

ORACLE 2026

**ORACLE 2026: Workshop on Open Reasoning Across
Cultures & Languages**

Proceedings of the Workshop

October 29, 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 979-8-89176-483-5

ORACLE 2026

Reasoning lies at the core of human intelligence: it allows us to connect entities, events, and quantities into coherent arguments. While the underlying logical capacity is universal, the ways in which people reason are shaped by the languages they speak and the cultural practices they adopt—whether in how numbers are conceptualised, how space is oriented, or how categories are formed. Despite such significance, applying contemporary Generative AI and Large Language Models (LLMs) to complex reasoning across diverse languages faces severe limitations. ORACLE represents a forum for discussing new ideas and directions to advance Open Reasoning Across Cultures & Languages through transparent methods, reusable resources, and reproducible evaluation, welcoming novel contributions on training strategies, architectures, evaluation methods and downstream applications.

Recent advances in Natural Language Processing (NLP) offer the opportunity to support complex reasoning capabilities at scale. The deployment of these methods can broaden technological access and address recurring gaps where typological variation, script change, and cultural context are currently poorly represented. Previous work demonstrates the efficacy of multilingual pre-training and post-training strategies in aligning capabilities across languages, mitigating cultural bias, and handling data scarcity in low-resource settings.

Nevertheless, technical challenges persist, particularly concerning the availability of robust evaluation protocols, model performance degradation under code-switching, and the comprehension of culturally mediated meanings. Critical factors such as calibration, robustness, and faithfulness continue to face linguistic barriers, necessitating rigorous and reproducible solutions to assess cross-lingual fidelity accurately.

For additional details about ORACLE 2026, please visit the website: <https://sites.google.com/view/workshoporacle/>.

Organizing Committee

General Chair

Leonardo Ranaldi, University of Edinburgh and University of Rome Tor Vergata

Program Chairs

Giulia Pucci, University of Aberdeen

Federico Ranaldi, University of Rome Tor Vergata

Marco Valentino, University of Sheffield

Alexandra Birch, University of Edinburgh

Fabio Massimo Zanzotto, University of Rome Tor Vergata and Almagest Labs

Program Committee

Program Committee

Leonardo Arcangeli
Krishna Aswani
Rita Berrada
Alessia Fidelibus
Carlotta Foschini
Svetlana Gorovaia
Volha Harytskaya
Muhammad Junaid Iqbal
Alina Kapanova
Madhav Kapoor
Anna Kołos
Julia Kreutzer
Jeyashree Krishnan
Wojciech Kusa
Heung-No Lee
Yebeen Lee
Luana Magnante
Daniil Makovetskii
Niva Manchanda
Ajay Mandyam Rangarajan
Gabriele Maraia
Valentin Noël
Hiroki Okuyama
Gaia Palazzi
Sahil Pardasani
Tejas Patil
Alexander Piperski
Maria Cristina Popescu
Federico Ranaldi
Fabio Massimo Zanzotto

Table of Contents

<i>Empathy or Escalation? A Multilingual Framework for Assessing Anxiety-Inducing Language in LLMs</i> Leonardo Arcangeli, Carlotta Foschini, Ilaria Melis and Giulia Pucci	1
<i>The Sycophancy Matrix: Measuring and Steering Ethos-Driven Sycophancy Across Languages and Cultures</i> Gabriele Maraia, Micol Menchi, Alessia Fidelibus, Luana Magnante, Irene Orlando, Gaia Palazzi and Fabio Massimo Zanzotto	22
<i>EvenVLM: A Text and Multimodal Dataset for the Even Language with Vision-Language Fine-Tuning</i> Daniil Makovetskii and Ivan Bondarenko	40
<i>Jako Tako or Fluent? Presenting PoVisLE: A Polish Vision-Language Evaluation</i> Anna Kołos, Grzegorz Statkiewicz, Karolina Seweryn, Katarzyna Kowol, Karolina Piosek and Wojciech Kusa	47
<i>Chain-of-Thought in a Low-Resource Language: A Factorial Study of Exemplar and Reasoning Language Effects in Urdu</i> Ateeb Ahmed	75
<i>Same SQL, Different Decision: Aggregate Risk Hides Cross-Lingual Instability in Text-to-SQL Verification</i> Avinash Goutham Aluguvelly	84
<i>On-Policy Delta Distillation for Multilingual Math Reasoning</i> Byeongho Heo, Jaehui Hwang, Sangdoo Yun and Dongyoon Han	99
<i>Cross Lingual Transfer in Tulu Legal Comprehension: Script-Dependent Improvement and RAG-Induced Knowledge Conflict</i> Sindhu Shetty, Spurthi Setty, Alina Kapanova and Natan Vidra	107
<i>Language-Triggered Value Inconsistencies in Multilingual LLMs</i> Mauricio Neubecker, Maximilian Eder, Tim Würfl, Jason Hoelscher-Obermaier, Jeyashree Krishnan and Ajay Mandyam Rangarajan	134
<i>MedRECT: A Bilingual Medical Reasoning Benchmark for Error Correction in Clinical Texts</i> Naoto Iwase, Hiroki Okuyama and Junichiro Iwasawa	150
<i>Country Is Not Enough: Recognition Action Gaps in Cultural Answerability</i> Aditya Channa, Madhav Kapoor and Jastej Singh	166
<i>One Vector, Thirty-Eight Languages: Benchmark-Derived Anti-Sycophancy Steering Works Worst Where Sycophancy Is Worst</i> Avinash Goutham Aluguvelly	177
<i>Does Safety Reasoning Travel? A Multilingual Measurement of Prompt-Injection Attacks and Runtime Monitors in Tool-Using LLM Agents</i> Ruiyang Zhang	194
<i>Cultural Alignment of Open-Weight LLMs on the Inglehart–Welzel Map</i> Shav Vimalendiran	205
<i>How Well Do Language Models Handle Romanian Diacritics? A Comparison of Monolingual and Multilingual Architectures</i> Alexandra Toma and Maria Cristina Popescu	234

