

SURGeLLM 2026

**The First Workshop on Structured Understanding, Retrieval,
and Generation in the LLM Era**

Proceedings of the Workshop

July 3, 2026

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

Sponsor



©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 979-8-89176-406-4

Introduction

Welcome to the First Workshop on Structured Understanding, Retrieval, and Generation in the LLM Era (SURGeLLM 2026), co-located with ACL 2026 and held on July 3, 2026 in San Diego, California, United States.

SURGeLLM brings together researchers and practitioners from natural language processing, information retrieval, data management, and visualization to advance structure-aware understanding, retrieval, generation, and evaluation. Structured artifacts, such as tables, databases, charts, maps, flowcharts, diagrams, workflows, and code, are central to many scientific, enterprise, and everyday applications. At the same time, modern large language model systems still face important challenges in faithfully reasoning over, retrieving from, and generating such artifacts. The workshop aims to provide a cross-community forum for new methods, datasets, benchmarks, systems, analyses, and lessons learned, including work on tabular understanding and question answering, natural language interfaces to structured data, structure-aware retrieval, structured generation, agentic analysis systems, robustness, faithfulness, reproducibility, and human-centered evaluation.

This year we received 79 eligible direct submissions, excluding withdrawn and desk-rejected papers, and accepted 30 of them (38.0%). Among these accepted direct submissions, 21 appear in these proceedings and 9 are presented as non-archival work. We also received 5 ACL Rolling Review commitments; drawing on their ARR reviews, meta-reviews, and scores, we accepted all 5, with 3 appearing in these proceedings and 2 presented as non-archival work. In addition, we accepted 6 non-archival presentations of recently accepted or published work solicited through a Google Form. In total, the workshop program features 41 accepted presentations: 24 archival papers in these proceedings and 17 non-archival presentations. The program also includes invited talks by Heng Ji (University of Illinois Urbana-Champaign), Dan Roth (University of Pennsylvania and Oracle AI), Hamid Palangi (Google and University of Washington), and Lisa Amini (IBM Research); an industry panel featuring Kamalika Das (Intuit), Yunyao Li (Adobe), Frederic Sala (Snorkel AI), Lisa Amini (IBM Research), and Sujith Ravi (Oracle AI); and opportunities for community interaction.

We are grateful to all authors who submitted work to SURGeLLM 2026 and to all presenters for contributing to the workshop program. We thank the Program Committee and additional reviewers for their thoughtful reviews, careful discussions, and commitment to building a high-quality program on a tight timeline. We also thank our invited speakers for sharing their perspectives with the community, and we appreciate the guidance and support of the ACL 2026 workshop chairs, publication chairs, and organizing team.

We hope these proceedings help support continued work on making structured artifacts first-class citizens in modern language technologies, and we look forward to the discussions, collaborations, and new ideas that will grow from this workshop.

Amit Agarwal, Shuaichen Chang, Xiusi Chen, Kaize Ding, Michael Glass, Vivek Gupta, Oktie Hassanzadeh, Harsha Kokel, Yuan Luo, Kavitha Srinivas, Yu Wang, Yu Zhang, Yue Zhao, and Qi Zhu
The SURGeLLM 2026 Organizing Committee

Organizing Committee

General Chairs

Amit Agarwal, Oracle
Shuaichen Chang, Amazon
Xiusi Chen, University of Illinois Urbana-Champaign
Kaize Ding, Northwestern University
Michael Glass, IBM Research
Vivek Gupta, Arizona State University
Oktie Hassanzadeh, IBM Research
Harsha Kokel, IBM Research
Yuan Luo, Northwestern University
Kavitha Srinivas, IBM Research
Yu Wang, University of Oregon
Yu Zhang, Texas A&M University
Yue Zhao, University of Southern California
Qi Zhu, Amazon

Program Committee

Reviewers

Hamidreza Almasi, Ekant Muljibhai Amin, Mohsen Arjmandi

Yuyang Bai, Fenil Bardoliya, Leslie Barrett, Kushagra Bhushan, Shivangi Bithel

Rajmohan C, Aman Chadha, Abhijit Chakraborty, Mohit Singh Chauhan, Kuang Ming Chen, Xi Cheng, Benoit Crabbé

