

EACL 2024

**The 18th Conference of the European Chapter of the
Association for Computational Linguistics**

Proceedings of the Student Research Workshop

March 21-22, 2024

The EACL organizers gratefully acknowledge the support from the following sponsors.

Kindly supported by



©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 979-8-89176-090-5

Preface of the EACL SRW 2024 by the Workshop Co-chairs

Welcome to the EACL 2024 Student Research Workshop!

The EACL 2024 Student Research Workshop (SRW) is a forum for students in the field of Computational Linguistics and Natural Language Processing to come together to discuss and advance their research with help from more experienced researchers from both academia and industry.

Following the tradition of the previous student research workshops, we have two tracks: research papers and thesis proposals. The research paper track is a venue for PhD students, Master's students, and advanced undergraduate students to describe completed work or work-in-progress along with preliminary results. The thesis proposal track is offered for students who have decided on a thesis topic and are interested in receiving feedback for their proposal with suggestions for both making the ideas achievable, as well as discussions related to future directions for their work.

The student research workshop has received considerable attention, and papers have addressed research questions in various areas. After excluding the desk-rejected (1) and withdrawn (1) submissions, we received 58 submissions in total: 6 thesis proposals and 52 research papers (divided into 50 direct submissions and 2 through ARR Commitment, where 1 ARR Commit withdrew). We accepted 3 thesis proposals and 29 research papers, resulting in an overall acceptance rate of around 55%. Excluding non-archival papers (3), 29 papers appear in these proceedings. All the accepted papers will be presented as part of the EACL conference. Moreover, similar to previous years, we organized in-person mentorship to enable the students to receive feedback and engage in discussions about their research directions, as well as improve the quality of the poster presentation of accepted papers.

We thank our program committee members for providing careful and comprehensive reviews for the papers, and all of our mentors for donating their time to provide feedback to our student authors. Thanks to our faculty advisors, Steffen Eger and Parisa Kordjamshidi, for the essential advice and suggestions, and to the EACL 2024 organizing committee for their support in the entire process. Finally, we would like to thank all the authors whose participation has made the workshop a success!

Program Committee

Program Chairs

Neele Falk, University of Stuttgart
Sara Papi, University of Trento and Fondazione Bruno Kessler
Mike Zhang, IT University of Copenhagen

Faculty Advisors

Steffen Eger, University of Mannheim
Parisa Kordjamshidi, Michigan State University

Program Committee

Gavin Abercrombie, Heriot-Watt University
Alan Akbik, Humboldt Universität Berlin
Sergi Alvarez-Vidal, Universitat Pompeu Fabra and Universitat Oberta de Catalunya
Parnia Bahar, RWTH Aachen
Valerio Basile, University of Turin
Pierpaolo Basile, University of Bari
Shaily Bhatt, Carnegie Mellon University
Chris Biemann, University Hamburg
Rusla Biswas, Hasso Plattner Institute
Cristina Bosco, University of Turin
Aoife Cahill, Dataminr
Ruken Cakici, Middle East Technical University
Tommaso Caselli, University of Groningen
Silvia Casola, University of Turin, University of Padua and Fondazione Bruno Kessler
Giovanni Cassani, Tilburg University
Asli Celikyilmaz, FAIR
Tanise Ceron, Universität Stuttgart
Grzegorz Chrupała, Tilburg University
Danilo Croce, University of Roma
Chris Develder, Universiteit Gent
Antoine Doucet, University of La Rochelle
Esra Dönmez, Universität Stuttgart
Annerose Eichel, Universität Stuttgart
Desmond Elliott, University of Copenhagen
Ali Emami, Brock University
Agnieszka Falenska, Interchange Forum for Reflecting on Intelligent Systems, University of Stuttgart
Dayne Freitag, SRI International
Simona Frenda, University of Turin
Dennis Fucci, Fondazione Bruno Kessler
Richard Futrell, University of California, Irvine
Anmol Goel, Technische Universität Darmstadt
Dan Goldwasser, Purdue University

