

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 2026) 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 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 15 archival papers, 19 accepted non-archival submissions, and 9 papers accepted at other venues. The workshop also features 3 invited talks from experts in academia and industry.

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.

MME 2026 Organizers

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 (ETH Zurich)
Hanxu Hu (University of Zurich)
Simran Khanuja (Carnegie Mellon University)
Wenhao Zhu (ByteDance)

Reviewers

Niyati Bafna (Johns Hopkins University)
Houda Bouamor (Carnegie Mellon University, Qatar)
Laurie Burchell (Common Crawl Foundation)
Yogen Vilas Chaudhari (PowerSchool)
Pinzhen Chen (Queen's University Belfast)
Koel Dutta Chowdhury (Saarland University)
Peter Devine (University of Edinburgh)
Ona de Gibert (University of Helsinki)
Shehenaz Hossain (Munster Technological University)
Songbo Hu (University of Cambridge)
Xu Huang (Nanjing University)
Aishwarya Jadhav (University of California, Berkeley)
Ahrii Kim (Soongsil University)
Mateusz Klimaszewski (Warsaw University of Technology)
Junxiao Liu (Nanjing University)
Bhavivya Malik (University of Edinburgh)
Nikita Moghe (Amazon)
Dayyán O'Brien (University of Edinburgh)
Stefano Perrella (Sapienza University of Rome)
Esther Ploeger (Aalborg University)
Lorenzo Proietti (Sapienza University of Rome)
Shenbin Qian (University of Oslo)
A B M Ashikur Rahman (King Fahad University of Petroleum and Minerals)
Jacqueline Rowe (University of Edinburgh)
Sherrie Shen (University of Edinburgh)
Ashok Urlana (International Institute of Information Technology, Hyderabad)
Zhijun Wang (Nanjing University)
Nikita Kiran Yeole (Virginia Tech)
Zheng Zhao (University of Edinburgh)
Vilém Zouhar (ETH Zurich)

Invited Talks

Invited Talks

Sebastian Ruder (Meta)

Freda Shi (University of Waterloo)

Shaomu Tan (University of Amsterdam)

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 Escoffier, Luis Martí, Djamé 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