

GroundLM 2026

**The 1st Workshop on Grounding Language Models:
Learning Faithfully and Efficiently**

Proceedings of the Workshop

October 29, 2026

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

Tencent



RBC Borealis

RBC BOREALIS

©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 979-8-89176-478-1

Preface

Welcome to the proceedings of the 1st Workshop on Grounding Language Models: Learning Faithfully and Efficiently (GroundLM 2026), held in conjunction with EMNLP 2026. GroundLM brings together researchers and practitioners interested in the principles, methods, benchmarks, and applications that help language models use evidence faithfully and communicate the basis of their outputs.

The workshop addresses grounding in a broad sense, including retrieval- augmented generation, multimodal and visual evidence, scientific and technical reasoning, tool use, attribution, uncertainty, and the evaluation of factuality and reliability. A central goal is to connect advances in model capability with auditable behavior: systems should not only produce useful answers, but should also be able to identify, use, and justify the evidence on which those answers depend.

We received 147 submissions across the GroundLM tracks. The regular main-paper track accepted 56 papers, corresponding to an acceptance rate of approximately 38.1%. These proceedings also contain 24 Findings papers and 10 shared-task system reports, which represent additional contributions selected through their respective evaluation processes. Together, the papers in this volume reflect a wide range of perspectives, from new datasets and benchmarks to modeling techniques, empirical studies, and practical systems for grounded language understanding and generation.

The program is complemented by invited talks from Freda Shi (University of Waterloo and Vector Institute), Lei Li (Carnegie Mellon University), Krzysztof Czarnecki (University of Waterloo), Valentin Maiorca (Institute of Science and Technology Austria), Danae Sánchez (University of Copenhagen), Anya Belz (Dublin City University), Iryna Gurevych (Technical University of Darmstadt), Ivan Titov (University of Edinburgh), and Sina Zarrieß (University of Bielefeld). We are also grateful to the panelists for the discussion of open challenges in grounding: Sina Zarrieß (University of Bielefeld), Carolin Lawrence (NEC Laboratories Europe), Amin Shabani (RBC Borealis), Dirk Hovy (Bocconi University), and Krisztian Balog (University of Stavanger and Google DeepMind). Their participation makes the workshop a place not only to present results, but also to exchange ideas across the language, vision, information retrieval, and scientific machine learning communities.

This program would not have been possible without the time and expertise of our area chairs and program committee members. We sincerely thank them for their careful deliberation, detailed reviews, and constructive guidance throughout the evaluation process. We are equally grateful to all authors for contributing their work and for engaging with the workshop community.

We gratefully acknowledge the generous support of Tencent and RBC Borealis. Their sponsorship has helped make the workshop and these proceedings possible. We also thank our invited speakers, panelists, organizing committee, and all volunteers who contributed to the preparation of the program and the final publication.

We hope that these proceedings will serve as a useful reference for the community and inspire further work toward language models that are capable, grounded, faithful, transparent, and auditable.

GroundLM 2026 Workshop Organizers

Organizing Committee

Workshop Organizers

Yimu Wang, University of Waterloo

Yee Man Choi, University of Waterloo

Di Wu, University of Amsterdam

Siqi Ouyang, Carnegie Mellon University

Mozhgan Nasr Azadani, University of Waterloo

Yi R. (May) Fung, Hong Kong University of Science and Technology

Program Committee

Area Chairs

Anamika Kumari, Microsoft
Eshaan Jain, University of Southern California
Kartheek Kumar Reddy Nareddy, German Aerospace Center (DLR)
Praveenkumar Kondikoppa, Research, Google
Qinglong Cao, Shanghai Artificial Intelligence Laboratory
Rishabh Agrawal, Amazon
Sachin Gupta, University of Colorado at Boulder and ebay Inc.
Saksham Kapoor, University of Maryland, College Park
Sarthak Ahuja, Amazon
Sourav Panda, Pennsylvania State University
Srijith Ravikumar, Amazon
Yuelyu Ji, University of Pittsburgh
Zian (Andy) Zheng, University of Waterloo

