

# **4th Workshop on Multilingual Representation Learning (MRL 2024)**

Miami, Florida, USA  
16 November 2024

ISBN: 979-8-3313-0848-3

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571

**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2024) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>SambaLingo: Teaching Large Language Models New Languages</i> Zoltan Csaki, Bo Li, Jonathan Lingjie Li, Qiantong Xu, Pian Pawakapan, Leon Zhang, Yun Du, Hengyu Zhao, Changran Hu and Urmish Thakker . . . . .	1
<i>What an Elegant Bridge: Multilingual LLMs are Biased Similarly in Different Languages</i> Viktor Mihaylov and Aleksandar Shtedritski . . . . .	22
<i>Adapting Open-Source Generative Large Language Models for Low-Resource Languages: A Case Study for Turkish</i> Cagri Toraman . . . . .	30
<i>An Efficient Approach for Studying Cross-Lingual Transfer in Multilingual Language Models</i> Fahim Faisal and Antonios Anastasopoulos . . . . .	45
<i>Are You Sure? Rank Them Again: Repeated Ranking For Better Preference Datasets</i> Peter Devine . . . . .	93
<i>Tagengo: A Multilingual Chat Dataset</i> Peter Devine . . . . .	106
<i>Language and Task Arithmetic with Parameter-Efficient Layers for Zero-Shot Summarization</i> Alexandra Chronopoulou, Jonas Pfeiffer, Joshua Maynez, Xinyi Wang, Sebastian Ruder and Priyanka Agrawal . . . . .	114
<i>Modeling Bilingual Sentence Processing: Evaluating RNN and Transformer Architectures for Cross-Language Structural Priming</i> Demi Zhang, Bushi Xiao, Chao Gao, Sangpil Youm and Bonnie J Dorr . . . . .	127
<i>Recipe for Zero-shot POS Tagging: Is It Useful in Realistic Scenarios?</i> Zeno Vandenbulcke, Lukas Vermeire and Miryam De Lhoneux . . . . .	137
<i>Gender-specific Machine Translation with Large Language Models</i> Eduardo Sánchez, Pierre Andrews, Pontus Stenetorp, Mikel Artetxe and Marta R. Costa-jussà	148
<i>Jina-ColBERT-v2: A General-Purpose Multilingual Late Interaction Retriever</i> Han Xiao, Bo Wang and Rohan Jha . . . . .	159
<i>Cross-Lingual Named Entity Recognition for Low-Resource Languages: A Hindi-Nepali Case Study Using Multilingual BERT Models</i> Dipendra Yadav, Sumaiya Suravee, Tobias Strauß and Kristina Yordanova . . . . .	167
<i>Parameter-efficient Adaptation of Multilingual Multimodal Models for Low-resource ASR</i> Abhishek Gupta, Amruta Parulekar, Sameep Chattopadhyay and Preethi Jyothi . . . . .	175
<i>Towards Cross-Linguistic Semantic Grounding using Dictionary Graph Analysis</i> Ethan Eschrich and Zoey Liu . . . . .	186
<i>Vikhr: Constructing a State-of-the-art Bilingual Open-Source Instruction-Following Large Language Model for Russian</i> Aleksandr Nikolich, Konstantin Korolev, Sergei Bratchikov, Igor Kiselev and Artem Shelmanov	189

<i>Mitigating the Linguistic Gap with Phonemic Representations for Robust Cross-lingual Transfer</i> Haeji Jung, Changdae Oh, Joeon Kang, Jimin Sohn, Kyungwoo Song, Jinkyu Kim and David R Mortensen .....	200
<i>Leveraging Adapters for Improved Cross-lingual Transfer for Low-Resource Creole MT</i> Marcell Richard Fekete, Ernests Lavrinovics, Nathaniel Romney Robinson, Heather Lent, Raj Dabre and Johannes Bjerva.....	212
<i>Evaluating Multilingual Long-Context Models for Retrieval and Reasoning</i> Ameeta Agrawal, Andy Dang, Sina Bagheri Nezhad, Rhitabrat Pokharel and Russell Scheinberg	216
<i>Community OSCAR: A Community Effort for Multilingual Web Data</i> Manuel Brack, Malte Ostendorff, Pedro Ortiz Suarez, José Javier Saiz, Iñaki Lacunza Castilla, Jorge Palomar-Giner, Alexander Shvets, Patrick Schramowski, Georg Rehm, Marta Villegas and Kristian Kersting .....	232
<i>Leveraging LLMs for Translating and Classifying Mental Health Data</i> Konstantinos Skianis, A. Seza Dođruöz and John Pavlopoulos .....	236
<i>Bridging the Bosphorus: Advancing Turkish Large Language Models through Strategies for Low-Resource Language Adaptation and Benchmarking</i> Emre Can Acikgoz, Mete Erdogan and Deniz Yuret.....	242
<i>Unsupervised Text Representation Learning via Instruction-Tuning for Zero-Shot Dense Retrieval</i> Qiu hai Zeng, Zimeng Qiu, Dae Yon Hwang, Xin He and William M. Campbell .....	269
<i>Language Bias in Multilingual Information Retrieval: The Nature of the Beast and Mitigation Methods</i> Jinrui Yang, Fan Jiang and Timothy Baldwin.....	280
<i>Representational Isomorphism and Alignment of Multilingual Large Language Models</i> Di Wu, Yibin Lei, Andrew Yates and Christof Monz.....	293
<i>Generalization Measures for Zero-Shot Cross-Lingual Transfer</i> Saksham Bassi, Duygu Ataman and Kyunghyun Cho .....	298
<i>Detecting and Translating Language Ambiguity with Multilingual LLMs</i> Behrang Mehrparvar and Sandro Pezzelle .....	310
<i>MLT-DR: Multi-Lingual/Task Demonstration Retrieval</i> <i>An Attempt towards Generalized Retriever for In-Context Learning</i> Kazuma Hashimoto, Arjun Reddy Akula, Karthik Raman and Michael Bendersky .....	324
<i>McGill NLP Group Submission to the MRL 2024 Shared Task: Ensembling Enhances Effectiveness of Multilingual Small LMs</i> Senyu Li, Hao Yu, Jessica Ojo and David Ifeoluwa Adelani.....	346
<i>CUNI and LMU Submission to the MRL 2024 Shared Task on Multi-lingual Multi-task Information Retrieval</i> Katharina Hämmerl, Andrei-Alexandru Manea, Gianluca Vico, Jindřich Helcl and Jindřich Libovický .....	357
<i>Findings of the 2nd Shared Task on Multi-lingual Multi-task Information Retrieval at MRL 2024</i> Francesco Tinner, Raghav Mantri, Mammad Hajili, Chiamaka Chukwuneke, Dylan Massey, Benjamin A. Ajibade, Bilge Deniz Kocak, Abolade Dawud, Jonathan Atala, Hale Sirin, Kayode Olaleye, Anar Rzayev, David Adelani and Duygu Ataman .....	365