Challenges and Opportunities of Table Representation Learning

Carsten Binnig*[1], Julian Martin Eisenschlos*[2], Madelon Hulsebos*[3], and Frank Hutter*[4]

1   TU Darmstadt, DE. `carsten.binnig@cs.tu-darmstadt.de`
2   Google Research – Zürich, CH. `eisenjulian@gmail.com`
3   CWI – Amsterdam, NL. `madelonhulsebos@gmail.com`
4   Prior Labs – Freiburg, DE & ELLIS Institute Tübingen, DE & Universität Freiburg, DE. `fh@cs.uni-freiburg.de`

──── **Abstract** ────

The growing volume and importance of structured data have sparked increasing interest in Table Representation Learning (TRL), an emerging field that leverages neural models to learn abstract, general-purpose representations for tabular data to support a wide range of downstream tasks such as tabular prediction, table question answering, tabular data cleaning, and many more. This seminar gathered the different communities (ML, NLP, IR, DB) who work on this topic to discuss the challenges & long-term vision of this field.

From the organizers: Carsten Binnig, Julian Eisenschlos, Madelon Hulsebos, Frank Hutter.

1   Executive Summary

Carsten Binnig (TU Darmstadt, DE)
Julian Martin Eisenschlos (Google Research – Zürich, CH)
Madelon Hulsebos (CWI – Amsterdam, NL)
Frank Hutter (Prior Labs – Freiburg, DE & ELLIS Institute Tübingen, DE & Universität Freiburg, DE)

The Dagstuhl Seminar 25182, held from April 27 to May 2, 2025, brought together researchers from machine learning, natural language processing, information retrieval, and databases to discuss the challenges and vision for Table Representation Learning (TRL). As structured data continues to grow in volume and importance, TRL aims to build representations that enable downstream tasks such as prediction, question answering, data preparation. The seminar served as a forum to share long-term visions, highlight challenges, and discuss research directions that bridge across these diverse communities.

────────

* Editor / Organizer

We opened the seminar with a series of opinionated tutorials that laid out the research landscape. Carsten Binnig, taking a database systems perspective, argued that TRL could help eliminate what he termed the "data tax," "query tax," and "tuning tax" that currently burden users of relational databases. By automating tasks such as query authoring, data cleaning, and performance optimization, TRL could significantly reduce entry barriers to database use. Julian Eisenschlos followed with an NLP-centered overview of table question answering. He reviewed benchmarks and encoding strategies, emphasizing the trade-offs between interpretability and computational cost. He highlighted how pre-training tasks and generation-based understanding link table reasoning with broader modalities such as charts and infographics, while also pointing to the challenges of training models in a post-LLM landscape. Madelon Hulsebos shifted the focus to table semantics, stressing that much of TRL happens "before generating insights" She argued that understanding table- and column-relationships is foundational for data preparation, search and retrieval, and predictive modeling. Her talk questioned whether billion-parameter general-purpose models are necessary for tabular tasks, or whether modular, specialized systems might be more effective. Finally, Frank Hutter presented recent advances in deep learning for tabular prediction. After reviewing early methods, he discussed TabPFN and its extensions, which aim to overcome previous limitations. His talk underscored TRL's growing ability to rival or surpass traditional tabular learning methods while also pointing out open challenges in scaling, context integration, and generalization.

The opinionated tutorials were complemented by shorter impulse talks that addressed specific aspects of TRL. Paolo Papotti reviewed transformer-based adaptations for tabular data, covering innovations in inputs, internals, outputs, and pretraining. Gaël Varoquaux analyzed architectural challenges in building table foundation models, particularly around heterogeneity and invariances. Michael Cochez argued for tighter links between graph learning and TRL to handle complex relations and incomplete data. Xue Effy Li emphasized that meaningful table representation requires integrating contextual information such as metadata, documentation, and world knowledge. Gerardo Vitagliano demonstrated the promise of multimodal pipelines that integrate genomic, imaging, textual, and tabular data for scientific discovery, while exposing current limitations. Andreas Müller shared a vision for deep integration of human feedback, LLMs, databases and table foundation models into agentic systems. Shuaichen Chang discussed the challenges of deploying Text-to-SQL in real-world settings, where ambiguity, noise, and mixed data require robust solutions. Finally, Fatma Özcan showed how long-context reasoning and multi-agent systems can improve Text-to-SQL robustness.