Reviewers

Aaditya Chauhan, Amazon
Aaditya Pai, Columbia University
Aastha Agrawal, Google
Abel Yagubyan, C3.ai
Adam Moffat, Cornell University and Vusion
Adi Shnaidman, DeepKeep AI
Aditya Channa, Maharaja Agrasen Institute Of Technology
Aditya Dubey, SRM Institute of Science and Technology
Advitya Gemawat, Microsoft
Ahsan Bilal, University of Oklahoma
Amitha Gopidi, Tualai
Anamika Kumari, Microsoft
Anay Dongre, California Polytechnic State University, Pomona
Anneswa Ghosh, Microsoft
Arghyadeep Das, University of Massachusetts at Amherst
Bala Krishna S Vegesna, Cohere AI Community
Chandan Prakash, Tata Consultancy Services Limited, India
Chirag Arora, Adobe Systems
Dmytro Negodiuk, Negodiuk LLC
Eshaan Jain, University of Southern California
Grégoire Cattani, International Business Machines
Guquan You, Guangzhou University
Harshit Jain, Adobe Systems
Harshita Khandelwal, UCLA Computer Science Department, University of California, Los Angeles
Huan Xu, Georgia Institute of Technology
Hyeonseop Yoon, maum.ai
Jiawei Li, Nanyang Technological University
Joel Vazquez Lopez, Uppsala University
Kartheek Kumar Reddy Nareddy, German Aerospace Center (DLR)

Khalid Hussain Ahmad, Federal University Lokoja
 Khashayar Khajavi, Simon Fraser University
 Kriti Faujdar, Yelp
 Kun Zhang, University of Leeds
 Lifan Hu, national university of singaore, National University of Singapore
 Lucky Verma, Eccalon LLC
 Md. Asaduzzaman Shuvo, United International University
 Md. Faiyaz Abdullah Sayeedi, BRAC University
 Menghao Huo, Santa Clara University
 Min-Oh Heo, HYUNDAI AUTOEVER
 Mogana Kumaran, IEEE
 Mohammadali Khodabandehlou, University of Southern California
 Noor Islam S. Mohammad, Istanbul Technical University
 Parth Suresh, Datology AI
 Pin Qian, Facebook
 Praveenkumar Kondikoppa, Research, Google
 Priyanshi Garg, University of Washington
 Pulkit Arya, Pointer
 Qian Chen, Tencent America
 Qinglong Cao, Shanghai Artificial Intelligence Laboratory
 Qingyang Xu, LinkedIn
 Qinwu Xu, Facebook
 Qirui Wang, Xi'an Jiaotong University
 R S Pavithra, Anurag University
 Rajeel Ahmad Ansari, dylog.ai
 Rohith Reddy Bellibatlu, Independent Living Systems
 Shashi Bhushan Tyamagondlu Nagabhushan, Dialpad
 Sachin Gupta, University of Colorado at Boulder and ebay Inc.
 Samarth Vinayaka, University of Florida
 Satveer Singh, Amazon
 Shankar Venkitachalam, Adobe Systems
 Sheela Ramprakash Goenka, Google Deepmind
 Shuhuai Lin, Google
 Siddharth Vohra, Amazon Web Services
 Siming Zhao, Google
 Smit Kadvani, Facebook
 Spandana Balumuri, Apple
 Srijith Ravikumar, Amazon
 Sriram Selvam, Microsoft
 Su Wang, Carnegie Mellon University
 Suneet Malhotra, Motorola Solutions
 Suryadeep Singh Deswal, Indian Institute of Technology, Roorkee
 Tautik Agrahari, Delve
 Tianyu Ding, Amazon
 Varun Kotte, Adobe Systems
 Xiankun Lin, SUN YAT-SEN UNIVERSITY
 Yash Ganpat Sawant, Wethos AI
 Yashdeep Thorat, Department of Computer Science, University of Illinois at Urbana-Champaign
 Yerang Kim, Korea Advanced Institute of Science & Technology
 Yu Wang, Xi'an Jiaotong University
 Yuelyu Ji, University of Pittsburgh

