RIHAN HAI

Email: R.Hai@tudelft.nl
Personal webpage: https://rihanhai.com/
Team webpage: https://infinidata-team.github.io/

RESEARCH INTERESTS

AI in Data Lakes: in-database ML, synthetic tabular data generation using diffusion models, energy-efficient tabular learning

Quantum Data Management: quantum databases, data-intensive quantum simulators

EDUCATION

RWTH Aachen University
Dr. rer. nat. in Computer Science (**summa cum laude**)

Tsinghua University
Master of Engineering in Control Engineering

Tsinghua University
Bachelor of Engineering in Automation

EXTERNAL FUNDING (> 1 MILLION €)

NWO Veni (Dutch prestigious personal grants) 2024–2028

- Understanding Implicit Dataset Relationships for Machine Learning
- **Top 3%** in the *Science* domain

HORIZON-CL4-2022-DATA-01 2023–2025

- ExtremeXP — EXPerimentation driven and user eXPerience oriented analytics for eXtremely Precise outcomes and decisions
- Role: work package lead, task lead

HORIZON-JU-SNS-2024-STREAM-B-01-08 2025–2027

- 6G-DALI: 6G Data and ML operations automation via an end to end AI framework
- Role: task lead

CURRENT EMPLOYMENT

Delft University of Technology (Delft, Netherlands) January 2021 – Now
Assistant Professor (Tenured, promoted to UD1)

ORGANIZATION

TU Delft March 2024 - September 2025

- Admission Committee of Data Science and AI Technology (DSAIT) Master's Programme (September 2025 - Now)
- Coordinator of DSAIT Master's Programme (March 2024 - September 2025)
- Coordinator of DSAIT theme (*Data and Information Management)*

AWARDS

- Distinguished PC member of VLDB 2024, 2025
- Best Demo Runner-Up, ACM International Conference on Management of Data (SIGMOD), 2025
- Best Demo Award, International Conference on Extending Database Technology (EDBT), 2023.
- Nomination for the best dissertation of the Faculty of Mathematics, Computer Science, and Natural Sciences of RWTH Aachen University (Friedrich Wilhelm Award 2020)

RESEARCH LEADERSHIP, INVITED TALKS & SERVICE

- **Conference Reviewer:** SIGMOD, VLDB, ICDE, EDBT, WWW
- **Journal Reviewer:** TODS, TKDE, VLDBJ, SIGMOD Record, Information Systems, JMLR, TPDS
- **Editorial Service:** Guest Editor, *Frontiers in Big Data* (Special Issue on *Synergies of Data Management and Machine Learning: Methods, Challenges, and Future*)
- **Community Service:** Registration Chair of VLDB 2026; Registration Chair of ICDE 2024

Selected Invited Talks

AI in Data Lakes

- Invited talk at *EDBT 2025 PhD Workshop*
- Invited talk at *International Workshop on Data-Centric AI (DATAI)*, co-located with VLDB 2025
- Invited panelist at *ICDE 2023 Special Track Paper Session*

Quantum Data Management

- Keynote at the Qiskit Fall Fest 2025 (IBM & TU Delft)
- Keynote at the *Second International Workshop on Quantum Data Science and Management (QDSM)*, co-located with VLDB 2024
- Invited tech talk at *SUE* (a Netherlands-based consultancy in cloud, data, and AI), 2024
- Invited talk at *Quantum Seminar, Yonsei University*, Korea, 2024

Workshop Organization

- **Founder:** *International Workshop on Databases and Machine Learning (DBML@ICDE)* 2022, 2023
- **Founder:** *International Workshop on Quantum Data and Machine Learning (QDML@ICDE)* 2025
- **Organizer:** *International Workshop on Quality in Databases (QDB@VLDB)* 2016

SELECTED PUBLICATIONS

Full publication list can be found at my google scholar. Below are representative works led by me or my InfiniData team.

