

REALM 2026

The 2nd Workshop for Research on Agent Language Models

Proceedings of the Workshop

October 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 979-8-89176-485-9

Introduction

We are excited to welcome you to REALM 2026, the Second Workshop on Research on Agent Language Models, held in conjunction with EMNLP 2026 on October 29, 2026, in Budapest, Hungary.

Recent developments and incidents involving frontier AI systems have highlighted both the rapid progress of AI agents and the growing challenges surrounding their reliability, safety, and real-world deployment. Given the increasing importance and complexity of agentic systems, the REALM Workshop aims to bring together researchers, practitioners, and thought leaders from the ACL community to discuss the current landscape, key challenges, and future directions of AI agents.

This year’s program includes 7 keynote talks, 1 panel discussion, 6 oral presentations, and 2 poster sessions. The workshop features a mix of remote and live presentations and discussions, allowing for a diverse and inclusive exchange of ideas.

This year, we received 264 submissions. Every submission was reviewed by at least three reviewers. In selecting papers for the program, we carefully considered the reviews and comments provided by the reviewers. We are grateful to the members of our Program Committee for their thoughtful and thorough reviews, which played an essential role in shaping a strong and balanced technical program.

We accepted 164 papers (acceptance rate=62%): 6 oral (top 2%), 9 spotlight (top 6%), and 149 poster presentations.

A successful workshop requires advice, support, and enthusiastic participation of many parties, and we have a big ‘thank you’ to say to all of them. Regarding the program, we thank our seven invited speakers, Maria Brbić (EPFL), Vishrav Chaudhary (Meta Superintelligence Labs), Marc-Alexandre Côté (Microsoft Research), Graham Neubig (OpenHands / Carnegie Mellon University), Manling Li (Northwestern University), Baptiste Rozière (Mistral AI), and Perouz Taslakian (McGill University / Mila) for their inspiring talks.

We extend special thanks to EMNLP Workshop chairs, ACL staff, and volunteers for their work in coordinating and supporting the many workshops held alongside the main conference. We also thank EMNLP for hosting our workshop.

We once again thank our program committee members for generously contributing their time and expertise to the review process and helping us select an excellent technical program.

Finally, we thank all the authors who submitted their work to REALM 2026 and all the participants joining us in Budapest and remotely. Your contributions and participation make the workshop a success and continue to advance the growing research areas of LLM Agents.

We hope you enjoy REALM 2026!

Ehsan Kamalloo, Nicolas Gontier, Nikolai Rozanov, Maryam Hashemzadeh, and Akhil Arora
— *Program Co-Chairs*

Organizing Committee

Program Chair

Ehsan Kamaloo, ServiceNow AI Research
Nicolas Gontier, Cohere
Nikolai Rozanov, Imperial College London
Maryam Hashemzadeh, Université de Montréal / Mila
Akhil Arora, Aarhus University

Advisory Board

Siva Reddy, McGill University / Mila
Yu Su, NeoCognition / Ohio State University

Program Committee

Program Chairs

Ehsan Kamalloo, ServiceNow AI Research
Nicolas Gontier, Cohere
Nikolai Rozanov, Imperial College London
Maryam Hashemzadeh, Université de Montréal / Mila
Akhil Arora, Aarhus University

Reviewers

Amirhossein Abaskohi, Ankush Agarwal, Parth Agarwal, Kushagra Agrawal, Shiva Ahir, Istiak Ahmed, Sainadh Ainala, Marjan Alirezaie, Avinash Goutham Aluguvelly, Riasad Alvi, Michael Andreev, Rajeel Ahmad Ansari, Rucha Apte, Pulkit Arya, Gagan Aryan

Hankyul Baek, Mohammad Taha Bahadori, Neeraj Baji, Spandana Balumuri, Krishna Chaitanya Balusu, Rista Baral, Ziyad Benomar, Sharat Bhat, Deniz Bilgin, Laurent Bindschaedler, Umesh Bodhwani, Angie Boggust, Mikolaj Boronski

Mingsheng Cai, Siddharth Chauhan, Tisha Chawla, Yihang Chen, Gaurav Chodwadia, Peter Choi, Dominic Colligan, Arnau Rodon Comas

