

EMNLP 2026

**The 2026 Conference on Empirical Methods in Natural
Language Processing**

Proceedings of the System Demonstrations

October 24-29, 2026

©2026 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 979-8-89176-461-3

Introduction

Welcome to the proceedings of the System Demonstration Track of the 2026 Conference on Empirical Methods in Natural Language Processing (EMNLP 2026), held from October 24–29, 2026, in Budapest, Hungary.

The EMNLP 2026 System Demonstration Track provides a platform for papers describing system demonstrations, ranging from early-stage prototypes to mature, production-ready systems. We particularly encouraged submissions presenting open-source or open-access systems, as well as industrial systems featuring significant technological innovation. In addition to a paper describing and evaluating the proposed system, submissions were required to include a video and a live demonstration website or an installable package.

For the EMNLP 2026 System Demonstration Track, we received a record 270 submissions. After carefully checking all submitted contributions, 258 submissions were validated and advanced to the reviewing stage. Based on a total of 781 reviews provided by 235 reviewers, together with the recommendations of 43 area chairs, 86 papers were ultimately accepted to the program, resulting in an acceptance rate of 33.3%, in line with previous years. Among the accepted papers, four were shortlisted for the Best System Demonstration Award.

We sincerely thank everyone who contributed to making the EMNLP 2026 System Demonstration Track possible. In particular, we are grateful to the members of the award committee—Zheng Yuan, Meng Jiang, and Mayank Kejriwal—for their invaluable contributions to the selection of the Best System Demonstration, as well as to the area chairs and reviewers for their time, expertise, and dedication.

Sweta Agrawal and Alberto Abad
EMNLP 2026 Demonstration Chairs

Best System Demonstration Award candidates

Simulstream: Open-Source Toolkit for Evaluation and Demonstration of Streaming Speech-to-Text Translation Systems

Marco Gaido, Sara Papi, Mauro Cettolo, Matteo Negri and Luisa Bentivogli 87

AgentDebugX: An Open-Source Toolkit for Failure Observability, Attribution, and Recovery in LLM Agents

Kunlun Zhu, Xuyan Ye, Zhiguang Han, Yuchen Zhao, Bingxuan Li, Weijia Zhang, Muxin Tian, Xiangru Tang, Pan Lu, James Zou, Jiaxuan You and Heng Ji 537

CoDA: A Human-LM Collaborative Codebook Development Assistant for Domain-Specific Text Annotation

Jaspreet Ranjit, Swaminathan Chellappa, Augusto Rivas Costante, Chandan Manjunath, Laura Onasch-Vera, Hailey Winetrobe Nadel, Eric Rice and Swabha Swayamdipta 547

WonderEcon: An Interactive LLM-Agent System for Simulating Economies

Jing Qian, Boyu Liu, Wenyuan Gu, Hang Ruan, Cheng Huo, Junjie Yang, Mingyu Deng, Jiale Han and Benyou Wang 818

Program Committee

Program Chairs

Sweta Agrawal, Google DeepMind
Alberto Abad, INESC-ID - Instituto Superior Técnico

Area Chairs

Abhilasha Sancheti, Amazon
Archiki Prasad, University of North Carolina at Chapel Hill
Ashutosh Modi, IIT Kanpur
Ben Peters, INESC-ID
Chanjun Park, Soongsil University
Chi Han, University of Wisconsin - Madison
Chongsheng Zhang, Ludwig-Maximilians-Universität München and Henan Univeristy
Chung-Chi Chen, NII, Tokyo Institute of Technology
Chunyang Wang, East China Normal University
Eleftheria Briakou, Google
Elizabeth Salesky, Google DeepMind
Emily Allaway, University of Edinburgh
Emma Strubell, Carnegie Mellon University
Esther Ploeger, Utrecht University
Hiba Arnaout, Technische Universität Darmstadt
Hua Wei, Arizona State University and Arizona State University
Hyeju Jang, Indiana University
Jipeng Qiang, Yangzhou University
John Mendonça, Sword Health, Inc
José Pombal, Sword Health, Inc and Instituto Superior Técnico
Jungyeul Park, Korea Advanced Institute of Science & Technology
Long Bai, Institute of Computing Technology, Chinese Academy of Sciences
Marco Gaido, Fondazione Bruno Kessler
Mayank Kejriwal, University of Southern California
Meng Fang, University of Liverpool and Eindhoven University of Technology
Meng Jiang, University of Notre Dame
Nikolay Nikitin, ITMO University
Patrick Fernandes, Carnegie Mellon University - Instituto Superior Técnico
Qingyun Wang, College of William and Mary
Radin Hamidi Rad, University of Toronto
Rui Zhang, Pennsylvania State University
Sachin Kumar, Ohio State University, Columbus
SeongKu Kang, Korea University
Shumin Deng, Zhejiang University
Syrielle Montariol, Université Pierre et Marie Curie - Paris 6, Sorbonne Université
Tirthankar Ghosal, Oak Ridge National Laboratory
Vilém Zouhar, Department of Computer Science, ETHZ - ETH Zurich
Vivek Gupta, Arizona State University
Weiyang Zhang, East China University of Science and Technology
Yaojie Lu, Institute of Software, Chinese Academy of Sciences
Yuntian Deng, University of Waterloo

