

ACL 2024

**The 62nd Annual Meeting of the Association for
Computational Linguistics**

Proceedings of the Student Research Workshop

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

Welcome to the ACL 2024 Student Research Workshop! The **ACL 2024 Student Research Workshop (SRW)** is a forum for student researchers in *computational linguistics* and *natural language processing*. The workshop provides an invaluable opportunity for students to present their work, receive constructive feedback, and engage with the broader research community.

Following the tradition of previous SRWs, we accept two types of submissions: **research papers** and **thesis proposals**. Authors also have the option to choose between *archival* (included in the conference proceedings) and *non-archival* (for presentation only) submissions. This encourages students to share work in progress, with the possibility of submitting the archival version to future conferences. Importantly, both archival and non-archival submissions are treated equally in terms of mentorship and reviewing. At the heart of the SRW is the **mentorship program**, which offers students a unique opportunity to receive feedback from experienced mentors. Before the submission deadline, many student participants took part in the **pre-submission mentoring program**, designed to help improve the writing and presentation of their submissions.

The ACL 2024 SRW received a total of **50 submissions**, including both research papers and thesis proposals. Out of these, **12 papers were non-archival** and **38 papers were archival**. All accepted papers will be presented during the conference, either in person or virtually. The presentations will include both poster sessions and oral presentations for selected papers.

We are proud to introduce the student chairs for the ACL 2024 SRW:

- **Eve Fleisig**, UC Berkeley, USA
- **Xiyang Fu**, Heidelberg University, Germany

We are deeply grateful for the guidance and support from our faculty advisors:

- **Ekapol Chuangsuwanich**, Chulalongkorn University, Thailand
- **Yuval Pinter**, Ben-Gurion University, Israel

A heartfelt thanks to all the mentors who contributed their time and expertise to guide our student authors: Mihir Kale, Youssef Al Hariri, Devang Kulshreshtha, Yifan Hou, Sandeep Mathias, Kemal Kurniawan, Surangika Ranathunga, Manish Shrivastava, Kiet Van Nguyen, Nihal V. Nayak, Prashant Kodali, Ritam Dutt, Sunny Rai, Sashank Santhanam, Andrea Varga, Silvia Casola, Sara Papi, Sowmya Vajjala, Manling Li, Francesca Franzon, Aina Garí Soler, Abulhair Saparov, Yangjun Zhang, Bonnie Webber, Ekapol Chuangsuwanich, Ratul Ghosh, Lucy H. Lin, Anne Beyer, Hyundong Justin Cho, Ben Peters, Elena Kochkina, Alok Debnath, Mascha Kurpicz-Briki, Philipp Sadler, Aditya Shah, Te Rutherford.

We are immensely thankful to our sponsors, who provided funding for travel grants, enabling many authors to attend the conference. We would also like to express our gratitude to the program committee members for their thorough reviews and feedback. Special thanks to the ACL 2024 organizing committee for their invaluable support and advice throughout this process. Lastly, we thank all the authors for their enthusiastic participation in SRW 2024.

Organizing Committee

Student Research Workshop Chairs

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

Student Research Faculty Advisors

Ekapol Chuangsuwanich, Chulalongkorn University, Thailand
Yuval Pinter, Ben-Gurion University, Israel

Program Committee

Program Chairs

Eve Fleisig

Xiyan Fu

Program Committee

Piush Aggarwal, Fernuniversität Gesamthochschule Hagen and Fernuniversität Gesamthochschule Hagen

Chandan Akiti, Samsung

Abeer Aldayel, King Saud University

Miguel A. Alonso, Universidade da Coruña

Begoña Altuna

Sergi Alvarez-Vidal, Universitat Pompeu Fabra and Universitat Oberta de Catalunya

Bharat Ram Ambati

Evelin Amorim, Inesc tec

Tatiana Anikina, German Research Center for AI

Raghav Arora

Parsa Bagherzadeh

Ahsaas Bajaj

Valerio Basile, University of Turin

Anil Batra, University of Edinburgh

Tim Baumgärtner, TU Darmstadt

Rachel Bawden, Inria

Shabnam Behzad

Gabriel Bernier-Colborne, National Research Council Canada

Anne Beyer, Universität Potsdam

Isil Poyraz Bilgin

Shikha Bordia

Meng Cao, McGill University

Arlene Casey, University of Edinburgh, University of Edinburgh

Silvia Casola, University of Turin, University of Padua and Fondazione Bruno Kessler

