Madelon Hulsebos

Amsterdam, NL | madelon@cwi.nl | madelonhulsebos.com | trl-lab-github.io

My research involves machine learning. natural language processing, information retrieval and data management. I particularly focus on **representation learning and generative models for structured data** (e.g. relational tables) to surface insights from structured data; accurate, easy, and transparent.

EMPLOYMENT

Centrum Wiskunde & Informatica – Amsterdam, Netherlands
Tenure-track Researcher (on-track for early tenure)   Sep 2024 – present
- Established the Table Representation Learning lab focused on democratizing insights from tabular data.
- Received a $1M NWO AiNed Fellowship grant, besides external funding from industry.
- Initiated the *Table Representation Learning research theme* at the ELLIS unit Amsterdam.

University of California, Berkeley – Berkeley, United States
Postdoctoral Scholar, EECS and BIDS   Nov 2023 – Aug 2024
- Worked on methods and benchmarks for retrieval-augmented generation and search over tabular.

University of Amsterdam – Amsterdam, Netherlands
PhD Researcher (external), Informatics Institute   Jan 2023 – Nov 2023
Guest Researcher, Informatics Institute   Aug 2020 – Jan 2023
- Initiated and pursued a research agenda on neural models for structured data, and applications thereof.
- Thesis title: Table Representation Learning.

Sigma Computing – San Francisco, United States
Full-time PhD Researcher, previously Research Intern (Summer 2021)   June 2021 – Dec 2022
- Developed a system for adaptive neural table models for applications like data cleaning and search.
- Contributed to the design and implementation of an interactive intelligent ML tool (Decision Studio).

KPN/HEINEKEN – Rotterdam/Amsterdam, Netherlands
Data Scientist   Mar 2019 – May 2021
- Built ML tools for financial forecasting and marketing analyses using e.g. Bayesian models.
- Mentored data analysts, led a reading group and gave tutorials on e.g. data validation, transfer learning.

Massachusetts Institute of Technology – Cambridge, United States
Visiting Collaborator, MIT Media Lab   Aug 2018 – Mar 2019
- Led research on semantic type detection in tables (Sherlock). Sherlock is well adopted in industry and benchmarks in research. Also contributed to a data visualization benchmarking project (VizNet).

Delft University of Technology – Delft, Netherlands
Graduate TA, Pattern Recognition & Web Information Systems groups   Sep 2017 – Feb 2018
- TA for the MSc courses: Pattern Recognition (IN4085), Web Science & Engineering (IN4252).
- Supported 250+ graduate students in labs and projects, and evaluated student assignments.

Aalto University – Helsinki, Finland
Research and Teaching Assistant, Machine Learning for Big Data   July - Oct 2017
- Developed material for a BSc course on Machine Learning, in which 500+ students participated.
- Conducted experiments for semi-supervised learning over networks, and presented at ICASSP.

EDUCATION

University of Amsterdam – Amsterdam, Netherlands
PhD Computer Science   Sep 2020 – Nov 2023

Delft University of Technology – Delft, Netherlands
MSc Computer Science   Sep 2016 – July 2018
BSc Technology, Policy and Management   Sep 2011 – July 2015

ACADEMIC SERVICE

Organizing Committees

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
Co-chair Tutorial Track @ **VLDB**	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

Program Committees

SIGMOD	2026
VLDB Journal	2025
NOVAS Workshop @ SIGMOD	2025
PVLDB	2024-2025
NeurIPS (Datasets & Benchmarks track)	2021-2025
aiDM Workshop @ SIGMOD	2024
ICDE (Industry track)	2024
Data-centric Machine Learning Research Workshop @ ICML	2023-2024
DBML Workshop @ ICDE	2023-2024
PhD Workshop @ VLDB	2023
EDBT (Industry track)	2022-2023
TheWebConf (Industry track)	2022-2023
SemTab @ ISWC	2021- 2023
AIDB Workshop @ VLDB	2022

Editorial

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

MENTORSHIP

Advising

Wojciech Kosiuk, University of Amsterdam, MSc	2025-
Nikola Selic, TU Munich, MSc	2025-
Effy (Xue) Li, CWI, Postdoc	2025-
Cornelius Wolff, CWI, PhD	2025-
Jan Henrik Bertrand, University of Amsterdam (ELLIS Honours program), MSc	2025-
Liang Telkamp, University of Amsterdam, MSc	2025-
Daniel Gomm, CWI, PhD	2024-
Wenjing Lin, UC Berkeley, MSc	2024-
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

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

BOARD MEMBERSHIPS (PRO BONO)

UniPartners Delft – Delft, Netherlands
Supervisory Board Member — May 2017 – Dec 2023
- Supervised the strategic position of the student consulting company, and implemented structured financial control, KPI oriented leadership and a supervision cycle.
- Consulted on software management projects.

Executive Board Member Feb 2015 – Feb 2016
- Controlled and optimized the quality of products & processes, and moderated the CRM system.
- Daily management of projects, contributing to a revenue of over €100K.

