14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL 2025)

Short Papers

Mumbai, India
20-24 December 2025

ISBN: 979-8-3313-3243-3

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© (2025) by the Asian Federation of Natural Language Processing and 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

Is OpenVLA Truly Robust? A Systematic Evaluation of Positional Robustness
Yiran Pang, Yiheng Zhao, Zhuopu Zhou, Tingkai Hu and Ranxin Hou 1

Defending Large Language Models against Jailbreak Attacks via Semantic Smoothing
Jiabao Ji, Bairu Hou, Alexander Robey, George J. Pappas, Hamed Hassani, Yang Zhang, Eric Wong and Shiyu Chang 7

Assessing the Macro and Micro Effects of Random Seeds on Fine-Tuning Large Language Models
Nghia Tuan Bui, Guergana K Savova and Lijing Wang 41

How Well Does First-Token Entropy Approximate Word Entropy as a Psycholinguistic Predictor?
Christian Clark, Byung-Doh Oh and William Schuler 47

Speaking the Right Language: The Impact of Expertise (Mis)Alignment in User-AI Interactions
Shramay Palta, Nirupama Chandrasekaran, Rachel Rudinger and Scott Counts 58

To Labor is Not to Suffer: Exploration of Polarity Association Bias in LLMs for Sentiment Analysis
Jiyu Chen, Sarvnaz Karimi, Diego Molla and Cecile Paris 70

Seeing isn't Hearing: Benchmarking Vision Language Models at Interpreting Spectrograms
Tyler Loakman, Joseph James and Chenghua Lin 79

A Formal Analysis of Chain-of-Thought Prompting via Turing Reductions
S M Rafiuddin and Muntaha Nujat Khan 87

Speak & Spell: LLM-Driven Controllable Phonetic Error Augmentation for Robust Dialogue State Tracking
Jihyun Lee, Solee Im, Wonjun Lee and Gary Lee 99

Are Relational Triple Extraction Frameworks Sufficient for Hyper-relational Facts ?
Pratik Saini, Chayan Sarkar and Tapas Nayak 112

IFEval-Audio: Benchmarking Instruction-Following Capability in Audio-based Large Language Models
Yiming Gao, Bin Wang, Chengwei Wei, Shuo Sun and AiTi Aw 120

IndoPref: A Multi-Domain Pairwise Preference Dataset for Indonesian
Vanessa Rebecca Wiyono, David Anugraha, Ayu Purwarianti and Genta Indra Winata 128

Large Language Models Exhibit Limited Reasoning Ability on Coding Problems
Jinyoung Jo, Jonah Engelmann and Sean Choi 139

Documentation Retrieval Improves Planning Language Generation
Renxiang Wang and Li Zhang 147

Does Synthetic Data Help Named Entity Recognition for Low-Resource Languages?
Gaurav Kamath and Sowmya Vajjala 159

From Facts to Folklore: Evaluating Large Language Models on Bengali Cultural Knowledge
Nafis Chowdhury, Moinul Haque, Anika Ahmed, Nazia Tasnim, Md. Istiak Hossain Shihab, Sajjadur Rahman and Farig Sadeque 168

Are ASR foundation models generalized enough to capture features of regional dialects for low-resource languages?
Tawsif Tashwar Dipto, Azmol Hossain , Rubayet Sabbir Faruque, Md. Rezuwan Hassan, Kanij Fatema, Tanmoy Shome, Ruwad Naswan, Md.Foriduzzaman Zihad, Mohaymen Ul Anam, Nazia Tasnim, Hasan Mahmud, Md Kamrul Hasan, Md. Mehedi Hasan Shawon, Farig Sadeque and Tahsin Reasat178

Gatsby without the 'E': Creating Lipograms with LLMs
Nitish Gokulakrishnan, Rohan Balasubramanian, Syeda Jannatus Saba and Steven Skiena . . . 189

