OLGA OVCHARENKO

Website: https://olgaovcharenko.github.io | GitHub: https://github.com/OlgaOvcharenko

Emails: ovcharenko.folga@gmail.com | ovcharenko@tu-berlin.de

Profile

I am a first-year Ph.D. student at BIFOLD and TU Berlin in the Data Engineering for ML (DEEM) Lab, researching machine learning and data management. My Ph.D. advisor is Sebastian Schelter. Currently, I am working on two research directions: (1) cross-modal error detection and repair and (2) single-cell data representation learning. Before my Ph.D., during my master's at ETH Zürich, I worked with Valentina Boeva and Julia Vogt and focused on benchmarking single-cell data integration methods. Additionally, I worked and have publications on visualizations, federated machine learning (ML), and systems for ML.

Experience

Research:

Mar 2025 – now	**Ph.D. Student** at DEEM Lab, BIFOLD & TU Berlin
	Advisor: Prof. Sebastian Schelter
Dec 2024 – Mar 2025	**Research Assistant** at Boeva Lab, ETH Zürich
	Benchmarking single-cell data integration frameworks.
May – Nov 2024	**Master's Thesis** at MDS Lab**,** ETH Zürich
	Contrastive Representation Learning for Single-Cell Multi-Omics Data Integration.
	Advisors: Prof. Julia Vogt, Prof. Valentina Boeva
Sep 2023 – Mar 2024	**Semester Project** with Prof. Dr. Valentina Boeva, ETH Zürich
	Development of a novel non-linear dimensionality reduction visualization.
Sep 2023 – Dec 2023	**Semester Project** with Immigration Policy Lab, ETH Zürich
	Fine-tuned transformers and LLMs for the toxic/hate speech detection.
Sep 2022 – Jan 2023.	**Research Assistant** with Prof. Theo Rekatsinas
	Implementation of the distributed data preparation pipeline.
Sep 2023 – Dec 2023	**Research Assistant** in the ExDRA Project
	Implementation of federated components (linear algebra operations, low-level instructions, and build-in algorithms) via contributions to Apache SystemDS.
Sep 2023 – Dec 2023	**IT Intern** (AIT, Angewandte Informationstechnik Forschungsgesellschaft mbH)
	Rhinodiagnost project. Pressure and volume flow calculation using the simulator.

Teaching Assistant:

Feb – Jun 2023	Data Modeling and Databases (Systems Group, ETH Zurich)
Feb – Aug 2022	Quality Assurance in Software Development (IST, TU Graz)
Oct 2019 – Jun 2020	Data Management (ISDS, TU Graz)

Education

Sep 2022 – Dec 2024 MSc Data Science, ETH Zürich (Zürich, Switzerland)

Mar 2019 – Aug 2022 BSc Computer Science *with distinction*, Graz University of Technology (Graz, Austria)

Publications

2025 **DataWorld Workshop ICML.** Towards Cross-Modal Error Detection with Tables and Images
2025 **ICML *Spotlight*.** scSSL-Bench: Benchmarking Self-Supervised Learning for Single-Cell Data [paper]
2025 **ICDE DAIS Workshop.** Towards Regaining Control over Messy ML Pipelines [paper]
2024 **SSL Workshop NeurIPS.** Benchmarking Self-Supervised Learning for Single-Cell Data [paper]
2024 **IEEE VIS.** Feature Clock: High-Dimensional Effects in Two-Dimensional Plots [library]
2022 **CIKM.** Federated Data Preparation, Learning, and Debugging in Apache SystemDS [paper]
2021 **SIGMOD.** ExDRa: Exploratory Data Science on Federated Raw Data [paper]

Awards & Honors

2025 scSSL-Bench [paper] is a *Spotlight* paper at ICML 2025
2024 Apache SystemDS Project Management Committee (PMC) member
2021 Dean's List 2021 (top 5% student in 4th semester), Graz University of Technology
2021 Apache SystemDS committer

Reviewing

2025 Shadow Program Committee. VLDB 2026
2025 Reviewer. DataWorld Workshop ICML 2025
2024 Availability and Reproducibility Reviewer. SIGMOD 2025

Key Skills

Python, R, Java & Scala.
Pandas, numpy, scikit-learn, Pytorch, Apache Spark.
SQL, MongoDB, RumbleDB, JSONIQ.