

MME 2026

Workshop on Multilingual and Multicultural Evaluation

Proceedings of the Workshop

March 28, 2026

©2026 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 979-8-89176-368-5

Introduction

The First Workshop on Multilingual and Multicultural Evaluation (MME) will be held on March 28, 2026, in Rabat, Morocco, co-located with the 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026).

While large language models (LLMs) are reshaping language technology across language and cultural barriers, the scientific community lacks a comprehensive understanding of their performance across languages and cultural contexts. Simple approaches such as counting languages in training data or relying on translated evaluation benchmarks are not sufficient.

The MME workshop is established to tackle these issues by creating a dedicated venue for multilingual evaluation resources and metrics, standardizing practices to ensure accuracy and fairness, and integrating cultural and social perspectives into NLP and LLM evaluation.

We received 27 archival and 20 non-archival submissions. Archival papers underwent double-blind peer review, while non-archival submissions received a lightweight relevance check. The program committee accepted 15 archival papers for publication in the ACL Anthology, representing a 55.6% acceptance rate. In total, the workshop comprises 43 presentations, including the archival papers, 19 accepted non-archival submissions, and 9 papers accepted at other venues. The workshop also features three invited talks from experts in the field: Sebastian Ruder (Meta), Freda Shi (University of Waterloo), and Wenda Xu (Google DeepMind).

We thank all authors for their scholarly contributions and the program committee for their reviewing efforts. We extend our gratitude to the EACL 2026 Organizing Committee, particularly the Workshop Chairs, for their support. Finally, we thank all participants for making the inaugural MME workshop a success.

The MME Organizers

Pinzhen Chen, Vilém Zouhar, Hanxu Hu, Simran Khanuja, Wenhao Zhu, Barry Haddow, Alexandra Birch, Alham Fikri Aji, Rico Sennrich, and Sara Hooker

Organizing Committee

Organizers

Pinzhen Chen, Queen's University Belfast

Vilém Zouhar, ETH Zurich

Hanxu Hu, University of Zurich

Simran Khanuja, Carnegie Mellon University

Wenhao Zhu, ByteDance

Barry Haddow, University of Edinburgh

Alexandra Birch, University of Edinburgh

Alham Fikri Aji, Mohamed bin Zayed University of Artificial Intelligence

Rico Sennrich, University of Zurich

Sara Hooker, Adaptable Intelligence

Program Committee

Program Chairs

Pinzhen Chen, Queen's University Belfast
Vilém Zouhar, Department of Computer Science, ETHZ - ETH Zurich
Hanxu Hu, Microsoft Research
Simran Khanuja, CMU, Carnegie Mellon University
Wenhao Zhu, ByteDance Inc.

Reviewers

Niyati Bafna, Houda Bouamor, Laurie Burchell
Yogen Vilas Chaudhari, Pinzhen Chen
Peter Devine, Koel Dutta Chowdhury, Ona de Gibert
Shehenaz Hossain, Songbo Hu, Xu Huang
Aishwarya Jadhav
Ahrii Kim, Mateusz Klimaszewski
Junxiao Liu
Bhavivya Malik, Nikita Moghe
Dayyán O'Brien
Stefano Perrella, Esther Ploeger, Lorenzo Proietti
Shenbin Qian
A.b.m. Ashikur Rahman, Jacqueline Rowe
Sherrie Shen
Ashok Urlana
Zhijun Wang
Nikita Kiran Yeole
Zheng Zhao, Vilém Zouhar

Table of Contents

<i>LLMs as Span Annotators: A Comparative Study of LLMs and Humans</i> Zdeněk Kasner, Vilém Zouhar, Patrícia Schmidtová, Ivan Kartáč, Kristýna Onderková, Ondrej Platek, Dimitra Gkatzia, Saad Mahamood, Ondrej Dusek and Simone Balloccu	1
<i>On the Credibility of Evaluating LLMs using Survey Questions</i> Jindřich Libovický	23
<i>An improved Code-Switching Detection System for some Indic Languages</i> Karan Bhanushali and Fritz Hohl	35
<i>Vinclat: Evaluating Reasoning, Cognition and Culture in One Game</i> Marc Pàmies, Javier Aula-Blasco, Aitor Gonzalez-Agirre and Marta Villegas	49
<i>Conceptual Cultural Index: A Metric for Cultural Specificity via Relative Generality</i> Takumi Ohashi and Hitoshi Iyatomi	67
<i>The Anthropology of Food: How NLP can Help us Unravel the Food cultures of the World</i> Arij Riabi, Sougata Saha and Monojit Choudhury	76
<i>LLM-as-a-qualitative-judge: automating error analysis in natural language generation</i> Nadezhda Chirkova, Tunde Oluwaseyi Ajayi, Seth Aycok, Zain Muhammad Mujahid, Vladana Perlić, Ekaterina Borisova and Markarit Vartampetian	99
<i>Cross-Lingual Stability of LLM Judges Under Controlled Generation: Evidence from Finno-Ugric Languages</i> Isaac Chung and Linda Freienthal	133
<i>Cross-lingual and cross-country approaches to argument component detection: a comparative study.</i> Cecilia Graiff, Chloé Clavel and Benoît Sagot	149
<i>UNSC-Bench: Evaluating LLM Diplomatic Role-Playing Through UN Security Council Vote Prediction</i> Ayush Nangia, Aman Gokrani and Ruggero Marino Lazzaroni	162
<i>Leveraging Wikidata for Geographically Informed Sociocultural Bias Dataset Creation: Application to Latin America</i> Yannis Karmim, Renato Pino, Hernan Contreras, Hernan Lira, Sebastian Cifuentes, Simon Escofier, Luis Martí, Djamá Seddah and Valentin Barriere	177
<i>Whom to Trust? Analyzing the Divergence Between User Satisfaction and LLM-as-a-Judge in E-Commerce RAG Systems</i> Arif Türkmen and Kaan Efe Keleş	189
<i>Query-Following vs Context-Anchoring: How LLMs Handle Cross-Turn Language Switching</i> Kyuhee Kim, Chengheng Li Chen and Anna Sotnikova	196
<i>Generating Difficult-to-Translate Texts</i> Vilém Zouhar, Wenda Xu, Parker Riley, Juraj Juraska, Mara Finkelstein, Markus Freitag and Daniel Deutsch	204
<i>'A Woman is More Culturally Knowledgeable than A Man?': The Effect of Personas on Cultural Norm Interpretation in LLMs</i> Mahammed Kamruzzaman, Hieu Minh Nguyen, Nazmul Hassan and Gene Louis Kim	220