<i>Mining Motifs in Fairy Tales</i>	
Saba Shahsavari, Alessa Mayer and Fabian M. Suchanek	246
<i>Retrieval-Augmented Generation for Symbolic Folk-Art Interpretation: Comparing Dense and Graph-Based Retrieval Using Korean Minhwa</i>	
Minjun Song, Sumin Kim, Junsung Kim, Yebeen Lee and Heung-No Lee	257
<i>We're Cooked! – Probing LLM Political Alignment Via Conflict-Framed Recipe Translation</i>	
Svetlana Gorovaia, Angelica Henestrosa and Ivan P. Yamshchikov	276
<i>UrduBench: An Urdu Reasoning Benchmark using Contextually Ensembled Translations with Human-in-the-Loop</i>	
Muhammad Ali Shafique, Areej Mehboob, Layba Fiaz, Muhammad Usman Qadeer and Hamza Farooq	286
<i>Probing the Think–Say Gap on Sensitive Concepts in Chinese and English Open LLMs</i>	
Yu-Chieh Wang, Lang-Ching Yeh, Guan-Yu Tseng and Shu-Kai Hsieh	304
<i>Why Calques Look Rare in Hindi Model Output: A Human-Anchored Measurement Study</i>	
Abdur Rahman, Vasu Sharma and Mohammad Saifullah	316
<i>CultAct-ML: Measuring Cultural Value Defaults in Multilingual LLM Agents Under Action</i>	
Ahmad Fahmid, Jannatul Ferdous Mim and Md Adnan Arefeen	329
<i>Sample-Matched Comparisons Overstate Test-Time Compute Gains</i>	
Srinivas Harish	339
<i>Tokenization Sets the Measurement Scale for Cross-Lingual Diagnostics</i>	
Arij Riabi and Valentin Noël	354
<i>Preserving Identity, Losing Practice: Religious Schema Transfer in Culturally Grounded Multi-Turn LLM Advice</i>	
Muhammad Junaid Iqbal, Fabio Massimo Zanzotto, Usman Nawaz and Tahir Alyas	381
<i>SQARB: A Benchmark of Sophisticated Question Answering and Reasoning in Belarusian</i>	
Vladislav Poritski, Oksana Volchek, Maksim Aparovich, Volha Harytskaya and Pavel Smrz	392
<i>It Knows More Than It Can Say: Self-Evolving Prompts for Extremely Low-Resource Languages</i>	
Gaurav Sarkar	406
<i>Beyond Surface Form: Probing Script and Discourse-Marker Sensitivity in Marathi LLM Reasoning</i>	
Tejas Patil and Barnali Chetia	412
<i>Investigating the Role of Reasoning-Language Alignment in Monolingual Retrieval-Augmented Generation</i>	
Oliver Hauck, Mario Sanz-Guerrero and Katharina von der Wense	420
<i>The IOL-AI Challenge: An Open Challenge towards Advancing Linguistic Reasoning</i>	
Eduardo Sánchez, Rita Berrada, Dan-Mircea Mirea, Sara Rajaei, Alexander Piperski, Ana Meta Dolinar, Boris Iomdin, Andrey Nikulin, Mariya Shmatova, Marzieh Fadaee and Julia Kreutzer	433
<i>Recognized but Not Produced: A Generation Benchmark for Culturally Specific Kinship Terms</i>	
Sahil Pardasani and Madhusudan Singh	459
<i>Knowledge-Graph Based Augmentation versus Retrieval Augmented Generation for Cultural-Related Question Answering</i>	
Pablo Poulenard, Yannis Karmim and Valentin Barrière	468

<i>Do Multilingual Models Represent Colour Differently Across Languages? The Gap Is in the Vocabularies</i>	
Rong Wang	476
<i>Naturalistic Measure of Social Norms Alignment</i>	
Yevhen Kostiuk, Kenneth Enevoldsen, Peter Bjerregaard Vahlstrup, Márton Kardos and Kristoffer Nielbo	484
<i>Prompt as Stratified Interface: Decomposing Label-Mediated Effects in Chinese-English NLI</i>	
Hua Zhao, Zhiqing Yang and Michelle Mingyue Gu	515
<i>A Vision for Cultural Alignment: Opportunities and Safety Imperatives for AI in Mental Health Support</i>	
Ratna Kandala, Akshata Kishore Moharir, Niva Manchanda and Samantha Adorno	534