio can be found at https://www.tghat.com/gebrekirstos-gebreselassie-gebremeskel/

Keynote Talk: Next generation of perspective: Multilingual large language models and combating online harassment

Daniel Borkan, Lucas Dos Santos, Alyssa Lees, and Rachel Rosen
Google Jigsaw

Abstract: We explore two developments in Google Jigsaw's Perspective API. First, we describe a new multilingual, token-free, Charformer model infrastructure that is applicable across a range of languages, domains, and tasks. This architecture was extensively evaluated on an array of tasks and enabled Perspective API launches in 10 new languages, including Arabic, Chinese, Indonesian, Korean, and Japanese. We also discuss how we leveraged Perspective API to create Harassment Manager, an open-source web application that enables users to document and take action on abuse targeted at them on online platforms. The tool allows users to consolidate their experiences of online harassment into a story, complete with context and examples.

Bio:

- Daniel Borkan attended UCSC where he graduated with a BSc in computer science. He joined Jigsaw in 2014 to build the Outline tool to bypass repressive censorship. Daniel now works on the Perspective API to combat online toxicity, where he focuses on internationalization, bias mitigation, and new model development.
- Lucas Dos Santos attended Pomona College where he graduated with a BA in computer science. He joined the Conversation AI team at Jigsaw in 2018 and focuses on efforts around combatting online harassment, machine learning model development, and API infrastructure.
- Alyssa Lees attended Brown University and NYU where she received BSc/MS/PhD degrees in statistics and computer science while cultivating interests in AI, cooking, architecture and fine art. Alyssa has worked in various capacities at Google Jigsaw including developing the next generation of the Perspective API and currently as lead combatting disinformation. Her research interests include ML Fairness, NLP and Knowledge Acquisition.
- Rachel Rosen attended NYU where she graduated with a BA for a joint computer science and math major. She completed two Google internships while studying at NYU and began working for Google full time after graduating in 2014. She joined the ConversationAI team at Jigsaw in 2016 where she began working on solutions for countering toxic speech and online harassment.

Table of Contents

Program

Thursday, July 14, 2022

08:40 - 08:30 *Welcome + Opening Remarks*

09:25 - 08:40 *Keynote 1 - Mona Diab*

10:10 - 09:25 *Keynote 2 - Murali Shanmugavalan*

10:30 - 10:10 *Morning Break*

11:15 - 10:30 *Keynote 3 - Gebre Gebremeskel*

12:00 - 11:15 *Poster Session (1)*

13:30 - 12:00 *Lunch Break*

14:15 - 13:30 *Poster Session (2)*

15:00 - 14:15 *Keynote 4 - Daniel Borkan, Lucas Dos Santos, Alyssa Lees, and Rachel Rosen*

15:05 - 15:00 *Closing Remarks*