After the talks, we divided into working groups to explore research problems in more depth. The Multi-Modal Data Analysis group examined how to query and process data spanning text, images, genomics, and audio, emphasizing the need for new operators, adaptive indexing, and cost-aware query planning. The Predictive ML and Context group discussed how to integrate metadata, external knowledge, and domain expertise into statistical tabular prediction, proposing hybrid architectures that combine foundation models with agentic systems. The Conversational Analytics group envisioned natural language interfaces that go beyond text-to-SQL by supporting explanations, causal reasoning, and iterative dialogue with human oversight. The Architectures for Table Foundation Models group debated whether a universal foundation model for tables is achievable, weighing trade-offs between adaptability, semantic grounding, and efficiency.

Reflecting on the seminar's discussions, we identified a few cross-cutting themes that seem promising for future research. First, context, whether in the form of metadata, domain knowledge, or multimodal signals, was consistently identified as crucial for making TRL robust and useful. Second, the limitations of current benchmarks remain a bottleneck, as they fail to capture the complexity of real-world tabular reasoning. Finally, participants questioned the pursuit of monolithic "one-size-fits-all" tabular models, favoring instead modular or hybrid systems that can flexibly combine database principles, machine learning, and human expertise. We concluded the seminar with a shared recognition that while TRL has achieved significant progress, its long-term promise lies in deeply integrating methods across disciplines to build more adaptive, interpretable, and context-aware tabular intelligence.

2 Table of Contents

3   Overview of Talks

3.1   An opinionated talk on TRL from a database lens OR Rethinking databases in the Age of LLMs/AI

Carsten Binnig (TU Darmstadt, DE)

Relational databases have been a success story since their incarnation in the 1970s. However, using these systems has extreme high overhead for users. First, extracting data from potentially unstructured or unclear sources and transform those in clean structured schema cause high overheads (= data tax). Moreover, authoring queries using SQL has similarly high overheads as queries (query tax) involve many tables and SQL requires from users to be highly precise in how tables are combined, which attributes are queries, etc.

Finally, relational databases require high overhead of managed tuning to provide high performance (tuning tax).

In this talk, I explore the opportunities of TRL to tackle these overheads and help users with query authoring (text-to-SQL), data engineering (e.g. automated cleaning) and learned tuning.

3.2   Impulse Talk: Text-to-SQL and Table Analysis in the Wild.

Shuaichen Chang (Amazon Web Services – New York, US)

In this talk, we explore the challenges faced by Text-to-SQL systems in real-world applications. These systems must operate effectively over complex and large-scale databases and abstract tasks.

We highlight key issues such as robustness to semantically equivalent variations in databases and user query, and the need to handle noisy user queries, including ambiguous and unanswerable questions.

Additionally, we discuss a particularly realistic setting where structured tables are intertwined with unstructured text, requiring integrated reasoning across both data types for effective data analysis. Crucially, there is no one-size-fit-all solution in such setting. Effective system design must be tailored to the data characteristics.

Looking forward, we envision that AI agent automatically build such systems based on specific data and task.

3.3   Impulse talk: Relational learning and reasoning for table representation leaning

Michael Cochez (VU Amsterdam, NL)

When data becomes complex, including complex relations exist between entities, it becomes more natural to represent as a graph, Besides, in the graph representation learning field, people have experience with constraints. Tasks in this field include attribute and missing link prediction, but also more complex query answering. Especially in cases where the graph is incomplete, which in real world settings is a ways the case, this is interesting. In this setting, we also answer questions like "Give me all scientists that have worked A in 3 or more universities, in at least 2 countries", which requires joins, counting, iterating, and aggregation. They have also limitations regarding the quality of benchmarks and scalability of methods. Overall, I want to see more collaborations between the graph and the table so fields.

3.4   An Opinionated Overview of Deep Learning for Table Question Answering: An NLP Perspective

Julian Martin Eisenschlos (Google Research – Zürich, CH)

Through this talk I will define visual language and TR utility, with a focus on table understanding. I will introduce the main benchmarks and the ways the community has tackled them over the last couple of years. We'll deep dive into the way tables are encoded through embedding, transfer, and the semantics of the answer trading off interpretability and cost of acquisition. We will also touch on the more trend pre-training tasks: how generation-driven understanding and how understanding table helps in chart and infographic understanding. To close, we will highlight what are the training challenges in a world post LLMs.