Zheng Yuan, University of Sheffield
Zhisong Zhang, City University of Hong Kong
Zhiyuan Wen, Hong Kong Polytechnic University
Zixuan Li, Institute of Computing Technology, Chinese Academy of Sciences
Ziyang Luo, Amazon
Zongyang Ma, Xi'an Jiaotong-Liverpool University

Reviewers

Abhijit Chakraborty, Abhishek Kumar, Aditya Sarkar, Alay Dilipbhai Shah, Albert Sawczyn, Al-da Mari, Alina Borisovna Zhidkovskaya, Alireza Bayat Makou, Amit Surana, Anand Kumar Rai, Andreas Säuberli, Andrey Getmanov, Andrey Sakhovskiy, Anjishnu Kumar, Anuj Diwan, Aravind Mandiga, Arda Yüksel, Aria Chen, Aria Nourbakhsh, Aritra Mazumder, Avinash Trivedi

Benjamin Mamut, Bing Yan, Bo Chen

Chadi Helwe, Chang Ma, Chengfeng Zhao, Chengzu Li, Chenyang Zhao, Chunhua Hu, Cuixi Li

Daniil Orel, Dawei Li, Di Lu, Diego Zafferani, Dingcheng Huang, Dingjie Song, Dongchan Shin, Dongsheng Zhu

Elena Tutubalina, Elisabetta Rocchetti

Fabian Lukassen, Firoj Alam, Frolov Anatolii

Gaetano Cimino, Georgeta Bordea, Giacomo De Luca, Gosuddin Kamaruddin Siddiqi, Goun Pyeon, Guanghui Wang, Guanhua Chen, Guochen Zhou, Guozhao Mo

Han Meng, Hannah Kim, Hao He, Hao Yu, Haolei Xu, Harshit Taneja, Hirokazu Kiyomaru, Hongdong Lee, Hongchao Du, Hongru Cai, Huy V. Nguyen

Ilia Karpov, Innokentiy S. Humonen

Ja Young Lee, Jamshid Mozafari, Jan Nietzard, Jaspreet Ranjit, Jerzy Kamiński, Jiacheng Ye, Jialun Zhong, Jianchang Su, Jianda Chen, Jiangming Liu, Jiannan Mao, Jiaxuan Luo, Jie Sun, Jinghan Zhang, Jinyu Ye, Jize Jiang, John Philip McCrae, Jonas Wallat, Jonathan Drechsel, Josh Zhang, José Sacadura Menezes, Jun Zhang, Jun-En Ding, Junhao Ruan, Jyotika Singh

KaShun Shum, Ke Xu, Keerthana Murugaraj, Kenan Alkiek, Kim Yoonseo, Komang Rinatha, Kshitij P Fadnis, Kunlun Zhu

Lars Bungum, Laura De Santis, Lecheng Yan, Linfan Zhang, Linh The Nguyen, Liting Zhang, Long Bai, Longchao Da, Lucas Moeller, Lucia Quirke, Luigi Di Caro, Luigi Petrucco