Parag Pravin Dakle, Niall Dalton, Quang Dao, Bidit Das, Shreeya Dhakal, Amit Dhanda, Harsh Rao Dhanyamraju, Aditya Dhavala, Heejin Do, Ashish Dsa, Boxin Du, Zenghao Duan, Arka Dutta

Kenneth Eaton, Behnam Ensan

Younma Farag, Kriti Faujdar, Shaoting Feng

Jian Gao, Mitul Garg, Advitya Gemawat, Melat Ghebreselassie, Aniruddha Ghosal, Yusi Gong, Amitha Gopidi, Prashant Govindarajan, Matej Gradoš, Praneeth Gubbala, Itai Guez, Penghong Guo, Yongjian Guo, Honey Gupta, Kushagra Gupta, Rahul Gupta, Sachin Gupta

Syed Ali Haider, Sina Hajimiri, Krishna Harish, Srinivas Harish, Bella He, Jianfeng He, Yifan He, Yifeng He, Omid Reza Heidari, Tung Thanh Hoang, Anton Ivanov Hristov, Jingzhuo Hu, Tianyi Hu, Jerry Huang, Menghao Huo, Juwan Hwang

Shin-nosuke Ishikawa

Siddharth Jain, Haris Jalal, Minjun Jee, Daeun Jeong, Yulong Ji, Leqi Jia, Xiaochong Jiang, Ishan Jindal, Dahuin Jung

Ratna Kandala, Saksham Kapoor, Vijay Ananth Karunanithi, Purvi Kathalkar, Imene Kerboua, Matin Khajavi, Harshita Khandelwal, Prakhar Khatri, Hyunseo Kim, Jungwhan Kim, Lars Henning Klein, Lakshman Kolasani, Praveenkumar Kondikoppa, Tamara Kostova, Susheem Koul, Markus Kreft, Dhairya Kuchhal, Ajay Anil Kumar, Akshay Kumar, Anoop Kumar, Dhruv Kumar, Tejaswini Kumar, Vishwanath Kumaraswamy, Gaurav kukreja

Juntae Lee, Woongkyu Lee, Man-Fai Leung, Qiang Li, Shangao Li, Xiaoxuan Li, Yibo Li, Yixuan Li, Yueshen Li, Yuxuan Li, Zhenhao Li, Zhuofeng Li, Baohao Liao, Duoduo Liao, Peng Lin, Chang Liu, Enze Liu, Jiacheng Liu, Yunan Lu, Yutian Lu, deli liu

Yu-Seung Ma, Suneet Malhotra, Amit Maraj, Sara Vera Marjanovic, Florian Matthes, Parsa Mazaheri, Aditya Mazumdar, Rishabh Mehan, Md Muntaqim Meherab, Manan Mehta, Ruben Cuevas Menendez, Wissam Merhi, Safa Messaoud, Rohan Mishra, Suyash Mishra, Shilpa Mitra, Bardia Mohammadi, Swastik Mohapatra, Sromana Mukhopadhyay, Kunal Mulwani, Rushab Munot

RUTHUPARNA NAIKAR, Mohamed Nabeel, Sai Vinod Nagabhairava, Shashi Bhushan Tyamagondlu Nagabhushan, Lakshmi Nair, Christina Nasika, Poli Nemkova, Loc Nguyen, Liang Nie, NohMyongSung NohMyongSung, Aarati Andrea Noronha

Kevin O’Sullivan

Jacqueline J. Pang, Seungbeom Park, Santosh V. Patapati, Neelraj J. Patil, Srikanth Patil, Sanjana Pedada, Punitha Ponnuraj, Lauren Pothuru, Akshara Prabhakar, Ioannis Prokopiou, Usukhbayar Purevdorj, Partha Sarathi Purkayastha

SHAOBO QIAO, Pin Qian, Junhao Qiu, Xinyu Qiu, Xingyu Qu

Hossein A. Rahmani, Ishan Rai, Sathyamoorthy Rajendran, Sai Krishna Rallabandi, Dhanesh Ramachandram, Vallabhan Kudlu Ramakrishnan, Ninaad Rao, Selvan Sunitha Ravi, Soham Ray, Shaina Raza, Vikas Reddy, Nazia Riasat, Carson Conception Rodrigues

