

ACL 2025

**The 63rd Annual Meeting of the Association for
Computational Linguistics**

Proceedings of the Conference (System Demonstrations)

July 27 - August 1, 2025

©2025 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 979-8-89176-253-4

Introduction

Welcome to the proceedings of the System Demonstration Track of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025), held from July 27 – August 1, 2025 in Vienna, Austria.

The ACL 2025 System Demonstration Track provides a platform for papers describing system demonstrations, ranging from early prototypes to mature, production-ready systems. We are particularly interested in publicly available open-source or open-access systems.

For the ACL 2025 System Demonstration Track, we received a record 187 submissions, of which 178 papers were valid with required materials. We carefully checked all submitted reviews. Based on these reviews, we have accepted 64 papers, resulting in an acceptance rate of 34.22%, in line with previous years.

From the accepted papers, we short-listed 7 papers for the Best System Demonstration award. We sincerely thank the members of the award committee for their invaluable contributions in determining the best system demonstration: Christopher Hidey, Junxian He, Rui Zhang, Milad Alshomary, and Phu Mon Htut.

Pushkar Mishra

Smaranda Muresan

Tao Yu

ACL 2025 Demonstration Chairs

Program Committee

Program Chairs

Pushkar Mishra, Google DeepMind and University of Cambridge
Smaranda Muresan, Amazon and Columbia University
Tao Yu, The University of Hong Kong

Area Chairs

Tariq Alhindi
Milad Alshomary
Lora Aroyo, Google DeepMind
Shizhe Diao, NVIDIA
Ramy Eskander, Google
Junxian He, Hong Kong University of Science and Technology
Christopher Hidey, Google
Phu Mon Htut, AWS AI Labs
Ji-Ung Lee, Saarland University, Universität des Saarlandes
Aleksandra Piktus
Charvi Rastogi, Google
Sara Rosenthal, International Business Machines
Simra Shahid, University of Virginia, Charlottesville and Adobe Systems
Huajie Shao, College of William and Mary
Shannon Zejiang Shen
Anthony Sicilia, Northeastern University
Hua Wei, Arizona State University
Zhaozhuo Xu, Stevens Institute of Technology
Rui Zhang, Pennsylvania State University
Denghui Zhang, Stevens Institute of Technology

Reviewers

Sharada S Acharya, Pegah Ahadian, Sharifa Alghowinem, Amith Ananthram, Neha Anna John, Mario Ezra Aragon

Long Bai, Georgeta Bordea

Sky CH-Wang, Yuxuan Cao, Yekun Chai, Chung-Chi Chen, Jixuan Chen, Kehai Chen, Shiqi Chen, Tiejin Chen, Xinchu Chen, Mao Cheng, Zhoujun Cheng, Bonaventura Coppola

Longchao Da, Nicholas Deas, Luigi Di Caro, Zi-Yi Dou

Georgina Evans

Anthony Ferritto, Yeqiao Fu

Ashley Gao, Debanjan Ghosh, Debanjan Ghosh, Voula Giouli, Ziwei Gong, Yi Gu, Renxiang Guan

Ivan Habernal, Shanshan Han, Chaitra Hegde, Arpana Hosabettu, Chan-Wei Hu, Dayu Hu, Yue Huang, Yuzhen Huang

Saki Imai, Mert Inan, Avitej Iyer

Anubhav Jangra, Hyewon Jeong, Xueying Jia, Xiaomeng Jin

Yoshihide Kato, Jeonghwan Kim, Maarit Koponen

Yuhang Lai, Yuxuan Lai, Hoang Anh Duy Le, Chengzu Li, Haifang Li, Ryan Li, Shuo Li, Weijiang Li, Xiaochang Li, Xinjin Li, Xintong Li, Zixuan Li, Yueqian Lin, Jiangming Liu, Junteng Liu, Tianrui Liu, Wei Liu, Xiao Liu, Xiaoou Liu, Yuhang Liu, Zeyu Leo Liu, Zheng Liu, Dunjie Lu, Xinyu Lu, Yaojie Lu, Yuxuan Lu