Talks and Lectures

Towards open-domain querying over structured data

AI Research Retreat, SAP	July 2025
Business Analytics Seminar, University of Amsterdam	July 2025
Seminar on AI, Semantics, and Tabular Data, IBM	May 2025

Representation Learning and Generative Models for Tabular Data (guest lecture)

BIFOLD, TU Berlin	Jan 2025

What Table Representation Learning brings to Data Systems

COMPASS Seminar, ETH, Zürich, Switzerland	Dec 2024
CWI Lectures, CWI, Amsterdam, Netherlands	Nov 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	Oct 2024
Berkeley EECS EPIC Retreat, Berkeley, USA	Oct 2024
Harnessing AI for Relational Data: Industry and Research Perspectives, SF, USA	Oct 2024

Advances, challenges, and opportunities in Table Representation Learning

Microsoft Gray Systems Lab, USA	July 2024
University of Washington Seattle, Seattle, USA	May 2024
Snowflake, USA	May 2024
Google, Sunnyvale, USA	Apr 2024
UC Berkeley, Berkeley, USA	Mar 2024
Transformers at Work 2023, Zeta Alpha, Amsterdam, Netherlands	Sep 2023

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

INRIA-Saclay, Paris, France	Apr 2023
ML for Systems and Systems for ML Workshop @ BTW, Dresden, Germany	Mar 2023
Hasso Plattner Institute, Berlin, Germany	Mar 2023
TU Darmstadt, Darmstadt, Germany	June 2022

GitTables: a large corpus of relational tables

Tabular Data Analysis workshop, VLDB, Vancouver, Canada	Aug 2023
Database Architectures group, CWI, Amsterdam, Netherlands	Feb 2022

Awards and Funding

Internal call for research collaboration CWI x Inria, €627K, CWI	2025
Table Representation Learning Theme in ELLIS unit Amsterdam, €50K, SAP	2024
Table Representation Learning Research Grant, €300K, SAP	2024
AiNed Fellowship Grant, €993K, NWO	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

Publications

2025

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., <u>Hulsebos, M.</u>, Hutter, F.

Beyond PII: Contextual Sensitive Data Detection in Tabular Datasets, **In Review**
Telkamp, L., <u>Hulsebos, M.</u>

How well do LLMs reason over tabular data, really?, **Table Representation Learning workshop at ACL**
Wolff, D., <u>Hulsebos, M.</u>

Rethinking Dataset Discovery with DataScout, **User Interface Software and Technology (UIST)**
Lin, R., Chopra, B.., Lin, W. Shankar, S. <u>Hulsebos, M.,</u> Parameswaran, A.

Metadata Matters in Dense Table Retrieval, **ELLIS workshop on RL and GM for Structured Data**
Gomm, D., <u>Hulsebos, M.</u>

TARGET: Benchmarking Table Retrieval for Generative Tasks, **In Review**
Li, X., Glenn, P., Parameswaran, A., <u>Hulsebos, M.</u>

*Towards Accurate and Efficient Document Analytics with Large Language Models***, ICDE**
Lin, Y., <u>Hulsebos, M.,</u> 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., <u>Hulsebos, M.</u>

"It Took Longer than I was Expecting:" Why is Dataset Search Still so Hard?, **HILDA@SIGMOD**
<u>Hulsebos, M.,</u> Lin, W., Shankar, S., Parameswaran, A.

SchemaPile: A Large Collection of Relational Database Schemas, **SIGMOD**
Doehmen, T., Geacu, R., <u>Hulsebos, M.,</u> Schelter, S.

SPADE: Synthesizing Assertions for Large Language Model Pipelines, **Proceedings of VLDB**
Shankar, S., Li, H., Asawa, P., <u>Hulsebos, M.,</u> 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**
<u>Hulsebos, M.,</u> Shankar, S., Interlandi, M.

2023
AdaTyper: Adaptive Semantic Type Detection, **ArXived**
<u>Hulsebos, M.,</u> 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., <u>Hulsebos, M.</u>

Observatory: Characterizing Embeddings of Relational Tables, **Proceedings of VLDB**
Cong, T., <u>Hulsebos, M.,</u> Sun., Z. Groth, P., Jagadish, H.V.

Models and Practice of Neural Table Representations [tutorial], **SIGMOD**
<u>Hulsebos, M.,</u> Deng, X., Sun, H., Papotti, P.

Seventh Workshop on Data Management for End-to-End Machine Learning (DEEM), **SIGMOD**
Boehm, M., <u>Hulsebos, M.,</u> Shankar, S., Varma, P.

2022
GitTables: A Large-Scale Corpus of Relational Tables, **SIGMOD**
<u>Hulsebos, M.,</u> Demiralp, C., Groth, P.

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

Results of SemTab 2022, **Proceedings of SemTab @ ISWC**
Abdelmageed, N., Chen, J., Cutrona, V., Efthymiou, V., Hassanzadeh, O., <u>Hulsebos, M.,</u> 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., Zgraggen, 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., Zgraggen, 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.