William Chen, University of California, Berkeley

Canyu Chen

Hyundong Justin Cho, USC/ISI

Sukmin Cho

Juhwan Choi

Arijit Ghosh Chowdhury, Amazon

Ekapol Chuangsuwanich, Chulalongkorn University

Rodolfo Corona

Amanda Cercas Curry

Aryia Dattamajumdar, Meta AI

Brian Davis, Dublin City University

Alok Debnath, Trinity College, Dublin

Ayse Ayyuce Demirbas, Firat (Euphrates) University

Nesat Dereli, Adyen

Chris Develder, Universiteit Gent

Noshin Tabassum Dina

Bonaventure F. P. Dossou
Zi-Yi Dou, University of California, Los Angeles
Ritam Dutt, Carnegie Mellon University
Aditi Dutta
Micha Elsner, Ohio State University
Neele Falk, University of Stuttgart, University of Stuttgart
António Farinhas, Instituto Superior Técnico
Fauzan Farooqui
Eve Fleisig
Francesca Franzon, Universitat Pompeu Fabra
Dayne Freitag, SRI International
Xiyan Fu
Diana Galvan, University of Cambridge
Shubham Garg, Amazon
Aina Garí Soler, Télécom-Paris
Ratul Ghosh, Amazon
Kristina Gligoric, Stanford University
Anmol Goel, Technische Universität Darmstadt
Satya Kesav Gundabathula, Google
Ankita Gupta
Ivan Habernal, Ruhr-Universität Bochum
Youssef Al Hariri, Edinburgh University, University of Edinburgh
Yifan Hou, Department of Computer Science, Swiss Federal Institute of Technology
Phu Mon Htut, AWS AI Labs
Tiancheng Hu, University of Cambridge
Zhongzhan Huang, Sun Yat-Sen University
Robert L. Logan IV
Glorianna Jagfeld, LEP AG
Labiba Jahan, Southern Methodist University
Thanmay Jayakumar, Indian Institute of Technology, Madras, Dhirubhai Ambani Institute Of Information and Communication Technology
Jyoti Jha, EY GDS India
Kenan Jiang, Emory University
Kristen Johnson, Michigan State University
Jad Kabbara, Massachusetts Institute of Technology
Patrick Kahardipraja, Fraunhofer HHI
Mihir Kale, Google
Amr Keleg, University of Edinburgh, University of Edinburgh
Yova Kementchedjhieva, Mohamed bin Zayed University of Artificial Intelligence
Mohammad Aflah Khan
Harshita Khandelwal, UCLA Computer Science Department, University of California, Los Angeles
Anirudh Khatry
Elena Kochkina, J.P. Morgan Chase
Prashant Kodali, International Institute of Information Technology Hyderabad
Mamoru Komachi, Hitotsubashi University
Fajri Koto, Mohamed bin Zayed University of Artificial Intelligence
Pranoy Kovuri, Google
Devang Kulshreshtha, Amazon
Kemal Kurniawan, University of Melbourne
Mascha Kurpicz-Briki, BFH - Bern University of Applied Sciences

Alexandra Lavrentovich
Amirhossein Layegh
Messi H.j. Lee
Yoonsang Lee, Seoul National University
Manling Li, Northwestern University
Sha Li, University of Illinois, Urbana Champaign
Jia Li
Siyan Li
Baohao Liao
Jasy Suet Yan Liew, Universiti Sains Malaysia
Zheng Wei Lim
Peerat Limkonchotiwat, AI Singapore
Lucy H. Lin, Spotify
Tatiana Litvinova
Wei Liu, Heidelberg University
Li Lucy, Allen Institute for Artificial Intelligence and University of California Berkeley
Bohan Lyu
Bolei Ma, Ludwig-Maximilians-Universität München
Valentin Malykh
Lara J. Martin, University of Maryland, Baltimore County
Sandeep Mathias
Shubhanshu Mishra, shubhanshu.com
Sai Ajay Modukuri, Samsung Research America
Hongbin Na, University of Technology Sydney
Murali Nandan Nagarapu, Nucleus AI
Sachin Narayan Nagargoje
Masaaki Nagata, NTT Corporation
Pakawat Nakwijit, Queen Mary, University of London
Nihal V. Nayak, Brown University
Kota Shamanth Ramanath Nayak, Concordia University
Mariana Neves, German Federal Institute for Risk Assessment
Kiet Van Nguyen, University of Information Technology, VNU-HCM
Ratna Nirupama, University of Kansas
Naoki Otani, Megagon Labs
Kun Ouyang
Maria Leonor Pacheco, University of Colorado at Boulder
Jiayi Pan, University of California, Berkeley
Artemis Panagopoulou, Salesforce.com and University of Pennsylvania, University of Pennsylvania
Sara Papi
Tanmay Parekh
Ajay Patel
Naiara Perez, University of the Basque Country (UPV/EHU)
Gabriele Pergola, University of Warwick
Ben Peters, Instituto de Telecomunicações, Portugal and Instituto Superior Técnico
Shubham Milind Phal, Google
Van-Thuy Phi, RIKEN
Dina Pisarevskaya
Dongqi Pu, Universität des Saarlandes
Rajkumar Pujari, Purdue University
Yusu Qian, Apple

