ACL-SRW 2023

The 61st Annual Meeting of the Association for Computational Linguistics

Proceedings of the Student Research Workshop

July 10-12, 2023

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

Gold

Google

Bronze

In cooperation with

©2023 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-959429-69-2

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 A further 9 submissions were desk rejected and 2 submissions were withdrawn by the authors prior to the completion of the review process.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
Sedeeq Al-khazraji, 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 Baumgärtner, 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

Neele Falk, University of Stuttgart
Antonio Farinhas, Instituto de Telecomunicacoes, IST
Francesca Franzon, Universitat Pompeu Fabra
Jie Fu, Beijing Academy of Artificial Intelligence
Yoshinari Fujinuma, AWS AI Labs
Rena Gao, The University of Melbourne
Iker García-ferrero, HiTZ Basque Center for Language Technologies - Ixa NLP Group, University of the Basque Country UPV/EHU
Aina Garí Soler, LTCI, Télécom-Paris, Institut Polytechnique de Paris
Esam Ghaleb, University of Amsterdam
Anmol Goel, IIIT Hyderabad
Matt Grenander, University of Edinburgh
Jonas Groschwitz, University of Amsterdam (UvA)
Ivan Habernal, Technical University of Darmstadt
Xudong Han, The university of Melbourne
Dirk Hovy, Bocconi University
Dieuwke Hupkes, Facebook AI Research
Joseph Marvin Imperial, National University
Tunazzina Islam, Purdue University
Jeff Jacobs, Columbia University
Labiba Jahan, Augustana College
Jyoti Jha, Ernst and Young
Lavender Jiang, New York University
Abhinav Joshi, Indian Institute of Technology Kanpur
Nitish Joshi, New York University
Jad Kabbara, MIT
Patrick Kahardipraja, University of Potsdam
Ehsan Kamalloo, University of Waterloo
Haoqiang Kang, University of Washington
Sarvnaz Karimi, CSIRO
Borhan Kazimipour, Australian Taxation Office
Anirudh Vishal Khatry, Microsoft
Prashant Kodali, IIIT Hyderabad
Mamoru Komachi, Hitotsubashi University
Fajri Koto, MBZUAI
Maria Kunilovskaya, Saarland University
Mascha Kurpicz-briki, Bern University of Applied Sciences
Alexandra Lavrentovich, Amazon Alexa
Celine Lee, Cornell University
Richard Leibbrandt, Flinders University
Christoph Leiter, Bielefeld University - CITEC NLLG
Manling Li, UIUC
Jasy Suet Yan Liew, School of Computer Sciences, Universiti Sains Malaysia
Zheng Wei Lim, University of Melbourne
Lucy Lin, Spotify
Robert Litschko, LMU Munich
Tatiana Litvinova, Voronezh State Pedagogical University
Chunhua Liu, The University of Melbourne
Robert L Logan Iv, Dataminr
Nicholas Lourie, New York University
Qing Lyu, University of Pennsylvania

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

Sina Sheikholeslami, KTH Royal Institute of Technology
Manish Shrivastava, International Institute of Information Technology Hyderabad
Chenglei Si, University of Maryland, College Park
Katira Soleyman Zadeh, Ege University
Mayank Soni, ADAPT Centre, Trinity College Dublin
Cesare Spinoso-di Piano, McGill University
Tejas Srinivasan, University of Southern California
Vivek Srivastava, TCS Research
Marija Stanojevic, WinterLightLabs
Ashima Suvarna, UCLA
Stan Szpakowicz, EECS, University of Ottawa
Dage Särg, University of Tartu
Shabnam Tafreshi, UMD:ARLIS
Koichi Takeda, Nagoya University
Zeerak Talat, Simon Fraser University
Chuanqi Tan, Alibaba Group
Hao Tan, Adobe Research
Evgeniia Tokarchuk, University of Amsterdam
Thinh Hung Truong, The University of Melbourne
David Trye, University of Waikato
Sowmya Vajjala, National Research Council
Andrea Varga, Theta Lake
Francielle Vargas, University of São Paulo
Saranya Venkatraman, The Pennsylvania State University
Takashi Wada, The University of Melbourne
Zhiruo Wang, Carnegie Mellon University
Yuxia Wang, The University of Melbourne
Jun Wang, University of Melbourne
Bonnie Webber, University of Edinburgh
Steven Wilson, Oakland University
Patrick Wu, New York University
Zhuohan Xie, The University of Melbourne
Rui Xing, University of Melbourne
Yue Yang, University of Pennsylvania
Wenting Ye, ByteDance Inc
Hiyori Yoshikawa, Fujitsu Limited
Zheng Yuan, Alibaba Group
Zenan Zhai, University of Melbourne
Zixuan Zhang, University of Illinois Urbana-Champaign
Mike Zhang, IT University of Copenhagen
Ran Zhang, University of Bielefeld
Zheng Zhao, University of Edinburgh
Danna Zheng, University of Edinburgh
Zhong Zhou, Carnegie Mellon University
Rongxin Zhu, The University of Melbourne
Elena Zotova, Vicomtech