Maksim Savkin, Marco Simoni, Marek Šuppa, Maria Mahbub, Maria Molchanova, Marina Polukoshko, Mario Ezra Aragon, Matthias Aßenmacher, Matīss Rikters, Maya K. Nachesa, Md Mosaddek Khan, Menan Velayuthan, Meng Fang, Mengliang Zhang, Mengyang Qiu, Mengze Hong, Ming Zhang, Mirac Suzgun, Muge Zhang

Na Zhang, Nabil Ibtehaz, Navapat Nananukul, Neemias Buceli da Silva, Nicholas Deas, Nicolò

Penzo, Nikolay Nikitin, Nils Feldhus, Nishant Balepur

Omer Ehrlich

Patara Trirat, Pegah Ahadian, Peiqiong Chen, Peixian Ma, Peiyu Liu, Pin Ni, Pratinav Seth

Qiao Wang, Qin Liu, Qingkai Zeng, Qiyuan Peng, Quang Minh Nguyen

Ramya Keerthy Thatikonda, Ran Elgedawy, Ray Chen, Raymond Alavo, Renxi Wang, Renxiang Guan, Rik Koncel-Kedziorski, Roy Ka-Wei Lee, Ruoci Song, Ryan Li, Ryuji Hashimoto

Sachin Thukral, Sahand Sabour, Sai Tulasi Kolapudi, Sai Karthik Kosuri, Sai Rama Akash Varma Sagi, Sajjad Haider, Santhiya Rajan, Sara Rosenthal, Saurabh Yergattikar, Seid Muhie Yimam, Seonghyeon Cho, Shantanu Agarwal, Shenzhe Zhu, Shreya Ghosh, Shunfeng Zheng, Siqi Shen, Song-Ze Yu, Songming Zhang, Sourabrata Mukherjee, Stanislav Chumakov, Stephen Meisenbacher, Sunan Qian

Taiga Ikeda, Tampu Ravi Kumar, Theron S. Wang, Thomas Rolland, Tianlin Zhang, Tianrui Liu, Tiantian Zhu, Tiejin Chen, Tomoki Doi

Vijayasri Iyer, Voula Giouli, Vsevolod Karimov

Wanting Ning, Wei Cai, Wei Liu, Wei Wang, Wei Xu, Weijiang Li, Wenzhi Wang, Wessel Poelman, Wolfgang Otto

Xiaochang Li, Xiaoxia Cheng, Xiaoyi Gu, Xiaozhi Wang, Xinchu Chen, Xingbo Wang

Yalun Wu, Yanyi Luo, Yao Lu, Yao Yao, Ye He, Ye-eun Cho, Yegor Anichkov, Yen YU Shan, Yidan Sun, Yide Ran, Yige Luo, Yining Lu, Yixiong Xiao, Yiyang Du, Yongbin Choi, Yoshihide Kato, Youyang Ng, Yu Lei, Yu-Chien Tang, Yu-Shiang Huang, Yudong Tao, Yueqian Lin, Yueru Yan, Yufan Dang, Yufeng Han, Yuhang Lai, Yujia Hu, Yun Luo, Yuxuan Huang, Yuxuan Lai, Yuxuan Wan, Yuze Gao, Yuzhen Huang, Yuzhou Wang

Zaid Sheikh, Zeyu Leo Liu, Zhengzhao Ma, Zhennan Shen, Zhenyu Wu, Zheyuan Zhang, Zhi Qu, Zhongxi Wang, Zixuan Li, Zongyu Wu

