TRL 2025

The 4th Table Representation Learning Workshop at ACL 2025

Proceedings of the Workshop

July 31, 2025

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

©2025 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 979-8-89176-268-8

Introduction

We are excited to welcome you to TRL 2025, the Workshop on Table Representation Learning, held in conjunction with ACL 2025. This year, the workshop takes place on July 31st, 2025, in Vienna, Austria, and brings together researchers working on all aspects of modeling, understanding, and reasoning over tabular data.

TRL serves as a forum for recent advances in table representation learning, spanning a wide range of topics including pretraining and foundation models for tables, table-based retrieval and question answering, table-to-text generation, semantic parsing, and applications in both textual and multimodal settings.

This year, we received 30 submissions, reflecting the growing interest in this area. Following a thorough review process, 24 papers were accepted for presentation, including 6 selected for oral presentations. The overall acceptance rate was 80%. The accepted papers represent both the breadth and depth of current research, and we are excited to showcase them during the workshop.

We are honored to feature invited talks by Edward Choi, Ruoxi Sun, Tao Yu, and Jiani Zhang, whose insights help frame the ongoing challenges and opportunities in table representation learning. We thank them for accepting our invitation and for their generous contributions to the program.

The workshop program includes oral presentations, a poster session, and invited talks. The oral sessions are organized thematically to highlight emerging directions and shared challenges across subfields. We hope these discussions foster new connections and inspire future research.

We thank all authors for their contributions and all attendees for their participation. We are especially grateful to our Program Committee for their thoughtful and timely reviews—their efforts were essential in shaping a strong and balanced program.

Finally, we thank our sponsors, SAP and Snowflake, for their generous support in making TRL 2025 possible.

We hope you enjoy the workshop!

Madelon Hulsebos, Qian Liu, Shuaichen Chang, Wenhu Chen, Huan Sun Organizers of TRL 2025

Organizing Committee

Organizers

Madelon Hulsebos, Centrum voor Wiskunde en Informatica
Qian Liu, Tiktok
Shuaichen Chang, AWS AI Lab
Wenhu Chen, University of Waterloo
Huan Sun, The Ohio State University

Program Committee

Reviewers

Rajat Agarwal, Amazon
Simran Arora, Stanford University, University of Pennsylvania, University of Pennsylvania and The Wharton School, University of Pennsylvania
Sebastian Bordt, Eberhard-Karls-Universität Tübingen
Shuaichen Chang, AWS AI Lab
Jiajing Chen, New York University
Sharad Chitlangia, Amazon
Naihao Deng, University of Michigan
Yuntao Du, Purdue University
Till Döhmen, University of Amsterdam and Rheinisch Westfälische Technische Hochschule Aachen
Katharina Eggensperger, Eberhard-Karls-Universität Tübingen
Moonjung Eo, LG AI Research
Matthias Feurer, Ludwig Maximilian University of Munich
Yury Gorishniy, Moscow Institute of Physics and Technology and Yandex
Yiqun Hu, AWS AI Labs
Zezhou Huang, Columbia University
Mouxiao Huang, Huawei Technologies Ltd.
Xiangjian Jiang, University of Cambridge
Myung Jun Kim, Inria
Andreas Kipf, University of Technology Nuremberg
Aneta Koleva, Siemens
Gaurav Kumar, Moveworks
Kyungeun Lee, LG AI Research
Gyubok Lee, Korea Advanced Institute of Science and Technology
Alexander Hanbo Li, Amazon
Tianyang Liu, University of California, San Diego
Xinyuan Lu, national university of singaore, National University of Singapore
Sazan Mahbub, Carnegie Mellon University
Sascha Marton, Technical University of Clausthal
Amine Mhedhbi, École Polytechnique de Montréal, Université de Montréal
Akshata Kishore Moharir, Microsoft
Mira Moukheiber, Massachusetts Institute of Technology
Andreas C Mueller, Microsoft
Simone Papicchio, Politecnico di Torino
Paolo Papotti, Eurecom
Panupong Pasupat, Google
Yuxin Qiao, Northern Arizona University
Ivan Rubachev, Yandex Research
Maximilian Schambach, SAP
Sebastian Schelter, BIFOLD & TU Berlin
Ananya Singha, Research, Microsoft
Gerardo Vitagliano, Computer Science and Artificial Intelligence Laboratory, Electrical Engineering & Computer Science
Liane Vogel, Technische Universität Darmstadt
Tianshu Wang, Chinese Academy of Sciences

