

SURGe: Structured Understanding, Retrieval, and Generation

Abstract

We propose organizing Structured Understanding, Retrieval, and Generation (SURGe 2026), a new workshop at ACL that places structured data at the center of NLP research. Building on the recent momentum in table QA, Text-to-SQL, and structured retrieval, the workshop emphasizes treating structured artifacts - tables, charts, maps, flowcharts, and diagrams - as first-class citizens rather than secondary to text. SURGe will focus on three core challenges: enabling reliable reasoning over structured data, ensuring faithful generation of structured outputs, and advancing robust retrieval of structured artifacts. By bringing together communities across NLP, IR, DB, and visualization, we aims to consolidate progress, spotlight emerging challenges, and spark cross-disciplinary collaborations that shape future research.

Tagline: First-class NLP for tables, charts, maps, flowcharts and diagrams.

1 Workshop Topic and Content

Modern NLP systems are increasingly interacted with structured artifacts, tables, charts, maps, and diagrams. However, reasoning, generation, and retrieval over these data types remain fragmented across research communities like NLP, DB, IR and CV. There is a need to bring together researchers from these communities to create a shared space for interdisciplinary research.

While table QA, Text-to-SQL, and structured retrieval have seen strong progress, current models often treat structured data as secondary to text, leading to unreliable reasoning, limited faithfulness, and ad hoc evaluation. To close this gap, Structured Understanding, Retrieval, and Generation (SURGe 2026) aims to make structured artifacts first-class citizens in NLP research.

The workshop will focus on three themes. **1. Structured Understanding:** Work that focuses on inference over structured data, including Question Answering, claim verification over tables and DBs (Chen et al., 2020; Lu et al., 2023), attribution tasks (Mathur et al., 2024), and explainability tasks. **2. Structured Retrieval:** Retrieving relevant tables from text or columns (Chen et al., 2024; Kokel et al.), schema-linking (Li et al., 2023a;

Glass et al., 2025), retrieving charts or workflows with structure-aware evaluation. **3. Structure Generation:**

Going from Text to SQL (Lei et al.; Chang et al.; Li et al., 2023b), text to table (Jain et al., 2024; Ahuja et al., 2025), text to charts or code (Zhang et al., 2024), plan generation (Podubnyi and Dorodnykh), their evaluation metrics and approaches to ensure robustness and reliability (Somov and Tutubalina, 2025). The methods of interest include agentic LLM/VLM pipelines, direct prompting, fine-tuning, structure-specialized embeddings, reinforcement learning based, and decomposition based approaches. We would like to scope the workshop to include:

- Methods for tabular understanding and question answering, including multi-table reasoning, multi-hop inference, and integration with unstructured or multimodal data.
- Techniques for generating structured outputs (tables, charts, figures, code) from natural language, with attention to schema intent.
- Approaches to improve correctness, faithfulness, interpretability, and robustness in structured generation, including rubric-based evaluation.
- Retrieval methods for structured content, enabling access to cells, rows, map regions, chart elements, and code fragments from natural language.
- Synchronization of information across structures and modalities, such as cross-lingual tables, text-table-figure alignment, and temporal updates.
- Evaluation frameworks for structure-aware reasoning, retrieval, and generation, combining automated metrics with human-centered assessment.
- Applications of large language and vision-language models in structure-aware workflows, particularly in agentic pipelines, DataOps, and system design with governance.

We expect to attract researchers and practitioners from NLP, data management, and visualization communities, and to foster collaboration across academia and industry.

2 Invited Speakers

Lisa Amini ↗ IBM Research, USA (confirmed)
Dr. Lisa Amini leads IBM's Data & AI Platforms Research efforts globally with the mission to infuse generative and agentic AI throughout IBM's Data Platform, to make it more intelligent, self-service, and autonomous, and to optimize its performance on AI workloads. Her experience on enterprise platforms for data management can provide significant insights into what is required to operationalize agentic AI for large-scale, reliable, and secure data workflows in real-world enterprise environments. This could shape future research directions on structured understanding, retrieval and generation.

Dan Roth ↗ Univ. of Pennsylvania and Oracle AI, USA (confirmed)

Dr. Dan Roth is the Eduardo D. Glandt Distinguished Professor at the University of Pennsylvania and Chief AI Scientist at Oracle. He is a Fellow of the AAAS, ACM, AAAI, and ACL, and a recipient of the IJCAI John McCarthy Award "for major conceptual and theoretical advances in the modeling of natural language understanding, machine learning, and reasoning." His pioneering work on learning and reasoning over structured representations can guide the development of principled methods and evaluation frameworks.