Table of Contents

<i>ChatDev 2.0: A No-Code Multi-Agent Platform for Developing Everything</i> Yufan Dang, Shu Yao, Bowen Lai, Chenting Xu, Ruijie Shi, Wai-Shing Leung, Huatao Li, Chen Qian and Zhiyuan Liu	1
<i>HintEval: An Open-Source Python Toolkit for Hint Generation and Hint Evaluation</i> Jamshid Mozafari, Bhawna Piryani, Abdelrahman Abdallah and Adam Jatowt	13
<i>EasySteer: A Unified Framework for High-Performance and Extensible LLM Steering</i> Haolei Xu, Xinyu Mei, Yuchen Yan, Rui Zhou, Wenqi Zhang, Weiming Lu, Yueting Zhuang and Yongliang Shen	24
<i>ReasoningLens: Hierarchical Visualization and Diagnostic Auditing for Large Reasoning Models</i> Jun Zhang, Jiasheng Zheng, Boxi Cao, Yaojie Lu, Hongyu Lin, Jia Zheng, Xianpei Han and Le Sun	36
<i>Jiuge-Tuiqiao: An Interpretable Human-AI System for Classical Chinese Poetry Refinement</i> Yufeng Han, Lifan Deng, Cunliang Kong, Wenhao Li, Xin Cong, Yuzhuo Bai, Kangyang Luo and Maosong Sun	46
<i>CoLearn: An Agentic Tutor that Learns its Learner in a Human-AI Co-Learning Loop</i> Kailai He, Zhihao Wu, Linhai Zhang, Runcong Zhao, Yulan He and Jiazheng Li	56
<i>Pearmut: Human Evaluation of Translation Made Trivial</i> Vilém Zouhar and Tom Kocmi	66
<i>Simulstream: Open-Source Toolkit for Evaluation and Demonstration of Streaming Speech-to-Text Translation Systems</i> Marco Gaido, Sara Papi, Mauro Cettolo, Matteo Negri and Luisa Bentivogli	87
<i>HiMe: Real-Time Self-Hosted Personal Agent Platform for Health Insights with Wearable Devices</i> Wei Liu, Siya Qi, Linhai Zhang, Lorainne Tudor Car and Yulan He	98
<i>ARCCS: An Automated Regulatory Compliance Checking System</i> Giorgos Filandrianos, José Sacadura Menezes, Chrysoula Zerva and Alessandro Gianola	109
<i>RISE-NLP: Interactive Residual Distribution Diagnostics for Fairness Auditing in Text Classification</i> Ray Chen, Kevin Kurian and Christan Grant	120
<i>MangaLens: A Multi-Model Tool for Structured Manga Parsing</i> Jiannan Mao, Jin Tei, Zhi Qu, Hideki Tanaka and Masao Utiyama	128
<i>Drift Inspector: Exploring and Measuring Scientific Drift with Atomic Contribution Claims</i> Vsevolod Karimov, Stepan Ostarkov, Anastasia Poroshina, Frolov Anatolii and Alexander Panchenko	138
<i>LLMberjack: Guided Trimming of Debate Trees for Multi-Party Conversation Creation</i> Leonardo Bottona, Nicolò Penzo, Bruno Lepri, Marco Guerini and Sara Tonelli	149
<i>TraceFork: A System for Inspecting GUI-Agent Trajectory Forks</i> Yuzhang Zhong and Stephen Xia	160
<i>Who Is Asking? Personalized and Tiered Memory for Long-Lived AI Assistants</i> Yueqian Lin, Sathiya Kumaran Mani, Yiran Chen, Sujata Banerjee and Vaishnavi Ranganathan	167

