

62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)

Student Research Workshop

Bangkok, Thailand
11-16 August 2024

ISBN: 979-8-3313-0151-4

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. (2024)

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>Feriji: A French-Zarma Parallel Corpus, Glossary & Translator</i> Mamadou K. Keita, Elysabhete Amadou Ibrahim, Habibatou Abdoulaye Alfari and Christopher M Homan	1
<i>Pragmatic inference of scalar implicature by LLMs</i> Ye-eun Cho and ismkim99@skku.edu ismkim99@skku.edu	10
<i>Topic Modeling for Short Texts with Large Language Models</i> Tomoki Doi, Masaru Isonuma and Hitomi Yanaka	21
<i>Can LLMs substitute SQL? Comparing Resource Utilization of Querying LLMs versus Traditional Relational Databases</i> Xiang Zhang, Khatoon Khedri and Reza Rawassizadeh	34
<i>Speech-to-Speech Translation with Discrete-Unit-Based Style Transfer</i> Yongqi Wang, Bai Jionghao, Rongjie Huang, Ruiqi Li, Zhiqing Hong and Zhou Zhao	42
<i>InstructCoder: Instruction Tuning Large Language Models for Code Editing</i> Kaixin Li, Qisheng Hu, James Xu Zhao, Hui Chen, Yuxi Xie, Tiedong Liu, Michael Shieh and Junxian He	50
<i>BiasDPO: Mitigating Bias in Language Models through Direct Preference Optimization</i> Ahmed Allam	71
<i>MoExtend: Tuning New Experts for Modality and Task Extension</i> Shanshan Zhong, Shanghua Gao, Zhongzhan Huang, Wushao Wen, Marinka Zitnik and Pan Zhou	80
<i>On the Interpretability of Deep Learning Models for Collaborative Argumentation Analysis in Classrooms</i> Deliang Wang and Gaowei Chen	92
<i>Document Alignment based on Overlapping Fixed-Length Segments</i> Xiaotian Wang, Takehito Utsuro and Masaaki Nagata	103
<i>Automatically Suggesting Diverse Example Sentences for L2 Japanese Learners Using Pre-Trained Language Models</i> Enrico Benedetti, Akiko Aizawa and Florian Boudin	114
<i>Z-coref: Thai Coreference and Zero Pronoun Resolution</i> Poomphob Suwannapichat, Sansiri Tarnpradab and Santitham Prom-on	132
<i>ReMAG-KR: Retrieval and Medically Assisted Generation with Knowledge Reduction for Medical Question Answering</i> Sidhaarth Sredharan Murali, Sowmya Kamath S. and Supreetha R	140
<i>Plot Retrieval as an Assessment of Abstract Semantic Association</i> Shicheng Xu, Liang Pang, Jiangnan Li, Mo Yu, Fandong Meng, Huawei Shen, Xueqi Cheng and Jie Zhou	146
<i>Demystifying Instruction Mixing for Fine-tuning Large Language Models</i> Renxi Wang, Haonan Li, Minghao Wu, Yuxia Wang, Xudong Han, Chiyu Zhang and Timothy Baldwin	162

<i>Fine-Tuning ASR models for Very Low-Resource Languages: A Study on Mvskoke</i> Julia Mainzinger and Gina-Anne Levow	170
<i>Automating Qualitative Data Analysis with Large Language Models</i> Angelina Parfenova, alexander.denzler@hslu.ch alexander.denzler@hslu.ch and Jrgen Pfeffer	177
<i>ANHALTEN: Cross-Lingual Transfer for German Token-Level Reference-Free Hallucination Detection</i> Janek Herrlein, Chia-Chien Hung and Goran Glava	186
<i>Label-Aware Automatic Verbalizer for Few-Shot Text Classification in Mid-To-Low Resource Languages</i> Thanakorn Thaminkaew, Piyawat Lertvittayakumjorn and Peerapon Vateekul	195
<i>Vector Spaces for Quantifying Disparity of Multiword Expressions in Annotated Text</i> Louis Estve, Agata Savary and Thomas Lavergne	204
<i>Narratives at Conflict: Computational Analysis of News Framing in Multilingual Disinformation Campaigns</i> Antonina Sinelnik and Dirk Hovy	225
<i>Assessing In-context Learning and Fine-tuning for Topic Classification of German Web Data</i> Julian Schelb, Andreas Spitz and Roberto Ulloa	238
<i>Knowledge Editing of Large Language Models Unconstrained by Word Order</i> Ryoma Ishigaki, Jundai Suzuki, Masaki Shuzo and Eisaku Maeda	253
<i>Exploring the Effectiveness and Consistency of Task Selection in Intermediate-Task Transfer Learning</i> Pin-Jie Lin, Miaoran Zhang, Marius Mosbach and Dietrich Klakow	264
<i>Does the structure of textual content have an impact on language models for automatic summarization?</i> Eve Sauvage, Sabrina Campano, Lydia Ould Ouali and Cyril Grouin	280
<i>Action Inference for Destination Prediction in Vision-and-Language Navigation</i> Anirudh Reddy Kondapally, Kentaro Yamada and Hitomi Yanaka	286
<i>A Computational Analysis and Exploration of Linguistic Borrowings in French Rap Lyrics</i> Lucas Zurbuchen and Rob Voigt	294
<i>On Improving Repository-Level Code QA for Large Language Models</i> Jan Strich, Florian Schneider, Irina Nikishina and Chris Biemann	303
<i>Compromesso! Italian Many-Shot Jailbreaks undermine the safety of Large Language Models</i> Fabio Pernisi, Dirk Hovy and Paul Rttger	339
<i>Foundation Model for Biomedical Graphs: Integrating Knowledge Graphs and Protein Structures to Large Language Models</i> Yunsoo Kim	346
<i>ViMedQA: A Vietnamese Medical Abstractive Question-Answering Dataset and Findings of Large Language Model</i> Minh-Nam Tran, Phu-Vinh Nguyen, Long HB Nguyen and Dien Dinh	356
<i>Rescue: Ranking LLM Responses with Partial Ordering to Improve Response Generation</i> Yikun Wang, Rui Zheng, Haoming Li, Qi Zhang, Tao Gui and Fei Liu	365
<i>Basreh or Basra? Geoparsing Historical Locations in the Svoboda Diaries</i> Jolie Zhou, Camille Lyans Cole and Annie Chen	377

