

NAACL 2024

**The 2024 Conference of the North American Chapter of the
Association for Computational Linguistics**

Proceedings of the Demonstrations Session

June 16-21, 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 979-8-89176-116-2

Introduction

Welcome to the proceedings of the System Demonstration Track at NAACL 2024, taking place from June 16 to June 21.

The System Demonstration Track at NAACL 2024 serves as a platform for presenting papers that describe system demonstrations, ranging from early prototypes to mature, production-ready systems. Publicly available open-source or open-access systems are of special interest.

This year, we received 48 submissions, and 21 were selected for inclusion in the program, resulting in an acceptance rate of 43

We would like to thank the members of the program committee for their timely assistance in reviewing the submissions and to the authors who submitted their work to the demonstration track.

Best,
Kai-Wei Chang, En-Shiun Annie Lee, and Nazneen Rajani,
NAACL 2024 Demonstration Track Chairs

Program Committee

Program Chairs

Kai-Wei Chang, University of California, Los Angeles
En-Shiun Annie Lee, OntarioTech University, Oshawa, and University of Toronto, Toronto
Nazneen Rajani

Reviewers

Ahmed Abdelali, Hamad Bin Khalifa University
Rodrigo Agerri, University of the Basque Country
Zeynep Akkalyoncu, StackAdapt
Miguel A. Alonso, Universidade da Coruña
Rafael Anchiêta, Federal Institute of Piauí
Lisa Bauer, Amazon
Shubhang Bhatnagar, University of Illinois, Urbana-Champaign
Aljoscha Burchardt, German Research Center for AI
Daniel F Campos, Snowflake
Daniel Cer, Google
Wei-Rui Chen, University of British Columbia
Guanyi Chen, Central China Normal University
Jih-Jie Chen, National Tsinghua University
Chung-Chi Chen, AIST, National Institute of Advanced Industrial Science and Technology
Christos Christodoulopoulos, Amazon
Marina Danilevsky, International Business Machines
Michael Desmond, IBM
Shizhe Diao, Hong Kong University of Science and Technology
James Fan
Chi Han, University of Illinois Urbana-Champaign
Ales Horak, Masaryk University
Claudiu Daniel Hromei, University of Roma Tor Vergata
Ziniu Hu, Caltech
Xiaodan Hu, University of Illinois, Urbana-Champaign
Kuan-Hao Huang, University of Illinois Urbana-Champaign
Ali Hürriyetoglu, Wageningen University & Research
Vivek Iyer, University of Edinburgh, University of Edinburgh
Sneha Jha, Imperial College London
Pratik Joshi, Google DeepMind
Khalid Al Khatib, University of Groningen
Philipp Koehn, Johns Hopkins University
Mamoru Komachi, Hitotsubashi University
Tuan Lai, University of Illinois Urbana Champaign
Mark Last, Ben-Gurion University of the Negev
Xintong Li, Apple
Yanran Li
Bill Yuchen Lin, AI2
Ye Liu, Universität Ulm
Pan Lu, University of California, Los Angeles
Wolfgang Maier, Mercedes Benz Research & Development

Tao Meng, University of California, Los Angeles
Margot Mieskes, University of Applied Sciences Darmstadt
Aldrian Obaja Muis
Pierre Nugues, Lund University
Tsuyoshi Okita, Kyushu Institute of Technology
Eda Okur, Intel Labs
Tanmay Parekh, University of California, Los Angeles
Xutan Peng, Huawei Technologies Ltd.
Prokopis Prokopidis, Institute for Language and Speech Processing, Athena Research Center
Saurav Sahay, Intel
Jaydeep Sen, IBM
Sheng Shen, University of California Berkeley
Chenkai Sun, University of Illinois at Urbana-Champaign
Yufei Tian, University of California, Los Angeles
Natalia Vanetik, SCE
Andrea Varga, Theta Lake Ltd
Nikhil Verma, LG Toronto AI Lab
Xuan Wang, Virginia Polytechnic Institute and State University
Qingyun Wang, University of Illinois, Urbana Champaign
Jingjing Wang, soochow university
Yiwei Wang, University of California, Los Angeles
Xianchao Wu, NVIDIA
Di Wu, University of California, Los Angeles
Zhen Xu, PCG Tencent
Ashok Yadav, Indian Institute of Information Technology Allahabad
Seid Muhie Yimam, University Hamburg
Pengfei Yu, University of Illinois at Urbana-Champaign
Dian Yu, Google
Qi Zeng, Facebook
Wenjie Zi, Borealis AI
Imed Zitouni, Google

