

19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026)

Volume 2 (Short Papers)

St. Julian's, Malta
24-29 March 2026

ISBN: 979-8-3313-3450-5

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© (2026) 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

<i>Investigating the Multilingual Calibration Effects of Language Model Instruction Tuning</i> Jerry Huang, Peng Lu, Qiuhaio Zeng, Yusuke Iwasawa, Yutaka Matsuo, Sarath Chandar, Edison Marrese-Taylor and Irene Li	1
<i>Detecting Subtle Sense Shift with Polysemy-Aware Trends</i> Ondřej Herman and Pavel Rychlý	60
<i>Logic Haystacks: Probing LLMs' Long-Context Logical Reasoning (Without Easily Identifiable Unrelated Padding)</i> Damien Sileo	66
<i>When Does Auxiliary Modality Matter in Solving Geometric Problems? A Comprehensive Study of Textual, Formal, and Visual Modalities</i> Hyuk Namgoong, Jeesu Jung, Yerim Han and Sangkeun Jung	76
<i>Progressive Visual Refinement for Multi-modal Summarization</i> Ye Xiong, Hidetaka Kamigaito, Soichiro Murakami, Peinan Zhang, Hiroya Takamura and Manabu Okumura	93
<i>Benchmarking Temporal Reasoning and Alignment Across Chinese Dynasties</i> Zhenglin Wang, Jialong Wu, Pengfei LI, Yong Jiang and Deyu Zhou	105
<i>Training in Step-by-Step Formal Reasoning Improves Pronominal Reasoning in Language Models</i> Vagrant Gautam	121
<i>When Words Wear Masks: Detecting Malicious Intents and Hostile Impacts of Online Hate Speech</i> Priyansh Singhal and Piyush Joshi	136
<i>Lost in Activations: A Neuron-level Analysis of Encoders for Cross-Lingual Emotion Detection</i> Pranaydeep Singh, Orphee De Clercq and Els Lefever	154
<i>McMining: Automated Discovery of Misconceptions in Student Code</i> Erfan Al-Hossami and Razvan Bunescu	160
<i>Surprisal from Larger Transformer-based Language Models Predicts fMRI Data More Poorly</i> Yi-Chien Lin and William Schuler	179
<i>WebRollback: Enhancing Web Agents with Explicit Rollback Mechanisms</i> Zhisong Zhang, Tianqing Fang, Kaixin Ma, Wenhao Yu, Hongming Zhang, Haitao Mi and Dong Yu	187
<i>Hacking Neural Evaluation Metrics with Single Text</i> Hiroyuki Deguchi, Katsuki Chousa and Yusuke Sakai	198
<i>To Paraphrase or Not: Efficient Comment Detoxification with Unsupervised Detoxifiability Discrimination</i> Jing Ke, Zheyong Xie, Shaosheng Cao, Tong Xu and Enhong Chen	207
<i>Hey, wait a minute: on at-issue sensitivity in Language Models</i> Sanghee J. Kim and Kanishka Misra	214
<i>Exploring Cross-Lingual Voice Conversion Methods for Anonymizing Low-Resource Text-to-Speech</i> Shenran Wang, Aidan Pine and Mengzhe Geng	225

<i>Korean Canonical Legal Benchmark: Toward Knowledge-Independent Evaluation of LLMs' Legal Reasoning Capabilities</i>	
Hongseok Oh, Wonseok Hwang and Kyoung-Woon On	232
<i>Task-Level Instructions Induction for Audio Question Answering from Few Examples</i>	
Po-Chun Chen, Hen-Hsen Huang and Hsin-Hsi Chen	244
<i>Optical Character Recognition for the International Phonetic Alphabet</i>	
Shu Okabe, Dejvi Zelo and Alexander Fraser	265
<i>How DDAIR you? Disambiguated Data Augmentation for Intent Recognition</i>	
Galo Castillo-López, Alexis Lombard, Nasredine Semmar and Gaël de Chalendar	274
<i>Measuring Linguistic Competence of LLMs on Indigenous Languages of the Americas</i>	
Justin Vasselli, Arturo MP, Frederikus Hudi, Haruki Sakajo and Taro Watanabe	287
<i>Morpheme Matters: Morpheme-Based Subword Tokenization for Korean Language Models</i>	
DongHyeok Lee, Jeongyeon Park, Kyungbeen Cho and Jae Sung Lee	297
<i>Communication Enables Cooperation in LLM Agents: A Comparison with Curriculum-Based Approaches</i>	
Hachem Madmoun and Salem Lahlou	307
<i>Mind Your Special Tokens! On the Importance of Dedicated Sequence-End Tokens in Vision-Language Embedding Models</i>	
Elio Musacchio, Giovanni Semeraro and Goran Glavaš	322
<i>The Reasoning Lingua Franca: A Double-Edged Sword for Multilingual AI</i>	
Alan Saji, Raj Dabre, Anoop Kunchukuttan and Ratish Puduppully	329
<i>When Flores Bloomz Wrong: Cross-Direction Contamination in Machine Translation Evaluation</i>	
David Tan, Pinzhen Chen, Josef Van Genabith and Koel Dutta Chowdhury	345
<i>SAFARI: A Community-Engaged Approach and Dataset of Stereotype Resources in the Sub-Saharan African Context</i>	
Aishwarya Verma, Laud Ammah, Olivia Nercy Ndlovu Lucas, Andrew Zaldivar, Vinodkumar Prabhakaran and Sunipa Dev	359
<i>STREAM-ZH: Simplified Topic Retrieval Exploration and Analysis Module for Chinese Language</i>	
Hongyi Li, Jianjun Lian, Anton Frederik Thielmann and Andre Python	371
<i>MiSCHiEF: A Benchmark in Minimal-Pairs of Safety and Culture for Holistic Evaluation of Fine-Grained Image-Caption Alignment</i>	
Sagarika Banerjee, Tangatar Madi, Advait Swaminathan, Jolie Nguyen, Shivank Garg, Kevin Zhu and Vasu Sharma	384
<i>Exploring Generative Process Reward Modeling for Semi-Structured Data: A Case Study of Table Question Answering</i>	
Lei Tang, Wei Zhou and Mohsen Mesgar	407
<i>Out of Distribution, Out of Luck: Process Rewards Misguide Reasoning Models</i>	
Alexey Dontsov, Anton Korznikov, Andrey V. Galichin and Elena Tutubalina	421
<i>Learning to Ideate for Machine Learning Engineering Agents</i>	
Yunxiang Zhang, Kang Zhou, Zhichao Xu, Kiran Ramnath, Yun Zhou, Sangmin Woo, Haibo Ding and Lin Lee Cheong	436