Rob van der Goot, IT University of Copenhagen
Gerard I. Gállego, Universidad Politécnica de Catalunya
Darryl Hannan, Pacific Northwest National Laboratory and Drexel University
Filip Ilievski, Vrije Universiteit Amsterdam
Javier Iranzo-Sánchez, AppTek
Mete Ismayilzada, EPFL - EPF Lausanne
Amr Keleg, University of Edinburgh
Simon King, University of Edinburgh
Mare Koit, ETIS
Florian Kunneman, Utrecht University
Murathan Kurfali, Stockholm University
Moreno La Quatra, Kore University of Enna
Mirko Lai, University of Turin
Beomseok Lee, University of Trento, Fondazione Bruno Kessler and Naver Labs Europe
Inna Wanyin Lin, University of Washington
Danni Liu, Karlsruher Institut für Technologie
Natalia V Loukachevitch, Lomonosov Moscow State University
Shervin Malmasi, Amazon
Sepideh Mamooler, School of Computer and Communication Sciences, EPFL - EPF Lausanne
Bruno Martins, Instituto Superior Técnico
Louise McNally, Universitat Pompeu Fabra
Filip Miletic, University of Stuttgart
Prakanya Mishra, AMD AI
Aditya Mogadala, Amazon
Hoang H Nguyen, University of Illinois at Chicago
Dmitry Nikolaev, University of Manchester
Tae-Gil Noh, OMQ GmbH
Daniel Ortega, University of Stuttgart, Universität Stuttgart
Sebastian Padó, University of Stuttgart, Universität Stuttgart
Rohan Pandey, Department of Computer Science, University of Massachusetts at Amherst
Ted Pedersen, University of Minnesota, Duluth
Van-Thuy Phi, RIKEN
Prisca Piccirilli, Universität Stuttgart
Flor Miriam Plaza-del-Arco, Bocconi University
Marco Polignano, University of Bari Aldo Moro
Peter Polák, Charles University
Mathieu Roche, Centre de coopération internationale en recherche agronomique pour le développement
Michael Roth, University of Stuttgart
Andrea Santilli, Sapienza University of Rome
Beatrice Savoldi, Fondazione Bruno Kessler
Sabine Schulte Im Walde, University of Stuttgart
Rico Sennrich, University of Zurich and University of Edinburgh
Cheril Shah, University of Pune
Sina Sheikholeslami, KTH Royal Institute of Technology
A.b. Siddique, University of Kentucky
Max Silberztein, Université de Franche-Comté
Tarun Tater, Universität Stuttgart
Ashok Urlana, IIIT-Hyderabad
Sowmya Vajjala, National Research Council Canada
Bram Vanroy, KU Leuven

Bonnie L. Webber, Edinburgh University, University of Edinburgh
Amelie Wuehrl, University of Stuttgart, Universität Stuttgart
Shanshan Xu, Technische Universität München
Ziyu Yao, George Mason University

