1st Workshop on Language Models for Underserved Communities (LM4UC 2025)

Albuquerque, New Mexico, USA
4 May 2025

ISBN: 979-8-3313-1922-9

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© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

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

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
Web: www.proceedings.com

Table of Contents

Enhance Contextual Learning in ASR for Endangered Low-resource Languages
Zhaolin Li and Jan Niehues 1

Empowering Low-Resource Languages: TraSe Architecture for Enhanced Retrieval-Augmented Generation in Bangla
Atia Shahnaz Ipa, Mohammad Abu Tareq Rony and Mohammad Shariful Islam 8

ABDUL: A New Approach to Build Language Models for Dialects Using Formal Language Corpora Only
Yassine Toughrai, Kamel Smaïli and David Langlois 16

Untangling the Influence of Typology, Data, and Model Architecture on Ranking Transfer Languages for Cross-Lingual POS Tagging
Enora Rice, Ali Marashian, Hannah Haynie, Katharina Wense and Alexis Palmer 22

Serving the Underserved: Leveraging BARTBahnar Language Model for Bahnaric-Vietnamese Translation
Long Nguyen, Tran Le, Huong Nguyen, Quynh Vo, Phong Nguyen and Tho Quan 32

Caption Generation in Cultural Heritage: Crowdsourced Data and Tuning Multimodal Large Language Models
Artem Reshetnikov and Maria-Cristina Marinescu 42

Preserving Cultural Identity with Context-Aware Translation Through Multi-Agent AI Systems
Mahfuz Anik, Abdur Rahman, Azmine Wasi and Md Ahsan 51

Enhancing Small Language Models for Cross-Lingual Generalized Zero-Shot Classification with Soft Prompt Tuning
Fred Philippy, Siwen Guo, Cedric Lothritz, Jacques Klein and Tegawendé Bissyandé 61

Cognate and Contact-Induced Transfer Learning for Hamshentsnag: A Low-Resource and Endangered Language
Onur Keleş, Baran Günay and Berat Doğan 76

Nayana OCR: A Scalable Framework for Document OCR in Low-Resource Languages
Adithya Kolavi, Samarth P and Vyoman Jain 86

On Tables with Numbers, with Numbers
Konstantinos Kogkalidis and Stergios Chatzikyriakidis 104