Panel on Neural Relational Data: Tabular Foundation Models, LLMs... or both?

Paolo Papotti
papotti@eurecom.fr
EURECOM

Carsten Binnig
carsten.binnig@cs.tu-darmstadt.de
TU Darmstadt

ABSTRACT

Recent breakthroughs in artificial intelligence have produced *Large Language Models* (LLMs) and a new wave of *Tabular Foundation Models* (TFMs). Both promise to redefine how we query, integrate, and reason over relational data, yet they embody opposing philosophies: LLMs pursue broad generality through massive text-centric pre-training, whereas TFMs embed inductive biases that mirror table structure and relational semantics. This panel assembles researchers and practitioners from academia and industry to debate which path, specialized TFMs, ever stronger general-purpose LLMs, or a hybrid of the two, will most effectively power the next generation of data management systems. Panelists will confront questions of generality, accuracy, scalability, robustness, cost, and usability across core data management tasks such as Text-to-SQL translation, schema understanding, and entity resolution. The discussion aims to surface critical research challenges and guide the community's investment of effort and resources over the coming years.

PVLDB Reference Format:
Paolo Papotti and Carsten Binnig. Panel on Neural Relational Data: Tabular Foundation Models, LLMs... or both?. PVLDB, 18(12): 5523 - 5525, 2025. doi:10.14778/3750601.3760519

1 PANEL DESCRIPTION

While traditional relational database systems have long provided robust and efficient mechanisms for transactional query processing and structured data retrieval through SQL, the advent of powerful foundation models is now driving a significant transformation in how we envision new ways to interact with, understand, and leverage relational data. This evolution opens up frontiers for data exploration, insight generation, and a broader range of data management tasks that go beyond conventional querying paradigms. At the core of this shift lies a pivotal question: as we explore these novel applications, what is the optimal AI architecture for the sophisticated understanding and manipulation of relational data?

This panel will debate this question, contrasting two prominent approaches.

On one hand, we have Large Language Models (LLMs), such as GPT-4 [16] or Llama [21]. These are general-purpose AI systems pre-trained on vast quantities of text and code. They understand and generate human language, follow complex instructions, and perform a wide array of tasks with minimal task-specific training, often through in-context learning. When applied to data management, LLMs typically process tabular data by serializing it into text, leveraging their ability for tasks like translating natural language questions into SQL queries (Text-to-SQL).

On the other hand, we have an emerging category of specialized models: Tabular Foundation Models (TFMs) [1, 5, 7]. Unlike general-purpose LLMs, TFMs are designed ad hoc or specifically pre-trained with an understanding of the characteristics of structured, relational data. This might involve architectures that explicitly account for the row-columnar nature of tables, separate embeddings for cells and headers, or pre-training objectives tailored to learn statistical patterns and semantic relationships directly from large corpora of tables (e.g., TaBERT [22], TaPas [11], TabPFN [12]). The hypothesis is that such specialization can lead to greater efficiency, accuracy, and robustness when dealing directly with relational databases.

This panel will analyze the dichotomy between specialized TFMs and general LLMs in the context of tasks crucial to the database community, such as natural language querying (Text-to-SQL [8]), understanding and inferring schema semantics (metadata discovery [6], semantic type detection), and data integration (entity resolution [2], column matching). While the panel will primarily focus on the TFM vs. LLM comparison for data management tasks, it will also touch upon the role of AI agents in orchestrating these models and interacting with humans within this data ecosystem. The core question remains: for the future of data management, should we invest in building more sophisticated, data-aware TFMs, or can the ever-improving capabilities of general LLMs suffice with novel adaptations?