<i>MUSE: A Run-Centric Platform for Multimodal Unified Safety Evaluation of Large Language Models</i> Zhongxi Wang, Yueqian Lin, Jingyang Zhang, Zinuo Cheng, Hai Helen Li and Yiran Chen ..	178
<i>The GRADIEND Python Package: An End-to-End System for Gradient-Based Feature Learning</i> Jonathan Drechsel and Steffen Herbold	189
<i>e2a: A Secure Agent-Native Email Gateway with Phishing and Prompt-Injection Defense</i> Josh Zhang, Hongyuan Jia, Xueyan Zhang, Hanzhi Tan and Jingzhuo Hu	199
<i>ShieldMCP: An Interactive System for Runtime Security of Model Context Protocol Tool Calls</i> Saurabh Yergattikar	208
<i>OmegaUse-SOP: SOP Engineering for Professional Computer Use from Human Demonstrations</i> Yixiong Xiao, Lang An, Hucheng Yang, Pinxue Ma, Yongquan Chen, Jingjia Cao, Yusai Zhao, Ting Wang, Ting Liu, Siqi Bao, Jingbo Zhou and Hua Wu	215
<i>ToFu: A White-Box, Token-Efficient Agent Harness for Researchers</i> Junhao Ruan, Yuan Ge, Bei Li, Yongjing Yin, Yuchun Fan, Xin Chen, Jingang Wang, Chenglong Wang, JingBo Zhu and Tong Xiao	225
<i>Explainable Coarse-to-Fine Ancient Manuscript Duplicates Discovery</i> Chongsheng Zhang, Shuwen Wu, Yingqi Chen, Yi Men, Gaojuan Fan, Yubin Jiang, Matthias Aßenmacher and Christian Heumann	235
<i>CoEvoLoop: Closed-Loop RL Training of Customer-Service Agents by Co-Evolving Environment, Reward, and Policy</i> Hui Cai, Jianda Chen, Yuhong Zhou, Wei Ren, Choy Zhen Wen Marcus, Tan Li Hui Faith, Lyuxin Xue, Hao Chen, Jiawei Luo, Xuanwei Hu and Xin Zhang	242
<i>AgentCompass: A Unified Evaluation Infrastructure for Agent Capabilities</i> Dongsheng Zhu, Shufan Jiang, Wenhui Tian, Zixin Shang, Peiheng Zhou, Tianhao Liang, Mo Li, Qingqiu Li, Shudong Liu, Zonglin Li, Liwei Wu, Zichen Ding, Bowen Yang, Zehao Li, Zhenyu Wu, Jun XU, Zun Wang, Songyang Zhang, Dingbo Yuan, Ma Zerun and Kai Chen	253
<i>BNB Agent Studio: An Open Toolchain for Provisioning, Deploying, and Operating Economically Au- tonomous LLM Agents</i> H Zhao, Guanghui Wang, Xing Zhang, Yanwei Cui, Leroy Lim Hong Quan, NingNing Li, Keyu Deng, Gu Yang, Bowen Ai, Peiyang He, ChengHung Wu, Evan Chen, Jinsong Zhu, Chaoqun Sun, Zhao Beibei, Di Yang, Hongyue Wang, Yuxiang Zhai, Shuo Yang, Yolin and David. Z	264
<i>QQ: A Language Metadata Toolkit for Multilingual NLP</i> Wessel Poelman, Yiyi Chen and Miryam de Lhoneux	275
<i>GPTKB 2.0: Browsing, Querying, and Auditing a Disambiguated LLM-Derived Knowledge Base</i> Yujia Hu, Tuan-Phong Nguyen and Simon Razniewski	286
<i>RefactorPlatform: An Open-Source Harness for Controlled Evaluation of Repository-Scale Refactoring Agents</i> Aziz Ben Amor, Drish Mali, Mann Acharya, Vijayasri Iyer and Sébastien Bratières	293
<i>Alyas: Evidence-Grounded Synthetic Populations for Simulated Social and Market Research</i> Luigi Petrucco, Samuel Bortolin, Elena Leonelli, Eric Suardi, Elena Zorzi, Daniele Miorandi and Davide Piazza	302
<i>KDFlow: A User-Friendly and Efficient Knowledge Distillation Framework for Large Language Models</i> Songming Zhang, Xue Zhang, Tong Zhang, Bojie Hu, Yufeng Chen and Jinan Xu	310

