

C3NLP 2024

The 2nd Workshop on Cross-Cultural Considerations in NLP

Proceedings of the Workshop

August 16, 2024

The C3NLP organizers gratefully acknowledge the support from the following sponsors.

Platinum



Gold



©2024 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-146-9

Introduction

Natural Language Processing has seen impressive gains in recent years. This research includes the demonstration by NLP models to have turned into useful technologies with improved capabilities, measured in terms of how well they match human behavior captured in web-scale language data or through annotations. However, human behavior is inherently shaped by the cultural contexts humans are embedded in, the values and beliefs they hold, and the social practices they follow, part of which will be reflected in the data used to train NLP models, and the behavior these NLP models exhibit. Not accounting for this factor could cause incongruencies and misalignments between the cultural contexts that underpin the NLP model development process and the multi-cultural ecosystems they are expected to operate in. These misalignments may result in various harms, including barriers to those from under-represented cultures, violating cultural norms and values, and erasure of cultural knowledge.

While recent work in the field has started to acknowledge this issue, it is important to build a long-term research agenda for the NLP community around (1) deeper understanding of how global cultures and NLP technologies intersect, in a way that goes beyond multi-lingual and cross-lingual research, (2) how to detect, measure, and attempt to mitigate potential biases and harms in NLP technology in ways that reflect local cultures and values, and (3) how to build more cross-culturally competent NLP systems. This agenda requires looking beyond the NLP community, bringing in multi-disciplinary expertise to shape the inquiries in this important area.

We introduce the workshop on Cross-Cultural Considerations in NLP as a platform to bring together the growing number of NLP researchers interested in this topic, along with a community of scholars with multi-disciplinary expertise spanning linguistics, social sciences, and cultural anthropology. Our aim is to build this important inquiry within NLP on a solid basis of cultural theories from social sciences. To this end, the workshop program will focus on the following themes: Inclusivity and Representation of cultures in NLP, Cultural harms of NLP technologies, and Culture Sensitive lens on Social Biases and Harms in NLP.

In the interest of having a broad conversation, inclusive of different disciplinary norms, we invited submissions of different kinds. Authors were able to choose between: (1) archival papers which will be published in the C3NLP proceedings as well as presented during the workshop, and (2) non-archival papers which are not published in the proceedings but are given a presentation slot during the workshop. Archival papers may be long (up to 9 pages) or short (up to 5 pages), and went through mutually anonymous peer review by our program committee members or were already reviewed through ACL Rolling Review (ARR). Non-archival papers include extended abstracts which were also subjected to mutually anonymous peer review by our program committee, or papers that were already reviewed through ARR or accepted for publication at another peer-reviewed venue.

We received 27 direct submissions and 6 submissions through ARR. We accepted 16 of the direct submissions (8 short and 8 long, 9 archival and 7 non-archival), and 5 of the ARR submissions (1 short and 4 long, all of which were non-archival). In addition, our program includes presentations of selected papers on this topic accepted at other venues, two interdisciplinary panel discussions with six experts on various topics, as well as an industry panel discussion.

We welcome you to the 2rd Workshop on Cross-Cultural Considerations in NLP. We are grateful to our program committee for their in-depth and constructive reviews, and to our authors who sent impressive cutting-edge research on this topic to our workshop. We look forward to a day filled with thought-provoking discussions and seeds for future collaborations.

– Vinodkumar Prabhakaran, Sunipa Dev, Luciana Benotti, Daniel Hershcovich, Laura Cabello, Yong Cao, Ife Adebare, Li Zhou

Organizing Committee

Vinodkumar Prabhakaran, Google Research

Sunipa Dev, Google Research

Luciana Benotti, Universidad Nacional de Córdoba

Daniel Hershcovich, University of Copenhagen

Yong Cao, Huazhong University of Science and Technology

Li Zhou, The Chinese University of Hong Kong, Shenzhen

Laura Cabello, University of Copenhagen

Ife Adebara, The University of British Columbia

Program Committee

Luis Chiruzzo, Universidad de la República
Marie-Catherine De Marneffe, UCLouvain
Lucie-Aimée Kaffee, Hugging Face
François Yvon, ISIR, Sorbonne Université & CNRS
Valerio Basile, University of Turin
David Schlangen, University of Potsdam
Teresa Lynn, Mohamed bin Zayed University of Artificial Intelligence
Nisansa De Silva, University of Moratuwa
Agrima Seth, University of Michigan - Ann Arbor
Alice Oh, Korea Advanced Institute of Science and Technology
Shaily Bhatt, Carnegie Mellon University
Michael Bloodgood, The College of New Jersey
Nikola Ljubešić, Jožef Stefan Institute
Surangika Ranathunga, Massey University
Günter Neumann, German Research Center for AI
Roberto Navigli, Sapienza University of Rome
Steven R Wilson, Oakland University (Michigan)
Barbara Plank, Ludwig-Maximilians-Universität München and IT University of Copenhagen
Roman Klinger, Otto-Friedrich Universität Bamberg
Graeme Hirst, University of Toronto
Rada Mihalcea, University of Michigan
Aubrie Amstutz, Apple
Diana Maynard, University of Sheffield
Partha Talukdar, Google Research and Indian Institute of Science, Bangalore
Vivek Kulkarni, Grammarly
Laura Alonso Alemany, Universidad Nacional de Córdoba
Philipp Koehn, Johns Hopkins University
Akshita Jha, Virginia Tech
Mark Diaz, Google
Seyed Abolghasem Mirroshandel, University of Guilan
Siddhesh Milind Pawar, University of Copenhagen
Arnav Arora, University of Copenhagen

Keynote Talk

Alice Oh

Korea Advanced Institute of Science Technology



Bio: Alice Oh (Korea Advanced Institute of Science & Technology) is a Professor in the School of Computing at KAIST. She received her MS in 2000 from Carnegie Mellon University and PhD in 2008 from MIT. Her major research area is at the intersection of natural language processing (NLP) and computational social science. She collaborates with social scientists to study topics such as political science, education, and history, developing NLP models for various textual data including legislative bills, historical documents, news articles, social media posts, and personal conversations. She has served as a Tutorial Chair for NeurIPS 2019, Diversity & Inclusion Chair for ICLR 2019, Program Chair for ICLR 2021, Senior Program Chair for NeurIPS 2022, and General Chair for NeurIPS 2023.

