NAACL HLT 2018

Ethics in Natural Language Processing

Proceedings of the Second ACL Workshop

June 5, 2018
New Orleans, Louisiana

Platinum

Gold

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 978-1-948087-14-8

Introduction

Welcome to the second ACL Workshop on Ethics in Natural Language Processing! We are pleased to again have participants from a variety of backgrounds and perspectives, including social science, computational linguistics, and philosophy; academia as well as industry.

The workshop consists of invited talks, contributed papers, and a panel discussion. Based on the success of the first iteration, we decided to make more room for interactive sessions, and also present a science cafe. We would like to thank all authors, speakers, and panelists for their thoughtful contributions, as well as the large and supportive program committee, who have given their time to review. We are especially grateful for our sponsors (Bloomberg, Google, and HITS), who have helped making the workshop in this form possible. For the first time, we were able to also provide over $5000 in scholarships, wish enable several students to attend and add their perspective.

Our invited speakers include Amanda Stent (Bloomberg, USA), who is a NLP Architect at Bloomberg LP. Her background is in dialogue, discourse and natural language generation although she currently works on text analytics. She is president emeritus of SIGDial, is on the board of SIGGEN and is on the editorial board of the journal Dialogue and Discourse. Her research also includes work on factuality, and inclusiveness, and she is one of the two chairs of this year's NAACL conference.

Suresh Venkatasubramanian (University of Utah, USA), a professor of computer science at the School of Computing at the University of Utah. His research extends to the social ramifications of automated decision making and algorithmic fairness. He is also a founding member of the FAT* organization and workshop series, as well as a member of the last three FATML workshops.

Oisin Deery (Monash University, Australia) is a Lecturer in the Department of Philosophy at Monash University, in Melbourne, Australia. His research interests lie at the intersection of philosophy of mind and action, metaphysics, and ethics. He has published on free will and the impact of machine learning on ethical decisions.

Katherine Bailey (Acquia) is a researcher and team leader in industry. Her recent work has been on machine learning applications for natural language processing and other fields. She is pioneering a "few-shot learning" approach which promises greater efficiency in machine learning. Katherine has spoken at international conferences on both the technical details of artificial intelligence and the ethical issues that arise from its use in a variety of contexts.

Francien Dechesne (Leiden University, The Netherlands), is a researcher at the Center for Law and Digital Technologies (eLaw) of the Leiden Law School, and lecturer at TU Eindhoven. Her research lies at the intersection between societal and ethical issues of information and communication technologies, including the question of how to balance public, commercial, and individual interests in data-driven innovations. In particular, her research focuses on potential negative societal impact of decisions based on data-analytics, and the design of accountability mechanisms to address this impact.

We received fewer paper submissions than in the previous year, but present a large range of topics, addressing issues related to overgeneralization, dual use, privacy protection, bias in NLP models, underrepresentation, fairness, and more. Authors share insights about the intersection of NLP and ethics in academic, industrial, and clinical work. We selected three papers for oral presentation. Due to the involvement of different disciplines with differing publication traditions, we also offered a non-archival submission option, which means not all papers presented at the workshop are included here.

We are glad to see the continued interest in this important topic and hope that this workshop will help defining ethical issues in NLP, and raising awareness of ethical considerations throughout the community.

Mark Alfano, Dirk Hovy, Margaret Mitchell, and Michael Strube

Organizers:

Mark Alfano, Associate Professor, Delft University of Technology & Professor, Australian Catholic University
Dirk Hovy, Associate Professor, Bocconi University
Margaret Mitchell, Senior Research Scientist, Google
Michael Strube, Scientific Director, Heidelberg Institute for Theoretical Studies gGmbH

Program Committee:

Miguel Ballesteros, Solon Barocas, Daniel Bauer, Steven Bedrick, Adrian Benton, Steven Bethard, Rahul Bhagat, Yonatan Bisk, Michael Bloodgood, Matko Bosnjak, Chris Brockett, Ann Clifton, Kevin Cohen, Court Corley, Ryan Cotterell, Aron Culotta, Walter Daelemans, Munmun De Choudhury, Francien Dechesne, Steve DeNeefe, Mona Diab, Mark Dredze, Desmond Elliott, Micha Elsner, Katrin Erk, Raquel Fernandez, Sorelle Friedler, Spandana Gella, Oul Han, Graeme Hirst, Kristy Hollingshead, Anna Jobin, Anders Johannsen, David Jurgens, Brian Keegan, Roman Klinger, Ekaterina Kochmar, Philipp Koehn, Alexander Koller, Jonathan K. Kummerfeld, Brian Larson, Jochen L. Leidner, Alessandro Lenci, Dave Lewis, Maria Liakata, Nikola Ljubešić, Teresa Lynn, Nitin Madnani, Gideon Mann, Chandler May, Paola Merlo, Margot Mieskes, David Mimno, Alessandro Moschitti, Jason Naradowsky, Dong Nguyen, Brendan O'Connor, Sebastian Padó, Alexis Palmer, Carla Parra Escartín, Emily Pitler, Thierry Poibeau, Christopher Potts, Daniel Preoţiuc-Pietro, Nikolaus Pöchhacker, Will Radford, Siva Reddy, Luis Reyes-Galindo, Sebastian Riedel, Frank Rudzicz, Asad Sayeed, Frank Schilder, David Schlangen, Natalie Schluter, Tyler Schnoebelen, Djamé Seddah, Dan Simonson, Sameer Singh, Charese Smiley, Erin Smith Crabb, Vivek Srikumar, Pontus Stenetorp, Veselin Stoyanov, Simon Suster, Rachael Tatman, Ivan Titov, Sara Tonelli, Oren Tsur, Yulia Tsvetkov, Lyle Ungar, L. Alfonso Urena Lopez, Andreas van Cranenburgh, Janneke van der Zwaan, Benjamin Van Durme, Yannick Versley, Aline Villavicencio, Andreas Vlachos, Rob Voigt, Bonnie Webber, Joern Wuebker, Luke Zettlemoyer, Sanja Štajner

Invited Speakers:

Katherine Bailey, Researcher, Acquia, USA
Francien Dechesne, Researcher, Center for Law and Digital Technologies, Leiden University & Lecturer, TU Eindhoven, The Netherlands
Oisin Deery, Lecturer, Department of Philosophy, Monash University, Australia
Amanda Stent, NLP Architect, Bloomberg, USA
Suresh Venkatasubramanian, Professor, School of Computing, University of Utah, USA

Table of Contents

Workshop Program

Tuesday, 5th June 2018

9:00–10:30 Session 1

9:00–9:15 *Welcome*

9:15–9:40 *On the Utility of Lay Summaries and AI Safety Disclosures: Toward Robust, Open Research Oversight*
Allen Schmaltz

9:40–10:05 *#MeToo Alexa: How Conversational Systems Respond to Sexual Harassment*
Amanda Cercas Curry and Verena Rieser

10:05–10:30 *Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems*
Svetlana Kiritchenko and Saif Mohammad

10:30–11:00 *Coffee*

11:00–12:30 Session 2

11:00–11:45 *Invited Talk*

11:45–12:30 *Invited Talk*

12:30–14:00 *Lunch*

Tuesday, 5th June 2018 (continued)

14:00–15:30 Session 3

14:00–14:45 *Invited Talk*

14:45–15:30 *Invited Talk*

15:30–16:00 *Coffee Break*

16:00–17:00 ***Science cafe roundtable discussions***

17:00–17:15 ***Reaction to roundtable***

17:15–18:00 *Invited talk*