Interest to the VLDB Community. The database community is at a crossroads with the advent of foundation models [3, 7]. This topic is of interest to the database community for multiple reasons:

- **Core Data Management Tasks:** The panel addresses the future of fundamental database tasks such as querying via natural language [8, 19]), understanding and inferring schema semantics [6], integrating and cleaning data [2], and ensuring data quality.
- **Efficiency & Scalability:** A key question is whether LLMs, massive in size [18], can be efficient & scalable on large-scale structured data, which would cause extremely high latencies and costs on recent LLMs.
- **Accuracy & Robustness:** Do LLMs capture the nuances of relational algebra and schema constraints, or TFMs are needed for a more robust path for tasks demanding deep structural understanding [4, 20]? This includes robustness to schema and linguistic variations [15].
- **Database Interfaces:** The choice between TFMs and LLMs will shape how users, both technical and non-technical, interact with databases, and what new tools the community needs to build.

Proceedings of the VLDB Endowment, Vol. 18, No. 12 ISSN 2150-8097.
doi:10.14778/3750601.3760519

- **Research Directions:** The panel aims to stimulate debate on where research efforts should be concentrated: on building more versatile TFMs, or on adapting general LLMs for structured data.

Shifting Perspectives. This panel aims to engage the database community by:

- **Challenging the "One Model Fits All" Assumption:** While LLMs are powerful generalists, the panel will probe whether this generality comes at a cost when dealing with the specific constraints and logic of relational data.
- **Elevating TFMs as a Critical Research Area:** It will highlight the research challenges in designing TFMs that are not just predictive but also adept at a range of data management tasks, potentially leading to new architectural innovations.
- **Fostering a Nuanced View of LLMs for Data:** Instead of a binary "good" or "bad," the discussion will explore where LLMs excel, where they falter with tabular data, and what augmentation (e.g. fine-tuning, agentic wrappers [14]) is necessary.
- **Re-evaluating Benchmarking for Tabular AI:** The panel will touch upon the need for benchmarks that specifically test the capabilities of models on diverse data management tasks, moving beyond general NLP metrics or isolated Text-to-SQL evaluations [10, 17, 19].

These discussions will challenge the common view of LLMs as a universal solution to recognizing the need for a specialized landscape for foundation models designed for data management.

Prompts for Panelists' Opening Statements. Panelists will be asked to address:

(1) "For data management tasks like Text-to-SQL, semantic type detection, and entity resolution on relational data, do you believe the future lies primarily with specialized TFMs or with general-purpose LLMs? What is the single most compelling reason for your stance?"
(2) "Considering the current state-of-the-art, what specific data management task represents the biggest hurdle for your preferred model class (TFM or LLM), and what breakthrough is needed to overcome it?"
(3) "Several works envision human experts working alongside TFMs and agents. If we narrow the focus to the TFM vs. LLM debate for automating data management tasks, how does the need for human oversight and intervention differ between these two model classes?"

The moderators will also guide the discussion with questions like:

- "While LLMs excel at tasks like Text-to-SQL, they struggle with problems crucial for internal database operations, such as cardinality estimation. Do you foresee specialized TFMs becoming integral components within the DBMS for tasks like query optimization, potentially outperforming general LLMs in these internal roles?"
- "What advantages or disadvantages do TFMs (e.g., leveraging structural priors) have compared to LLMs (e.g., leveraging vast textual knowledge)?"
- "Scalability and efficiency are key for real-world database applications. Can LLMs realistically be deployed for interactive data management tasks, or do TFMs offer a more viable path in terms of latency and computational cost?"
- "How critical is the nature of pre-training data? Is pre-training on massive corpora of diverse tables [13] a 'moat' for TFMs, or can LLMs overcome this with instruction tuning or few-shot learning on specific database schemas?"
- "Could sophisticated agents effectively 'wrap' a general LLM to perform specialized tabular tasks, potentially negating the need for ad hoc TFMs? What are the limits of such agent-based adaptation?"
- "What about tasks beyond querying, like automated data cleaning, imputation, or even suggesting schema refinements? Are TFMs or LLMs better suited for these data management tasks?"
- "How do we evaluate these models beyond task-specific accuracy? What about their robustness to noisy data [9], their ability to explain their 'reasoning' over tables, or their susceptibility to data memorization and privacy leaks?"
- "If you had to invest $100M in research to revolutionize interaction with relational data in the next 5 years, would you bet on building the ultimate TFM or on adapting the most powerful LLM? Why?"