Katarzyna Woźnica, Warsaw University of Technology
Jiani Zhang, Google
Tianshu Zhang, The Ohio State University
Ye Zhang, University of Pittsburgh
Zecheng Zhang, Kumo.AI
Fuheng Zhao, University of California, Santa Barbara
Mingyu Zheng, University of Chinese Academy of Sciences
Shuhan Zheng, Hitachi, Ltd.
Mengyu Zhou, Microsoft Research
Fengbin Zhu, National University of Singapore

Invited Speakers

Edward Choi, KAIST
Ruoxi Sun, Google
Tao Yu, HKU
Jiani Zhang, Google

Table of Contents

Theme-Explanation Structure for Table Summarization using Large Language Models: A Case Study on Korean Tabular Data
TaeYoon Kwack, Jisoo Kim, Ki Yong Jung, DongGeon Lee and Heesun Park 1

Generating Synthetic Relational Tabular Data via Structural Causal Models
Frederik Hoppe, Astrid Franz, Lars Kleinemeier and Udo Göbel 13

Tables as Thought: Exploring Structured Thoughts in LLM Reasoning
Zhenjie Sun, Naihao Deng, Haofei Yu and Jiaxuan You 19

R^3: This is My SQL, Are You With Me?A Consensus-Based Multi-Agent System for Text-to-SQL Tasks
Hanchen Xia, Feng Jiang, Naihao Deng, Cunxiang Wang, Guojiang Zhao, Rada Mihalcea and Yue Zhang 34

SQLong: Enhanced NL2SQL for Longer Contexts with LLMs
Dai Quoc Nguyen, Cong Duy Vu Hoang, Duy Quang Vu, Gioacchino Tangari, Thanh Vu, Don Dharmasiri, Yuan-Fang Li and Long Duong 47

iTBLS: A Dataset of Interactive Conversations Over Tabular Information
Anirudh Sundar, Christopher Gordon Richardson, Larry Heck and Adar Avsian 56

Something's Fishy in the Data Lake: A Critical Re-evaluation of Table Union Search Benchmarks
Allaa Boutaleb, Bernd Amann, Hubert Naacke and Rafael Angarita 71

RITT: A Retrieval-Assisted Framework with Image and Text Table Representations for Table Question Answering
Wei Zhou, Mohsen Mesgar, Heike Adel and Annemarie Friedrich 86

Ask Me Like I'm Human: LLM-based Evaluation with For-Human Instructions Correlates Better with Human Evaluations than Human Judges
Rudali Huidrom and Anya Belz 98

Table Understanding and (Multimodal) LLMs: A Cross-Domain Case Study on Scientific vs. Non-Scientific Data
Ekaterina Borisova, Fabio Barth, Nils Feldhus, Raia Abu Ahmad, Malte Ostendorff, Pedro Ortiz Suarez, Georg Rehm and Sebastian Möller 109

Perspective: Leveraging Domain Knowledge for Tabular Machine Learning in the Medical Domain
Arijana Bohr, Thomas Altstidl, Bjoern Eskofier and Emmanuelle Salin 143

LLM-Mixer: Multiscale Mixing in LLMs for Time Series Forecasting
Md Kowsher, Md. Shohanur Islam Sobuj, Nusrat Jahan Prottasha, E. Alejandro Alanis, Ozlem Garibay and Niloofar Yousefi 156

TableKV: KV Cache Compression for In-Context Table Processing
Giulio Corallo, Elia Faure-Rolland, Miriam Lamari and Paolo Papotti 166

OrQA – Open Data Retrieval for Question Answering dataset generation
Giovanni Malaguti, Angelo Mozzillo and Giovanni Simonini 172

In-Context Learning of Soft Nearest Neighbor Classifiers for Intelligible Tabular Machine Learning
Mykhailo Koshil, Matthias Feurer and Katharina Eggensperger 182

Retrieval-Augmented Forecasting with Tabular Time Series Data
Zichao Li 192

Resolution-Alignment-Completion of Tabular Electronic Health Records via Meta-Path Generative Sampling
S Mehryar 200

Embeddings for Numerical Features Using tanh Activation
Bingyan Liu, Charles Elkan and Anil N. Hirani 208

Improving Table Retrieval with Question Generation from Partial Tables
Hsing-Ping Liang, Che-Wei Chang and Yao-Chung Fan 217

Sparks of Tabular Reasoning via Text2SQL Reinforcement Learning
Josefa Lia Stoisser, Marc Boubnovski Martell and Julien Fauqueur 229

How well do LLMs reason over tabular data, really?
Cornelius Wolff and Madelon Hulsebos 241