3.5   An opinionated talk on Table Representation Learning: on everything before you get the insight

Madelon Hulsebos (CWI – Amsterdam, NL)

Getting insights from structured data comprises many subsequent steps. In this talk I reflect on a critical construct that enables downstream insights coming from table semantics. Particular column semantics, regarding a single column, relationship across columns and tables, are instrumental for data integration, cleaning, search, as well as question answering and predictive modeling.

Reflecting on the performance of small and specialized models versus billion-parameter models for type prediction raises the general question whether we need high-capacity general-purpose tabular models for every task. Do we need to work towards these models, or think more about compound systems (with mixtures, modularity, etc.) for tabular tasks?

The talk continues with a building block for advancing TRL: data & metrics to measure realistic performance on real-world tasks & data. What are the right datasets, tasks, metrics to push TRL further? Early insights show that tabular analytical reasoning capabilities of LLMs and specialized tabular models are far behind.

I conclude with the shared vision of end-to-end tabular analysis: should this be multi-agent systems that abstract humans away? No. We need intersection with domain expertise. But it is a dot on the horizon...

3.6 An Opinionated Overview of Deep Learning for Tabular Prediction

Frank Hutter (Prior Labs – Freiburg, DE & ELLIS Institute Tübingen, DE & Universität Freiburg, DE)

In this talk, I will motivate the tabular prediction problem to the TRL community and discuss the state of deep learning for solving it. Specifically, I will briefly discuss early tabular deep learning methods, explain TabPFN in some detail, and then focus on the many extensions that have been introduced to move beyond the limitations of TabPFN.

3.7 Impulse Talk: Beyond Tables: Context Matters

Xue Li (CWI – Amsterdam, NL)

Tabular Data rarely exist in isolation – they are embedded in broader ecosysten of document-ations, meta data, domain knowledge, and even world knowledge. In this talk, I suggest that meaningful table representation requires incorporating global context for downstream tasks such as TableDA, Text. etc. Global context can include source documentation, organizing organizational processes, and even common-sense or parametric knowledge stored in large Language models. I highlight recent approaches including RAG, Graph RAG. and prompting – based techniques for retrieving and integrating external context, and discuss why these solutions are far from sufficient.. I outline key technical challenger, meta- questions about context retrieval and utilization, and opportunities towards more holistic table understanding.

3.8 Impulse Talk: Transformer-based Table Representation Learning

Paolo Papotti (EURECOM – Biot, FR)

In the last few years, the NLP community witnessed advances in neural representations of free-form text with transformer-based architectures for producing language models (LMs). Given the importance of knowledge in relational tables, recent efforts extended LMs by developing neural representations for tabular data. In this talk, I present these proposals covering extensions to the original transformer in terms of input, internals (such as attention), output, and pre-training task.

3.9   Impulse Talk: Architectures for tabular learning

Gaël Varoquaux (INRIA Saclay – Île-de-France, FR)

If we want to solve tabular learning problems, what are the constraints and advances in architectures for table foundation models? We want to construct representations that can be reused to facilitate learning and/or transfer learning. For this goal, we need to think about representations of tables and how to represent mixed data types, as well as how to learn with these representations.

3.10   Impulse talk: Beyond tables, Multimodal pipelines

Gerardo Vitagliano (MIT – Cambridge, US)

Real-world data science is often complex and insights or outputs derive from data collected in a wide variety of modalities. In this talk, we present an example of insightful science based on a multimodal research scenario: how do researchers find and characterize tumor progression using cancer patient's big clinical records?

We discuss how to process this information, recalling process data in a variety of modalities: genomic, imaging (MRI included), tabular, and free text formats. Through several examples of data processing, we outline the limitations of state-of-the-art multimodal processing systems, as well as processing approximations (approximate query processing, data summarization, embeddings, indexing).

We outline three macro-areas of research: multimodal embeddings, retrieval strategies, integration and processing joint modalities.

Within these systems we envision all of these components will play a fundamental role to enable complex pipelines and integrative analysis of multimodal data. State-of-the-art AI models process data of various modalities, including text, image and sound. However, the context size of these models is severely limited, making it impossible to directly apply them to large data collections with millions of entries. Even if it is possible to directly apply models to large data collections, it is often prohibitively expensive.

Our goal was to create data processing systems that scale up multimodal data processing to large data collections, leveraging techniques such as approximate processing (processing a carefully selected data subset to obtain an approximated result), caching (reusing results from prior queries to answer new queries more efficiently), and compression. Compressing data to make processing with AI models cheaper. In addition, we plan to create a benchmark that allows us to evaluate such systems for multimodal data processing, according to metrics such as result quality and monetary processing fees.