Table of Contents

TOPICAL: TOPIC Pages Automagically

John Michael Giorgi, Amanpreet Singh, Doug Downey, Sergey Feldman and Lucy Lu Wang . . . 1

Low-code LLM: Graphical User Interface over Large Language Models

Yuzhe Cai, Shaoguang Mao, Wenshan Wu, Zehua Wang, Yaobo Liang, Tao Ge, Chenfei Wu, WangYou WangYou, Ting Song, Yan Xia, Nan Duan and Furu Wei 12

EdTec-QBuilder: A Semantic Retrieval Tool for Assembling Vocational Training Exams in German Language

Alonso Palomino, Andreas Fischer, Jakub Kuzilek, Jarek Nitsch, Niels Pinkwart and Benjamin Paassen 26

DIALIGHT: Lightweight Multilingual Development and Evaluation of Task-Oriented Dialogue Systems with Large Language Models

Songbo Hu, Xiaobin Wang, Moy Yuan, Anna Korhonen and Ivan Vulić 36

RTSUM: Relation Triple-based Interpretable Summarization with Multi-level Saliency Visualization

Seonglae Cho, Myungha Jang, Jinyoung Yeo and Dongha Lee 53

Edu-ConvoKit: An Open-Source Library for Education Conversation Data

Rose E Wang and Dorottya Demszky 61

jp-evalb: Robust Alignment-based PARSEVAL Measures

Jungyeul Park, Junrui Wang, Eunkyul Leah Jo and Angela Yoonseo Park 70

OpinionGPT: Modelling Explicit Biases in Instruction-Tuned LLMs

Patrick Haller, Ansar Aynetdinov and Alan Akbik 78

ATLAS: A System for PDF-centric Human Interaction Data Collection

Alexa Siu, Zichao Wang, Joshua Hoefflich, Naman Kapasi, Ani Nenkova and Tong Sun 87

BeLeaf: Belief Prediction as Tree Generation

John Murzaku and Owen Rambow 97

QueryExplorer: An Interactive Query Generation Assistant for Search and Exploration

Kaustubh Dhole, Shivam Bajaj, Ramraj Chandradevan and Eugene Agichtein 107

LMFlow: An Extensible Toolkit for Finetuning and Inference of Large Foundation Models

Shizhe Diao, Rui Pan, Hanze Dong, KaShun Shum, Jipeng Zhang, Wei Xiong and Tong Zhang
116

DOCMaster: A Unified Platform for Annotation, Training, & Inference in Document Question-Answering

Alex Nguyen, Zilong Wang, Jingbo Shang and Dheeraj Mekala 128

RedCoast: A Lightweight Tool to Automate Distributed Training of LLMs on Any GPU/TPUs

Bowen Tan, Yun Zhu, Lijuan Liu, Hongyi Wang, Yonghao Zhuang, Jindong Chen, Eric P. Xing and Zhiting Hu 137

Concept Over Time Analysis: Unveiling Temporal Patterns for Qualitative Data Analysis

Tim Fischer, Florian Schneider, Robert Geislinger, Florian Helfer, Gertraud Koch and Chris Bie-
mann 148

<i>pyvene: A Library for Understanding and Improving PyTorch Models via Interventions</i> Zhengxuan Wu, Atticus Geiger, Aryaman Arora, Jing Huang, Zheng Wang, Noah Goodman, Christopher D Manning and Christopher Potts	158
<i>Newspaper Signaling for Crisis Prediction</i> Prajvi Saxena, Sabine Janzen and Wolfgang Maass	166
<i>FastFit: Fast and Effective Few-Shot Text Classification with a Multitude of Classes</i> Asaf Yehudai and Elron Bandel	174
<i>AgentQuest: A Modular Benchmark Framework to Measure Progress and Improve LLM Agents</i> Luca Gioacchini, Giuseppe Siracusano, Davide Sanvito, Kiril Gashteovski, David Friede, Roberto Bifulco and Carolin Lawrence	185
<i>ZhuJiu-Knowledge: A Fairer Platform for Evaluating Multiple Knowledge Types in Large Language Models</i> Pengfan DU, Sirui Liang, Baoli Zhang, Pengfei Cao, Yubo Chen, Kang Liu and Jun Zhao ...	194
<i>Unitxt: Flexible, Shareable and Reusable Data Preparation and Evaluation for Generative AI</i> Elron Bandel, Yotam Perlitz, Elad Venezian, Roni Friedman, Ofir Arviv, Matan Orbach, Shachar Don-Yehiya, Dafna Sheinwald, Ariel Gera, Leshem Choshen, Michal Shmueli-Scheuer and Yoav Katz	207