NLPerspectives 2025

The 4th Workshop on Perspectivist Approaches to NLP

Proceedings of the Workshop

November 8, 2025

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

Silver

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 979-8-89176-350-0

Introduction

The NLPerspectives workshop brings together researchers and practitioners to explore how Natural Language Processing (NLP) systems can better capture, represent, and engage with diverse perspectives. Building on the prior workshops of NLPerspectives, this year's 4th edition workshop centered on the challenges and opportunities of representing multiple, and sometimes conflicting, viewpoints in annotation, modeling, and evaluation. Contributions spanned theoretical frameworks, empirical studies, and methodological innovations, highlighting both crowd-truth approaches to data annotation and novel techniques for integrating plurality into NLP models. The program featured research papers, a position paper, and hosted a shared task called Learning with Disagreement. This proceedings volume documents the range of ideas and findings presented at NLPerspectives, including the shared task, offering insight into the state of the field and charting directions for future work at the intersection of language technology, social values, and human-centered AI.

Until recently, language resources supporting many tasks in Natural Language Processing (NLP) and other areas of Artificial Intelligence (AI) have been based on the assumption of a single 'ground truth' label sought via aggregation, adjudication, or statistical means. However, the field is increasingly focused on subjective and controversial tasks, such as quality estimation or abuse detection, in which multiple points of view may be equally valid; subjectivity, indeed, is considered one of the main causes of Human Variation.

The fourth edition of the workshop builds upon these ideas to explore interdisciplinary synergies throughout the programme, starting with the keynote speaker Dr. Jose Camacho-Collados. In particular, this event is aimed at widening the discussed methodology to include not only current and ongoing work on collecting and labelling non-aggregated datasets, but also approaches to mining, modelling and inclusion of diverse perspectives in data, evaluation, and applications of multi-perspective Machine Learning models. In addition, it involved techniques from social science and Human-Computer Interaction, such as participatory approaches and how they can be implemented at all stages of the supervised learning pipeline.

Working with language as data poses unique challenges. Words and their definitions evolve over time, and even within the same time period, interpretation of language is contextual, influenced by the cultural, geographic and linguistic environment of people. Bringing interdisciplinary approaches from Feminist theories, Critical Discourse Analysis, and indigenous epistemologies, among others, to Computational Linguistics can provide valuable methods for working with the subjectivity and uncertainty of language data.

Besides the community of data perspectivism and human label variation, we expect the workshop to attract researchers and industry practitioners, as happened in previous editions, interested in learning from disagreement, personalization, and participatory design. For the first time, the workshop has hosted a shared task, Learning with Disagreement (Le-Wi-Di), which explores new approaches to modelling and evaluation of perspectivist data.

The Shared Task on Learning with Disagreement (LeWiDi) aims to raise the visibility of the challenge of variation in interpretation, and to encourage the community to engage with the problem. The objective of the shared task is to provide a unified testing framework for learning from disagreements and evaluating with such datasets. Two editions of the shared task were organized as part of SEMEVAL: in 2021, focused on ambiguity in language and vision; and in 2023, focused on disagreement in subjective tasks. These shared tasks created benchmarks that have since been widely downloaded, and started an ongoing discussion on how to evaluate perspectivist models. This new edition, co-located within the workshop, differs from the previous ones in: the inclusion of two new classification tasks in which disagreements are prevalent, one on irony, and one on Natural Language Inference; the addition of two generative tasks, paraphrasing and summarization; testing new approaches to evaluation, inspired by recent research.

In its fourth edition, the workshop accepted 14 submissions—13 research papers and one stance paper, with one designated as non-archival. Additionally, 9 teams submitted successful entries to the shared task.

One of the primary motivations for the pursuit of perspectivism is expanding the number of voices represented in NLP datasets, modelling, and evaluation. The workshop therefore has diversity and inclusion as one of its central tenets.

As at previous events, this edition of NLPerspectives will host presenters, panellists, and keynote speakers representing diverse demographics, research backgrounds, and career stages. Our organizing and programme committee also includes a balance of geographical locations, genders, and career stages.

For access, the most important contributions and available resources are listed and reported in the manifesto page of perspectivist data (https://pdai.info/) and proceedings and available slides of previous editions are online on our website: https://nlperspectives.di.unito.it/.