<i>DIASENTINEL: An Auditable Multi-Agent System for Guideline-Grounded Diabetes Risk Screening</i>	
Yung Wei Shueh, Zhi-Jie Chen, Chia-Hsuan Hsu, Hsin-Ling Hsu, Donghua Zhang, Chenwei Wu, Jun-En Ding, Tongze Zhang, Shihao Yang, Pengfei Hu, Fang Ming Hung and Feng Liu	320
<i>ZSERVE: An Adaptive Framework for Optimizing Large Language Models Serving in Heterogeneous Infrastructures</i>	
Taosong Fang, Zhengzhao Ma, Jiawei Chen, Yushen Yang, Kaijie Zhou, Jia Zheng, Xianpei Han, Le Sun, Hongyu Lin and Yaojie Lu	330
<i>RAGAPHENE A RAG Annotation Platform with Human ENhancements and Edits</i>	
Kshitij P Fadnis, Sara Rosenthal, Maeda Hanafi, Yannis Katsis and Marina Danilevsky	338
<i>Predicting Collective Behavior with Large Language Models: The EBI Framework for Response-Side Inference</i>	
Yen YU Shan, Li-Wun Chang and Lun-Wei Ku	349
<i>SCOPE: Source Credibility and Outlet Profiling Engine</i>	
Ilyas Nagiyev, Miras Baisbay, Daniil Orel and Preslav Nakov	359
<i>TraceForge: A Human-Gated System for Auto-Hardening Tool-Using LLM Agents into Immutable Workflow DAGs</i>	
Santhiya Rajan	369
<i>Open, Reliable, and Collective: A Community-Driven Framework for Tool-Using AI Agents</i>	
Hy Dang, Quang Dao and Meng Jiang	380
<i>PrivBench: A Holistic and Modular Benchmarking Platform for Evaluating Text-to-Text Privatization</i>	
Stephen Meisenbacher, Andreea-Elena Bodea, Ahmet Bilal Akin, Alexandra Klymenko, Jana Diesner and Florian Matthes	391
<i>WikiSTAR: A System for Shedding Light on the Hidden History of Scientific Wikipedia Articles</i>	
Omer Ehrlich, Nitzan Barzilay, Rona Aviram and Tom Hope	414
<i>MemTools: A Unified Research Framework for Interoperable Agent Memory</i>	
Chengfeng Zhao, Jinhui Chen, Sirui Liang, Shizhu He, Yequan Wang, Jun Zhao and Kang Liu	424
<i>string2string Studio: An Interactive, In-Browser Platform for String-to-String Algorithms</i>	
Mirac Suzgun, James Zou, Stuart Shieber and Dan Jurafsky	435
<i>Murano: Design, Run, and Reproduce Mechanistic Interpretability Experiments as Composable Pipelines</i>	
Alireza Bayat Makou, Emirhan Böge, Phu Gia Hoang, Federico Tiblias, Jingcheng Niu, Subhabrata Dutta, Richard Eckart de Castilho and Iryna Gurevych	445
<i>Anamnesis: An Open-Source Platform for Large-Scale Backstory-Conditioned Survey Simulation</i>	
Song-Ze Yu, Joseph Suh, Serina Chang and David M. Chan	456
<i>CALIBER: Calibrated Assessment via LLM for Interview Behavioral Evaluation and Rating</i>	
Katherine Kappaz, Raymond Alavo, Sofia Calderon and Ahmed Abbasi	466
<i>CAT: A CounterArgument Tutoring System based on Logical Structure Analysis</i>	
Wenzhi Wang, Thiemo Wambsganss, Chenyang Wang, Paul Reisert, Léane Wettstein, Katja Pott, Kenshi Yamaguchi, Machi Shimmei, Roman Rietsche, Surawat Pothong, Naoya Inoue and Kentaro Inui	476

