TrustNLP 2023

The Third Workshop on Trustworthy Natural Language Processing

Proceedings of the Workshop (TrustNLP 2023)

July 14, 2023

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

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 978-1-959429-86-9

Introduction

We welcome all participants of TrustNLP 2023, the third Workshop on Trustworthy Natural Language Processing. This year, we are embracing a hybrid format for the workshop, scheduled for July 14, 2023, and is co-located with ACL 2023.

Recent advances in Natural Language Processing, and the emergence of pretrained Large Language Models (LLM) specifically, have made NLP systems omnipresent in various aspects of our everyday life. In addition to traditional examples such as personal voice assistants, recommender systems, etc, more recent developments include content-generation models such as ChatGPT, text-to-image models (Dall-E), and so on. While these emergent technologies have an unquestionable potential to power various innovative NLP and AI applications, they also pose a number of challenges in terms of their safe and ethical use.

In response to these challenges, NLP researchers have formulated various objectives, e.g., intended to make models more fair, safe, and privacy-preserving. However, these objectives are often considered separately, which is a major limitation since it is often important to understand the interplay and/or tension between them. For instance, meeting a fairness objective might require access to users' demographic information, which creates tension with privacy objectives. The goal of this workshop is to move toward a more comprehensive notion of Trustworthy NLP, by bringing together researchers working on those distinct yet related topics, as well as their intersection.

Our agenda features four keynote speeches, a panel session, a presentation session, and a poster session. This year, we were delighted to receive 57 submissions, out of which 41 papers were accepted. Among these, 28 have been included in our proceedings. These papers span a wide array of topics including fairness, robustness, factuality, privacy, explainability, and model analysis in NLP.

We would like to express our gratitude to all the authors, committee members, keynote speakers, panelists, and participants. We also gratefully acknowledge the generous sponsorship provided by Amazon.

Program Committee

Organizers

Ninareh Mehrabi, Amazon Alexa AI
Anaelia Ovalle, University of California Los Angeles
Trista Cao, University of Maryland
Jwala Dhamala, Amazon Alexa AI
Apurv Verma, Amazon Alexa AI
Anoop Kumar, Amazon Alexa AI
Yada Pruksachatkun, Infinitus Systems
Aram Galystan, University of Southern California, Amazon Alexa AI
Rahul Gupta, Amazon Alexa AI
Kai-Wei Chang, University of California Los Angeles, Amazon Alexa AI

Program Committee

Griffin Adams, Columbia University
Stefan Arnold, FAU Erlangen-Nürnberg
Connor Baumler, University of Maryland
Keith Burghardt, USC Information Sciences Institute
Yang Trista Cao, University of Maryland
Jwala Dhamala, Amazon Alexa AI-NLU
Jacob Eisenstein, Google
Katja Filippova, Google
Aram Galstyan, USC Information Sciences Institute
Umang Gupta, University of Southern California
Devamanyu Hazarika, Amazon
Zihao He, University of Southern California
William Held, Georgia Tech
Qian Hu, Amazon.com
Fatemah Husain, Kuwait University
Anoop Kumar, Amazon
Sasha Luccioni, Hugging Face
Pranav Narayanan Venkit, Pennsylvania State University
Isar Nejadgholi, National Research Council Canada
Aishwarya Padmakumar, Amazon
Ashwinee Panda, Princeton University
Anirudh Raju, Amazon, Alexa
Anthony Rios, University of Texas at San Antonio
Robik Shrestha, RIT
Anna Sotnikova, University of Maryland
Arjun Subramonian, University of California, Los Angeles
Jialu Wang, University of California, Santa Cruz
Chhavi Yadav, UCSD
Kiyoon Yoo, Seoul National University

Table of Contents

Program

Friday, July 14, 2023

09:00 - 09:10 *Opening Remarks*

09:10 - 09:50 *Keynote 1 Hal Daume III*

09:50 - 10:30 *Keynote 2 Ramprasaath R. Selvaraju*

10:30 - 11:00 *Break*

11:00 - 11:40 *Keynote 3 Rachel Rudinger*

11:40 - 12:30 *Poster Session*

12:30 - 14:00 *Lunch*

14:00 - 14:40 *Keynote 4 Sunipa Dev*

14:40 - 15:30 *Panel Discussion*

15:30 - 16:00 *Break*

16:00 - 17:10 *Oral Presentation*

17:10 - 17:30 *Closing Session/Best Paper Announcement*