Quantum Data Management

1. (Poster) Giancarlo Gatti and **Rihan Hai**. Private quantum database. The 29th Annual Quantum Information Processing Conference, January 2026
2. **Rihan Hai**, Shih-Han Hung, Tim Coopmans, Tim Littau, and Floris Geerts. Quantum data management in the NISQ era. *PVLDB*, 18(6):1720–1729, 2025
3. Tim Littau and **Rihan Hai**. Qymera: Simulating quantum circuits using RDBMS. In *SIGMOD*, pages 179–182, 2025

4. **Rihan Hai**, Shih-Han Hung, and Sebastian Feld. Quantum data management: From theory to opportunities. In *ICDE*, pages 5376–5381, 2024

AI in Data Lakes

2026

1. Wenbo Sun, Qiming Guo, Wenlu Wang, and **Rihan Hai**. TranSQL+: Serving large language models with SQL on low-resource hardware. *SIGMOD*, 3(6):1–27, December 2025
2. Aditya Shankar, Lydia Chen, Arie van Deursen, and **Rihan Hai**. WaveStitch: Flexible and fast conditional time series generation with diffusion models. *SIGMOD*, 3(6):1–25, December 2025

2025

3. Wenbo Sun, Ziyu Li, and **Rihan Hai**. Database as runtime: Compiling LLMs to SQL for in-database model serving. In *SIGMOD*, page 231–234, 2025 (*Best demo runner-up*)

2024

4. Ziyu Li, Wenbo Sun, Danning Zhan, Yan Kang, Lydia Chen, Alessandro Bozzon, and **Rihan Hai**. Amalur: The convergence of data integration and machine learning. *TKDE*, pages 1–14, 2024
5. Ziyu Li, Hilco Van Der Wilk, Danning Zhan, Megha Khosla, Alessandro Bozzon, and **Rihan Hai**. Model selection with model zoo via graph learning. In *ICDE*, pages 1296–1309, 2024
6. Andra Ionescu, Kiril Vasilev, Florena Buse, **Rihan Hai**, and Asterios Katsifodimos. Autofeat: Transitive feature discovery over join paths. In *ICDE*, pages 1861–1873. IEEE, 2024
7. Aditya Shankar, Hans Brouwer, **Rihan Hai**, and Lydia Chen. Silofuse: Cross-silo synthetic data generation with latent tabular diffusion models. In *ICDE*, pages 110–123, 2024
8. Ziyu Li, Wenjie Zhao, Asterios Katsifodimos, and **Rihan Hai**. LLM-PQA: LLM-enhanced prediction query answering. In *CIKM*, pages 5239–5243, 2024
9. Andra Ionescu, Zeger Mouw, Efthimia Aivaloglou, **Rihan Hai**, and Asterios Katsifodimos. Human-in-the-loop feature discovery for tabular data. In *CIKM*, pages 5215–5219, 2024

2023

10. **Rihan Hai**, Christos Koutras, Christoph Quix, and Matthias Jarke. Data lakes: A survey of functions and systems. *TKDE*, 35(12):12571–12590, 2023 (*Citation number: 239.*)
11. **Rihan Hai**, Christos Koutras, Andra Ionescu, Ziyu Li, Wenbo Sun, Jessie van Schijndel, Yan Kang, and Asterios Katsifodimos. Amalur: Data integration meets machine learning. In *ICDE*, pages 3729–3739, 2023

During PhD

12. **Rihan Hai** and Christoph Quix. Rewriting of Plain SO Tgds into Nested Tgds. In *VLDB*, pages 1526–1538, 2019
13. **Rihan Hai**, Sandra Geisler, and Christoph Quix. Constance: An intelligent data lake system. In *SIGMOD*, pages 2097–2100, 2016 (*It is one of the **most cited** data lake publications. Citation number: 355.*)
14. (Book) Mathieu D'Aquin, Alessandro Adamou, Nicolas Arnaud, Houssem Chihoub, Christine Collet, **Rihan Hai**, et al. *Data Lakes.* ISTE, 2020