Aashrith Madasu, Manuel Mager, Maria Mahbub, Lingyuan Meng, Nicolo Micheletti, Amita Misra

Youyang Ng, Huy V. Nguyen, Pin Ni

Yanzhou Pan, Ajay Patel, Yevgeniy Puzikov

Sibo Qi, Cheng Qian

Yide Ran, Eti Rastogi, Valeria Ruscio

Arkadiy Saakyan, Meghdut Sengupta, Alay Dilipbhai Shah, Zhennan Shen, Tongyue Shi, Yucheng Shi, Dongchan Shin, Yiheng Shu, Gosuddin Kamaruddin Siddiqi, Jyotika Singh, Lucas Spangher, Miao Su, Melanie Subbiah, Chengjie Sun, Marek Suppa, Gözde Gül Şahin

Shangqing Tu, Marco Turchi

Enes Yavuz Ugan

Ivo Verhoeven

Yuxuan Wan, Junli Wang, Qingyun Wang, Wei Wang, Yiwei Wang, Zihan Wang, Isadora White, Olivia Winn, Xueqing Wu, Zongyu Wu

Shuo Xing, Meilong Xu, Xin Xu, Ziyun Xu

Fan Yang, Hao Yu, Yige Yuan

Karina Zainullina, Qingkai Zeng, Zhichen Zeng, Yuheng Zha, Qiusi Zhan, Jinghan Zhang, Junlei Zhang, Linfan Zhang, Tianlin Zhang, Xuanming Zhang, Zhenyu Zhang, Jian Zheng, Wenliang Zheng, Xianrui Zhong, Fan Zhou, Sizhe Zhou, Kaiwen Zuo

Table of Contents

<i>MapQaTor: An Extensible Framework for Efficient Annotation of Map-Based QA Datasets</i> Mahir Labib Dihan, Mohammed Eunus Ali and Md Rizwan Parvez	1
<i>PEIRCE: Unifying Material and Formal Reasoning via LLM-Driven Neuro-Symbolic Refinement</i> Xin Quan, Marco Valentino, Danilo Carvalho, Dhairya Dalal and Andre Freitas	11
<i>MERaLiON-AudioLLM: Advancing Speech and Language Understanding for Singapore</i> Yingxu He, Zhuohan Liu, Geyu Lin, Shuo Sun, Bin Wang, Wenyu Zhang, Xunlong Zou, Nancy F. Chen and AiTi Aw	22
<i>NameTag 3: A Tool and a Service for Multilingual/Multitagset NER</i> Jana Straková and Milan Straka	31
<i>Multi-Programming Language Sandbox for LLMs</i> Shihan Dou, Jiazheng Zhang, Jianxiang Zang, Yunbo Tao, Weikang Zhou, Haoxiang Jia, Shichun Liu, Yuming Yang, Shenxi Wu, Zhiheng Xi, Muling Wu, Rui Zheng, Changze Lv, Limao Xiong, Shaoqing Zhang, Lin Zhang, Wenyu Zhan, Rongxiang Weng, Jingang Wang, Xunliang Cai, Yueming Wu, Ming Wen, Yixin Cao, Tao Gui, Xipeng Qiu, Qi Zhang and Xuanjing Huang	40
<i>FlagEvalMM: A Flexible Framework for Comprehensive Multimodal Model Evaluation</i> Zheqi He, Yesheng Liu, Jing-Shu Zheng, Xuejing Li, Jin-Ge Yao, Bowen Qin, Richeng Xuan and Xi Yang	51
<i>My Climate CoPilot: A Question Answering System for Climate Adaptation in Agriculture</i> Vincent Nguyen, Willow Hallgren, Ashley Harkin, Mahesh Prakash and Sarvnaz Karimi	62
<i>SPOT: Bridging Natural Language and Geospatial Search for Investigative Journalism</i> Lynn Khellaf, Ipek Baris Schlicht, Tilman Mirass, Julia Bayer, Tilman Wagner and Ruben Bouwmeester	71
<i>Textagon: Boosting Language Models with Theory-guided Parallel Representations</i> John P. Lalor, Ruiyang Qin, David Dobolyi and Ahmed Abbasi	82
<i>GigaChat Family: Efficient Russian Language Modeling Through Mixture of Experts Architecture</i> Valentin Mamedov, Evgenii Kosarev, Gregory Leleytner, Ilya Shchuckin, Valeriy Berezovskiy, Daniil Smirnov, Dmitry Kozlov, Sergei Averkiev, Lukyanenko Ivan, Aleksandr Proshunin, Ainur Israfilova, Ivan Baskov, Artem Chervyakov, Emil Shakirov, Mikhail Kolesov, Daria Khomich, Daria Lortortseva, Sergei Porkhun, Yury Fedorov, Oleg Kutuzov, Polina Kudriavtseva, Sofiia Soldatova, Kolodin Egor, Stanislav Pyatkin, Dzmitry Menshykh, Grafov Sergei Iurevich, Eldar Damirov, Vladimir Karlov, Ruslan Gaitukiev, Arkadiy Shatenov, Alena Fenogenova, Nikita Savushkin and Fedor Minkin	93
<i>Unifying Language Agent Algorithms with Graph-based Orchestration Engine for Reproducible Agent Research</i> Qianqian Zhang, Jiajia Liao, Heting Ying, Yibo Ma, Haozhan Shen, Jingcheng li, Peng Liu, Lu Zhang, Chunxin Fang, Kyusong Lee, Ruochen Xu and Tiancheng Zhao	107
<i>Abacus-SQL: A Text-to-SQL System Empowering Cross-Domain and Open-Domain Database Retrieval</i> Keyan Xu, Dingzirui Wang, Xuanliang Zhang, Qingfu Zhu and Wanxiang Che	118
<i>Tulun: Transparent and Adaptable Low-resource Machine Translation</i> Raphael Merx, Hanna Suominen, Lois Yinghui Hong, Nick Thieberger, Trevor Cohn and Ekaterina Vylomova	129