<i>Homophone2Vec: Embedding Space Analysis for Empirical Evaluation of Phonological and Semantic Similarity</i>	
Sophie Wu, Anita Zheng and Joey Chuang	391
<i>Trace-of-Thought Prompting: Investigating Prompt-Based Knowledge Distillation Through Question Decomposition</i>	
Tyler McDonald and Ali Emami	397
<i>Can LLMs Augment Low-Resource Reading Comprehension Datasets? Opportunities and Challenges</i>	
Vinay Samuel, Houda Aynaou, Arijit Ghosh Chowdhury, Karthik Venkat Ramanan and Aman Chadha	411
<i>Automatic Derivation of Semantic Representations for Thai Serial Verb Constructions: A Grammar-Based Approach</i>	
Vipasha Bansal	422
<i>Seed-Free Synthetic Data Generation Framework for Instruction-Tuning LLMs: A Case Study in Thai</i>	
Parinthapat Pengpun, Can Udomcharoenchaikit, Weerayut Buaphet and Peerat Limkonchotiwat	438
<i>Bridging Distribution Gap via Semantic Rewriting with LLMs to Enhance OOD Robustness</i>	
Manas Madine	458
<i>CoVoSwitch: Machine Translation of Synthetic Code-Switched Text Based on Intonation Units</i>	
Yeeun Kang	469
<i>An Analysis under a Unified Formulation of Learning Algorithms with Output Constraints</i>	
Mooho Song and Jay-Yoon Lee	482
<i>Beyond Abstracts: A New Dataset, Prompt Design Strategy and Method for Biomedical Synthesis Generation</i>	
James O’Doherty, Cian Nolan, Yufang Hou and Anya Belz	499
<i>Improving Sentence Embeddings with Automatic Generation of Training Data Using Few-shot Examples</i>	
Soma Sato, Hayato Tsukagoshi, Ryohei Sasano and Koichi Takeda	519
<i>Curriculum Learning for Small Code Language Models</i>	
Marwa Nar, Kamel Yamani, Lynda Said Lhadj and Riyadh Baghdadi	531
<i>Question-Analysis Prompting Improves LLM Performance in Reasoning Tasks</i>	
Dharunish Yugeswardeenoo, Kevin Zhu and Sean O’Brien	543
<i>An Individualized News Affective Response Dataset</i>	
Tiancheng Hu and Nigel Collier	555
<i>How Well Do Vision Models Encode Diagram Attributes?</i>	
Haruto Yoshida, Keito Kudo, Yoichi Aoki, Ryota Tanaka, Itsumi Saito, Keisuke Sakaguchi and Kentaro Inui	564
<i>CheckersGPT: Learning World Models through Language Modeling</i>	
Abhinav Joshi, Vaibhav Sharma and Ashutosh Modi	576
<i>In-Context Symbolic Regression: Leveraging Large Language Models for Function Discovery</i>	
Matteo Merler, Katsiaryna Haitsiukevich, Nicola Dainese and Pekka Marttinen	589
<i>STEP: Staged Parameter-Efficient Pre-training for Large Language Models</i>	
Kazuki Yano, Takumi Ito and Jun Suzuki	607