

Madelon Hulsebos

Amsterdam, Netherlands | madelon@cw.nl | madelonhulsebos.com

I'm a computer scientist working on the democratization of insights from tabular data (e.g. relational tables and spreadsheets) through table representation learning (TRL). My research involves machine learning, natural language processing, information retrieval and data management.

EMPLOYMENT

Centrum Wiskunde & Informatica – Amsterdam, Netherlands <i>Tenure-track Researcher, lead of the TRL Lab</i>	2024–present
University of California, Berkeley – Berkeley, United States <i>Postdoctoral Scholar, EECS and BIDS</i>	2023–2024
University of Amsterdam – Amsterdam, Netherlands <i>PhD Researcher (external), Informatics Institute</i>	2023
<i>Guest Researcher, Informatics Institute</i>	2020–2023
Sigma Computing – San Francisco, United States <i>Full-time PhD Researcher, previously Research Intern (Summer 2021)</i>	2021–2022
KPN/HEINEKEN – Rotterdam/Amsterdam, Netherlands <i>Data Scientist</i>	2019–2021
Massachusetts Institute of Technology – Cambridge, United States <i>Research Visitor, MIT Media Lab</i>	2018–2019
Delft University of Technology – Delft, Netherlands <i>Graduate TA, Pattern Recognition & Web Information Systems groups</i>	2017–2018
Aalto University – Helsinki, Finland <i>Research and Teaching Assistant, Machine Learning for Big Data</i>	2017

EDUCATION

University of Amsterdam – Amsterdam, Netherlands <i>PhD Computer Science</i>	2020–2024
Delft University of Technology – Delft, Netherlands <i>MSc Computer Science</i>	2016–2018
<i>BSc Technology, Policy and Management</i>	2011–2015

ANCILLARY ACTIVITIES

PriorLabs – Berlin/Freiburg, Germany <i>Scientific Advisory Board Member</i>	2025–present
UniPartners Delft – Delft, Netherlands <i>Supervisory Board Member</i>	2017–2023
<i>Executive Board Member</i>	2015–2016

ACADEMIC SERVICE

Chairing	
Co-chair, AI for Tabular Data workshop @ EurIPS	2025
Co-chair Tutorial Track @ VLDB	2025
Co-chair, New Researcher Symposium @ SIGMOD	2025
Co-chair Seminar on Challenges and Opportunities in Table Representation Learning @ Dagstuhl	2025
Co-chair Representation Learning and Generative Models for Structured Data @ ELLIS Amsterdam	2025
Initiator, Table Representation Learning research theme @ ELLIS Amsterdam	2025

Founder and co-chair, Table Representation Learning workshop (TRL) @ NeurIPS, ACL	2022-2025
Co-chair, Data Management for End-to-End ML workshop (DEEM) @ SIGMOD	2023-2025
Steering Committee, Tabular Data Analysis workshop (TaDA) @ VLDB	2023, 2024
Co-chair, SemTab challenge @ ISWC	2021-2023

Conference/Journal Reviewing

Conference of Innovative Data Systems Research (CIDR)	2026
SIGMOD	2026
VLDB Journal	2025-
PVLDB	2024-2025
NeurIPS (Datasets & Benchmarks track)	2021-2025
ICDE (Industry track)	2024
EDBT (Industry track)	2022-2023
TheWebConf (Industry track)	2022-2023

Workshop Reviewing

NOVAS @ SIGMOD (2025), aiDM @ SIGMOD (2024), DMLR @ ICML (2023, 2024), DBML @ ICDE (2023, 2024), PhD workshop @ VLDB (2023), SemTab @ ISWC (2021 - 2023), AIDB @ VLDB (2022).

Editorial

Assistant Editor, Journal of Systems Research (JSys)	2022 - 2023
--	-------------

ADVISING AND COMMITTEES**Advising postdoc/PhD**

Rohith Prabakaran, CWI, PhD	2026-
Aécio Santos, CWI, Postdoc	2025-
Xue (Effy) Li, CWI, Postdoc	2025-
Cornelius Wolff, CWI, PhD	2025-
Daniel Gomm, CWI, PhD	2024-

Advising (under)graduate

Hongqian Xia, University of Amsterdam, MSc	2026-
Yme Kingma, Radboud University, MSc	2026-
Lisa van Oosten, University of Amsterdam, MSc	2026-
Louis Gehringer, University of Amsterdam, MSc	2025-
Ahmed Omar, Utrecht University, MSc	2025-
Daniel De Dios, TU Delft, BSc	2025-
Wojciech Kosiuk, University of Amsterdam, MSc	2025-
Jan Henrik Bertrand, University of Amsterdam, MSc	2025-
Liang Telkamp, University of Amsterdam, MSc (🏆 Amsterdam AI Thesis Award)	2025
Nikola Selic, TU Munich, MSc	2025
Wenjing Lin, UC Berkeley, MSc	2024-2025
Rachel Lin, UC Berkeley, MSc	2024-2025
Xingyu (Carl) Ji, UC Berkeley, BSc	2023-2025
Teun Mathijssen, University of Amsterdam, MSc	2023