Harshdip Saha, Adib Sakhawat, Kacper Saks, Alessio Saladino, Nevasini Sasikumar, Yash Ganpat Sawant, Krishna Sayana, Md. Faiyaz Abdullah Sayeedi, Jonas Schewior, Sriram Selvam, Kannav Sethi, Muthaheera Yasmeena Belgur Shamiullah, Dingyan Shang, Matthew Shardlow, Jai Kumar Sharma, Yining She, Bonan Shen, Artem Shitov, Bronislav Sidik, Rishi N. Simhadri, Atin Kumar Singh, Prasannjeet Singh, Satveer Singh, Vedant Singh, Abhishek Singhanian, Prateek Singhi, Soumen Sinha, Yash Sinha, Himaneesh Sompalle, Dinghong Song, Junhao Song, Zirui Song, Alessandro Sosso, Malaiappan Srinivasan, Arunabh Srivastava, Mikayel Sukiasyan, Jeewoo Sul, Deshan Sumanathilaka, Haoran Sun, Lifan Sun, Yan Sun, Yiqi Sun, Yuqi Sun, Jinli Suo, Aakash Suresh, Naveen Suresh, Parth Suresh

Minh Ngoc Ta, Fatemeh Taheri Dezaki, Hovhannes Tamoyan, Taleef Tamsal, Lijuan Tang, Noam Tarshish, Shivnath Tathe, Nishal Thomas, Ketan Kumar Todi, Ankur Tomar, Tommy Tran

Rana Muhammad Usman

BADRI PRASAD V R, Kinjal Abhishek Vaishnav, Dipam Vasani, Bala Krishna S Vegesna, Filippos Ventirozos, Robert Vetter, Samarth Vinayaka, Amogh Vinaykumar, Ishitwa Viranchi, Khang Nhat Hoang Vo, Siddharth Vohra

Jicheng Wang, Jinlin Wang, Ruyi Wang, Su Wang, Ye Wang, Yifan Wang, Yiman Wang, Yinzhi Wang, Youting Wang, Yufeng Wang, Yujie Wang, Jiaming Wei, Jingxi Wei, Zhang Weizhou, Isadora White, Eugene Wu, Hang Wu

Huan Xu, Zixiang Xu

Rakshath Raj Y, Jiangnan YU, Abel Yagubyan, Fan Yang, Tong Yang, Xi Yang, Xinyi Yang,

Zhongyu Yang, Junyi Yao, Anisha Ramakrishna Yarlapati, Saurabh Yergattikar, Joeun Yook, Hao-ran Yu, Yue Yu, Yihao yang

Sicheng Zeng, Erdni Zergetaev, Andrew Zhang, Enze Zhang, Jiameng Zhang, Jiaquan Zhang, Luyang Zhang, Ruiyang Zhang, Yao Zhang, Yifan Zhang, Yuxuan Zhang, Siming Zhao, Zelin Zhao, Yuemeng Zheng, Houliang Zhou, Yinghao Zhou, Elizaveta Zinovyeva