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

Native Language Prediction from Gaze: a Reproducibility Study
Lina Skerath, Paulina Toborek, Anita Zielińska, Maria Barrett and Rob Van Der Goot 152

MedTem2.0: Prompt-based Temporal Classification of Treatment Events from Discharge Summaries
Yang Cui, Lifeng Han and Goran Nenadic 160

Sudden Semantic Shifts in Swedish NATO discourse
Brian Bonafilia, Bastiaan Bruinsma, Denitsa Saynova and Moa Johansson 184

Building a Buzzer-quiz Answering System
Naoya Sugiura, Kosuke Yamada, Ryohei Sasano, Koichi Takeda and Katsuhiko Toyama 194

Probing for Hyperbole in Pre-Trained Language Models
Nina Schneidermann, Daniel Hershcovich and Bolette Pedersen 200

Towards Efficient Dialogue Processing in the Emergency Response Domain
Tatiana Anikina 212

I already said that! Degenerating redundant questions in open-domain dialogue systems.
Long Mai and Julie Carson-berndsen 226

Is a Knowledge-based Response Engaging?: An Analysis on Knowledge-Grounded Dialogue with Information Source Annotation
Takashi Kodama, Hirokazu Kiyomaru, Yin Jou Huang, Taro Okahisa and Sadao Kurohashi . . 237

Choosing What to Mask: More Informed Masking for Multimodal Machine Translation
Julia Sato, Helena Caseli and Lucia Specia 244

Combining Tradition with Modernness: Exploring Event Representations in Vision-and-Language Models for Visual Goal-Step Inference
Chong Shen and Carina Silberer 254

Data Selection for Fine-tuning Large Language Models Using Transferred Shapley Values
Stephanie Schoch, Yangfeng Ji and Ritwick Mishra 266

Distractor Generation for Fill-in-the-Blank Exercises by Question Type
Nana Yoshimi, Tomoyuki Kajiwara, Satoru Uchida, Yuki Arase and Takashi Ninomiya 276

Moral Mimicry: Large Language Models Produce Moral Rationalizations Tailored to Political Identity
Gabriel Simmons 282

LECO: Improving Early Exiting via Learned Exits and Comparison-based Exiting Mechanism
Jingfan Zhang, Ming Tan, Pengyu Dai and Wei Zhu 298

Authorship Attribution of Late 19th Century Novels using GAN-BERT
Kanishka Silva, Burcu Can, Frédéric Blain, Raheem Sarwar, Laura Ugolini and Ruslan Mitkov 310

How-to Guides for Specific Audiences: A Corpus and Initial Findings
Nicola Fanton, Agnieszka Falenska and Michael Roth 321

When Words Fail, Emojis Prevail: A Novel Architecture for Generating Sarcastic Sentences With Emoji Using Valence Reversal and Semantic Incongruity
Faria Binte Kader, Nafisa Hossain Nujat, Tasmia Binte Sogir, Mohsinul Kabir, Hasan Mahmud and Md Kamrul Hasan 334

Semantic Accuracy in Natural Language Generation: A Thesis Proposal
Patricia Schmidtova 352

Math Word Problem Solving by Generating Linguistic Variants of Problem Statements
Syed Rifat Raiyan, Md Nafis Faiyaz, Shah Md. Jawad Kabir, Mohsinul Kabir, Hasan Mahmud and Md Kamrul Hasan 362