3.11 Impulse Talk: Text-to-SQL – Long Context

Fatma Özcan (Google – San Jose, US)

Text-to-SQL is challenging in that a natural language question is inherently ambiguous, while SQL generation requires a precise understanding of complex data schema and semantics. One solution is to provide sufficient contextual information. In this talk, I present a detailed study on the performance and the latency trade-offs long context offered by Google's Gemini model, showing the impact of various contextual information, including example column values, user-provided hints, in-context examples, SQL documentation, and relevant schema. I then describe CHASE-SQL, a novel multi-agent solution that uses LLMs to generate diverse SQL candidates using three different approaches, choosing the final answer by a selection agent.

4 Working groups

4.1 Working group: Multi-Modal Data Analysis

Michael Cochez (VU Amsterdam, NL), Tianji Cong (University of Michigan – Ann Arbor, US), Andreas Kipf (TU Nürnberg, DE), Olga Ovcharenko (TU Berlin, DE), Fatma Özcan (Google – San Jose, US), Shivam Sharma (TU Darmstadt, DE), Immanuel Trummer (Cornell University – Ithaca, US), and Gerardo Vitagliano (MIT – Cambridge, US)

Multi-modal data analysis addresses the challenge of querying and processing data across diverse modalities – numbers, text, images, audio, video, graphs, genomics, and time series – often within relational tables. Multimodality is not simply a matter of semantic variation; it requires system-level design to handle distinct representations and access patterns. Modern representation learning allows late fusion, reducing the need for early manual integration.

A multi-modal database system must address storage, querying, and processing. Storage may involve tables with references to external modalities, adaptive indexing, and multimodal database cracking. Querying may take the form of SQL, natural language, or operator pipelines, and requires specialized operations such as cross-modal joins, transformations (for example, speech-to-text), and multimodal search. Because multimodal analysis is computationally expensive, optimization strategies are essential: approximations, pruning redundant modalities, batching, caching, and cost-aware query planning.

Applications span e-commerce recommendations, bioacoustics, healthcare analytics, multimedia search, and scientific domains. Benchmarks are crucial, ideally combining real datasets with diverse query types and measuring not only accuracy but also cost, latency, and scalability. The overarching insight is that multimodal relational learning is achievable, but only with new operators, optimization techniques, and dedicated benchmarks that acknowledge the trade-offs between cost and accuracy.

4.2   Working group: Predictive ML and Context

Frank Hutter (Prior Labs – Freiburg, DE & ELLIS Institute Tübingen, DE & Universität Freiburg, DE), Katharina Eggensperger (Universität Tübingen, DE), Myung Jun Kim (INRIA Saclay – Île-de-France, FR), Xue Li (CWI – Amsterdam, NL), Lennart Purucker (Universität Freiburg, DE), and Sebastian Schelter (TU Berlin, DE)

Predictive machine learning on tabular data raises the question of how best to incorporate context and world knowledge. Current approaches exist along a spectrum: at one end, large language models can be fine-tuned on company data and prompted with customer information; at the other, traditional pipelines engineer features from available sources and train standard supervised models. Both approaches show limitations – pre-trained models often cannot disambiguate column semantics, while purely tabular methods struggle without interaction and ignore implicit meanings. In-between approaches such as CARTE, TARTE, and TabPFN attempt to combine graph or probabilistic structures with tabular input.

Context encompasses many forms: metadata, additional text, provenance, external knowledge bases, and domain-specific constraints. Agents can act as junior data scientists, augmenting tables, searching for relevant data, and performing feature engineering. A hybrid vision emerges: a world model provides semantic grounding, a statistical foundation model delivers predictive accuracy, and agentic systems serve as the glue between them. While the long-term goal may be a unified end-to-end model, a modular design currently appears more viable. Benchmarks and interfaces remain underdeveloped, highlighting a research challenge in marrying contextual world knowledge with robust statistical learning.

4.3   Working group: Converstaional Analytics

Andreas Müller (Microsoft Corp. – Mountain View, US), Carsten Binnig (TU Darmstadt, DE), Shuaichen Chang (Amazon Web Services – New York, US), Madelon Hulsebos (CWI – Amsterdam, NL), and Anupam Sanghi (TU Darmstadt, DE)

Conversational analytics seeks to enable dialogue-driven data analysis in natural language, extending far beyond text-to-SQL translation. The aim is not only to retrieve results but also to provide explanations, reason about causes, and support human decision-making. A central question concerns whether systems should rely on explicit planning – through a formal algebra of operations – or on step-by-step reasoning by large models. Planning and orchestration resemble challenges in robotics, where reinforcement learning and tool management are key.