Table of Contents

<i>Shared Selective Persistent Memory for Agentic LLM Systems</i> Sanjana Pedada, Aditya Dhavala and Neelraj J. Patil	1
<i>Hydra: Adaptive Adversarial Attacks for Dynamic Multi-Stage Edge Intelligence Services</i> Linhan Li, Chuanqi Jiang, Wenqian Weng, Jacqueline J. Pang and Zhiguo Tao	14
<i>Silent Locale Corruption: Diagnosing and Mitigating Chinese-Locale Semantic Failures in LLM Tool Calling</i> Ruiyang Zhang	27
<i>Self-Correcting Large Language Models: Generation vs. Multiple Choice</i> Hossein A. Rahmani, Satyapriya Krishna, Xi Wang, Mohammadmehdi Naghiaei and Emine Yilmaz	38
<i>Do Context Files Help Coding Agents? A Two-Agent Ablation Study on Real Repositories</i> Prakhar Khatri	57
<i>Causal Decomposition of LLM Agent Failure via Oracle-State Intervention</i> Leqi Jia, Muhammad Saadat, Marlin Wong, Adithya Venkatesh and Kiran Nijjer	68
<i>Measuring the Serving Stack Instead of the Model: Hidden Confounds in Local Tool-Use Evaluation</i> Lijuan Tang and Yuemeng Zheng	86
<i>Authentic but Not Neutral: Selective Evidence in LLM-Agent Leaderboards</i> Xiaoxuan Li, Bangbang Liu and Ying Zhao	95
<i>GRASP: Generating, Revising, and Assessing for Strategic Planning with Agentic AI</i> Arunabh Srivastava, Mohammad A. Khojastepour, Srimat Chakradhar and Sennur Ulukus ..	109
<i>Answer or Compute? Direct Inference vs. Code Generation for LLM Game-State Understanding in RTS Games</i> Chang Liu	133
<i>Foundation First: Stabilizing SLM Agents via Pre-execution Logical Scaffolding</i> Juntae Lee, Jihwan Bang, Seunghan Yang and Sungha Choi	146
<i>Agent Step Value: Auditing Evaluation-Channel Sensitivity in Black-Box Agent Traces</i> Andrew Zhang and Chengzhan Li	161
<i>FUSE: Failure-Guided Scenario Synthesis and Weighted Refinement for LLM Function Calling</i> Zhiguo Tao, Linhan Li, Lifan Sun and Zichen Yuan	177
<i>Is Progressive Disclosure All You Need for Long-Context Agents?</i> Yifeng He, Yinzhe Zhao, Jicheng Wang and Hao Chen	187
<i>From Credit Assignment to Repair: A Leakage-Controlled Diagnostic Study of Tool-Using Agents</i> Shaobo Qiao	204
<i>When Should LLM Agents Forget? Memory Filtering Under Query-Critical Resolver Errors</i> Samarth Vinayaka	212
<i>STAC: When Innocent Tools Form Dangerous Chains for LLM Agents</i> Jing-Jing Li, Jianfeng He, Chao Shang, Devang Kulshreshtha, Xun Xian, Yi Zhang, Hang Su, Sandesh Swamy and Yanjun Qi	226

<i>Do Query-Agnostic Memory Structures Earn Their Budget? A 10M-Token Audit with Exact Read Limits</i> Amit Maraj and Kannav Sethi	263
<i>Explicit State Elicitation Is Not Enough: A Controlled Audit of Memory-Policy Classification</i> Yihang Chen, Pin Qian, Su Wang, Chong Peng, Huan Xu, Shuaiting Li and Yiqi Sun	271
<i>Retrieval is Enough: A Memory-Augmented Agent for Algorithm Selection</i> Zhang Weizhou and Lixing Lyu	295
<i>Decomposing Self-Reflection in LLM Agents: Diagnostic Scaffolds, Uncertainty Labels, and Action Diversity</i> Poli Nemkova and Haeshitha Indukuri	306
<i>Agent-Readiness of Synthesized API Tools: Parametric Gravity and the Limits of Documentation Grounding</i> Teodora Stereciu, Baohao Liao, Christian Herold, Shahram Khadivi and Christof Monz	323
<i>CAPSafe: A Benchmark for Contextual Action and Plan Safety Reasoning in Embodied Agents</i> Younna Farag, Svetlana Stoyanchev, Simon Keizer, Mohan Li and Rama Doddipatla	341
<i>Where the Tokens Go: A Measured Anatomy of LLM Agent Trajectory Costs</i> Tommy Tran	365
<i>Feedback Blindness: Diagnosing Silent Failures in Closed-Loop Agentic Data-Generation Pipelines</i> Loc Nguyen, Nguyen Khanh Long, Vu Gia Nam and Nguyen Tran Minh Nhat	374
<i>Tool Quality, Not Framework Complexity: Evaluating RLM Vision, Code, and Memory Pipelines for Radiology Applications</i> Jonas Schewior, Ansari Hamid Raja Gulam, Mohamed Ashfaq Anwerdeen and Andreas Wendemuth	382
<i>Autonomous Event Driven Multi-Agent Orchestration for Enterprise AI at Scale</i> Harsh Rao Dhanyamraju, Leonidas Raghav and Aaron Lee	398
<i>ToolMisuseBench: An Offline Deterministic Benchmark for Tool Misuse and Recovery in Agentic Systems</i> Rista Baral and Akshey Sigdel	435
<i>Same Answers, Different Active-Store States: An Identification Audit of FactConsolidation Single-Hop</i> Daeun Jeong and Yongjun Zhu	443
<i>Externalizing Requirement-to-Repair Artifacts as Observable Traces for LLM-Based Program Repair</i> Zewen Tao and Shin-nosuke Ishikawa	451
<i>Flight to Mediocrity: Why LLM Agent Selectors Under-Hire Experts When Quality Claims Are Unverified</i> Soham Wasmatar	461
<i>Translating the Translator: Decomposing the Cost of English-Forced Inter-Agent Communication</i> Kushagra Agrawal, Yuming Feng and Man-Fai Leung	473
<i>Mo' Models, Mo' Problems: How to best select model pools when designing Multi-Agent Systems</i> Sara Vera Marjanovic, Jiacheng Xu, Aleksandr Laptev, Grigor Nalbandyan, Erik Arakelyan and Evelina Bakhturina	484
<i>Be Careful Who You Trust: Coordination Dynamics under Corrupted Communication in LLM Multi-Agent Games</i> Xuanyi Liu, Niall Dalton, Hairi Amin, Xiyuan Yin and Lydia Lim	508