PhD Thesis Committees

Félix Lefebvre, Inria, PhD	2026
Margherita Martorana, VU Amsterdam, PhD	2025
Vaishali Pal, University of Amsterdam, PhD	2025
Tianji Cong, University of Michigan, PhD	2023-2025

AWARDS AND FUNDING

Research grant for collaboration CWI x Inria on TRL, €627K, CWI	2025
Compute credit grants \$42K (Google, OpenAI, Cohere)	2025
Table Representation Learning Research Grant, €300K, SAP	2024
AiNed Fellowship Grant, €993K, NWO	2024

Table Representation Learning Theme in ELLIS unit Amsterdam, €50K, SAP	2024
Postdoctoral Fellowship, \$150K, Accenture-BIDS	2023
Best Reviewer Award, PhD Workshop, VLDB	2023
Travel Award, \$2.5K, VLDB Endowment	2022
Honorable mention GitTables, SemTab challenge	2021

TALKS, PANELS AND LECTURES

Talks

What are we asking from Tabular Data?

Keynote AI for Tabular Data workshop, EurIPS 2025 2025

Towards robust open-domain querying over structured data

KIK AI seminar, Amsterdam University Medical Center 2025

ELLIS NL Scientific Day 2025

AI Research Retreat, SAP 2025

Business Analytics Seminar, University of Amsterdam 2025

Seminar on AI, Semantics, and Tabular Data, IBM 2025

What Table Representation Learning brings to Data Systems

COMPASS Seminar, ETH, Zürich, Switzerland 2024

CWI Lectures, CWI, Amsterdam, Netherlands 2024

Retrieval Systems for Structured Data: the missing piece for grounding LLM-powered query interfaces in factual data

Berkeley Institute for Data Science Seminar, Berkeley, USA 2024

Berkeley EECS EPIC Retreat, Berkeley, USA 2024

Harnessing AI for Relational Data: Industry and Research Perspectives, SF, USA 2024

Advances, challenges, and opportunities in Table Representation Learning

Microsoft Gray Systems Lab, USA 2024

University of Washington Seattle, Seattle, USA 2024

Snowflake, USA 2024

Google, Sunnyvale, USA 2024

UC Berkeley, Berkeley, USA 2024

Transformers at Work 2023, Zeta Alpha, Amsterdam, Netherlands 2023

Towards Table Representation Learning for end-to-end data management and analysis

INRIA-Saclay, Paris, France 2023

ML for Systems and Systems for ML Workshop @ BTW, Dresden, Germany 2023

Hasso Plattner Institute, Berlin, Germany 2023

TU Darmstadt, Darmstadt, Germany 2022

GitTables: a large corpus of relational tables

Tabular Data Analysis workshop, VLDB, Vancouver, Canada 2023

Database Architectures group, CWI, Amsterdam, Netherlands 2022

(Guest) Lectures

Representation Learning and Generative Models for Tabular Data

University of Amsterdam 2026

BIFOLD, TU Berlin 2025

Panel discussions

"Neural Relational Data: Tabular Foundation Models, LLMs... or both?", VLDB, London 2025

Transformers at Work 2023, Zeta Alpha, Amsterdam, Netherlands 2023

PUBLICATIONS

2025

SQaLe: A Large Text-to-SQL Corpus Grounded in Real Schemas, AI for Tabular Data workshop at EurIPS 2025

Wolff, C., Gomm, D., Hulsebos, M.

Are We Asking the Right Questions? On Ambiguity in Natural Language Queries for Tabular Data Analysis, **AI for Tabular Data workshop at EurIPS 2025**

Gomm, D., Wolff, C., [Hulsebos, M.](#)

Unlocking the Full Potential of Data Science Requires Tabular Foundation Models, Agents, and Humans, **In Review**
Cong, T., Eisenschlos, J., Gomm, D., Grinsztajn, L., Müller, A., Sanghi, A., et al., Binnig, C., [Hulsebos, M.](#), Hutter, F.

Towards contextual sensitive data detection, **In Review**

Telkamp, L., [Hulsebos, M.](#)

How well do LLMs reason over tabular data, really?, **Table Representation Learning workshop at ACL**
Wolff, D., [Hulsebos, M.](#)

Rethinking Dataset Discovery with DataScout, **User Interface Software and Technology (UIST)**

Lin, R., Chopra, B., Lin, W. Shankar, S. [Hulsebos, M.](#), Parameswaran, A.