A conversational algebra can capture the steps of analysis: retrieving data, identifying granularity, selecting measures, reshaping data, choosing analytical methods, applying guardrails, and validating outcomes. Human involvement is essential for reviewing intermediate results, correcting assumptions, and guiding decisions, while automated components orchestrate tools and workflows. Risks include hallucinated results, invalid causal inferences, and misleading reasoning traces, underscoring the need for transparency and verifiability.

Existing benchmarks such as Spider 2.0 and DiscoveryBench offer starting points but fail to capture the full complexity of reasoning. Progress in this area requires richer benchmarks and formal abstractions, combined with human-in-the-loop oversight to ensure reliability and explainability.

4.4 Working group: Architectures for Table Foundation Models

Gaël Varoquaux (INRIA Saclay – Île-de-France, FR), Vadim Borisov (tabularis.ai – Tübingen, DE), Julian Martin Eisenschlos (Google Research – Zürich, CH), Floris Geerts (University of Antwerp, BE), Filip Gralinski (Snowflake – Warsaw, PL), Tassilo Klein (SAP SE – Walldorf, DE), Paolo Papotti (EURECOM – Biot, FR), and Liane Vogel (TU Darmstadt, DE)

Table foundation models (TFMs) sit at the intersection of multiple research traditions: statistical estimation in machine learning, relational structures in databases, and semantic grounding in NLP. The guiding question is whether a single general-purpose foundation model can cover all tasks involving tabular data, or whether domain-specific variants will remain necessary. Trade-offs span adaptability versus cost, the diversity of supported inputs, and the role of world knowledge alongside statistical learning.

Architectural choices extend beyond transformer-based approaches to include graph neural networks, symbolic reasoning components, and modules specialized for numbers and time series. Key issues include which invariances should be respected, such as row and column order, and how to scale computation for very large tables. Pretraining raises questions of data sourcing: synthetic datasets may capture structural patterns, while real and domain-specific data provide semantic grounding. Evaluation of TFMs requires a multidimensional benchmark framework covering adaptability, scalability, modality handling, semantic understanding, and efficiency. The vision is not to treat TFMs as scaled-up language models but as flexible, principled architectures that integrate statistical inference, semantic reasoning, and database knowledge.

Participants

- Carsten Binnig
  TU Darmstadt, DE
- Vadim Borisov
  tabularis.ai – Tübingen, DE
- Shuaichen Chang
  Amazon Web Services –
  New York, US
- Michael Cochez
  VU Amsterdam, NL
- Tianji Cong
  University of Michigan –
  Ann Arbor, US
- Katharina Eggensperger
  Universität Tübingen, DE
- Julian Martin Eisenschlos
  Google Research – Zürich, CH
- Floris Geerts
  University of Antwerp, BE
- Filip Gralinski
  Snowflake – Warsaw, PL

- Madelon Hulsebos
  CWI – Amsterdam, NL
- Frank Hutter
  Prior Labs – Freiburg, DE &
  ELLIS Institute Tübingen, DE &
  Universität Freiburg, DE
- Myung Jun Kim
  INRIA Saclay –
  Île-de-France, FR
- Andreas Kipf
  TU Nürnberg, DE
- Tassilo Klein
  SAP SE – Walldorf, DE
- Xue Li
  CWI – Amsterdam, NL
- Andreas Müller
  Microsoft Corp. –
  Mountain View, US
- Fatma Özcan
  Google – San Jose, US
- Olga Ovcharenko
  TU Berlin, DE

- Paolo Papotti
  EURECOM – Biot, FR
- Lennart Purucker
  Universität Freiburg, DE
- Anupam Sanghi
  TU Darmstadt, DE
- Sebastian Schelter
  TU Berlin, DE
- Shivam Sharma
  TU Darmstadt, DE
- Immanuel Trummer
  Cornell University – Ithaca, US
- Gaël Varoquaux
  INRIA Saclay –
  Île-de-France, FR
- Gerardo Vitagliano
  MIT – Cambridge, US
- Liane Vogel
  TU Darmstadt, DE