

ACL 2024

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

Proceedings of the Conference

August 11-16, 2024

©2024 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-097-4

Introduction

The **ACL Student Research Workshop (SRW) 2024** will be held in conjunction with **ACL 2024**. The SRW provides student researchers in *Computational Linguistics* and *Natural Language Processing* the opportunity to present their work and receive constructive feedback and mentorship from experienced members of the ACL community.

Similar to past SRWs, the **ACL 2024 SRW** offers two submission options: *archival* (appears in proceedings) and *non-archival* (only for presentation). Authors can choose to submit both research papers and thesis proposals as non-archival, explicitly encouraging work in progress that can later be submitted to a future (archival) conference. From a mentorship and reviewing standpoint, archival and non-archival submissions are treated equally.

Organizing Committee

Student Research Workshop Chairs

Eve Fleisig, UC Berkeley, America
Xiyang Fu, Heidelberg University, Germany

Student Research Workshop Chairs

Eve Fleisig, UC Berkeley, America
Xiyang Fu, Heidelberg University, Germany

Program Committee

General Chair

Claire Gardent, CNRS and Université de Lorraine

Program Chairs

Lun-Wei Ku, Academia Sinica

Andre Martins, Instituto Superior Técnico / Instituto de Telecomunicações / Unbabel

Vivek Srikumar, University of Utah

Local Organization

Thepchai Supnithi, NECTEC and AIAT

Prachya Bookwan, NECTEC and AIAT

Thanaruk Theeramunkong, SIIT and AIAT

Workshop Chairs

Xipeng Qiu, Fudan University

Eunsol Choi, The University of Texas at Austin

Tutorial Chairs

Luis Chiruzzo, Universidad de la República

Hung-yi Lee, National Taiwan University

Leonardo Ribeiro, Amazon Alexa Seattle

Demonstration Chairs

Yixin Cao, Singapore Management University

Yang Feng, Chinese Academy of Science

Deyi Xiong, Tianjin University

Student Research

Xiyan Fu, Heidelberg University

Eve Fleisig, UC Berkeley

Student Research: Faculty Advisor

Ekapol Chuangsuwanich, Chulalongkorn University

Yuval Pinter, Ben Gurion University

Publicity and Social Media Chairs

Yuki Arase, Osaka University

Jing Jiang, Singapore Management University

Dimitra Gkatzia, Napier University

Publication Chairs

Miruna Clinciu, University of Edinburgh
Bing Liu, Meta AI
Zhiyu Zoey Chen, University of Texas at Dallas
Chen Liang, Google DeepMind

Handbook Chairs

Pierre Colombo, Université Paris Saclay
Loic Barrault, Meta AI

Technical Open Review

Taro Watanabe, Nara Institute of Science and Technology (NAIST)
Thiago Castro, Federal University of Minas Gerais

Diversity and Inclusion

Jing Li, Hong Kong Polytechnic University
Aparna Garimella, Adobe Research
Steven Wilson, Oakland University
Lin Gui, King's College London

Ethic Committee

Alice Oh, KAIST
Aurélie Névéol, CNRS and Université Paris Saclay

Internal Communications

Claudia Borg, University of Malta
Valentina Pyatkin, Allen Institute for AI
Yannick Parmentier, Université de Lorraine

Student Volunteer

Margot Mieskes, University of Applied Science, Darmstadt
Hao Fei, National University of Singapore
Liangming Pan, University of California, Santa Barbara

Virtual Infrastructure

Gözde Gül, Koç University
Gael Guibon, University of Lorraine
Rachada Kongkrachantra, Thammasat University

Website and Conference App

Yun-Nung (Vivian) Chen, National Taiwan University
Vipas Sutantayawalee, Artificial Intelligence Entrepreneur Association of Thailand

Sponsorship Chairs

Lluis Marquez, Amazon
Kobkrit Viriyayudhakorn, iAPP Co., Ltd

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

Program

Monday, August 12, 2024

16:00 - 17:30 *Posters: Session 5 (Posters C)*

Tuesday, August 13, 2024

10:30 - 12:00 *Oral presentations: Session 8 oral*

16:00 - 17:30 *Panel: Session 12*