The ACL-SRW organizers gratefully acknowledge the support from the following sponsors.

Gold

Google

Bronze

In cooperation with

[NSF logo]
Introduction

Welcome to the ACL 2023 Student Research Workshop!

The ACL 2023 Student Research Workshop (SRW) is a forum for student researchers in computational linguistics and natural language processing. The workshop provides a great opportunity for student participants to take part in a mentorship program, present their work and receive valuable feedback from the international research community.

Following the tradition of the previous student research workshops, we accept three kinds of submissions: long and short research papers as well as thesis proposals. The research paper track is a venue for students to describe completed work or work-in-progress along with preliminary results. The thesis proposal track is offered for Ph.D. students who have decided on a thesis topic and are interested in getting feedback on their proposal and ideas about future directions for their work.

Mentoring is at the heart of the SRW. In keeping with previous years, we had a pre-submission mentoring program before the submission deadline. A total of 58 papers participated in the pre-submission mentoring program. This program offered students the opportunity to receive feedback from a mentor to improve the writing style and presentation of their submissions.

The Student Research Workshop again attracted a very large number of submissions this year. We received 145 submissions including 135 research papers (81 long papers and 54 short papers) and 10 thesis proposals. Out of these, 5 were ACL Findings papers whose authors wished to present their work at the SRW. A further 9 submissions were desk rejected and 2 submissions were withdrawn by the authors prior to the completion of the review process. A total of 50 submissions (5 Findings, 2 Thesis Proposals, 28 long papers and 15 short papers) were accepted. 46 of the accepted papers will be presented in person and/or virtually in the poster sessions of the main conference and 4 will be presented as oral presentations.

We are deeply grateful to our sponsors for providing funds for the travel grants that we make available to paper authors. We thank our program committee members for their careful reviews of each paper and all of our mentors for donating their time to provide feedback to our student authors. We are deeply thankful to our faculty advisors, Ivan Vulic and Lu Wang, and to the ACL 2023 organizing committee for their advice and support throughout the process. Finally, we thank each and every one of the authors for their enthusiastic participation!
Organizing Committee

Student Chair
  Vishakh Padmakumar, New York University, USA

Student Chair
  Gisela Vallejo, University of Melbourne, Australia

Student Chair
  Yao Fu, University of Edinburgh, UK

Faculty Advisor
  Ivan Vulic, University of Cambridge, UK

Faculty Advisor
  Lu Wang, University of Michigan, USA
Program Committee

Chairs

Yao Fu, The University of Edinburgh
Vishakh Padmakumar, New York University
Gisela Vallejo, The University of Melbourne

Program Committee

Omri Abend, The Hebrew University of Jerusalem
Ibrahim Abu Farha, University of Sheffield
Piush Aggarwal, FernUniversität in Hagen, Computational Linguistics
Youssef Al Hariri, University of Edinburgh
Sedeq Al-khazraj, Rochester Institute of Technology
Abeer Aldayel, King Saud University
Begoña Altuna, Universidad del País Vasco - Euskal Herriko unibertsitatea
Fernando Alva-Manchego, Cardiff University
Evelin Amorim, Porto University
Tatiana Anikina, DFKI / Saarland Informatics Campus
Parsa Bagherzadeh, McGill University
Ahsaas Bajaj, University of Massachusetts Amherst
Valerio Basile, University of Turin
Tim Baumgartner, TU Darmstadt
Rachel Bawden, Inria
Daniel Beck, University of Melbourne
Shabnam Behzad, Georgetown University
Jonas Belouadi, Bielefeld University
Uri Berger, The Hebrew University of Jerusalem, University of Melbourne
Gabriel Bernier-colborne, National Research Council Canada
Terra Blevins, University of Washington
Shikha Bordia, Verisk Analytics
Agostina Calabrese, The University of Edinburgh
Jon Ander Campos, University of the Basque Country (UPV/EHU)
Ronald Cardenas, University of Edinburgh
Silvia Casola, Fondazione Bruno Kessler, University of Padua
Yanda Chen, Columbia University
Jiangjie Chen, Fudan University
Hyundong Cho, USC, Information Sciences Institute
Alessandra Teresa Cignarella, Computer Science Department - University of Turin
Mike Conway, University of Utah
Xiang Dai, CSIRO Data61
Brian Davis, Dublin City University
Alok Debnath, Trinity College, Dublin
Chris Develder, Ghent University
Bonaventure F. P. Dossou, Mila Quebec AI Institute, McGill University
Gabriel Doyle, San Diego State University
Ritam Dutt, Carnegie Mellon University
Micha Elsner, The Ohio State University
Carlos Escolano, Universitat Politècnica de Catalunya
Brielen Madureira, University of Potsdam
Rahmad Mahendra, Universitas Indonesia
Lara Martin, University of Maryland, Baltimore County
Sandeep Mathias, Presidency University
Puneet Mathur, University of Maryland College Park
Arya D. Mccarthy, Johns Hopkins University
Nikhil Mehta, Purdue University
William Merrill, New York University
Tsvetomila Mihaylova, Aalto University
Gosse Minnema, University of Groningen
Fatemehsadat Mireshghallah, UC San Diego
Shubhanshu Mishra, shubhanshu.com
Amita Misra, Amazon
Adib Mosharrof, University of Kentucky
Sachin Nagargoje, Microsoft
Masaaki Nagata, NTT Corporation
Nihal V. Nayak, Brown University
Mariana Neves, German Federal Institute for Risk Assessment
Kiet Nguyen, University of Information Technology, VNU-HCM
Vincent Nguyen, Australian National University
CSIRO Data61
Aitor Ormazabal, University of the Basque Country
Maria Leonor Pacheco, University of Colorado Boulder / Microsoft Research
Xiang Pan, New York University
Artemis Panagopoulou, University of Pennsylvania
Sara Papi, Fondazione Bruno Kessler
Tanmay Parekh, University of California Los Angeles
Gabriele Pergola, University of Warwick
Ben Peters, Instituto de Telecomunicações
Ian Porada, Mila, McGill University
Adithya Pratapa, Carnegie Mellon University
Yiyuan Pu, the University of Melbourne
Rajkumar Pujari, purdue.edu
Yusu Qian, Apple
Sunny Rai, University of Pennsylvania
Sree Harsha Ramesh, UMass Amherst
Surangika Ranathunga, university of moratuwa
Shruti Rijhwani, Google
Hayley Ross, Harvard University
Hossein Rouhizadeh, University of Geneva
Shamik Roy, Purdue University
Vlada Rozova, RMIT University
Andreas Rücklé, Amazon
Philipp Sadler, University of Potsdam
Oscar Sainz, University of the Basque Country (UPV/EHU)
Sashank Santhanam, University of North Carolina at Charlotte/ Apple
Abulhair Saparov, New York University
Ryohei Sasano, Nagoya University
Tatjana Scheffler, Ruhr University Bochum
Michael Sejr Schlichtkrull, University of Cambridge
Viktoria Schram, University of Melbourne
# Table of Contents