<i>Trust the Context, Not the Weights: Faithful Multi-Hop QA via Synthetic Mid-Training</i>	
Maksim Savkin, Mikhail Goncharov, Alexander Gambashidze, Alla Chepurova, Dmitrii Tarasov, Timur Ionov, Nikita Andriianov, Daria Pugacheva, Ivan Oseledets, Vasily Konovalov and Andrey V. Galichin	488
<i>SAFER: Speech-First Auditable Follow-Up with Evidence-Gated Routing</i>	
Boyu Zhang, Juntao Su, Zihua Wang and Chunhua Hu	498
<i>EviDAG: Auditable Causal DAG Authoring with Biomedical Literature</i>	
Yi-han Sheu, Michael R. Steigman, Yu Zhou, Bo Wang, Fan-Yu Yen and Jordan W. Smoller .	506
<i>CAFE: A Compound-AI Factorial Evaluation Framework</i>	
Fabian Lukassen, Christoph Weisser, Thomas Kneib and Alexander Silbersdorff	516
<i>RailVaani: An Optimized Bilingual ASR Pipeline for Entity-Rich Railway Announcements</i>	
Anand Kumar Rai, Swar Desai, Shshank Mittal and Animesh Mukherjee	528
<i>AgentDebugX: An Open-Source Toolkit for Failure Observability, Attribution, and Recovery in LLM Agents</i>	
Kunlun Zhu, Xuyan Ye, Zhiguang Han, Yuchen Zhao, Bingxuan Li, Weijia Zhang, Muxin Tian, Xiangru Tang, Pan Lu, James Zou, Jiaxuan You and Heng Ji	537
<i>CoDA: A Human-LM Collaborative Codebook Development Assistant for Domain-Specific Text Annotation</i>	
Jaspreet Ranjit, Swaminathan Chellappa, Augusto Rivas Costante, Chandan Manjunath, Laura Onasch-Vera, Hailey Winetrobe Nadel, Eric Rice and Swabha Swayamdipta	547
<i>ShareChatInsights: An Automated Pipeline for Analyzing Large-Scale Human–LLM Conversations</i>	
Yueru Yan, Yilin Gong, Siqi Wu and Thai Le	558
<i>Karma: A Composable Framework for Evaluating LLM Agents on Stateful Kubernetes Operations</i>	
Junhan Ouyang, Shaw Liu, Jinlang Wang and Jiawei Zhou	570
<i>Bergson: An Open Source Library for Data Attribution</i>	
Lucia Quirke, Louis Jaburi, David O. Johnston, William Z. Li, Gonalo Paulo, Guillaume Martres, Girish Gupta, Nora Belrose and Stella Biderman	582
<i>TreeThink: A Modular Tree Search Library for Mathematical Reasoning with LLMs</i>	
Burak Sina Akbudak, Zeynel Abidin Uluan, Salih Can Erer and Gzde Gl ahin	593
<i>hoBIT: A Profile-Aware Retrieval-Augmented Chatbot for University Academic Advising</i>	
Kim Yoonseo, Seongmin Lee, Joongheon Kim and SeongKu Kang	604
<i>TechMatch: Matching Enterprise Technology Demands to Academic Resources with Graph-Augmented Retrieval</i>	
Rui Chai, Shiyun Huang, Yubing Su, Yuhan Zhen, Yi Sun, Maozhe Zhao, Shengzhong Liu, Rui Shen, Ruijie Wang and Jianxin Li	614
<i>SentinelEdge: A Lightweight, Streaming, Auditable Scam-Call Detection System for the Edge</i>	
Mayank Vyas, Shiven Agarwal, Prisha Nag, Ishita Rajesh Sharma, Abhijit Chakraborty and Vivek Gupta	625
<i>Neurobiber: Fast and Interpretable Linguistic Feature Extraction</i>	
Kenan Alkiek, Anna Wegmann, Jian Zhu and David Jurgens	639
<i>CorpusLens: LLMs as Data Analysis Facilitators in the Social, Humanistic, and Language Sciences</i>	
Blake Vente, Nicholas Deas, Desmond U. Patton, Shana Kleiner, Jessica A Grieser and Kathleen McKeown	651