Table of Contents

<i>AutoAugment Is What You Need: Enhancing Rule-based Augmentation Methods in Low-resource Regimes</i>	
Juhwan Choi, Kyohoon Jin, Junho Lee, Sangmin Song and YoungBin Kim	1
<i>Generating Diverse Translation with Perturbed kNN-MT</i>	
Yuto Nishida, Makoto Morishita, Hidetaka Kamigaito and Taro Watanabe	9
<i>The KIND Dataset: A Social Collaboration Approach for Nuanced Dialect Data Collection</i>	
Asma Z. Yamani, Raghad Alziyady, Reem AlYami, Salma A. Albelali, Leina A. Albelali, Jawhara Almulhim, Amjad K. Alsulami, Motaz Alfarraj and Rabeah A. Al-Zaidy	32
<i>Can Stanza be Used for Part-of-Speech Tagging Historical Polish?</i>	
Maria Irena Szawerna	44
<i>Toward Zero-Shot Instruction Following</i>	
Renze Lou and Wenpeng Yin	50
<i>UnMASKed: Quantifying Gender Biases in Masked Language Models through Linguistically Informed Job Market Prompts</i>	
Iñigo Parra	61
<i>Distribution Shifts Are Bottlenecks: Extensive Evaluation for Grounding Language Models to Knowledge Bases</i>	
Yiheng Shu and Zhiwei Yu	71
<i>AttriSage: Product Attribute Value Extraction Using Graph Neural Networks</i>	
Rohan Potta, Mallika Asthana, Siddhant Yadav, Nidhi Goyal, Sai Amrit Patnaik and Parul Jain	89
<i>HypoTermQA: Hypothetical Terms Dataset for Benchmarking Hallucination Tendency of LLMs</i>	
Cem Uluoglakci and Tugba Taskaya Temizel	95
<i>Arabic Synonym BERT-based Adversarial Examples for Text Classification</i>	
Norah F Alshahrani, Saied Alshahrani, Esmā Wali and Jeanna Matthews	137
<i>A Hypothesis-Driven Framework for the Analysis of Self-Rationalising Models</i>	
Marc Braun and Jenny Kunz	148
<i>Align before Attend: Aligning Visual and Textual Features for Multimodal Hateful Content Detection</i>	
Eftekhār Hossain, Omar Sharif, Mohammed Moshīul Hoque and Sarah Masud Preum	162
<i>Topic-guided Example Selection for Domain Adaptation in LLM-based Machine Translation</i>	
Seth Aycock and Rachel Bawden	175
<i>Reforging : A Method for Constructing a Linguistically Valid Japanese CCG Treebank</i>	
Asa Tomita, Hitomi Yanaka and Daisuke Bekki	196
<i>Thesis Proposal: Detecting Agency Attribution</i>	
Igor Ryazanov and Johanna Björklund	208
<i>A Thesis Proposal ClaimInspector Framework: A Hybrid Approach to Data Annotation using Fact-Checked Claims and LLMs</i>	
Basak Bozkurt	215

<i>Large Language Models for Mathematical Reasoning: Progresses and Challenges</i> Janice Ahn, Rishu Verma, Renze Lou, Di Liu, Rui Zhang and Wenpeng Yin	225
<i>Representation and Generation of Machine Learning Test Functions</i> Souha Ben Hassine and Steven R. Wilson	238
<i>The Generative AI Paradox in Evaluation: What It Can Solve, It May Not Evaluate</i> Juhyun Oh, Eunsu Kim, Inha Cha and Alice Oh	248
<i>Generative Data Augmentation using LLMs improves Distributional Robustness in Question Answering</i> Arijit Ghosh Chowdhury and Aman Chadha	258
<i>Japanese-English Sentence Translation Exercises Dataset for Automatic Grading</i> Naoki Miura, Hiroaki Funayama, Seiya Kikuchi, Yuichiroh Matsubayashi, Yuya Iwase and Kentaro Inui	266
<i>The Impact of Integration Step on Integrated Gradients</i> Masahiro Makino, Yuya Asazuma, Shota Sasaki and Jun Suzuki	279
<i>GesNavi: Gesture-guided Outdoor Vision-and-Language Navigation</i> Aman Jain, Teruhisa Misu, Kentaro Yamada and Hitomi Yanaka	290
<i>Can docstring reformulation with an LLM improve code generation?</i> Nicola Dainese, Alexander Ilin and Pekka Marttinen	296
<i>Benchmarking Diffusion Models for Machine Translation</i> Yunus Demirag, Danni Liu and Jan Niehues	313
<i>Forged-GAN-BERT: Authorship Attribution for LLM-Generated Forged Novels</i> Kanishka Silva, Ingo Frommholz, Burcu Can, Fred Blain, Raheem Sarwar and Laura Ugolini	325
<i>Thesis Proposal: Detecting Empathy Using Multimodal Language Model</i> Md Rakibul Hasan, Md Zakir Hossain, Aneesh Krishna, Shafin Rahman and Tom Gedeon ..	338
<i>Toward Sentiment Aware Semantic Change Analysis</i> Roksana Goworek and Haim Dubossarsky	350
<i>Dynamic Task-Oriented Dialogue: A Comparative Study of Llama-2 and Bert in Slot Value Generation</i> Tiziano Labruna, Sofia Brenna and Bernardo Magnini	358

Program

Tuesday, March 19, 2024

- 14:00 - 15:30 *Poster Session (7)*
- 16:00 - 16:10 *Opening Remarks*
- 16:15 - 17:15 *Panel*
- 17:20 - 17:55 *Best Student Papers Session*
- 17:55 - 18:00 *Closing Remarks*