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  Fernado 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 Madnani  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 Schuler
Mohit Bansal Laura Fichtner Kathy McKeown  H. Andrew Schwartz
Solon Barocas Karén Fort Paola Merlo  Hinrich Schütze
Daniel Bauer Victoria Fossum David Minno  Djamé Seddah
Eric Bell Lily Frank Shachar Mirkin  Dan Simonson
Steve 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 Stanetorp
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’Seaghdha  Ivan Titov
Joana J. Bryson Oul Han Miles Osborne  Sara Tonelli
Ryan Calo Graeme Hirst Jahan 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 Pitter  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  Martin Warner
Aron Culotta Roman Kummerfeld Vinod Prabhakaran  Zeerak Waseem
Walter Daelemans Ekaterina Kochmar Daniel Preotiuc  Bonnie Webber
Dipanjan Das Philipp Koehn Nikolaus Pöchhacker  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

**A Quantitative Study of Data in the NLP community**  
Margot Mieskes .......................... 1

**Ethical by Design: Ethics Best Practices for Natural Language Processing**  
Jochen L. Leidner and Vassilis Plachouras ................................. 8

**These are not the Stereotypes You are Looking For: Bias and Fairness in Authorial Gender Attribution**  
Corina Koolen and Andreas van Cranenburgh .......................... 19

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

**Building Better Open-Source Tools to Support Fairness in Automated Scoring**  
Nitin Madnani, Anastassia Loukina, Alina von Davier, Jill Burstein and Aoife Cahill .......... 41

**Gender and Dialect Bias in YouTube’s Automatic Captions**  
Rachael Tatman .......................... 53

**Integrating the Management of Personal Data Protection and Open Science with Research Ethics**  
Dave Lewis, Joss Moorkens and Kaniz Fatema .......................... 60

**Ethical Considerations in NLP Shared Tasks**  
Carla Parra Escartín, Wessel Reijers, Teresa Lynn, Joss Moorkens, Andy Way and Chao-Hong Liu 66

**Social Bias in Elicited Natural Language Inferences**  
Rachel Rudinger, Chandler May and Benjamin Van Durme .......................... 74

**A Short Review of Ethical Challenges in Clinical Natural Language Processing**  
Simon Suster, Stephan Tulkens and Walter Daelemans .......................... 80

**Goal-Oriented Design for Ethical Machine Learning and NLP**  
Tyler Schnoebelen .......................... 88

**Ethical Research Protocols for Social Media Health Research**  
Adrian Benton, Glen Coppersmith and Mark Dredze .......................... 94

**Say the Right Thing Right: Ethics Issues in Natural Language Generation Systems**  
Charese Smiley, Frank Schilder, Vassilis Plachouras and Jochen L. Leidner .......................... 103
Workshop Program

Tuesday, 4 April, 2017

9:30–11:00  Morning Session I - Invited Talks

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 II - Data and Design, Invited Talk

11:30–11:45  A Quantitative Study of Data in the NLP community
Margot Mieskes

11:45–12:00  Ethical by Design: Ethics Best Practices for Natural Language Processing
Jochen L. Leidner and Vassilis Plachouras

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 I - Gender, Invited Talk

14:30–15:05  Invited Talk
Jason Baldridge

15:05–15:20  These are not the Stereotypes You are Looking For: Bias and Fairness in Authorial Gender Attribution
Corina Koolen and Andreas van Cranenburgh

15:20–15:35  Gender as a Variable in Natural-Language Processing: Ethical Considerations
Brian Larson

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

16:00–17:00  Afternoon Session II - 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 - Panel**

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

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