Kirushikesh DB, Sarthak Dash, Naihao Deng, Srimonti Dutta

Luis Garcés Erice, Anand Eswaran

Xi Fang, Raul Castro Fernandez

Jugal Gajjar, Shuyu Gan, Neelamadhav Gantayat, Sambit Ghosh, Shlok Gilda, Ioana Giurgiu, Parker Glenn, Pranay Goel, Suranjan Goswami, Zhijin Guo, A Said Gurbuz

Asela Hevopathige, Phillip Howard, Dan C Hsu, Cassie Huang, Seongyun Hwang

Eun Woo Im

Rakesh Jain, Chengtao Jian

Tanvi Kaple, Meher Gitika Karumuri, Vivek Hruday Kavuri, Ceyhun Efe Kayan, Phani Rohitha Kaza, MD Mahdi Mosleh Khan, Daniel Khashabi, Aamod Khatiwada, Daehui Kim, Jayoung Kim, Philipp Koehn, Aneta Koleva, Achintha I Kondarage, Yuta Koreeda, Tanay Kulkarni, Gokul Karthik Kumar, Tampu Ravi Kumar, Vineet Kumar

Corey Lammie, Jeongwoo Lee, Xuechen Li, Zhuochun Li, Zhuofeng Li, Yashu Pura Lingaraju, Leyuan Liu, Muxin Liu, Zefang Liu, Sixing Lu, Xinyuan Lu

Rakesh Chowdary Machineni, Simon Malberg, Sumanth Manduru, Sascha Marton, Tobias Materzok, Aritra Mazumder, Samuel Mensah, Rudra Murthy

Zoltn Arnold Nagy, Gaurav Najpande, Krishnasuri Narayanam, Sumit Neelam, Wuguang Ni

Robin Oester

Maitrik Patel, Nhan H Pham, Biswas Poudel, Khussal Pradhan

Zhisheng Qi, Lijun Qian, Yinzhu Quan

Arushi Rai, Nishat Raihan, Gaetano Rossiello, Sangwon Ryu

Utkarsh Sahu, Madhavan Seshadri, Abhilash Shankarampeta, Sola Shirai, Nehaa Shri, Adarsh Singh, Ankit Pratap Singh, Mukul Singh, Usneek Singh, Ranjan Sinha, Marko Sterbentz, Alexander Sternfeld, Kamalasanankari Subramaniakuppasamy

Shir Ashury Tahan, Haixiang Tang, Prasham Titiya, Ioannis Tzachristas

Rajendra Ugrani, Mansi Uniyal, Ritam Upadhyay

Rohith D Vallam, Josh Viktorov, Rahul Vimalkanth, Gerardo Vitagliano

Isaac Waweru Wambugu, Hongzhi Wang, Ziyu Wang, Zirui Wei, Fei Fan Wu

Qiushui Xu, Ruiyao Xu, Xueqiang Xu

Youngsik Yun

Yiming Zeng, Sheng Zhang, Xin Zhang, Shuhan Zheng, Hangxiao Zhu, Henghui Zhu, Chunsheng Zuo