ChatGPT vs Human-authored Text: Insights into Controllable Text Summarization and Sentence Style Transfer  
Dongqi Pu and Vera Demberg ................................................................. 1

Multi-Dialectal Representation Learning of Sinitic Phonology  
Zhibai Jia ................................................................................................. 19

Prompt-based Zero-shot Text Classification with Conceptual Knowledge  
Yuqi Wang, Wei Wang, Qi Chen, Kaizhu Huang, Anh Nguyen and Suparna De .............. 30

How do different tokenizers perform on downstream tasks in scriptio continua languages?: A case study in Japanese  
Takuro Fujii, Koki Shibata, Atsuki Yamaguchi, Terufumi Morishita and Yasuhiro Sogawa ... 39

Semantic-Aware Dynamic Retrospective-Prospective Reasoning for Event-Level Video Question Answering  
Chenyang Lyu, Tianbo Ji, Yvette Graham and Jennifer Foster ........................................ 50

Jamp: Controlled Japanese Temporal Inference Dataset for Evaluating Generalization Capacity of Language Models  
Tomoki Sugimoto, Yasumasa Onoe and Hitomi Yanaka .................................................. 57

Constructing Multilingual Code Search Dataset Using Neural Machine Translation  
Ryo Sekizawa, Nan Duan, Shuai Lu and Hitomi Yanaka .................................................. 69

Multimodal Neural Machine Translation Using Synthetic Images Transformed by Latent Diffusion Model  
Ryoya Yuasa, Akihiro Tamura, Tomoyuki Kajiwara, Takashi Ninomiya and Tsuneo Kato .... 76

Enhancing Ancient Chinese Understanding with Derived Noisy Syntax Trees  
Ping Wang, Shitou Zhang, Zuchao Li and Jingrui Hou ....................................................... 83

The Turing Quest: Can Transformers Make Good NPCs?  
Qi Chen Gao and Ali Emami ................................................................. 93

Making the Most Out of the Limited Context Length: Predictive Power Varies with Clinical Note Type and Note Section  
Hongyi Zheng, Yixin Zhu, Lavender Jiang, Kyunghyun Cho and Eric Oermann ............... 104

Intriguing Effect of the Correlation Prior on ICD-9 Code Assignment  
Zihao Yang, Chenkang Zhang, Muru Wu, Xujin Liu, Lavender Jiang, Kyunghyun Cho and Eric Oermann ................................................................. 109

Classical Out-of-Distribution Detection Methods Benchmark in Text Classification Tasks  
Mateusz Baran, Joanna Baran, Mateusz Wójcik, Maciej Zieba and Adam Gonczarek ........ 119

Can LMs Store and Retrieve 1-to-N Relational Knowledge?  
Haruki Nagasawa, Benjamin Heinzerling, Kazuma Kokuta and Kentaro Inui ............... 130

Theoretical Linguistics Rivals Embeddings in Language Clustering for Multilingual Named Entity Recognition  
Sakura Imai, Daisuke Kawahara, Naho Orita and Hiromune Oda ..................................... 139