Yusu Qian, Apple
Sunny Rai, School of Engineering and Applied Science, University of Pennsylvania
Surangika Ranathunga, Massey University
Pretam Ray
Frederick Riemenschneider, Ruprecht-Karls-Universität Heidelberg
Shruti Rijhwani, Google DeepMind
Hossein Rouhizadeh, University of Geneva
Markian Rybchuk
Sangwon Ryu, Pohang University of Science and Technology
Farig Sadeque, BRAC University
Philipp Sadler, University of Potsdam
Oscar Sainz, University of the Basque Country (UPV/EHU)
Nicholas Sanders
Sashank Santhanam, Apple
Abulhair Saparov, New York University
Ryohei Sasano, Nagoya University
Tatjana Scheffler, Ruhr-Universität Bochum
Fabian Schmidt
Sebastian Schuster, University College London, University of London
Indira Sen, Rheinisch Westfälische Technische Hochschule Aachen
Sheikh Shafayat, KAIST
Aditya Shah
Roshan Sharma, Google
Sina Sheikholeslami, KTH Royal Institute of Technology
Kejian Shi
Manish Shrivastava, International Institute of Information Technology Hyderabad, India
Chenglei Si, Stanford University
Pushdeep Singh, Tata Consultancy Services Limited, India
Kevin Small, Amazon
Cesare Spinoso-Di Piano, , McGill University
Richard Sproat, Google
Marija Stanojevic, Ellipsis Health
Mihai Surdeanu, University of Arizona
Dage Särg, institute of computer science, University of Tartu
Dage Särg, institute of computer science, University of Tartu
Shabnam Tafreshi, University of Maryland: ARLIS
Zeeraq Talat, Mohamed bin Zayed University of Artificial Intelligence
Alberto Testoni, University of Amsterdam
Nicholas Tomlin, University of California Berkeley
Son Quoc Tran
Can Udomcharoenchaikit, Vidyasirimedhi Institute of Science and Technology (VISTEC)
Sowmya Vajjala, National Research Council Canada
Andrea Varga, Theta Lake Ltd (<https://thetalake.com/>)
Francielle Vargas
Deeksha Varshney
David Vilar, Google
Hao Wang
Mingyang Wang
Bonnie Webber, Edinburgh University, University of Edinburgh
Alexander Wettig, Princeton University
Chenglin Wu, DeepWisdom

Xinyi Wu, Google DeepMind
Qi Xin
Rui Xing, Mohamed bin Zayed University of Artificial Intelligence and University of Melbourne
Fuzhao Xue, National University of Singapore
Dixi Yao
Michael Miller Yoder, University of Pittsburgh
Olga Zamaraeva, Universidad de La Coruña
Zenan Zhai
Haoran Zhang, Facebook
Yangjun Zhang
Mike Zhang
Zhehao Zhang, Dartmouth College
Chenhao Zhang, Shanghai Artificial Intelligence Laboratory, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences and Huazhong University of Science and Technology
Yiqun Zhang, Northeastern University
Yao Zhang, Nankai University
Yu Zhao, University of Edinburgh
Xin Zhao
Wenjie Zhao
Siheng Zhao, University of Southern California
Xinyi Zheng, Google
Shanshan Zhong
Naitian Zhou, University of California, Berkeley
Rongxin Zhu, University of Melbourne
Elena Zotova

Table of Contents

<i>Feriji: A French-Zarma Parallel Corpus, Glossary</i> <i>Translator</i> Mamadou K. Keita, Elysabhete Amadou Ibrahim, Habibatu 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>Speech-to-Speech Translation with Discrete-Unit-Based Style Transfer</i> Yongqi Wang, Bai Jionghao, Rongjie Huang, Ruiqi Li, Zhiqing Hong and Zhou Zhao	34
<i>BiasDPO: Mitigating Bias in Language Models through Direct Preference Optimization</i> Ahmed Allam	42
<i>Document Alignment based on Overlapping Fixed-Length Segments</i> Xiaotian Wang, Takehito Utsuro and Masaaki Nagata	51
<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.	62
<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	68
<i>Fine-Tuning ASR models for Very Low-Resource Languages: A Study on Mvskoke</i> Julia Mainzinger and Gina-Anne Levow	76
<i>Automating Qualitative Data Analysis with Large Language Models</i> Angelina Parfenova, alexander.denzler@hslu.ch alexander.denzler@hslu.ch and Jrgen Pfeffer	83
<i>ANHALTEN: Cross-Lingual Transfer for German Token-Level Reference-Free Hallucination Detection</i> Janek Herrlein, Chia-Chien Hung and Goran Glava	92
<i>Label-Aware Automatic Verbalizer for Few-Shot Text Classification in Mid-To-Low Resource Languages</i> Thanakorn Thaminkaew, Piyawat Lertvittayakumjorn and Peerapon Vateekul	101
<i>Vector Spaces for Quantifying Disparity of Multiword Expressions in Annotated Text</i> Louis Estve, Agata Savary and Thomas Lavergne	110
<i>Narratives at Conflict: Computational Analysis of News Framing in Multilingual Disinformation Campaigns</i> Antonina Sinelnik and Dirk Hovy	131
<i>Assessing In-context Learning and Fine-tuning for Topic Classification of German Web Data</i> Julian Schelb, Andreas Spitz and Roberto Ulloa	144
<i>Knowledge Editing of Large Language Models Unconstrained by Word Order</i> Ryoma Ishigaki, Jundai Suzuki, Masaki Shuzo and Eisaku Maeda	159