Yufeng Wang, Netflix
Yuichi Kitagawa, Hitachi, Ltd.
Yusi Gong, Fauna Robotics
Yuxin Wang, Dartmouth College
Zhan Shi, Amazon
Zian (Andy) Zheng, University of Waterloo
Zida Yang, Peking University
Prashanth Jamalapurapu, Friendzy LLC
Amlaan Bhoi, Apple
Rob Sneiderman, Independent and Statistics Canada
Mohan Premchand Bhambhani, LinkedIn
Riddhimaan Senapati
Yihang Chen, Georgia Institute of Technology
Atta UL Asad
Nishi Doshi, University of Southern California
Yan Sun, Pennsylvania State University
Zixiang Xu, University of Southern California

Keynote Talk

Freda Shi

University of Waterloo, Vector Institute

Keynote Talk

Lei Li

Carnegie Mellon University

Keynote Talk

Krzysztof Czarnecki

University of Waterloo

Keynote Talk

Valentino Maiorca

Institute of Science and Technology Austria

Keynote Talk

Danae Sánchez

University of Copenhagen

Keynote Talk

Anya Belz

Dublin City University

Keynote Talk

Iryna Gurevych

Technical University of Darmstadt

Keynote Talk

Ivan Titov
University of Edinburgh

Keynote Talk

Sina Zarrieß
University of Bielefeld

Panel

Panel Discussion: Open Challenges in Grounding

Thu, October 29, 2026 – Room: Budapest, Hungary

Panelists:

Sina Zarrieß, University of Bielefeld
Carolin Lawrence, NEC Laboratories Europe
Amin Shabani, RBC Borealis
Dirk Hovy, Bocconi University
Krisztian Balog, University of Stavanger, Google DeepMind

Table of Contents

<i>MedReason: Eliciting Factual Medical Reasoning Steps in LLMs via Knowledge Graphs</i> Juncheng Wu, Wenlong Deng, Xingxuan Li, Sheng Liu, Taomian Mi, Yifan Peng, Ziyang Xu, Yi Liu, Hyunjin Cho, Chang-In Choi, Yihan Cao, Hui Ren, Xiang Li, Xiaoxiao Li and Yuyin Zhou	1
<i>AAgiSphere-CORE: A Multi-Dimensional Co-Optimization Framework for Autonomous Prescriptive Maintenance in Smart Manufacturing</i> Pinghao Luo and Zhiguo Tao	18
<i>MSRA: A Parameter-Efficient Spectral Correction for Local Attention — Where It Helps, and Where It Fails</i> Tian Zihan	33
<i>Trust-Aware Multi-Axis Reliability Auditing for NLP Classifiers: A Theoretical and Empirical Framework</i> Noor Islam S. Mohammad	43
<i>VERITYGATE: A Four-Gate Schema-Level Faithfulness Framework and Paired Benchmark for Grounded LLM Narrations over Structured Evidence</i> Sachin Gupta	56
<i>EVO-RAG: Evolving Retrieval-Augmented Agents for Efficient Multi-Hop Query Optimization</i> Yuelyu Ji, Rui Meng, Zhuochun Li and Daqing He	72
<i>Retrieval-Reasoning Processes for Multi-hop Question Answering: A Four-Axis Design Framework and Empirical Trends</i> Yuelyu Ji, Zhuochun Li, Rui Meng and Daqing He	84
<i>SkillNav: A Multi-Agent Blackboard for Hallucination-Resistant Zero-Shot Object Navigation</i> Zhi Yang, Chengpu Yu and Yipin Zheng	104
<i>Riemannian Geometry for Pre-trained Language Model Embeddings</i> Szczepean Konior, Alexandre Quemy, Przemysław Klocek, Bartłomiej Sobieski and Grégoire Cattani	113
<i>Trust but Verify: Introducing DAVinCI - A Framework for Dual Attribution and Verification in Claim Inference for Language Models</i> Vipula Rawte, Ryan A. Rossi, Franck Dernoncourt and Nedim Lipka	132
<i>Semi-Supervised Learning to Improve Detection of Factual and Causal Errors in Language Models' Explanations</i> Mingxuan Liu, Han Yuan, Yilin Wu, Li Zhang and Zheng Ma	141
<i>Is Safety Grounded or Detachable? Removal-Cost Audits for Open-Weight Language Models</i> Nedim Mutlu Sezer, Ömer Faruk Aksoy, Feyzi Arda Salihoğlu, Yağız Ekrem Dalar and Ahmet Rıfat Öztürk	148
<i>CacheWeaver: Cache-Aware Evidence Ordering for Efficient Grounded RAG Inference</i> Kaizhen Tan, Rong Gu and Mingyuan Li	161
<i>Grounding or Lexical Overlap? A Matched-Pair Study of Internals-Based Faithfulness Probes</i> Kun Zhang	171
<i>PARSE: Provenance-Aware Retrieval Sanitization for Professional Domain LLM Agents</i> Aaditya Pai	185