<i>ReasonGraph: Visualization of Reasoning Methods and Extended Inference Paths</i>	
Zongqian Li, Ehsan Shareghi and Nigel Collier	140
<i>Dia-Lingle: A Gamified Interface for Dialectal Data Collection</i>	
Jiugeng Sun, Rita Sevastjanova, Sina Ahmadi, Rico Sennrich and Mennatallah El-Assady ...	148
<i>Token Level Routing Inference System for Edge Devices</i>	
Jianshu She, Wenhao Zheng, Zhengzhong Liu, Hongyi Wang, Eric P. Xing, Huaxiu Yao and Qirong Ho	159
<i>The ShareLM Collection and Plugin: Contributing Human-Model Chats for the Benefit of the Community</i>	
Shachar Don-Yehiya, Leshem Choshen and Omri Abend	167
<i>OLMoTrace: Tracing Language Model Outputs Back to Trillions of Training Tokens</i>	
Jiacheng Liu, Taylor Blanton, Yanai Elazar, Sewon Min, Yen-Sung Chen, Arnavi Chheda-Kothary, Huy Tran, Byron Bischoff, Eric Marsh, Michael Schmitz, Cassidy Trier, Aaron Sarnat, Jenna James, Jon Borchardt, Bailey Kuehl, Evie Yu-Yen Cheng, Karen Farley, Taira Anderson, David Albright, Carrissa Schoenick, Luca Soldaini, Dirk Groeneveld, Rock Yuren Pang, Pang Wei Koh, Noah A. Smith, Sophie Lebrecht, Yejin Choi, Hannaneh Hajishirzi, Ali Farhadi and Jesse Dodge	178
<i>AutoAlign: Get Your LLM Aligned with Minimal Annotations</i>	
Xinyu Lu, Dong Xu, Chunkang Zhang, Xinyan Guan, Junxiang Wang, Qingyu Zhang, Pengbo Wang, Yingzhi Mao, Hao Xiang, Xueru Wen, Zichao Li, Yaojie Lu, Hongyu Lin, Le Sun and Xianpei Han	189
<i>Know-MRI: A Knowledge Mechanisms Revealer&Interpreter for Large Language Models</i>	
Jiaxiang Liu, Boxuan Xing, Chenhao Yuan, ChenxiangZhang ChenxiangZhang, Di Wu, Xiusheng Huang, Haida Yu, Chuhan Lang, Pengfei Cao, Jun Zhao and Kang Liu	199
<i>AI4Reading: Chinese Audiobook Interpretation System Based on Multi-Agent Collaboration</i>	
Minjiang Huang, Jipeng Qiang, Yi Zhu, Chaowei Zhang, Xiangyu Zhao and Kui Yu	211
<i>DEEP: an automatic bidirectional translator leveraging an ASR for translation of Italian sign language</i>	
Nicolas Tagliabue, Elisa Colletti, Francesco Roberto Dani, Roberto Tedesco, Sonia Cenceschi and Alessandro Trivilini	221
<i>VenusFactory: An Integrated System for Protein Engineering with Data Retrieval and Language Model Fine-Tuning</i>	
Yang Tan, Chen Liu, Jingyuan Gao, Wu Banghao, Mingchen Li, Ruilin Wang, Lingrong Zhang, Huiqun Yu, Guisheng Fan, Liang Hong and Bingxin Zhou	230
<i>GenGO Ultra: an LLM-powered ACL Paper Explorer</i>	
Sotaro Takeshita, Tornike Tsereteli and Simone Paolo Ponzetto	242
<i>SpatialWebAgent: Leveraging Large Language Models for Automated Spatial Information Extraction and Map Grounding</i>	
Shunfeng Zheng, Meng Fang and Ling Chen	252
<i>DocSpiral: A Platform for Integrated Assistive Document Annotation through Human-in-the-Spiral</i>	
Qiang Sun, Sirui Li, Tingting Bi, Du Q. Huynh, Mark Reynolds, Yuanyi Luo and Wei Liu ...	267
<i>ADEPT-SQL: A High-performance Text-to-SQL Application for Real-World Enterprise-Level Databases</i>	
Yongnan Chen, Zhuo Chang, Shijia Gu, Yuanhang Zong, Zhang Mei, Shiyu Wang, Hezixiang Hezixiang, Hongzhi Chen, Jin Wei and Bin Cui	275