<i>Weighted Memory Tree: Remembering What Matters for Long-Horizon LLM Agents</i>	
Quang Dao, Purvi Kathalkar and Kenneth Eaton	524
<i>ACIES: A Self-Evolving Agentic Contract-Based Framework for Cyber-Physical System Control Synthesis</i>	
Yifeng Xiao, Zhiyu Ni, Dinghong Song, Mi Zhou and Pierluigi Nuzzo	541
<i>Capability Is Not Complementarity: A Routing Feasibility Study in Financial Question Answering</i>	
Rishi N. Simhadri	551
<i>Generation and Verification: Human-Agent Collaboration for Safety-Critical Automotive Software Engineering</i>	
Elizaveta Zinovyeva, Aiham Taleb, Nikita Kozodoi, Daniel Schleicher, Amir Mahdi Namazi, Martin Kraus, Levent Kent and Brian Jensen	557
<i>Two Thirds of the Best-of-K Oracle Band in Tool Agents Is Unreachable by Reward-Free Selection</i>	
Srinivas Harish	571
<i>AIDG: A Formal Decomposition of Information Extraction and Containment Asymmetries in Multi-Turn LLM Dialogue</i>	
Adib Sakhawat, Fardeen Sadab and Rakin Shahriar	579
<i>V-RAM: A Visual Retrieval-Augmented Multi-Agent Reasoning Framework for Medical Dialogue Diagnosis</i>	
Shukraditya Bose, Soumen Sinha, Ananya R Bhat and Rahul Roy	599
<i>AgenticLOCI: A Multi-Agent Memory Framework for Conversational AI with Longitudinal Personalization</i>	
Jonas Schewior, Deniz Bilgin, Gonzalo Haefner and Andreas Wendemuth	618
<i>Capture Is Easy, Forgetting Is Hard: Benchmarking Prospective Memory in Conversational Agents</i>	
Deniz Bilgin, Jonas Schewior and Andreas Wendemuth	639
<i>When Hiring Becomes Agent-Mediated: Evaluating Access and Recurrence in Two-Agent Résumé Screening</i>	
Jian Gao and Hang Jiang	655
<i>Research Agents Feed Doubts, Not Beliefs: Auditing Belief Framing in Live-Web Search</i>	
Siddharth Vohra and Min Xu	664
<i>When Failures Propagate: Causal Failure Attribution in Agentic Retrieval-Augmented Generation</i>	
Lauren Pothuru	675
<i>What Agent Benchmarks Don't Tell You: Recoverability and Credit Assignment in Agentic Retrieval Evaluation</i>	
Sai Krishna Rallabandi, Parag Pravin Dakle and Preethi Raghavan	687
<i>When the API Speaks the Wrong Language: Revisiting Post-Training for Multilingual Tool Use</i>	
Siddharth Chauhan, Thomas Butler, Abhishek Singhanian, Pankaj Kumar Porwal and Honey Gupta	702
<i>Beating a Random-Direction Bank Does Not Establish Policy Improvement: An Exact-Value Audit of Activation Steering for Agents</i>	
Avinash Goutham Aluguvelly	724
<i>DySAS: Dynamic State-Aware Student Simulation Beyond Static Personas</i>	
Partha Sarathi Purkayastha and Heejin Do	738

