

EACL 2017

Ethics in Natural Language Processing

Proceedings of the First ACL Workshop

April 4th, 2017
Valencia, Spain

Sponsors:



Heidelberg Institute for
Theoretical Studies



Bloomberg LP

© 2017 The Association for Computational Linguistics

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 978-1-945626-47-0

Introduction

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

The workshop consists of invited talks, contributed discussion papers, posters, demos, and a panel discussion. Invited speakers include **Graeme Hirst**, a Professor in NLP at the University of Toronto, who works on lexical semantics, pragmatics, and text classification, with applications to intelligent text understanding for disabled users; **Quirine Eijkman**, a Senior Researcher at Leiden University, who leads work on security governance, the sociology of law, and human right; **Jason Baldridge**, a co-founder and Chief Scientist of People Pattern, who specializes in computational models of discourse as well as the interaction between machine learning and human bias; and **Joanna Bryson**, a Reader in artificial intelligence and natural intelligence at the University of Bath, who works on action selection, systems AI, transparency of AI, political polarization, income inequality, and ethics in AI.

We received paper submissions that span a wide range of topics, addressing issues related to overgeneralization, dual use, privacy protection, bias in NLP models, underrepresentation, fairness, and more. Their authors share insights about the intersection of NLP and ethics in academic work, industrial work, and clinical work. Common themes include the role of tasks, datasets, annotations, training populations, and modelling. We selected 4 papers for oral presentation, 8 for poster presentation, and one for demo presentation, and have paired each oral presentation with a discussant outside of the authors' areas of expertise to help contextualize the work in a broader perspective. All papers additionally provide the basis for panel and participant discussion.

We hope this workshop will help to define and raise awareness of ethical considerations in NLP throughout the community, and will kickstart a recurring theme to consider in future NLP conferences. We would like to thank all authors, speakers, panelists, and discussants for their thoughtful contributions. We are also grateful for our sponsors (Bloomberg, Google, and HITS), who have helped making the workshop in this form possible.

The Organizers

Margaret, Dirk, Shannon, Emily, Hanna, Michael

Organizers:

Dirk Hovy, University of Copenhagen (Denmark)
 Shannon Spruit, Delft University of Technology (Netherlands)
 Margaret Mitchell, Google Research & Machine Intelligence (USA)
 Emily M. Bender, University of Washington (USA)
 Michael Strube, Heidelberg Institute for Theoretical Studies (Germany)
 Hanna Wallach, Microsoft Research, UMass Amherst (USA)

Program Committee:

Gilles Adda	Fernando Diaz	Nikola Ljubesic	Molly Roberts
Nikolaos Aletras	Benjamin Van Durme	Adam Lopez	Tim Rocktäschel
Mark Alfano	Jacob Eisenstein	L. Alfonso Urena Lopez	Frank Rudzicz
Jacob Andreas	Jason Eisner	Teresa Lynn	Alexander M. Rush
Isabelle Augenstein	Desmond Elliott	Nitin Madhani	Derek Ruths
Tim Baldwin	Micha Elsner	Gideon Mann	Asad Sayeed
Miguel Ballesteros	Katrin Erk	Daniel Marcu	David Schlangen
David Bamman	Raquel Fernandez	Jonathan May	Natalie Schluter
Mohit Bansal	Laura Fichtner	Kathy McKeown	H. Andrew Schwartz
Solon Barocas	Karèn Fort	Paola Merlo	Hinrich Schütze
Daniel Bauer	Victoria Fossum	David Mimno	Djamé Seddah
Eric Bell	Lily Frank	Shachar Mirkin	Dan Simonson
Steven Bethard	Sorelle Friedler	Alessandro Moschitti	Sameer Singh
Rahul Bhagat	Annemarie Friedrich	Jason Naradowsky	Vivek Srikumar
Chris Biemann	Juri Ganitkevich	Roberto Navigli	Sanja Stajner
Yonatan Bisk	Spandana Gella	Arvind Neelakantan	Pontus Stenetorp
Michael Bloodgood	Kevin Gimpel	Ani Nenkova	Brandon Stewart
Matko Bosnjak	Joao Graca	Dong Nguyen	Veselin Stoyanov
Chris Brockett	Yvette Graham	Brendan O'Connor	Anders Søgaard
Miles Brundage	Keith Hall	Diarmuid O'Seaghda	Ivan Titov
Joana J. Bryson	Oul Han	Miles Osborne	Sara Tonelli
Ryan Calo	Graeme Hirst	Jahna Otterbacher	Oren Tsur
Marine Carpuat	Nathan Hodas	Sebastian Padó	Yulia Tsvetkov
Yejin Choi	Kristy Hollingshead	Alexis Palmer	Lyle Ungar
Munmun De Choudhury	Ed Hovy	Martha Palmer	Suresh Venkatasubramanian
Grzegorz Chrupala	Georgy Ishmaev	Michael Paul	Yannick Versley
Ann Clifton	Jing Jiang	Ellie Pavlick	Aline Villavicencio
Kevin B. Cohen	Anna Jobin	Emily Pitler	Andreas Vlachos
Shay B. Cohen	Anders Johannsen	Barbara Plank	Rob Voigt
Court Corley	David Jurgens	Thierry Poibeau	Svitlana Volkova
Ryan Cotterell	Brian Keegan	Chris Potts	Martijn Warnier
Aron Culotta	Roman Klinger	Vinod Prabhakaran	Zeera Waseem
Walter Daelemans	Ekaterina Kochmar	Daniel Preotiuc	Bonnie Webber
Dipanjan Das	Philipp Koehn	Nikolaus Pöschhacker	Joern Wuebker
Hal Daumé III	Zornitsa Kozareva	Will Radford	François Yvon
Steve DeNeefe	Jayant Krishnamurthy	Siva Reddy	Luke Zettlemoyer
Francien Dechesne	Jonathan K. Kummerfeld	Luis Reyes-Galindo	Janneke van der Zwaan
Leon Derczynski	Vasileios Lampos	Sebastian Riedel	
Aliya Deri	Angeliki Lazaridou	Ellen Riloff	
Mona Diab	Alessandro Lenci	Brian Roark	