LCDS: A Logic-Controlled Discharge Summary Generation System Supporting Source Attribution and Expert Review

Cheng Yuan, Xinkai Rui, Yongqi Fan, Yawei Fan, Boyang Zhong, Jiacheng Wang, Weiyan Zhang and Tong Ruan 284

Revealing the Unwritten: Visual Investigation of Beam Search Trees to Address Language Model Prompting Challenges

Thilo Spinner, Rita Sevastjanova, Rebecca Kehlbeck, Tobias Stähle, Daniel A. Keim, Oliver Deussen, Andreas Spitz and Mennatallah El-Assady 295

Scholar Inbox: Personalized Paper Recommendations for Scientists

Markus Flicke, Glenn Angrabeit, Madhav Iyengar, Vitalii Protsenko, Illia Shakun, Jovan Cicvaric, Bora Kargi, Haoyu He, Lukas Schuler, Lewin Scholz, Kavayanjali Agnihotri, Yong Cao and Andreas Geiger 307

The Open Argument Mining Framework

Debela Gemechu, Ramon Ruiz-Dolz, Kamila Górski, Somaye Moslemnejad, Eimear Maguire, Dimitra Zografistou, Yohan Jo, John Lawrence and Chris Reed 318

Bel Esprit: Multi-Agent Framework for Building AI Model Pipelines

Yunsu Kim, Ahmedelmogtaba Abdelaziz, Thiago Castro Ferreira, Mohamed Al-Badrashiny and Hassan Sawaf 329