Table of Contents

<i>UNJOIN: Enhancing Multi-Table Text-to-SQL Generation via Schema Simplification</i>	
Poojah Ganesan, Rajat Aayush Jha, Dan Roth and Vivek Gupta 1
<i>The Mighty ToRR: A Benchmark for Table Reasoning and Robustness in LLMs</i>	
Shir Ashury-Tahan, Yifan Mai, Rajmohan C, Ariel Gera, Yotam Perlitz, Asaf Yehudai, Elron Bandel, Leshem Choshen, Eyal Shnarch, Percy Liang and Michal Shmueli-Scheuer 16
<i>Ontology-Free General-Domain Knowledge Graph-to-Text Generation Dataset Synthesis using Large Language Model</i>	
Daehui Kim, Deokhyung Kang, Sangwon Ryu and Gary Lee 52
<i>Output-Space Search: Targeting LLM Generations in a Frozen Encoder-Defined Output Space</i>	
Tobias Materzok 70
<i>TreeDiff: AST-Guided Code Generation with Diffusion LLMs</i>	
Yiming Zeng, Jinghan Cao, Zexin Li, Yiming Chen, Tao Ren, Zhuochun Li, Dawei Xiang, Xidong Wu, Shangqian Gao and Tingting Yu 93
<i>Chart-RL: Generalized Chart Comprehension via Reinforcement Learning with Verifiable Rewards</i>	
Xin Zhang, Xingyu Li, Rongguang Wang, Ruizhong Miao, Zheng Wang, Yuying Wang, Dan Roth and Chenyang Li 107
<i>RSAT: Structured Attribution Makes Small Language Models Faithful Table Reasoners</i>	
Jugal Gajjar and Kamalaskankari Subramaniakuppusamy 119
<i>Framework of Thoughts: A Foundation Framework for Dynamic and Optimized Reasoning based on Chains, Trees, and Graphs</i>	
Felix Fricke, Simon Malberg and Georg Groh 132
<i>TabGuard: Agentic LLM Orchestration for Adaptive Tabular Anomaly Detection via Dynamic Validator Selection and Generation</i>	
Srihari Unnikrishnan and Minghua Ma 152
<i>StructSurvey: Structured Agentic Retrieval for Automated Survey Paper Generation</i>	
Paolo Pedinotti and Enrico Santus 162
<i>Modeling LLM Agent Reviewer Dynamics in Elo-Ranked Review System</i>	
Hsiang-Wei Huang, Junbin Lu, Kuang-Ming Chen, Jianxu Shangguan and Jenq-Neng Hwang 182
<i>DSMentor: Curriculum-Guided Inference with Online Memory for Data-Science LLM Agents</i>	
He Wang, Alexander Hanbo Li, Yiqun Hu, Sheng Zhang, Hideo Kobayashi, Jiani Zhang, Henghui Zhu, Chung-Wei Hang and Patrick Ng 190
<i>Asking language models how to represent data for fine-tuning</i>	
Usneek Singh, Ananya Singha, Abhijeet Awasthi, Sumit Gulwani, Aditya Kanade, Vu Le, Mukul Singh and Gust Verbruggen 209
<i>TabBridge: Bridging Structure and Context for Accurate Table Reasoning</i>	
Jeongwoo Lee, Eunsoo Lee and Jihie Kim 219
<i>Map-of-Actions: Deliberate Reasoning over Multi-Labeled Graphs</i>	
Wuguang Ni and Kai Yang 241

<i>Routing End User Queries to Enterprise Databases</i>	
Saikrishna Sudarshan, Tanay Kulkarni, Manasi Patwardhan, Lovekesh Vig, Ashwin Srinivasan and Tanmay Tulsidas Verlekar 260
<i>SchemaScope: How Join-Hop Depth Breaks Text-to-SQL in Large Language Models, and a Decomposition-Based Remedy</i>	
Kaustubh S. Bukkapatnam and Rayan Malik 269
<i>Generalization in Graph Reasoning: A Systematic Comparison of LLM Training Approaches</i>	
Sola Shirai, Kavitha Srinivas, Julian Dolby, Michael Katz, Shirin Sohrabi and Horst Samulowitz 275
<i>Can LLMs Self-Correct Table Reasoning Errors?</i>	
Farseen Shaikh 298
<i>Mixed-Policy GRPO for Text-to-SQL with Off-Policy Data Generation</i>	
Marko Sterbentz, Michael Glass, Nhan H Pham, Dharmashankar Subramanian and Kristian J Hammond 313
<i>TabFaith: Benchmarking and Improving Structural Faithfulness in LLM Table Summarization</i>	
Kaustubh S. Bukkapatnam and Sohum Mehta 326
<i>StructHallu-Drift: Benchmarking Structured Hallucinations Under Schema Evolution in LLMs</i>	
Mujtaba Hasan 333
<i>Reasoners or Translators? Contamination-aware Evaluation and Neuro-Symbolic Robustness on Tax Law</i>	
Parisa Kordjamshidi, Samer Aslan, Madhavan Seshadri, Leslie Barrett and Enrico Santus 344
<i>More Than Efficiency: Embedding Compression Improves Domain Adaptation in Dense Retrieval</i>	
Chunsheng Zuo and Daniel Khashabi 361