<i>Does Quantization Erode Grounding? A Faithfulness Audit of On-Device LLM Compression</i>	
Neeraj Sai Venkat Movva and Sathvik Loke	195
<i>The Citation-Faithfulness Illusion: Measuring Whether Grounded Answers Preserve Source Disagreement</i>	
Rob Sneiderman	201
<i>Lifecycle Faithfulness for Grounded Retrieval-Augmented Generation: Benchmarks, Evaluation, and the Vera-Life Controller</i>	
Zida Yang, Guquan You, Jiexin Zheng and Yijie Peng	211
<i>Do LLM Attribution Metrics Transfer? Auditing Retrieval-Augmented Generation Evaluation Across Datasets and Constructs</i>	
Tianyu Ding, Aditya Nannapaneni and Juan Pablo De La Cruz Weinstein	241
<i>EviScope: Paired Counterfactual Evidence Diagnostics for Faithful and Efficient Grounded Language Models</i>	
Suryadeep Singh Deswal	257
<i>Does RAG Know When Retrieval Is Wrong? Diagnosing Context Compliance under Knowledge Conflict</i>	
Yihang Chen, Pin Qian, Su Wang, Sipeng Zhang, Huan Xu, Shuhuai Lin and Xinpeng Wei ..	263
<i>Self-Detect-Refine: Iterative Training Label Refinement Method for Clinical Note Generation</i>	
Gin-kyeng Lee, Serim Kim and Harin Jun	277
<i>Graph Evidence Is Not Text Evidence: Serialization Faithfulness as Structural Invariance in Graph-Grounded Language Models</i>	
Lifan Hu	305
<i>CoSched-Cert: Faithful Multi-Turn Scheduling Dialogue via Solver-Verified Certificates</i>	
Zida Yang, Minhao Zou and Yijie Peng	310
<i>The Citation Expansion Trap: Operating-Point Control for Fixed-Answer Multi-Hop RAG Citation Repair</i>	
Jiawei Li	322
<i>Staged Linguistic Seeding: Grounded Query Expansion for Verified-Unit QA in AI Contact Centers</i>	
Hyeonseop Yoon and Jeong-eun Park	330
<i>Structured Pruning and Lightweight Adaptation of Foundation Models for Resource-Efficient Earth Observation</i>	
Aniket Ramchandra Munde, Kartheek Kumar Reddy Nareddy and Julia Niebling	343
<i>E-SENS: Exclusion-Sensitive Penalization for Negative-Constraint Retrieval</i>	
Yerang Kim, Jiyeon Myung and Joohyung Han	353
<i>How Far Can You Get Without a GPU? A Systematic Benchmark of Lightweight Hallucination Detection Across Question Answering, Dialogue, and Summarisation</i>	
Kriti Faujdar and Smit Kadvani	367
<i>Memory Makes the Difference: Evaluating How Different Memory Roles Shape Conversational Agents</i>	
Yuxin Wang, Paul Thomas, Zhiwei Yu, Yuan Gao, Saeed Hassanpour, Soroush Vosoughi, Robert Sim and Nick Craswell	378
<i>Relational Grounding Failures in Zero-Shot Cross-Lingual Dialogue Summarization with Small Language Models</i>	
Yunwoo Lee, Ziwa Li and Jennifer Huang	402