Invited Speakers:

Graeme Hirst, University of Toronto (Canada)
 Quirine Eijkman, Leiden University (Netherlands)
 Jason Baldridge, People Pattern (USA)
 Joanna Bryson, University of Bath (UK)

Table of Contents

<i>Gender as a Variable in Natural-Language Processing: Ethical Considerations</i> Brian Larson	1
<i>These are not the Stereotypes You are Looking For: Bias and Fairness in Authorial Gender Attribution</i> Corina Koolen and Andreas van Cranenburgh.....	12
<i>A Quantitative Study of Data in the NLP community</i> Margot Mieskes	23
<i>Ethical by Design: Ethics Best Practices for Natural Language Processing</i> Jochen L. Leidner and Vassilis Plachouras.....	30
<i>Building Better Open-Source Tools to Support Fairness in Automated Scoring</i> Nitin Madnani, Anastassia Loukina, Alina von Davier, Jill Burstein and Aoife Cahill	41
<i>Gender and Dialect Bias in YouTube’s Automatic Captions</i> Rachael Tatman	53
<i>Integrating the Management of Personal Data Protection and Open Science with Research Ethics</i> Dave Lewis, Joss Moorkens and Kaniz Fatema	60
<i>Ethical Considerations in NLP Shared Tasks</i> Carla Parra Escartín, Wessel Reijers, Teresa Lynn, Joss Moorkens, Andy Way and Chao-Hong Liu 66	
<i>Social Bias in Elicited Natural Language Inferences</i> Rachel Rudinger, Chandler May and Benjamin Van Durme	74
<i>A Short Review of Ethical Challenges in Clinical Natural Language Processing</i> Simon Suster, Stephan Tulkens and Walter Daelemans.....	80
<i>Goal-Oriented Design for Ethical Machine Learning and NLP</i> Tyler Schnoebelen	88
<i>Ethical Research Protocols for Social Media Health Research</i> Adrian Benton, Glen Coppersmith and Mark Dredze	94
<i>Say the Right Thing Right: Ethics Issues in Natural Language Generation Systems</i> Charese Smiley, Frank Schilder, Vassilis Plachouras and Jochen L. Leidner.....	103

Workshop Program

Tuesday, 4 April, 2017

9:30–11:00 Morning Session 1

9:30–9:40 *Welcome, Overview*
Dirk, Margaret, Shannon, Michael

9:40–10:15 *Invited Talk*
Graeme Hirst

10:15–10:50 *Invited Talk*
Joanna Bryson

11:00–11:30 Coffee Break

11:30–13:00 Morning Session 2 - Gender

11:30–11:45 *Gender as a Variable in Natural-Language Processing: Ethical Considerations*
Brian Larson

11:45–12:00 *These are not the Stereotypes You are Looking For: Bias and Fairness in Authorial Gender Attribution*
Corina Koolen and Andreas van Cranenburgh

12:00–12:25 *Paper Discussion*
Authors and Discussant

12:25–13:00 *Invited Talk*
Quirine Eijkman

Tuesday, 4 April, 2017 (continued)

13:00–14:30 Lunch Break

14:30–16:00 Afternoon Session 1 - Data and Design

14:30–15:05 *Invited Talk*
Jason Baldrige

15:05–15:20 *A Quantitative Study of Data in the NLP community*
Margot Mieskes

15:20–15:35 *Ethical by Design: Ethics Best Practices for Natural Language Processing*
Jochen L. Leidner and Vassilis Plachouras

15:35–16:00 *Paper Discussion*
Authors and Discussant

16:00–17:00 Afternoon Session 2 - Coffee and Posters

16:00–17:00 *Building Better Open-Source Tools to Support Fairness in Automated Scoring*
Nitin Madnani, Anastassia Loukina, Alina von Davier, Jill Burstein and Aoife Cahill

16:00–17:00 *Gender and Dialect Bias in YouTube’s Automatic Captions*
Rachael Tatman

16:00–17:00 *Integrating the Management of Personal Data Protection and Open Science with Research Ethics*
Dave Lewis, Joss Moorkens and Kaniz Fatema

16:00–17:00 *Ethical Considerations in NLP Shared Tasks*
Carla Parra Escartín, Wessel Reijers, Teresa Lynn, Joss Moorkens, Andy Way and Chao-Hong Liu

16:00–17:00 *Social Bias in Elicited Natural Language Inferences*
Rachel Rudinger, Chandler May and Benjamin Van Durme

16:00–17:00 *A Short Review of Ethical Challenges in Clinical Natural Language Processing*
Simon Suster, Stephan Tulkens and Walter Daelemans

Tuesday, 4 April, 2017 (continued)

16:00–17:00 *Goal-Oriented Design for Ethical Machine Learning and NLP*
Tyler Schnoebelen

16:00–17:00 *Ethical Research Protocols for Social Media Health Research*
Adrian Benton, Glen Coppersmith and Mark Dredze

16:00–17:00 *Say the Right Thing Right: Ethics Issues in Natural Language Generation Systems*
Charese Smiley, Frank Schilder, Vassilis Plachouras and Jochen L. Leidner

17:00–18:00 Evening Session

17:10–17:45 *Panel Discussion*
Panelists

17:45–18:00 *Concluding Remarks*
Dirk, Margaret, Shannon, Michael