Paolo Papotti ↗ EURECOM, France (confirmed)

Dr. Paolo Papotti is an associate professor in the Data Science department at EURECOM. His research is in the broad areas of scalable data management and NLP, with a focus on data integration and information quality. He has authored more than 140 publications, and his work has been recognized with two "Best of the Conference" citations, three best demo award, and two Google Faculty Research Award (2016, 2020). Paolo's recent work of improving Text2SQL with reasoning models would be a great addition to our program.

William Yang Wang ↗ Univ. of Santa Barbara, USA (confirmed)

Dr. William Wang is the Duncan and Suzanne Mellichamp Professor at UCSB. He has published more than 100 papers at leading NLP/AI/ML conferences and journals, and received best paper awards (or nominations) at ASRU 2013, CIKM 2013, EMNLP 2015, and CVPR 2019. His expertise in responsible and scalable AI research can

help shape the workshop's agenda toward building trustworthy and impactful systems for structured understanding, retrieval, and generation.

3 Workshop Size / Prior Events

This is a new workshop, proposed for the first time. However, we can get an estimate of the number of attendees from a similar workshop held last year at ACL, [Table Representation Learning](#). We expect > 40 papers and ≈ 200 attendees. We propose a **one-day** workshop with 4–5 invited talks, contributed oral and poster presentations, and a panel discussion. We will accept short and long papers, new and ARR submissions, on an OpenReview portal, and are not proposing any new shared task for the first iteration of the workshop.

Preferred Venue: The preferred venue is ACL due to location and logistics constraints.

4 Diversity and Inclusion

The workshop brings together a wide range of research topics from multiple communities, including NLP, databases, information retrieval, and computer vision. This breadth of expertise, combined with a shared focus on structured data, inherently fosters diversity in perspectives, approaches, and applications. However, we have made an effort to ensure inclusion of diverse representations in all aspects of the program.

The **organizers** represent a diversity of gender, seniority, and industry as well as academia, all the while representing strong expertise in the subject matter of the workshop. The **invited speakers** represent a diversity of research expertise (including NLP, Data Management, Knowledge Representation, and Planning), seniority, and industry/academia, all the while being active thinkers in the field of structured understanding, retrieval, and generation.

For building community and foster collaboration across communities, we will organize a student mentoring lunch session, where students can enroll to have lunch with a mentor. Finally, we will make our best attempt to encourage participation from marginalized groups and diverse affinity groups by sending our CFPs to as many channels as possible.

5 Workshop Organizers

Vivek Gupta ↗, Assistant Professor, Arizona State University (vgupt140@asu.edu)

Vivek Gupta is an Assistant Professor of Computer Science at Arizona State University, where he leads the CoRAL Lab on evidence-first, agentic AI for structured and multimodal data (tables, charts, maps). He is an honored recipient of MSR and Bloomberg fellowships and best-paper awards. He co-organized NAACL SRW and ICDM RARA, serves as AC/Senior PC, and for SURGE 2026 will help lead program design, invited talks/panel, and industry-academia outreach across NLP/IR/DB/vis.

Harsha Kokel , Research Scientist, IBM Research (harsha.kokel@ibm.com)

Harsha Kokel's research focuses on planning and reasoning with LLMs for structured data and sequential decision making. She completed her Ph.D. at The University of Texas at Dallas where her doctoral dissertation received the 2024 David Daniel Thesis Award. Previously, she has co-organized [PRL workshops](#) at IJCAI, AAAI, and ICAPS, and has also organized [PLAN-FM Bridge at AAAI](#).

Peter Chen , PhD student, MIT (peterbc@mit.edu)

Peter Chen's research aims to blend the strengths of data systems and LLM paradigms. Specifically, his work explores how fundamental optimization strategies from data systems can be applied to improve LLM-based retrieval and reasoning. His work has appeared at ACL, EMNLP, VLDB, and OSDI, including an Outstanding Paper award at the ACL KnowledgeLM workshop.

Amit Agarwal , Oracle Inc (amit.h.agarwal@oracle.com)

Amit specializes in unstructured/multilingual retrieval, Agentic RAG, & source-linked generation evaluation. His work spans IR, NLP, & CV on multilingual, multimodal agentic systems, with publications at ACL, NAACL, EMNLP, & ICCV. He has led cross-domain teams, mentored LatinX & African ML communities, organized workshops and hackathons at UT Austin and MIT.