ZeroSumEval: An Extensible Framework For Scaling LLM Evaluation with Inter-Model Competition

Hisham Abdullah Alyahya, Haidar Khan, Yazeed Alnumay, M Saiful Bari and Bulent Yener 340

DECAF: A Dynamically Extensible Corpus Analysis Framework

Max Müller-Eberstein, Rob Van Der Goot and Anna Rogers 351

Dialz: A Python Toolkit for Steering Vectors

Zara Siddique, Liam Turner and Luis Espinosa-Anke 363

FORG3D: Flexible Object Rendering for Generating Vision-Language Spatial Reasoning Data from 3D Scenes

Oscar Pang and Freda Shi 376

RT-VC: Real-Time Zero-Shot Voice Conversion with Speech Articulatory Coding

Yisi Liu, Chenyang Wang, Hanjo Kim, Raniya Khan and Gopala Anumanchipalli 385

Live Football Commentary System Providing Background Information

Yuichiro Mori, Chikara Tanaka, Aru Maekawa, Satoshi Kosugi, Tatsuya Ishigaki, Kotaro Funakoshi, Hiroya Takamura and Manabu Okumura 394

GreaterPrompt: A Unified, Customizable, and High-Performing Open-Source Toolkit for Prompt Optimization

Wenliang Zheng, Sarkar Snigdha Sarathi Das, Yusen Zhang and Rui Zhang 405

iPET: An Interactive Emotional Companion Dialogue System with LLM-Powered Virtual Pet World Simulation

Zheyong Xie, Shaosheng Cao, Zuozhu Liu, Zheyu Ye, Zihan Niu, Chonggang Lu, Tong Xu, Enhong Chen, Zhe Xu, Yao Hu and Wei Lu 416

LiDARR: Linking Document AMRs with Referents Resolvers

Jon Cai, Kristin Wright-Bettner, Zekun Zhao, Shafiuddin Rehan Ahmed, Abijith Trichur Ramachandran, Jeffrey Flanigan, Martha Palmer and James Martin 426

<i>SlimLM: An Efficient Small Language Model for On-Device Document Assistance</i>	
Thang M. Pham, Phat T. Nguyen, Seunghyun Yoon, Viet Dac Lai, Franck Dernoncourt and Trung Bui	436
<i>VeriMinder: Mitigating Analytical Vulnerabilities in NL2SQL</i>	
Shubham Mohole and Sainyam Galhotra	448
<i>DocAgent: A Multi-Agent System for Automated Code Documentation Generation</i>	
Dayu Yang, Antoine Simoulin, Xin Qian, Xiaoyi Liu, Yuwei Cao, Zhaopu Teng and Grey Yang	460
<i>DISPUTool 3.0: Fallacy Detection and Repairing in Argumentative Political Debates</i>	
Pierpaolo Goffredo, Deborah Dore, Elena Cabrio and Serena Villata	472
<i>MedDecXtract: A Clinician-Support System for Extracting, Visualizing, and Annotating Medical Decisions in Clinical Narratives</i>	
Mohamed Elgaar, Hadi Amiri, Mitra Mohtarami and Leo Anthony Celi	481
<i>CiteLab: Developing and Diagnosing LLM Citation Generation Workflows via the Human-LLM Interaction</i>	
Jiajun Shen, Tong Zhou, Yubo Chen, Kang Liu and Jun Zhao	490
<i>CodeArena: A Collective Evaluation Platform for LLM Code Generation</i>	
Mingzhe Du, Anh Tuan Luu, Bin Ji, Xiaobao Wu, Yuhao Qing, Dong Huang, Terry Yue Zhuo, Qian Liu and See-Kiong Ng	502
<i>Ai2 Scholar QA: Organized Literature Synthesis with Attribution</i>	
Amanpreet Singh, Joseph Chee Chang, Dany Haddad, Aakanksha Naik, Jena D. Hwang, Rodney Kinney, Daniel S Weld, Doug Downey and Sergey Feldman	513
<i>gec-metrics: A Unified Library for Grammatical Error Correction Evaluation</i>	
Takumi Goto, Yusuke Sakai and Taro Watanabe	524
<i>AI2Agent: An End-to-End Framework for Deploying AI Projects as Autonomous Agents</i>	
Jiaxiang Chen, Jingwei Shi, Lei Gan, Jiale Zhang, Qingyu Zhang, Dongqian Zhang, Pang Xin, Zhucong Li and Xu Yinghui	535
<i>Efficient Annotator Reliability Assessment with EffiARA</i>	
Owen Cook, Jake A Vasilakes, Ian Roberts and Xingyi Song	542
<i>TRANSLATIONCORRECT: A Unified Framework for Machine Translation Post-Editing with Predictive Error Assistance</i>	
Syed Mekael Wasti, Shou-Yi Hung, Christopher Collins and En-Shiun Annie Lee	551
<i>First-AID: the first Annotation Interface for grounded Dialogues</i>	
Stefano Menini, Daniel Russo, Alessio Palmero Aprosio and Marco Guerini	563
<i>Mergenetic: a Simple Evolutionary Model Merging Library</i>	
Adrian Robert Minut, Tommaso Mencattini, Andrea Santilli, Donato Crisostomi and Emanuele Rodolà	572
<i>FlagEval-Arena: A Side-by-Side Comparative Evaluation Platform for Large Language Models and Text-Driven AIGC</i>	
Jing-Shu Zheng, Richeng Xuan, Bowen Qin, Zheqi He, Tongshuai.ren Tongshuai.ren, Xuejing Li, Jin-Ge Yao and Xi Yang	583