We believe this panel will offer a timely and focused debate on a critical architectural choice facing the VLDB community, with direct implications for the future of data management.

2 MODERATORS

Carsten Binnig is a Full Professor in the Computer Science department at at TU Darmstadt and an Adjunct Associate Professor in the Computer Science department at Brown University. Carsten received his PhD at the University of Heidelberg in 2008. Afterwards, he spent time as a postdoctoral researcher in the Systems Group at ETH Zurich and at SAP, working on in-memory databases. Currently, his research focus is on redesigning databases and data management in an era of AI models and hardware. His work has been awarded with a Google Faculty Award, as well as multiple best paper and best demo awards for his research.

Paolo Papotti is an Associate Professor at EURECOM (France) since 2017. He got his PhD from Roma Tre University (Italy) in 2007 and had research positions at the Qatar Computing Research Institute (Qatar) and Arizona State University (USA). His research is focused on data management and information quality, with recent contributions in computational fact-checking and language models. He has authored more than 150 publications and his work has been recognized with best paper awards (CIKM 2024, ISWC 2024), best demo awards (SIGMOD 2015, DBA 2020, SIGMOD 2022), and Google Faculty Research awards (2016, 2020).

3 PANELISTS

We panel will include the following profiles:

- Floris Geerts (University of Antwerp)
- Johannes Hoffart (SAP)
- Madelon Hulsebos (CWI)
- Fatma Özcan (Google)
- Gael Varoquaux (INRIA)

This blend will ensure discussion draws from machine learning, theory, target applications, and fundamental database systems principles.

REFERENCES

[1] Gilbert Badaro, Mohammed Saeed, and Papotti Paolo. 2023. Transformers for Tabular Data Representation: A Survey of Models and Applications. *Transactions of the Association for Computational Linguistics* 11 (2023), 227–249. https://doi.org/doi.org/10.1162/tacl_a_00544

[2] Chandra Sekhar Bhagavatula, Thanapon Noraset, and Doug Downey. 2015. TabEL: Entity Linking in Web Tables. In *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part I (Lecture Notes in Computer Science)*, Marcelo Arenas, Óscar Corcho, Elena Simperl, Markus Strohmaier, Mathieu d'Aquin, Kavitha Srinivas, Paul Groth, Michel Dumontier, Jeff Heflin, Krishnaprasad Thirunarayan, and Steffen Staab (Eds.), Vol. 9366. Springer, 425–441. https://doi.org/10.1007/978-3-319-25007-6_25

[3] Joyce Cahoon, Alexandra Savelieva, Andreas C Mueller, Avrilia Floratou, Carlo Curino, Hiren Patel, Jordan Henkel, Markus Weimer, Nellie Gustafsson, Richard Wydrowski, et al. 2022. The need for tabular representation learning: An industry perspective. In *NeurIPS 2022 First Table Representation Workshop*.

[4] Tianji Cong, Madelon Hulsebos, Zhenjie Sun, Paul Groth, and H. V. Jagadish. 2023. Observatory: Characterizing Embeddings of Relational Tables. *Proceedings of the VLDB Endowment* 17, 4 (Dec. 2023), 849–862. https://doi.org/10.14778/3636218.3636237

[5] Xiang Deng, Huan Sun, Alyssa Lees, You Wu, and Cong Yu. 2020. TURL: Table Understanding through Representation Learning. *Proc. VLDB Endow.* 14, 3 (2020), 307–319. https://doi.org/10.5555/3430915.3442430

[6] Grace Fan, Jin Wang, Yuliang Li, Dan Zhang, and Renée Miller. 2023. Semantics-aware Dataset Discovery from Data Lakes with Contextualized Column-based Representation Learning. (2023). arXiv:cs.DB/2210.01922

[7] Xi Fang, Weijie Xu, Fiona Anting Tan, Jiani Zhang, Ziqing Hu, Yanjun Qi, Scott Nickleach, Diego Socolinsky, Srinivasan Sengamedu, and Christos Faloutsos. 2024. Large Language Models on Tabular Data–A Survey. *arXiv preprint arXiv:2402.17944* (2024).