To encourage people coming from various countries to join the discussion, the workshop is planned to be in a hybrid format. As in the previous editions, we will provide financial aid to scholars (especially students) who may not otherwise be able to attend.

Organizing Committee

Organizing Committee

Gavin Abercrombie, Heriot-Watt University
Valerio Basile, University of Turin
Davide Bernardi, Amazon Alexa
Shiran Dudy, Northeastern University
Simona Frenda, University of Turin
Sara Tonelli, Fondazione Bruno Kessler

Shared Task Organizers

Silvia Casola, LMU Munich, Germany
Elisabetta Fersini, University of Milan Bicocca, Italy
Elisa Leonardelli, Bruno Kessler Foundation, Italy
Maja Pavlovic, Queen Mary University, London, UK
Massimo Poesio, Queen Mary University, London, UK; University of Utrecht, Netherlands
Barbara Plank, LMU Munich, Germany
Alexandra Uma, Builder AI, UK
Siyao Peng, LMU Munich, Germany

Program Committee

Chairs

Gavin Abercrombie, Heriot Watt University
Valerio Basile, University of Turin
Shiran Dudy, Northeastern University
Simona Frenda, Heriot-Watt University
Sara Tonelli, FBK

Program Committee

Riza Batista - Navarro, Department of Computer Science, The University of Manchester
Davide Bernardi, Amazon
Agostina Calabrese, The University of Edinburgh
Silvia Casola, LMU Munich
Amanda Cercas Curry, Bocconi University
Teddy Ferdinan, Wrocław University of Science and Technology
Samuel Goree, Stonehill College
Marco Guerini, Fondazione Bruno Kessler
Nancie Gunson, Heriot-Watt University
Annette H a u t l i - J a n i s z, University of Passau
Cassandra L. Jacobs, University at Buffalo
Aiqi Jiang, Heriot Watt University
Kamil Kanclerz, Wrocław University of Science and Technology
Anna Koufakou, Florida Gulf Coast University
Sofie Labat, Ghent University
Marta Marchiori Manerba, Università di Pisa
Michele Mastromattei, University of Rome Tor Vergata
Massimo Poesio, Queen Mary University of London and University of Utrecht
Julia Romberg, GESIS – Leibniz Institute for the Social Sciences
Pratik Sachdeva, University of California, Berkeley
Manuela Sanguinetti, University of Cagliari, Department of Mathematics and Computer Science
Erhan Sezerer, Amazon Research
Tiago Timponi Torrent, Federal University of Juiz de Fora

Keynote Talk

Cultural Awareness in Multilingual Language Models - A Perspectivist Personal Perspective

Jose Camacho Collados
Cardiff University
2020-11-08 12:30:00 – Room: **Room 1**

Abstract: Language models have become ubiquitous in NLP and beyond. In particular, the new wave of large language models (LLMs) are increasingly used to communicate and solve practical problems in many languages and countries, and by an increasingly diverse set of users. However, even though there is no doubt that these models open up plenty of opportunities, there are important issues and research questions that arise when it comes to LLMs and their application in different languages and cultures. For instance, the language coverage in language models drastically decreases for less-resourced languages and as such, their performance. And not only the general performance is affected, but general-purpose LLMs may be implicitly biased to specific cultures and languages depending on their underlying training data.
In this talk, I will discuss how language models reflect on cultural diversity, including potential shortcomings and how language coverage and cultural awareness may be intrinsically intertwined. I will also share some lessons learned based on recent research in this area – in particular, I will focus on the development of BLEnD, a large effort to develop a cultural benchmark of everyday knowledge for dozens of languages and countries.

Bio: Jose Camacho-Collados is a UKRI Future Leaders Fellow and Professor at the School of Computer Science of Cardiff University, where he co-founded the Cardiff Natural Language Processing group (Cardiff NLP). Before joining Cardiff University, he completed his PhD in Sapienza University of Rome and was a Google AI PhD Fellow.
Jose has worked in multiple NLP areas with a particular focus on semantics, multilinguality and computational social science with an interdisciplinary perspective. In this area, he has been developing specialised and efficient NLP models for social media applications, such as TweetNLP and related efforts. His work has received several recognitions, including awards at top NLP conferences, or the 2023 AIJ Prominent Paper Award. He is also the co-author of the "Embeddings in Natural Language Processing" book.

Table of Contents