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
Xiyan Fu, Heidelberg University, Germany

Student Research Workshop Chairs

Eve Fleisig, UC Berkeley, America
Xiyan 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
- Loïc 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 Névélol, 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
- Gaël 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

*Feriji: A French-Zarma Parallel Corpus, Glossary & Translator*
Mamadou K. Keita, Elysabeth Amadou Ibrahim, Habibatou Abdoulaye Alfari and Christopher M Homan ................................................................. 1

*Pragmatic inference of scalar implicature by LLMs*
Ye-eun Cho and ismkim99@skku.edu ismkim99@skku.edu ........................................ 10

*Topic Modeling for Short Texts with Large Language Models*
Tomoki Doi, Masaru Isonuma and Hitomi Yanaka .......................................................... 21

*Can LLMs substitute SQL? Comparing Resource Utilization of Querying LLMs versus Traditional Relational Databases*
Xiang Zhang, Khatoon Khedri and Reza Rawassizadeh .................................................... 34

*Speech-to-Speech Translation with Discrete-Unit-Based Style Transfer*
Yongqi Wang, Bai Jionghao, Rongjie Huang, Ruiqi Li, Zhiqing Hong and Zhou Zhao .......... 42

*InstructCoder: Instruction Tuning Large Language Models for Code Editing*
Kaixin Li, Qisheng Hu, James Xu Zhao, Hui Chen, Yuxi Xie, Tiedong Liu, Michael Shieh and Junxian He ........................................................................... 50

*BiasDPO: Mitigating Bias in Language Models through Direct Preference Optimization*
Ahmed Allam .................................................................................................................. 71

*MoExtend: Tuning New Experts for Modality and Task Extension*
Shanshan Zhong, Shanghua Gao, Zhongzhan Huang, Wushao Wen, Marinka Zitnik and Pan Zhou 80

*On the Interpretability of Deep Learning Models for Collaborative Argumentation Analysis in Classrooms*
Deliang Wang and Gaowei Chen ..................................................................................... 92

*Document Alignment based on Overlapping Fixed-Length Segments*
Xiaotian Wang, Takehito Utsuro and Masaaki Nagata .................................................... 103

*Automatically Suggesting Diverse Example Sentences for L2 Japanese Learners Using Pre-Trained Language Models*
Enrico Benedetti, Akiko Aizawa and Florian Boudin ....................................................... 114

*Z-coref: Thai Coreference and Zero Pronoun Resolution*
Poomphob Suwannapichat, Sansiri Tampradab and Santitham Prom-on ......................... 132

*ReMAG-KR: Retrieval and Medically Assisted Generation with Knowledge Reduction for Medical Question Answering*
Sidhaarth Sredharan Murali, Sowmya Kamath S. and Supreetha R ............................... 140

*Plot Retrieval as an Assessment of Abstract Semantic Association*
Shicheng Xu, Liang Pang, Jiangnan Li, Mo Yu, Fandong Meng, Huawei Shen, Xueqi Cheng and Jie Zhou ................................................................. 146

*Demystifying Instruction Mixing for Fine-tuning Large Language Models*
Renxi Wang, Haonian Li, Minghao Wu, Yuxia Wang, Xudong Han, Chiyu Zhang and Timothy Baldwin .............................................................................. 162
Fine-Tuning ASR models for Very Low-Resource Languages: A Study on Mvskoke
Julia Mainzinger and Gina-Anne Levow ................................................................. 170

Automating Qualitative Data Analysis with Large Language Models
Angelina Parfenova, alexander.denzler@hslu.ch alexander.denzler@hslu.ch and Jrgen Pfeffer 177

ANHALTEN: Cross-Lingual Transfer for German Token-Level Reference-Free Hallucination Detection
Janek Herrlein, Chia-Chien Hung and Goran Glava .............................................. 186

Label-Aware Automatic Verbalizer for Few-Shot Text Classification in Mid-To-Low Resource Languages
Thanakorn Thaminkaew, Piyawat Lertvittayakumjorn and Peerapon Vateekul ................. 195

Vector Spaces for Quantifying Disparity of Multiword Expressions in Annotated Text
Louis Estve, Agata Savary and Thomas Lavergne ................................................. 204

Narratives at Conflict: Computational Analysis of News Framing in Multilingual Disinformation Campaigns
Antonina Sinelnik and Dirk Hovy ........................................................................ 225

Assessing In-context Learning and Fine-tuning for Topic Classification of German Web Data
Julian Schelb, Andreas Spitz and Roberto Ulloa ...................................................... 238

Knowledge Editing of Large Language Models Unconstrained by Word Order
Ryoma Ishigaki, Jundai Suzuki, Masaki Shuzo and Eisaku Maeda .............................. 253