<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	170
<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	186
<i>Action Inference for Destination Prediction in Vision-and-Language Navigation</i> Anirudh Reddy Kondapally, Kentaro Yamada and Hitomi Yanaka	192
<i>A Computational Analysis and Exploration of Linguistic Borrowings in French Rap Lyrics</i> Lucas Zurbuchen and Rob Voigt	200
<i>On Improving Repository-Level Code QA for Large Language Models</i> Jan Strich, Florian Schneider, Irina Nikishina and Chris Biemann	209
<i>Compromesso! Italian Many-Shot Jailbreaks undermine the safety of Large Language Models</i> Fabio Pernisi, Dirk Hovy and Paul Rttger	245
<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	252
<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	261
<i>Basreh or Basra? Geoparsing Historical Locations in the Svoboda Diaries</i> Jolie Zhou, Camille Lyans Cole and Annie Chen	273
<i>Homophone2Vec: Embedding Space Analysis for Empirical Evaluation of Phonological and Semantic Similarity</i> Sophie Wu, Anita Zheng and ching-i-chuang@mail.mcgill.ca ching-i-chuang@mail.mcgill.ca	287
<i>Trace-of-Thought Prompting: Investigating Prompt-Based Knowledge Distillation Through Question Decomposition</i> Tyler McDonald and Ali Emami	293
<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	307
<i>Automatic Derivation of Semantic Representations for Thai Serial Verb Constructions: A Grammar-Based Approach</i> Vipasha Bansal	318
<i>Bridging Distribution Gap via Semantic Rewriting with LLMs to Enhance OOD Robustness</i> Manas Madine	334
<i>CoVoSwitch: Machine Translation of Synthetic Code-Switched Text Based on Intonation Units</i> Yeeun Kang	345
<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	358
<i>Improving Sentence Embeddings with Automatic Generation of Training Data Using Few-shot Examples</i> Soma Sato, Hayato Tsukagoshi, Ryohei Sasano and Koichi Takeda	378

<i>Curriculum Learning for Small Code Language Models</i> Marwa Nair, Kamel Yamani, Lynda Saïd Lhadj and Riyadh Baghdadi	390
<i>Question-Analysis Prompting Improves LLM Performance in Reasoning Tasks</i> Dharunish Yugeswardeenoo, Kevin Zhu and Sean O'Brien	402
<i>CheckersGPT: Learning World Models through Language Modeling</i> Abhinav Joshi, Vaibhav Sharma and Ashutosh Modi	414
<i>In-Context Symbolic Regression: Leveraging Large Language Models for Function Discovery</i> Matteo Merler, Katsiaryna Haitsiukevich, Nicola Dainese and Pekka Marttinen.....	427
<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	445
<i>Can LLMs substitute SQL? Comparing Resource Utilization of Querying LLMs versus Traditional Relational Databases</i> Xiang Zhang, Khatoon Khedri and Reza Rawassizadeh	465
<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	473
<i>MoExtend: Tuning New Experts for Modality and Task Extension</i> Shanshan Zhong, Shanghua Gao, Zhongzhan Huang, Wushao Wen, Marinka Zitnik and Pan Zhou	494
<i>On the Interpretability of Deep Learning Models for Collaborative Argumentation Analysis in Classrooms</i> Deliang Wang and Gaowei Chen.....	506
<i>Automatically Suggesting Diverse Example Sentences for L2 Japanese Learners Using Pre-Trained Language Models</i> Enrico Benedetti, Akiko Aizawa and Florian Boudin.....	517
<i>Z-coref: Thai Coreference and Zero Pronoun Resolution</i> Poomphob Suwannapichat, sansiri.tarn@kmutt.ac.th sansiri.tarn@kmutt.ac.th and santitham.pro@kmutt.ac.th santitham.pro@kmutt.ac.th.....	535
<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.....	543
<i>Foundation Model for Biomedical Graphs: Integrating Knowledge Graphs and Protein Structures to Large Language Models</i> Yunsoo Kim.....	559

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*