<i>A Randomization-Control Diagnostic for Document Grounding in Self-Distillation Gates</i>	
Xiankun Lin, Lixin Liu, Xinyi Lin and Zhezheng Hao	412
<i>Safe Adaptive Context and TACER: Task-Aware Evidence Routing for Token-Efficient Retrieval-Augmented Generation</i>	
Joel Vazquez Lopez, Andreia Alexa, Kārlis Mārtiņš Auce, Simon Backhaus Brudzewski and Sheraz Ahmad	418
<i>When Should an Agent Distrust Its Own Context? Model-Dependent Outcomes of an Audited Reasoning Pipeline</i>	
Aditya Channa, Madhav Kapoor and Jastej Singh.....	434
<i>What Does Answer Change Rate Actually Measure? A Specificity Audit of Causal Intervention Tests for Explanation Faithfulness</i>	
Durgesh Rao and Dr. Shirshendu Layek	441
<i>Gradient-Aligned Iterative Fine-Tuning for Adaptive LLM Post-Training</i>	
Zhan Shi	453
<i>Geometry-Conditioned Turn Scoring for Conversational Memory Compression</i>	
Pranav Kompally and Sibi Chakkaravarthy S.....	463
<i>Faithfulness Is Not Free: Auditing Offline KV-Cache Quantization in Retrieval-Augmented Generation</i>	
Atta UL Asad, Ahsan Bilal, Muhammad Ali, Muhammad Haseeb and Dean F. Hougen.....	472
<i>Cross-Modal Grounding for Calibrated Confidence in Vision Language Models</i>	
Trilok Padhi, Ramneet Kaur, Adam D. Cobb, Manoj Acharya, Colin Samplawski, Brian Matejek, Alexander Michael Berenbeim, Nathaniel D. Bastian, Susmit Jha, Ugur Kursuncu and Anirban Roy.....	480
<i>Lost In Translation: Uneven Grounding Degradation Under Cross-Lingual Shift in LLM-Assisted Cyber Threat Intelligence</i>	
Lailah Denny	505
<i>RuleChef: Grounding LLM Task Knowledge in Human-Editable Rules</i>	
Ádám Kovács, Nadia Verdha and Gabor Recski	513
<i>Compression-Aware Abstention: Teaching LLMs to Refuse When KV-Compression Masks Remove Answer Evidence</i>	
Mohammadali Khodabandehlou and Bhaskar Krishnamachari	531
<i>Eval-Pair Matrix: Answer-Paired Meta-Evaluation of LLM Judges for Grounded RAG</i>	
Sriram Selvam and Anneswa Ghosh.....	550
<i>Beyond Document Grounding: Span-Level Hallucination Detection over Code, Tool Output, and Documents</i>	
Ádám Kovács, Bowei He, Xue Liu, István Boros, Szilveszter Tóth and Gabor Recski	566
<i>Does Retrieval Augmented Generation Collapse When It Retrieves AI-Generated Documents?</i>	
Riddhimaan Senapati, Gregory Druck, Chang Zeng, Fabeha Fatima, Rati Rastogi and Ozel Yilmazel	578
<i>Similar Accuracy, Unequal Evidence: Search APIs as Decision Surfaces for Tool-Using Agents</i>	
Sriram Selvam and Anneswa Ghosh.....	606
<i>Latent Truth Decoding: Training Free Mitigation of Sycophancy in Small Language Models</i>	
Srutanik Bhaduri, Neelesh Kumar Shukla and Viji Krishnamurthy	623