<i>MCPAgentBench: A Real-world Task Benchmark for Evaluating LLM Agent MCP Tool Use</i> Zixiang Liu, Wenrui Liu, Elsie Dai, Wenhan Yu, Lei Yu, Tong Yang, Yinjun Han and Hong Gao	764
<i>Grounding First: Auditing a Structured Recovery Wrapper for Local Open-Weight Reflexion Agents</i> Nimay Ballal and Badri Prasad V R	774
<i>What Drives Recovery in Agentic Text-to-Cypher? LAST-CQ: An LLM Agent Self-Refinement Framework</i> Ioannis Prokopiou, Athanasios Aidinis, Panagiotis-Christos Kyrmpatsos and Pantelis Vikatos	785
<i>Runtime Verification of Dimensional Semantics in Checkpointed Large Language Model Tool Workflows</i> Aditya Mazumdar	800
<i>Bayesian Belief Layer for Controllable Opinion Dynamics in LLM Agents</i> Hafsa Akbar, Daniel Platnick, Marjan Alirezaie, Hossein Rahnama and Alex Pentland	807
<i>Failing Agents Leave Traces: Telemetry Signatures for Multi-Tier Agent Evaluation</i> Muthaheera Yasmeena Belgur Shamiullah	816
<i>EARS: Explanatory Abstention for Reliable Sub-Agent Modeling in Large-Scale Multi-Agent Systems</i> Shuang Xie, Yunan Lu, Han Li and Lingyun Wang	832
<i>ExGQL: Self-Improving Natural Language-to-Graph Query Generation via Experience-Guided Learning</i> Penghong Guo, Boxin Du, Houliang Zhou, Chaman Gupta, Yina Gu, Rayssa Küllian and Shen Jiang	846
<i>Environment Steering: Using Data Flow Control to Improve Agent Utility and Safety</i> Charlie Summers, Prajwal Raghunath, Aaditya Pai, Mayur Kulkarni, Zhuo Zhang, Oliver Kennedy and Eugene Wu	856
<i>Role-Diverse Agent Forests for Neuroimaging Triage: When Does Prompt-Induced Role Diversity Help?</i> Tamara Kostova and Sonja Gievska	881
<i>The Audit Decides the Verdict: Instrument Effects Rival Demographic Bias in LLM Decision Audits</i> Siddharth Vohra and Manikandan Ravikiran	893
<i>SEDIMA: Cross-Run Hierarchical Insight Memory for Evolutionary Search Agents</i> Amirhossein Abaskohi, Mahdi Mostajabdaveh and Zirui Zhou	903
<i>When Critics Go Stale: Benchmarking Safety Guardrail Degradation Under Environment Drift</i> Fateme Taheri Dezaki, Lakshman Kolasani, Ankita Bhaumik, Siddhartha Sathyamoorthy Rao and Ayush Jain	922
<i>Are LLM Shopping Agents Behaviorally Distinguishable from Human Shoppers? A Session-Level Study</i> Arka Dutta, Virinchi Roy Surabhi, Jiayue Tong, Dihan Dai, Olcay Boz and Amila Weerasinghe	930
<i>BioSciReview-Bench: Benchmarking Coding Agents on Silent Scientific Bug Detection in Bioinformatics Code</i> Junhao Qiu	947
<i>Do Tool Schemas Fully Specify Argument Binding?</i> Jai Kumar Sharma and Peeyush Tapadiya	957