[8] Avrilia Floratou, Fotis Psallidas, Fuheng Zhao, Shaleen Deep, Gunther Hagleither, Wangda Tan, Joyce Cahoon, Rana Alotaibi, Jordan Henkel, Abhik Singla, Alex Van Grootel, Brandon Chow, Kai Deng, Katherine Lin, Marcos Campos, K. Venkatesh Emani, Vivek Pandit, Victor Shnayder, Wenjing Wang, and Carlo Curino. 2024. NL2SQL is a solved problem... Not!. In *CIDR*.

[9] Boris Glavic, Giansalvatore Mecca, Renée J. Miller, Paolo Papotti, Donatello Santoro, and Enzo Veltri. 2024. Similarity Measures For Incomplete Database Instances. In *Proceedings 27th International Conference on Extending Database Technology, EDBT 2024, Paestum, Italy, March 25 - March 28*. OpenProceedings.org, 461–473. https://doi.org/10.48786/EDBT.2024.40

[10] Jorge Osés Grijalba, L Alfonso Urena Lopez, Eugenio Martínez-Cámara, and Jose Camacho-Collados. 2024. Question Answering over Tabular Data with DataBench: A Large-Scale Empirical Evaluation of LLMs. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. 13471–13488.

[11] Jonathan Herzig, Pawel Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Eisenschlos. 2020. TaPas: Weakly Supervised Table Parsing via Pre-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 4320–4333. https://doi.org/10.18653/v1/2020.acl-main.398

[12] Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. 2025. Accurate Predictions on Small Data with a Tabular Foundation Model. *Nature* 637, 8045 (Jan. 2025), 319–326. https://doi.org/10.1038/s41586-024-08328-6

[13] Zhengbao Jiang, Yi Mao, Pengcheng He, Graham Neubig, and Weizhu Chen. 2022. OmniTab: Pretraining with Natural and Synthetic Data for Few-shot Table-based Question Answering. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, United States, 932–942. https://doi.org/10.18653/v1/2022.naacl-main.68

[14] Kezhi Kong, Jiani Zhang, Zhengyuan Shen, Balasubramaniam Srinivasan, Chuan Lei, Christos Faloutsos, Huzefa Rangwala, and George Karypis. 2024. OpenTab: Advancing Large Language Models as Open-domain Table Reasoners. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=Qa0ULgosc9

[15] Pingchuan Ma and Shuai Wang. 2021. MT-teql: evaluating and augmenting neural NLIDB on real-world linguistic and schema variations. *Proc. VLDB Endow.* 15, 3 (nov 2021), 569–582. https://doi.org/10.14778/3494124.3494139

[16] OpenAI. 2023. ChatGPT 3.5. (2023). https://openai.com/blog/chatgpt

[17] Simone Papicchio, Paolo Papotti, and Luca Cagliero. 2023. QATCH: Benchmarking SQL-centric tasks with Table Representation Learning Models on Your Data. In *Conf. on Neural Information Processing Systems (NeurIPS)*.

[18] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21, 140 (2020), 1–67. http://jmlr.org/papers/v21/20-074.html

[19] Cédric Renggli, Ihab F. Ilyas, and Theodoros Rekatsinas. 2025. Fundamental Challenges in Evaluating Text2SQL Solutions and Detecting Their Limitations. *CoRR* abs/2501.18197 (2025). https://doi.org/10.48550/ARXIV.2501.18197 arXiv:2501.18197

[20] Yuan Sui, Mengyu Zhou, Mingjie Zhou, Shi Han, and Dongmei Zhang. 2024. Table meets llm: Can large language models understand structured table data? a benchmark and empirical study. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 645–654.

[21] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).

[22] Pengcheng Yin, Graham Neubig, Wen-tau Yih, and Sebastian Riedel. 2020. TaBERT: Pretraining for joint understanding of textual and tabular data. *arXiv preprint arXiv:2005.08314* (2020).