Hint-Augmented Re-ranking: Efficient Product Search using LLM-Based Query Decomposition
Yilun Zhu, Nikhita Vedula and Shervin Malmasi 200

p²-TQA: A Process-based Preference Learning Framework for Self-Improving Table Question Answering Models
Wei Zhou, Mohsen Mesgar, Heike Adel and Annemarie Friedrich 217

PerMed-MM: A Multimodal, Multi-Specialty Persian Medical Benchmark for Evaluating Vision Language Models
Ali Khoramfar, Mohammad Javad Dousti and Heshaam Faili 232

An Analysis of Large Language Models for Simulating User Responses in Surveys
Ziyun Yu, Yiru Zhou, Chen Zhao and Hongyi Wen 242

Enhancing BERT Fine-Tuning for Sentiment Analysis in Lower-Resourced Languages
Jozef Kubík, Marek Suppa and Martin Takac 260

What am I missing here?: Evaluating Large Language Models for Masked Sentence Prediction
Charlie Wyatt, Aditya Joshi and Flora D. Salim 273

A Detailed Factor Analysis for the Political Compass Test: Navigating Ideologies of Large Language Models
Sadia Kamal, Lalu Prasad Yadav Prakash, S M Rafiuddin, Mohammed Rakib, Atriya Sen and Sagnik Ray Choudhury 284

EditGRPO: Reinforcement Learning with Post -Rollout Edits for Clinically Accurate Chest X-Ray Report Generation
Kai Zhang, Christopher Malon, Lichao Sun and Martin Renqiang Min 304

Enhancing Long Document Long Form Summarisation with Self-Planning
Xiaotang Du, Rohit Saxena, Laura Perez-Beltrachini, Pasquale Minervini and Ivan Titov 317

Faithful Transcription: Leveraging Bible Recordings to Improve ASR for Endangered Languages
Eric Le Ferrand, Cian Mohamed Bashar Hauser, Joshua Hartshorne and Emily Prud'hommeaux 333

Still Not There: Can LLMs Outperform Smaller Task-Specific Seq2Seq Models on the Poetry-to-Prose Conversion Task?
Kunal Kingkar Das, Manoj Balaji Jagadeeshan, Nallani Chakravartula Sahith, Jivnesh Sandhan and Pawan Goyal 343

Exploring the Performance of Large Language Models on Subjective Span Identification Tasks
Alphaeus Dmonte, Roland R Oruche, Tharindu Ranasinghe, Marcos Zampieri and Prasad Calyam 358

Broken Words, Broken Performance: Effect of Tokenization on Performance of LLMs
Sachin Pawar, Manoj Apte, Kshitij Jadhav, Girish Keshav Palshikar and Nitin Ramrakhiyani 372

ReGraph: Learning to Reformulate Graph Encodings with Large Language Models
Amir Hadifar, Christopher Ochs and Arjan Van Ewijk 386

LLMs Can Covertly Sandbag on Capability Evaluations Against Chain-of-Thought Monitoring
Chloe Li and Noah Y. Siegel 395

Meronymic Ontology Extraction via Large Language Models
Dekai Zhang, Simone Conia and Antonio Rago 423

Compositional Phoneme Approximation for L1-Grounded L2 Pronunciation Training
Jisang Park, Minu Kim, DaYoung Hong and Jongha Lee 434

Can Language Models Handle a Non-Gregorian Calendar? The Case of the Japanese wareki
Mutsumi Sasaki, Go Kamoda, Ryosuke Takahashi, Kosuke Sato, Kentaro Inui, Keisuke Sakaguchi and Benjamin Heinzerling 444

Modeling Contextual Passage Utility for Multihop Question Answering
Akriti Jain and Aparna Garimella 464

Improving LLM's Attachment to External Knowledge In Dialogue Generation Tasks Through Entity Anonymization
Hadi Sheikhi, Chenyang Huang and Osmar Zaiane 472

Agreement-Constrained Probabilistic Minimum Bayes Risk Decoding
Koki Natsumi, Hiroyuki Deguchi, Yusuke Sakai, Hidetaka Kamigaito and Taro Watanabe . . . 484