<i>IRIS: Interactive Research Ideation System for Accelerating Scientific Discovery</i>	
Aniketh Garikaparthi, Manasi Patwardhan, Lovekesh Vig and Arman Cohan	592
<i>ROGRAG: A Robustly Optimized GraphRAG Framework</i>	
Zhefan Wang, Huanjun Kong, Jie Ying, Wanli Ouyang and Nanqing Dong	604
<i>LECTURE4ALL: A Lightweight Approach to Precise Timestamp Detection in Online Lecture Videos</i>	
Viktoria Wrobel, Simon Kazemi, Frank Hammerschmidt, Torben Hannemann, Gregor Stange, Seid Muhie Yimam and Robert Geislinger	614
<i>FlexRAG: A Flexible and Comprehensive Framework for Retrieval-Augmented Generation</i>	
Zhang Zhuocheng, Yang Feng and Min Zhang	621
<i>ComfyUI-Copilot: An Intelligent Assistant for Automated Workflow Development</i>	
Zhenran Xu, Yangxue Yangxue, Yiyu Wang, Qingli Hu, Zijiao Wu, Baotian Hu, Longyue Wang, Weihua Luo and Kaifu Zhang	632
<i>PRAISE: Enhancing Product Descriptions with LLM-Driven Structured Insights</i>	
Adnan Qidwai, Srija Mukhopadhyay, Prerana Khatiwada, Dan Roth and Vivek Gupta	644
<i>ATGen: A Framework for Active Text Generation</i>	
Akim Tsvigun, Daniil Vasilev, Ivan Tsvigun, Ivan Lysenko, Talgat Bektlevov, Aleksandr Medvedev, Uliana Vinogradova, Nikita Severin, Mikhail Mozikov, Andrey Savchenko, Ilya Makarov, Grigorev Rostislav, Ramil Kuleev, Fedor Zhdanov and Artem Shelmanov	653
<i>Value Compass Benchmarks: A Comprehensive, Generative and Self-Evolving Platform for LLMs' Value Evaluation</i>	
Jing Yao, Xiaoyuan Yi, Shitong Duan, Jindong Wang, Yuzhuo Bai, Muhua Huang, Yang Ou, Scarlett Li, Peng Zhang, Tun Lu, Zhicheng Dou, Maosong Sun, James Evans and Xing Xie	666