Michael Glass , Research Scientist, IBM Research (mrglass@us.ibm.com)

Michael Glass focuses on natural language interfaces to structured data, particularly Text-to-SQL. He completed his Ph.D. at the University of Texas at Austin (2012). His recent accomplishments include leading the development of the ExSL+Granite system, which achieved top perfor-

mance on the BIRD benchmark. He has published and reviewed for a number of conferences and workshops, especially ACL, EMNLP, IJCAI and AAAI.

Kavitha Srinivas , Senior Research Scientist, IBM Research (kavitha.srinivas@ibm.com)

Dr. Kavitha Srinivas is a Research Staff Member at IBM Research, where she has made foundational contributions in tabular data discovery, automated planning, and semantic data management. She has worked in many different roles from a tenured professor in cognitive science to a manager at IBM Research, and a CTO in a startup targeting data integration. She has previously organized various workshop including [OAEI 2024](#), [AI for Math Reasoning](#), [SemTab](#) etc.

Oktie Hassanzadeh , Senior Research Scientist, IBM Research (hassanzadeh@us.ibm.com)

Oktie Hassanzadeh's research expertise lies in managing and exploiting structured data in data lakes, with a focus on developing LLM-based solutions for data discovery, semantic understanding, and natural language querying of complex enterprise datasets. He has successfully co-organized several prominent workshops, including the [Tabular Data Analysis \(TaDA\)](#) workshop at VLDB and [Ontology Matching \(OM\)](#) workshop.

Shuaichen Chang , Applied Scientist, AWS AI Lab (cshuaich@amazon.com)

Shuaichen Chen obtained his Ph.D. in Computer Science from The Ohio State University before joining AWS. His research focuses on natural language interfaces to heterogeneous data, including accessing unstructured corpora and relational databases through natural language queries, the robustness of tabular models, and the intersection of vision and language for structured data. Shuaichen has organized the 4th [Table Representation Learning workshop at ACL 2025](#), and served on the program committees for several conferences, including ACL, EMNLP, NAACL, COLING, NeurIPS, ICML, ICLR, and AAAI.

6 Program Committee

Here is an initial reviewer pool, which we can expand if necessary (based on COIs and availability). We will ensure each paper gets atleast three reviews.

Aamod Khatiwada (Microsoft) , Carsten Binnig (TU Darmstadt) , Jiaoyan Chen (University

of Manchester) , Vasilis Efthymiou (Harokopio University of Athens) , Sainyam Galhotra (Cornell University) , Katja Hose (TU Wien) , Karan Dua (Oracle) , Madelon Hulsebos (CWI Amsterdam) , Junkyu Lee (IBM) , Paolo Papotti (EURECOM) , Nhan Pham (IBM) , Srikant Panda (Oracle) , Davood Rafei (Univeristy of Alberta) , Fatma Ozcan (Google) , Horst Samulowitz (IBM) , Immanuel Trummer (Cornell University) , Mayukh Das (Microsoft Research) , Vraj Shah (IBM) , Hitesh Patel (Oracle) , Allaa Boutaleb (Sorbonne University) , Udayan Khurana (IBM) , Gaetano Rossiello (IBM) , Sarthak Dash (IBM) , Nandana Mihindukulasooriya (IBM) , Erkang Zhu (Microsoft Research) , Ernesto Jiménez-Ruiz (City, University of London) , Liane Vogel (TU Darmstadt) , Hansa Meghwani (Oracle) , Micah Goldblum (New York University) , Andreas Muller (Microsoft) , Xi Fang (Yale University) , Naihao Deng (University of Michigan) , Weijie Xu (Amazon) , Rajat Agarwal (Amazon) , Sharad Chitlangia (Amazon) , Lei Cao (University of Arizona) , Jaehyun Nam (KAIST) , Sascha Marton (University of Mannheim) , Tianji Cong (University of Michigan) , Aneta Koleva (University of Munich) , Gerardo Vitagliano (MIT) , Reynold Cheng (University of Hong Kong) , Till Döhmen (Mother-Duck / University of Amsterdam) , Ivan Rubachev (Higher School of Economics) , Raul Castro Fernandez (University of Chicago) , Peng Shi (Amazon) , Tianyang Liu (University of California, San Diego) , Tianbao Xie (University of Hong Kong) , Jintai Chen (University of Illinois at Urbana-Champaign) , Sebastian Bordt (Eberhard-Karls-Universität Tübingen) , Noah Hollmann (Albert-Ludwigs-Universität Freiburg) , Linyong Nan (Yale University) , Roman Levin (Amazon) , Gyubok Lee (KAIST) , Artem Babenko (Yandex)