<i>Infinity-MoE: Generalizing Mixture of Experts to Infinite Experts</i>	
Shota Takashiro, Takeshi Kojima, Shohei Taniguchi, Yusuke Iwasawa and Yutaka Matsuo . . .	448
<i>Beyond Tokens: Concept-Level Training Objectives for LLMs</i>	
Laya Iyer, Pranav Somani, Alice Guo, Dan Jurafsky and Chen Shani	457
<i>Persuasion Tokens for Editing Factual Knowledge in LLMs</i>	
Paul Youssef, Jörg Schlötterer and Christin Seifert	475
<i>Language Family Matters: Evaluating SpeechLLMs Across Linguistic Boundaries</i>	
Yuchen Zhang, Ravi Shekhar and Haralambos Mouratidis	487
<i>When Benchmarks Age: Temporal Misalignment through Large Language Model Factuality Evaluation</i>	
Xunyi Jiang, Dingyi Chang, Julian McAuley and Xin Xu	500
<i>On the Additive Compositionality of Task Vectors in Vision–Language Models</i>	
Yuting Shi, Houjing Wei and Naoya Inoue	513
<i>Funny or Persuasive, but Not Both: Evaluating Fine-Grained Multi-Concept Control in LLMs</i>	
Arya Labroo, Ivaxi Sheth, Vyas Raina, Amaani Ahmed and Mario Fritz	522
<i>Common Sense or Ableism? Rethinking Commonsense Reasoning Through the Lens of Disability</i>	
Karina H Halevy, Kimi Wenzel, Seyun Kim, Kyle Dean Bauer, Bruno Neira, Mona T. Diab and Maarten Sap	555
<i>Machine translation Evaluation Eng-Thai MQM Ranking dataset</i>	
Phichet Phuangrot, Natdanai Trintawat, Kanawat Vilasri, Yanapat Patcharawiwatpong, Pachara Boonsarngsuk, Nat Pavasant and Ekapol Chuangsuwanich	572
<i>Evaluation of Deontic Conditional Reasoning in Large Language Models: The Case of Wason’s Selection Task</i>	
Hirohiko Abe, Kentaro Ozeki, Risako Ando, Takanoobu Morishita, Koji Mineshima and Mitsuhiro Okada	588
<i>Confidence Leaps in LLM Reasoning: Early Stopping and Cross-Model Transfer</i>	
Pavel Tikhonov, Ivan Oseledets and Elena Tutubalina	602
<i>Exploring Fine-Tuning for In-Context Retrieval and Efficient KV-Caching in Long-Context Language Models</i>	
Francesco Maria Molfese, Momchil Hardalov, Rexhina Blloshmi, Bill Byrne and Adrià de Gispert	617
<i>Post-ASR Correction in Hindi: Comparing Language Models and Large Language Models in Low-Resource Scenarios</i>	
Rishabh Kumar, Amrith Krishna, Ganesh Ramakrishnan and Preethi Jyothi	636
<i>CHiRPE: A Step Towards Real-World Clinical NLP with Clinician-Oriented Model Explanations</i>	
Stephanie Fong, Zimu Wang, Guilherme C Oliveira, Xiangyu Zhao, Yiwen Jiang, Jiahe Liu, Beau-Luke Colton, Scott W. Woods, Martha Shenton, Barnaby Nelson, Zongyuan Ge and Dominic Dwyer	646
<i>LLMs Know More About Numbers than They Can Say</i>	
Fengting Yuchi, Li Du and Jason Eisner	659
<i>On the Mathematical Relationship Between Layer Normalization and Dynamic Activation Functions</i>	
Felix Stollenwerk	674

Semantic Token Clustering for Efficient Uncertainty Quantification in Large Language Models
Qi Cao, Andrew Gambardella, Takeshi Kojima, Yutaka Matsuo and Yusuke Iwasawa 682

Becoming Experienced Judges: Selective Test-Time Learning for Evaluators
Seungyeon Jwa, Daechul Ahn, Reokyoung Kim, Dongyeop Kang and Jonghyun Choi 697

Statistical Foundations of DIME: Risk Estimation for Practical Index Selection
Giulio D’Erasmus, Cesare Campagnano, Antonio Mallia, Pierpaolo Brutti, Nicola Tonellotto and
Fabrizio Silvestri 722