Keynote Talk

Diyi Yang
Stanford University



Bio: Diyi Yang (Stanford University) is an assistant professor in the Computer Science Department at Stanford University, also affiliated with the Stanford NLP Group, Stanford HCI Group and Stanford Human Centered AI Institute. Her research focuses on human-centered natural language processing and computational social science. She is a recipient of IEEE “AI 10 to Watch” (2020), Microsoft Research Faculty Fellowship (2021), NSF CAREER Award (2022), an ONR Young Investigator Award (2023), and a Sloan Research Fellowship (2024). Her work has received multiple paper awards or nominations at top NLP and HCI conferences, (e.g., Best Paper Honorable Mention at SIGCHI 2019 and Outstanding Paper at ACL 2022).

Keynote Talk

Kalika Bali

Microsoft Research Labs India



Bio: Kalika Bali (Microsoft Research Labs India) is a Principal Researcher at Microsoft Research Labs India, where she has dedicated nearly two decades to enhancing human-computer interactions through language technologies. Her focus lies in creating inclusive tech for a diverse range of languages and communities, especially those that are underrepresented. She is particularly interested in how Foundational Models like GPT can impact society, for better or worse. Her recent work navigates the crossroads of multilingual and multicultural AI. She was on the first (2023) TIME100 AI list for her continuing work on breaking down language barriers and fostering inclusivity in the AI sphere.

Keynote Talk

Luis Chiruzzo

Universidad de la República



Bio: Luis Chiruzzo (Universidad de la República) is an associate professor at Universidad de la República, Uruguay. He studied Computer Science Engineering at Universidad de la República, and has a MSc. and a PhD. in Computer Science from Pedeciba - Universidad de la República. He belongs to the Uruguayan National System of Researchers (SNI). His main research interests include NLP and machine translation for low-resource languages, in particular for the indigenous language Guaraní, sign language processing, uses of NLP in education, sentiment and humor analysis, and parsing. He has been a collaborator with the AmericasNLP initiative to promote NLP research for indigenous languages of the Americas since 2021, and co-organized the AmericasNLP workshop in 2024.

Keynote Talk

Shalom H. Schwartz
The Hebrew University



Bio: Shalom H. Schwartz (The Hebrew University) is Professor Emeritus of Psychology—the Hebrew University of Jerusalem and a past president of the International Association for Cross-Cultural Psychology. He has spent the last 40 years seeking to identify the basic human values that are recognized across cultures, to understand the principles that organize values into coherent systems, to develop cross-culturally valid instruments to measure values, and to uncover the many ways that values relate to human behavior and attitudes. His theory of basic values and various measurement instruments have been applied in research in more than 90 countries.

Keynote Talk

Xun Wu

Hong Kong University of Science and Technology



Bio: Xun Wu (Hong Kong University of Science and Technology) is a policy scientist with a strong interest in the linkage between policy analysis and public management. Trained in engineering, economics, public administration, and policy analysis, his research seeks to make contribution to the design of effective public policies in dealing emerging policy challenges related to the applications of disruptive technologies. His research interests include science and technology policy, policy innovations, water resource management, health policy reform, and anti-corruption. He is currently a professor at the Hong Kong University of Science and Technology (Guangzhou).

Table of Contents

<i>CDEval: A Benchmark for Measuring the Cultural Dimensions of Large Language Models</i> Yuhang Wang, Yanxu Zhu, Chao Kong, Shuyu Wei, Xiaoyuan Yi, Xing Xie and Jitao Sang	1
<i>Conformity, Confabulation, and Impersonation: Persona Inconstancy in Multi-Agent LLM Collaboration</i> Razan Baltaji, Babak Hemmatian and Lav R. Varshney	17
<i>Synchronizing Approach in Designing Annotation Guidelines for Multilingual Datasets: A COVID-19 Case Study Using English and Japanese Tweets</i> Kiki Ferawati, Wan Jou She, Shoko Wakamiya and Eiji Aramaki	32
<i>CRAFT: Extracting and Tuning Cultural Instructions from the Wild</i> Bin Wang, Geyu Lin, Zhengyuan Liu, Chengwei Wei and Nancy F. Chen	42
<i>Does Cross-Cultural Alignment Change the Commonsense Morality of Language Models?</i> Yuu Jinnai	48
<i>Do Multilingual Large Language Models Mitigate Stereotype Bias?</i> Shangrui Nie, Michael Fromm, Charles Welch, Rebekka Görge, Akbar Karimi, Joan Plepi, Nazia Afsan Mowmita, Nicolas Flores-Herr, Mehdi Ali and Lucie Flek	65
<i>Sociocultural Considerations in Monitoring Anti-LGBTQ+ Content on Social Media</i> Sidney Gig-Jan Wong	84
<i>Are Generative Language Models Multicultural? A Study on Hausa Culture and Emotions using Chat-GPT</i> Ibrahim Said Ahmad, Shiran Dudy, Resmi Ramachandranpillai and Kenneth Church	98
<i>Computational Language Documentation: Designing a Modular Annotation and Data Management Tool for Cross-cultural Applicability</i> Alexandra O’Neil, Daniel Glen Swanson and Shobhana Lakshmi Chelliah	107