<i>Stable Scores, Failed Retrieval: Diagnosing RAG Evaluation in Low-Resource Indic Languages</i>	
Priyanshi Garg, Jasmin Gustafson and Yingzhe Fan	637
<i>Skip One Layer? A Multilingual and Multi-Task Study of Single-Layer Skipping in Modern Language Models</i>	
Ayu Purwarianti, Antonio Natthan Krishna, Muhammad Risqi Firdaus, Hanif Muhammad Zhafran, Wilson Tansil, Achmad Imam Kistijantoro, Nur Ahmadi, Rahmat Mulyawan, Dessi Puji Lestari and Gregory Kielian	649
<i>Quality-Aware LLM Data Augmentation for Arabic Dialectal NER</i>	
Kahina Houda Saadaoui, Chaima Abdelaziz, Faiza Belbachir and Lynda Said Lhadj	666
<i>Robustness of Diffusion and Autoregressive Language Models to CheckList-Style Perturbations</i>	
Aarush Rathore, Tirtharaj Dash, Lovekesh Vig and Ashwin Srinivasan	685
<i>DTS: Discriminant Truthfulness Scoring for Gradient-Free Hallucination Detection</i>	
Kaustubh Atey, Shreeyam Bangera, Sandeep Kumar, Ankit Kumar and Amit Sethi	694
<i>Reset Is Not Recovery: Evaluating Recoverability from False Conversational Context via Sycophancy Hysteresis</i>	
Adi Shnaidman	712
<i>Groundedness Is Not Truth: Evidence-Chain Forensics for Retrieval-Augmented Generation</i>	
Samarth Vinayaka	732
<i>Human-Perception-Aware but Not Yet Physically Grounded: Strengths and Failure Modes of Frontier Multimodal LLMs and the GroundVQA Benchmark</i>	
Qinwu Xu, Xin Liu and Yifan Jiang	743
<i>Polite on the Surface, Broken in Practice: A Curated Dataset for Fixing Generation and Register Failures in Low-Resource Bangla Text Generation</i>	
Md. Asaduzzaman Shuvo, Mahedi Hasan, Md. Tashin Parvez, Azizul Haque Noman and Md. Shafayet Hossain Ovi	751
<i>Retrieve, Locate, Generate: An Oracle-Substitution Diagnostic for Literature-Grounded QA</i>	
Siddhesh More and Kunal Jadhav	767
<i>CITECHECK: Retrieval-Grounded Detection of LLM Citation Hallucinations in Scientific Text</i>	
Khashayar Khajavi, Shaghayegh Sadeghi, Rise Adhikari and Alexander Tessier	779
<i>GRACE: Generating Conflict-Aware Cases for Evaluating Policy Compliance in Enterprise Conversational AI</i>	
Jinwon Kim, Kyunghyun Lim, Gayeon Kim, Sookeun Yun and Min-Oh Heo	796
<i>StreamSense: A Content- and Resource-Aware Edge Agent for Adaptive Bitrate Streaming</i>	
Chuanqi Jiang, Linhan Li and Zhiguo Tao	823
<i>From Position Bias to Position-Stable Agreement: Testing What LLM Judges Evaluate</i>	
Lixin Liu, Xiankun Lin and Zhezheng Hao	839
<i>Preserving Evidence Anchors: Query-Rewriting Guardrails for Retrieval-Side Grounding</i>	
Varun Kotte and Chirag Arora	857
<i>Grounding Creative Query Suggestions, Faithfully and Efficiently: From Corpus-Grounded Supervision to Reinforcement Learning</i>	
Harshit Jain, Vedanth Subramaniam, Twisha Naik, Chirag Arora and Vipul Jain	869

<i>The Coin Flip Judge? Reliability and Bias in LLM-as-a-Judge Evaluation</i>	
Abel Yagubyan	882
<i>The Attribution–Compression Frontier in Retrieval-Augmented Generation</i>	
Deepanshu Mody	891
<i>Budget-Aware Compression Pipeline for Single-GPU LLM Inference: Methods, Trade-offs, and Coupling Effects</i>	
Hongyu Yu and Yifei Shen	911
<i>Leveraging Low-Level Symbolic Competences for Unsupervised Grounding in Hallucination Detection</i>	
Renato Vukovic, Hsien-chin Lin, Carel van Niekerk, Benjamin Matthias Ruppik, Michael Heck, Shutong Feng, Nurul Lubis and Milica Gasic	923
<i>The Coverage Illusion: From Pre-retrieval Routing Failure to Post-retrieval Cascades in a Production RAG System</i>	
Zafar Hussain and Kristoffer Nielbo	937
<i>When Models Defer to Wrong Answers: A Robustness Audit of Source-Attributed Cues in Multiple-Choice QA</i>	
Manikandan Ravikiran and Siddharth Vohra	954
<i>Does the Selected Object Reach the Reader? Auditing Identity Handoffs in Grounded Language-Model Pipelines</i>	
Siddharth Vohra, Runmin Jiang, Xiaomo Li and Min Xu	966
<i>Learning from Dialogue Traces: Retrieval-Guided Tool Use and Response Generation for Task-Oriented Dialogue Agents</i>	
Wiradee Imrattanatrat, Ken Fukuda and Masahiro Hamasaki	981
<i>Enhancing Small Language Models in RAG via Triplet-Guided In-Context Distillation</i>	
Ilia Tambovtsev, Maria Molchanova, Anna Mikhailova, Alexey Korotkov, Ivan Sgibnev, Iurii Vizilter and Andrey Savchenko	1001
<i>Why Your Deep Research Agent Fails? On Hallucination Evaluation in Full Research Trajectory</i>	
Yuhao Zhan, Tianyu Fan, Lingxuan Huang, Zirui Guo and Chao Huang	1016
<i>A Modular Approach to Conversational Object Grounding in 3D Scenes Using VLMs and Prompt Chaining</i>	
Simon Keizer, Mohan Li, Youmna Farag, Svetlana Stoyanchev and Rama Doddipatla	1049
<i>BAND: Bias-Aware Retrieval Debiasing for Demographic Robustness in Retrieval-Augmented Medical Question Answering</i>	
Harim Lee, Dasom Lee and Chaehyun Ryu	1075
<i>Agents Out of Bounds: A Taxonomy of Architectural Drift and a Deployment-Ready Guardrail Framework for LLM Coding Agents</i>	
Saurabh Yergattikar	1080
<i>When Targeted Grounded Gains Hide Complementary Abstraction Regression: DION-Bench and a Conjunctive Evaluation Protocol</i>	
Zida Yang, Guquan You, Jiexin Zheng and Yijie Peng	1091
<i>Evidence-Indexed Slide Narration With Heuristic Checks and Template Fallback</i>	
Sachin Gupta	1116