Metadata Matters in Dense Table Retrieval, **ELLIS workshop on RL and GM for Structured Data**

Gomm, D., [Hulsebos, M.](#)

TARGET: Benchmarking Table Retrieval for Generative Tasks, **In Review**

Li, X., Glenn, P., Parameswaran, A., [Hulsebos, M.](#)

Towards Accurate and Efficient Document Analytics with Large Language Models, **ICDE**

Lin, Y., [Hulsebos, M.](#), Ma, R., Shankar, S., Zeigham, S., Parameswaran, A. G., & Wu, E.

2024

TARGET: Benchmarking Table Retrieval for Generative Tasks, **TRL @ NeurIPS**

Li, X., Parameswaran, A., [Hulsebos, M.](#)

"It Took Longer than I was Expecting:" Why is Dataset Search Still so Hard?, **HILDA@SIGMOD**

[Hulsebos, M.](#), Lin, W., Shankar, S., Parameswaran, A.

SchemaPile: A Large Collection of Relational Database Schemas, **SIGMOD**

Doehmen, T., Geacu, R., [Hulsebos, M.](#), Schelter, S.

SPADE: Synthesizing Assertions for Large Language Model Pipelines, **Proceedings of VLDB**

Shankar, S., Li, H., Asawa, P., [Hulsebos, M.](#), Lin, Y., Zamfirescu-Pereira, J., Chase, H., Fu-Hinthorn, W., Parameswaran, A., Wu, E.

Eighth Workshop on Data Management for End-to-End Machine Learning (DEEM), **SIGMOD**

[Hulsebos, M.](#), Shankar, S., Interlandi, M.

2023

AdaTyper: Adaptive Semantic Type Detection, **ArXived**

[Hulsebos, M.](#), Groth, P., Demiralp, C.

Introducing the Observatory Library for End-to-End Table Embedding Inference, **TRL @ NeurIPS**

Cong, T., Sun., Z., Groth, P., Jagadish, H., [Hulsebos, M.](#)

Observatory: Characterizing Embeddings of Relational Tables, **Proceedings of VLDB**

Cong, T., [Hulsebos, M.](#), Sun., Z. Groth, P., Jagadish, H.V.

Models and Practice of Neural Table Representations [tutorial], **SIGMOD**

[Hulsebos, M.](#), Deng, X., Sun, H., Papotti, P.

Seventh Workshop on Data Management for End-to-End Machine Learning (DEEM), **SIGMOD**

Boehm, M., [Hulsebos, M.](#), Shankar, S., Varma, P.

2022

GitTables: A Large-Scale Corpus of Relational Tables, **SIGMOD**

[Hulsebos, M.](#), Demiralp, C., Groth, P.

GitSchemas: A Dataset for Automating Relational Data Preparation Tasks, **DBML @ ICDE**
Döhmen, T., Hulsebos, M., Beecks, C., Schelter, S.

Results of SemTab 2022, **Proceedings of SemTab @ ISWC**
Abdelmageed, N., Chen, J., Cutrona, V., Efthymiou, V., Hassanzadeh, O., Hulsebos, M., Jiménez-Ruiz, E.,
Sequeda, J. and Srinivas, K.

Making Table Understanding Work in Practice [abstract], **CIDR**
Hulsebos, M., Gathani, S., Gale, J., Dillig, I., Groth, P., Demiralp, C.

Augmenting Decision Making via Interactive What-If Analysis, **CIDR**
Gathani, S., Hulsebos, M., Gale, J., Haas, P. J., Demiralp, C.

2021

Results of SemTab 2021, **Proceedings of SemTab @ ISWC**
Cutrona, V., Chen, J., Efthymiou, V., Hassanzadeh, O., Jiménez-Ruiz, E., Sequeda, J., Srinivas, K.,
Abdelmageed, N., Hulsebos, M., Oliveira, D., Pesquita, C.

2020

Sato: Contextual semantic type detection in tables, **Proceedings of VLDB**
Zhang, D., Suhara, Y., Li, J., Hulsebos, M., Demiralp, C., Tan, W.

2019

Sherlock: A deep learning approach to semantic data type detection, **ACM SIGKDD**
Hulsebos, M., Hu, K., Bakker, M., Zraggen, E., Satyanarayan, A., Kraska, T., Demiralp, C., Hidalgo, C.

VizNet: Towards a large-scale visualization learning and benchmarking repository, **ACM CHI**
Hu, K., Gaikwad, N., Hulsebos, M., Bakker, M., Zraggen, E., Hidalgo, C., Kraska, T., Li, G., Satyanarayan,
A., Demiralp, C.

2018

The Network Nullspace Property for Compressed Sensing of Big Data Over Networks, **IEEE ICASSP**
Hulsebos, M., Jung, A.