<i>Testing Behavioral Adaptation in Repeated Interactions Between LLM Agents</i>	
Lixin Liu, Yixuan Li, Xiankun Lin, Yutian Lu and Shiguang NI	973
<i>CaseInterviewBench: Evaluating Agentic AI's Purpose-Reflective Autonomy to Uncover What the User Really Needs</i>	
Shin-nosuke Ishikawa, Masato Todo, Atsushi Yoshino and Hirotsugu Ohba	986
<i>A Case Study of On-Premise ReAct for Collaborative Robot Programming</i>	
Yu-Seung Ma and Sangyeop Yeo	1003
<i>Causal Episodic Memory for Feedback-Driven Agent Repair</i>	
Khang Nhat Hoang Vo, Tam Minh, Anh Trac Duc Dinh, Thuyen Vinh Ha Bui and Tho Quan	1011
<i>Chronicle: Cut-Point Replay for Regression Testing of LLM Agents</i>	
Tisha Chawla and Susheem Koul	1026
<i>Contract Before Code: Failure-Compiled Contracts for Long-Horizon Coding Agents</i>	
Jiangnan YU and Kisson Songqi Lin	1032
<i>DeepRewind: Predicting and Repairing Premature Commitments in Deep Research Agents</i>	
Amirhossein Abaskohi, Amirhossein Dabiriaghdam, Lele Wang, Peter West and Giuseppe Carenini	1042
<i>BirdsongChat: A Hybrid Multi-Agent Framework for Interactive Multimodal Embodied Behavior Simulation</i>	
Callie C. Liao, Duoduo Liao and Ellie L. Zhang	1058
<i>An Evaluation of Data Leakage Risks in Tool-Using LLM Agents in Realistic Scenarios</i>	
Hankyul Baek, Jaewon Noh, Sang Seo, Yongsu Kim, Gabriel Waikin Loh Matienzo, Young Il Kim, Ee Wei Seah and Akriti Vij	1074
<i>GLOBE: Evolving Structured Agentic Memory for E-commerce Product Optimization in Generative Search</i>	
Hyunseo Kim, Suil Choe, Sukkyoung So and Daeyeon Yi	1097
<i>COMPASS: Critique-guided Optimization of MetaPolicies for Agentic Systems with Reinforcement Learning</i>	
Krishna Sayana, Ketan Kumar Todi, Ambarish Jash and Sukhdeep Sodhi	1118
<i>From Trajectories to Workflows: Online Procedural Memory Abstraction for Language Agents</i>	
Swathi Shree Narashiman, Ishan Jindal and Mahesh Chandran	1144
<i>Prior-History Sensitivity in Agentic Safety Evaluations</i>	
Melat Ghebreselassie and Melissa Li	1160
<i>CodeGrep: An RL-Trained Retrieval Agent for LLM Coding Agents</i>	
Wuya Chen, Yihao Yang, Yang Cao and Yue Lin	1171
<i>SkillCombiner: Optimizing Agent Skills via Edit Combination Search</i>	
Seungbeom Park, Jaewon Chu and Hyunwoo J. Kim	1188
<i>When Better Turns Do Not Make Better Agents: Diagnosing the Gap Between Next-Turn Metrics and Workflow Success</i>	
Md Tahmid Rahman Laskar, Xue-Yong Fu, Gundeep Singh, Karol Chang, Kevin Sanders, Shi Zong, Tania Habib, Julien Bouvier Tremblay, Shayna Gardiner, Harsh Saini, Matthias Lee, Elena Khasanova, Quinten McNamara and Shashi Bhushan Tyamagondlu Nagabhushan	1198