<i>MUDIPA: A Human-in-the-Loop AI-Assisted Platform for Discourse Graph Annotation in Multi-party Dialogues</i>	
Laura De Santis, Gaetano Cimino, Nils Feldhus, Giuseppe Carenini and Vincenzo Deufemia	664
<i>AgentCheck: A Reproduce–Intervene–Mitigate Workbench for LLM Agents over MCP</i>	
Aritra Mazumder and Nusrat Jahan Lia	674
<i>GIMFlow: An LLM-Guided Dialogue-to-Music System for GIM-Inspired Well-Being Experiences</i>	
Ruijia Huang, Yao Yao and Hai Zhao	685
<i>Schematize: An Agentic System for Generating and Refining Information-Extraction Schemas for Legal Research</i>	
Albert Sawczyn, Jakub Binkowski, Kamil Tagowski, Lukasz Augustyniak, Berenika Kaczmarek-Templin and Tomasz Jan Kajdanowicz	696
<i>NotAI.AI: An Interactive Explainable System for AI-Generated Text Detection</i>	
Oleksandr Marchenko Breneur, Adelaide Danilov, Aria Nourbakhsh and Salima Lamsiyah	706
<i>Embedding Orrery: Visual Exploration of Embedding Spaces and Sparse Autoencoder Features</i>	
Giacomo De Luca	717
<i>AskChem: Claim-Centered Infrastructure for Chemistry Literature Synthesis</i>	
Bing Yan, Gregory Wolfe, Stefano Martiniani and Kyunghyun Cho	728
<i>A Composable Evaluation System for Reproducible Omni-Modal Foundation Model Evaluation</i>	
Hodong Lee, Sanghee Park, Dohoon Ryu, Jungwhan Kim, Junyeob Kim, Soyeon Kim and Geewook Kim	738
<i>GOD: Govern, Observe, and Direct — A Real-Time Control Room for Agent Societies</i>	
Yige Luo and Ran Guan	750
<i>EviStreams: Human-in-the-Loop AI Data Extraction for Systematic Reviews in Medicine</i>	
Sai Karthik Kosuri, Ankita Shashikant Bhosale, Michael Glick, Alonso Carrasco-Labra and Chris Callison-Burch	759
<i>SAFETUNE : A Unified, Faithful Library for Auditing and Repairing Safety Drift in Fine-Tuned LLMs</i>	
Pratinav Seth, Saisab Sadhu, Anshul Kaushal and Vinay Kumar Sankarapu	771
<i>ABRA: Agent-Bound UI for Human-Agent Communication beyond Chat with Reusable Views</i>	
Seon Gyeom Kim and Hannah Kim	786
<i>ImmersiveCFL: Multi-Agent Learning Path Generation for Scenario-Oriented Chinese Language Learning</i>	
Shanghai Yang, Zhixuan Wu, Yanan Zeng, Feixiang He, Chunyang Wang and Weining Qian	796
<i>PatientHub: A Unified Framework for Patient Simulation</i>	
Sahand Sabour, NG Tsz Yam and Minlie Huang	806
<i>WonderEcon: An Interactive LLM-Agent System for Simulating Economies</i>	
Jing Qian, Boyu Liu, Wenyan Gu, Hang Ruan, Cheng Huo, Junjie Yang, Mingyu Deng, Jiale Han and Benyou Wang	818
<i>Compile by Training: Turning Natural-Language Specifications into Local Neural Functions</i>	
Yuntian Deng, Pengyu Nie and Stuart Shieber	830
<i>PalmClaw: A Native On-Device Agent Framework for Mobile Phones</i>	
Hongru Cai, Yongqi Li, Ran Wei and Wenjie Li	840

<i>CLT-Forge: A Scalable Library for Cross-Layer Transcoders and Attribution Graphs</i>	
Florent Draye, Vedant Palit, Abir Harrasse, Tung-Yu Wu, Jiarui Liu, Punya Syon Pandey, Roderick Wu, Chih-Hao Hsu, Terry Jingchen Zhang, Zhijing Jin and Bernhard Schölkopf	850
<i>VISTA: A Controllable Platform for Generating and Auditing Egocentric Assistance Scenarios</i>	
Yu-Hsiang Liu, Yu-Chien Tang and An-Zi Yen	859
<i>Dr. Claw: An AI Scientist Workspace for Vibe Research</i>	
Dingjie Song, Hanrong Zhang, Dawei Liu, Yixin Liu, Zongxia Li, Zhengqing Yuan, Siqi Zhang, Henry Peng Zou, Zhiling Yan, Yuxuan Zhang, Yanfang Ye, Philip S. Yu and Lichao Sun	868
<i>RAGU: A Multi-Step GraphRAG Engine with a Compact Domain-Adapted LLM</i>	
Mikhail Komarov, Ivan Bondarenko, Stanislav Shtuka, Oleg Sedukhin, Roman Shuvalov, Yana Dementyeva, Matvey Solovyov and Nikolay Nikitin	880
<i>AgentGUI: An Interface for Observing and Steering Long-Running AI Agents</i>	
Xuan Zhao, Jiwoong Sohn, Qinyue Zheng and Michael Moor	892
<i>SentryLine: Evidence-Grounded Question Answering over Evolving Documents in Oncology Care</i>	
Tampu Ravi Kumar, Gaurav Najpande, Muhammad Ali Khan, Kaneez Zahra Rubab Khakwani, Karan Kathuria, Shorya Azriel Moses, Yuvraj Kalia, M Bassam Sonbol, Irbaz Bin Riaz and Vivek Gupta	901
<i>MineTRACE: An Evidence-Grounded Interactive Reasoning System for Mineral Prospectivity</i>	
Yiran Zhang, Jinwen Liu, Daniel Yitian Su, Yisu Chen, Qiang Sun, Christopher M. Gonzalez, Eun-Jung Holden, Marco L. Fiorentini, Wei Liu and Yihao Ding	912