Exploring the Effectiveness and Consistency of Task Selection in Intermediate-Task Transfer Learning
Pin-Jie Lin, Miaoran Zhang, Marius Mosbach and Dietrich Klakow ............................ 264

Does the structure of textual content have an impact on language models for automatic summarization?
Eve Sauvage, Sabrina Campano, Lydia Ould Ouali and Cyril Grouin .......................... 280

Anirudh Reddy Kondapally, Kentaro Yamada and Hitomi Yanaka .............................. 286

A Computational Analysis and Exploration of Linguistic Borrowings in French Rap Lyrics
Lucas Zurbuchen and Rob Voigt ........................................................................ 294

On Improving Repository-Level Code QA for Large Language Models
Jan Strich, Florian Schneider, Irina Nikishina and Chris Biemann .............................. 303

Compromesso! Italian Many-Shot Jailbreaks undermine the safety of Large Language Models
Fabio Pernisi, Dirk Hovy and Paul Rttger ............................................................... 339

Foundation Model for Biomedical Graphs: Integrating Knowledge Graphs and Protein Structures to Large Language Models
Yunsoo Kim ........................................................................................................ 346

ViMedAQA: A Vietnamese Medical Abstractive Question-Answering Dataset and Findings of Large Language Model
Minh-Nam Tran, Phu-Vinh Nguyen, Long HB Nguyen and Dien Dinh .......................... 356

Rescue: Ranking LLM Responses with Partial Ordering to Improve Response Generation
Yikun Wang, Rui Zheng, Haoming Li, Qi Zhang, Tao Gui and Fei Liu .......................... 365

Basreh or Basra? Geoparsing Historical Locations in the Svoboda Diaries
Jolie Zhou, Camille Lyans Cole and Annie Chen .................................................... 377
Homophone2Vec: Embedding Space Analysis for Empirical Evaluation of Phonological and Semantic Similarity
Sophie Wu, Anita Zheng and Joey Chuang .......................... 391

Trace-of-Thought Prompting: Investigating Prompt-Based Knowledge Distillation Through Question Decomposition
Tyler McDonald and Ali Emami ........................................... 397

Can LLMs Augment Low-Resource Reading Comprehension Datasets? Opportunities and Challenges
Vinay Samuel, Houda Aynaou, Arijit Ghosh Chowdhury, Karthik Venkat Ramanan and Aman Chadha ........................................ 411

Automatic Derivation of Semantic Representations for Thai Serial Verb Constructions: A Grammar-Based Approach
Vipasha Bansal ............................................................... 422

Seed-Free Synthetic Data Generation Framework for Instruction-Tuning LLMs: A Case Study in Thai
Parinthapat Pengpun, Can Udomcharoensaikitt, Weerayut Buaphet and Peerat Limkonchotiwat 438

Bridging Distribution Gap via Semantic Rewriting with LLMs to Enhance OOD Robustness
Manas Madine ............................................................... 458

CoVoSwitch: Machine Translation of Synthetic Code-Switched Text Based on Intonation Units
Yeeun Kang .................................................................. 469

An Analysis under a Unified Formulation of Learning Algorithms with Output Constraints
Mooho Song and Jay-Yoon Lee ........................................... 482

Beyond Abstracts: A New Dataset, Prompt Design Strategy and Method for Biomedical Synthesis Generation
James O’Doherty, Cian Nolan, Yufang Hou and Anya Belz ........................................ 499

Improving Sentence Embeddings with Automatic Generation of Training Data Using Few-shot Examples
Soma Sato, Hayato Tsukagoshi, Ryohei Sasano and Koichi Takeda ........................................ 519

Curriculum Learning for Small Code Language Models
Marwa Nar, Kamel Yamani, Lynda Said Lhadj and Riyadh Baghdadi ........................................ 531

Question-Analysis Prompting Improves LLM Performance in Reasoning Tasks
Dharunish Yugewardeenoo, Kevin Zhu and Sean O’Brien ........................................ 543

An Individualized News Affective Response Dataset
Tiancheng Hu and Nigel Collier ........................................... 555

How Well Do Vision Models Encode Diagram Attributes?
Haruto Yoshida, Keito Kudo, Yoichi Aoki, Ryota Tanaka, Itsumi Saito, Keisuke Sakaguchi and Kentaro Inui ........................................ 564

CheckersGPT: Learning World Models through Language Modeling
Abhinav Joshi, Vaibhav Sharma and Ashutosh Modi ........................................ 576

In-Context Symbolic Regression: Leveraging Large Language Models for Function Discovery
Matteo Merler, Katsiaryna Haitsiukevich, Nicola Dainese and Pekka Marttinen ........................................ 589

STEP: Staged Parameter-Efficient Pre-training for Large Language Models
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