<i>Default Zero-Shot LLM APIs Over-Flag Indian Financial SMS: A Policy-Aligned Benchmark and Template-Level Pre-Deployment Audit</i>	
Sachin Gupta	1124
<i>PageRecall: Measuring Page Selection in Literature-Grounded Question Answering</i>	
Aaditya Chauhan	1137
<i>LitTraceQA: A Benchmark for Multi-Stage Grounding and Verification in Scientific Question Answering</i>	
Xuye Liu, Yimu Wang, Peng Shi, Bo Xue, Xiangrui Ke, Songcheng Cai, Kath Choi, Di Wu, Freda Shi and Krzysztof Czarnecki	1146
<i>CoVeR-VQA: Multi-Stage Evidence Verification for GoldenViewVQA</i>	
Kun Wang, Yupeng Hu, Ruping Cao, Hao Liu, Zhiran Li, Qianlong Xiang and Harry Cheng	1154
<i>A Non-Visual Baseline for Multi-View Visual Question Answering on GoldenViewVQA</i>	
Aravind Mandiga	1162
<i>Combining RAG and Agentic Research for Evidence-Grounded Scientific Question Answering</i>	
Zhirui Zhu	1168
<i>Maddies at GroundLM 2026 Shared Tasks: Answer-Preserving Visual Re-grounding for Multi-View Driving VQA</i>	
Aditya Channa, Madhav Kapoor and Jastej Singh	1174
<i>TRACE: Target-Aware Retrieval, Attributed Evidence, and Contract-Constrained Extraction for LitTraceQA</i>	
Sachin Gupta and Divya Godara	1179
<i>SAGE: Slot-Aware Grounded Evidence Acquisition for LitTraceQA</i>	
Gang Xu, Ce Huang, Yushan Pan and Wei Wang	1187
<i>OdeD at GroundLM 2026 Shared Tasks: Reading the Scorer for Literature-Grounded QA</i>	
Omar Mokhtar Gaber Saqr and Mostafa Gaafar	1196
<i>Lit3R: Retrieve–Relate–Read for Evidence-Grounded Question Answering over Scientific Literature</i>	
Akira Ise, Kotaro Kumagai, Yuta Yamaguchi, Hisanori Ozaki, Yukio Uematsu and Ikuya Yamada	1205
<i>Ensar at GroundLM 2026 Shared Tasks: Symbolic Scene Serialisation for View-Level Evidence Identification</i>	
Mostafa Gaafar and Omar Mokhtar Gaber Saqr	1214
<i>Efficient Adaptation of Small Language Models for Reliable Mathematical Reasoning</i>	
Kanak Anil Malpani	1221
<i>Findings of the First GroundLM Shared Tasks: Evaluating Grounded Language Models Across Visual and Scientific Evidence</i>	
Yimu Wang, Mozghan Nasr Azadani, Xuye Liu, Yee Man Choi, Bo Xue and Krzysztof Czarnecki	1232