References

- Naman Ahuja, Fenil Denish Bardoliya, Chitta Baral, and Vivek Gupta. 2025. Map&make: Schema guided text to table generation. In *ACL (1)*, pages 30249–30262. Association for Computational Linguistics.
- Shuaichen Chang, Jun Wang, Mingwen Dong, Lin Pan, Henghui Zhu, Alexander Hanbo Li, Wuwei Lan, Sheng Zhang, Jiarong Jiang, Joseph Lilien, et al. Dr. spider: A diagnostic evaluation benchmark towards text-to-sql robustness. In *The Eleventh International Conference on Learning Representations*.
- Peter Baile Chen, Yi Zhang, and Dan Roth. 2024. Is table retrieval a solved problem? exploring join-aware multi-table retrieval. In *ACL (1)*, pages 2687–2699. Association for Computational Linguistics.
- Wenhu Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyou Zhou, and William Yang Wang. 2020. Tabfact: A large-scale dataset for table-based fact verification. In *ICLR*. OpenReview.net.
- Michael R. Glass, Mustafa Eyceoz, Dharmashankar Subramanian, Gaetano Rossiello, Long Vu, and Alfio Gliozzo. 2025. Extractive schema linking for text-to-sql. *CoRR*, abs/2501.17174.
- Parag Jain, Andreea Marzoca, and Francesco Piccinno. 2024. STRUCTSUM generation for faster text comprehension. In *ACL (1)*, pages 7876–7896. Association for Computational Linguistics.
- Harsha Kokel, Aamod Khatiwada, Tejaswini Pedapati, Oktie Hassanzadeh, Horst Samulowitz, and Kavitha Srinivas. Topjoin: A context-aware multi-criteria approach for joinable column search. *VLDB 2025 Workshop: Tabular Data Analysis (TaDA)*, 2150:8097.
- Fangyu Lei, Jixuan Chen, Yuxiao Ye, Ruisheng Cao, Dongchan Shin, Hongjin SU, ZHAOQING SUO, Hongcheng Gao, Wenjing Hu, Pengcheng Yin, et al. Spider 2.0: Evaluating language models on real-world enterprise text-to-sql workflows. In *The Thirteenth International Conference on Learning Representations*.
- Haoyang Li, Jing Zhang, Cuiping Li, and Hong Chen. 2023a. RESDSL: decoupling schema linking and skeleton parsing for text-to-sql. In *AAAI*, pages 13067–13075. AAAI Press.
- Jinyang Li, Binyuan Hui, Ge Qu, Jiaxi Yang, Binhua Li, Bowen Li, Bailin Wang, Bowen Qin, Ruiying Geng, Nan Huo, Xuanhe Zhou, Chenhao Ma, Guoliang Li, Kevin Chen-Chuan Chang, Fei Huang, Reynold Cheng, and Yongbin Li. 2023b. Can LLM already serve as A database interface? A big bench for large-scale database grounded text-to-sqls. In *NeurIPS*.
- Xinyuan Lu, Liangming Pan, Qian Liu, Preslav Nakov, and Min-Yen Kan. 2023. SCITAB: A challenging benchmark for compositional reasoning and claim verification on scientific tables. In *EMNLP*, pages 7787–7813. Association for Computational Linguistics.
- Puneet Mathur, Alexa Siu, Nedim Lipka, and Tong Sun. 2024. **MATSA: Multi-agent table structure attribution**. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 250–258, Miami, Florida, USA. Association for Computational Linguistics.
- Ivan A Poddubnyi and Nikita O Dorodnykh. Query plan generation for table question answering. *Proceedings of the VLDB Endowment. ISSN*, 2150:8097.
- Oleg Somov and Elena Tutubalina. 2025. Confidence estimation for error detection in text-to-sql systems. In *AAAI*, pages 25137–25145. AAAI Press.
- Songheng Zhang, Lei Wang, Toby Jia-Jun Li, Qiaomo Shen, Yixin Cao, and Yong Wang. 2024. Chartifytext: Automated chart generation from data-involved texts via LLM. *CoRR*, abs/2410.14331.