<i>Dynamics-Aware Navigation for Autonomous UI Agents on Directional-Input Devices</i>	
Ishitwa Viranchi, Akshintala Nagesh Kashyap and Mrunal Shah	1205
<i>Why Pay for Long Reasoning When Rewrite Do Trick? Causal Effects of Reasoning History Compaction on Future Generation and Agent Cost</i>	
Anton Ivanov Hristov	1213
<i>Diversity and Deliberation in Homogeneous LLM Prediction Polls</i>	
Jonas Gebele, Tristan Till and Florian Matthes	1229
<i>When Agreement Is Not Enough: Evidence-Risk Arbitration for Multi-Strategy Agentic QA</i>	
Yang He, Jinlin Wang, Yulong Ji and Jianwei Zhang	1236
<i>FlexSciEval: Structure-Grounded Multi-agent Evaluation for Flexible Electronics Research</i>	
Ke Li, Xinyi Yang, Runzhe Zhan, Qianying Hu, Derek F. Wong, Qingsong Li and Zhiyuan Liu	
1243	
<i>What People Almost Did: Evaluating LLM Social Simulations Beyond Behavioral Fit</i>	
JaeWon Kim and Angie Boggust	1263
<i>Lost in the Handoff: Contract-Aware Compliance Checking for Multi-Agent LLM Systems</i>	
Yang He, Jinlin Wang, Yulong Ji and Jianwei Zhang	1270
<i>SFT or RL for Tool-Calling Agents? A Controlled Study Across Data, Method, and Scale</i>	
Md Tahmid Rahman Laskar, Xue-Yong Fu and Shashi Bhushan Tyamagondlu Nagabhushan	1280
<i>D-Mem: A Dual-Process Memory System for LLM Agents</i>	
Zhixing You, Jiachen Yuan and Jason Cai	1287
<i>VISTA: A Versatile Interactive User Simulation Toolkit for Agent Evaluation</i>	
Yunan Lu, Ryan Shea, Yusen Zhang and Zhou Yu	1305
<i>CoDA-NAS: Pattern Knowledge-Based Cross-Task Neural Architecture Search with LLM Agents</i>	
Hao-Qun Huang, Kai-Chun Wu, Hsin-Yu Wang and Chun-Wei Tsai	1324
<i>Disentangling Topology and Diversity in Multi-Agent LLMs for Multilingual Low-Resource Emotion Detection</i>	
Ulugbek Shernazarov, Charitha Ruwansiri Weerakon Basnayake, Abdelkhaleq El Jarjini, Noel Crespi and Praboda Rajapaksha	1340
<i>Teaching Coding Assistants What to Change: Fine-Tuning from Pairwise Preferences</i>	
Hongmo Tao and Hiroyuki Kamata	1363
<i>Does The Way You Plan Matter? An Empirical Study of Planning Representations for LLM Web Agents</i>	
Alejandra Zambrano, Sara Vera Marjanovic, Imene Kerboua, Xing Han Lù and Leila Kosseim	
1375	
<i>Prompting Strategies for Conceptual Ideation in Problem Solving</i>	
Chris Bates and Ian Perera	1389
<i>Policy Loopholes in Agent Evaluation: When Policy Ambiguity Masquerades as Agent Error</i>	
Hongliu Cao	1404
<i>ToolFuse: Documentation-Intent Fusion for Training-Free Tool Retrieval</i>	
Shao-Cheng Lu, Shui-Hsiang Hsu, Chen-Jui Yu and Yao-Chung Fan	1420

<i>Trace2Ground: Scenario, Disambiguation, and Relation Traces for Web GUI Grounding</i>	
Ruozhen He, Zhongnan Qu, Sheng Zhang, Mesut Erhan Unal, Zhuowen Tu, Srikar Appalaraju, Jianbo Yuan, Shabnam Ghadar and Haofu Liao	1444
<i>Securing the Floor and Raising the Ceiling: A Merging-based Paradigm for Multi-modal Search Agents</i>	
Zhixiang Wang, Jingxuan Xu, Dajun Chen, Yunfang Wu, Wei Jiang and Yong Li	1471
<i>CoHyDE: Iterative Co-Training of LLM Rewriter & Dense Encoder for Tool Retrieval</i>	
Vaishali Senthil, Ashutosh Hathidara and Sebastian Schreiber	1482
<i>MATEO: A Multimodal Benchmark for Planning with Temporal Execution Ordering</i>	
Gabriel Roccabruna, Olha Khomyn and Giuseppe Riccardi	1501
<i>Rethinking Agentic Search with Piika: Is Lexical Retrieval Sufficient?</i>	
Tz-Huan Hsu, Jheng-Hong Yang, Yijun Ge and Jimmy Lin	1534
<i>CamPilot: A Multi-Agent Cinematic Assistant for Camera-Controlled Movie Generation</i>	
Yang Wu, Stefano Petrangeli, Ishita Dasgupta and Yu Shen	1551
<i>Whose Repair? A Multi-Loop Diagnostic of Interactive Alignment in Human–LLM Conversation</i>	
Wanqing Psyche He and Midam Kim	1565