

Tobias von Arx

tobiasvonarx@proton.me | +41 79 903 67 74 | github.com/tobiasvonarx | [LinkedIn](#) | avanis.xyz

Data Science Master's student at ETH Zurich, currently visiting the University of Cambridge, with published work in LLM security evaluation and pretraining, and experience in agentic and applied AI. I enjoy working on difficult problems.

EDUCATION

Visiting Student, **University of Cambridge**, Fitzwilliam College Mar 2026 – present

MSc Data Science, **ETH Zurich** (projected GPA: 5.7/6.0) Sep 2024 – present

- Selected Coursework: Reliable & Trustworthy AI, LLMs, Probabilistic AI, Network Security, Optimization for Data Science, Computational Statistics, Big Data, Algorithms Lab

BSc Computer Science, **ETH Zurich** (GPA: 5.22/6.0) Sep 2020 – Sep 2023

- Selected Coursework: Formal Methods and Functional Programming, Compiler Design, Quantum Physics, Approximation Algorithms, Intro to ML, Visual Computing, Theoretical Computer Science, Discrete Mathematics

RESEARCH EXPERIENCE

Master Thesis – University of Cambridge, EEG Lab (Prof. Srinivasan Keshav) Mar 2026 – present

- *Machine Learning for Optimal Solar Panel Placement*, supervised by Prof. Srinivasan Keshav
- Optimization and Vision Transformers for rooftop solar potential estimation from visual/geospatial data

Research in Data Science – ETH Zurich, SRI Lab (Prof. Martin Vechev) Mar 2025 – Dec 2025

- Developed AutoBaxBuilder, an LLM-agent system for generating web-backend code security benchmark tasks
- Implemented functional tests, exploit checks, plausibility filters, and the public AutoBaxBench release

Data Science Lab – ETH Zurich, AI Center (Dr. Imanol Schlag) Sep 2025 – Dec 2025

- Ran controlled 100B+ token pretraining experiments on GH200 GPUs with Megatron-LM and Slurm
- Analyzed left-to-right vs. Fill-in-the-Middle (FIM) memorization via extraction and attention patterns

Bachelor Thesis – ETH Zurich, Boeva Lab (Prof. Valentina Boeva) Feb 2023 – Jun 2023

- *Neural Networks for Signal Deconvolution in Cancer Transcriptomics Data*, [report](#)
- Developed autoencoder and variational autoencoder models for cell-type deconvolution from bulk RNA-seq data

PUBLICATIONS

AutoBaxBuilder: Bootstrapping Code Security Benchmarking [ICML 2026](#)

Tobias von Arx, Niels Mündler, Mark Vero, Maximilian Baader, Martin Vechev

Memorization Dynamics of Fill-in-the-Middle Pretraining [MemFM @ ICML 2026](#)

Tobias von Arx, Tanguy Dieudonné (equal contribution)

WORK EXPERIENCE

Cyber Security Risk Intern – UBS, Third Party Risk Jan 2024 – Jul 2024

- Built LLM-based pipeline to assist human review and reduce effort in grading cyber security questionnaires
- Developed and supported internal cyber security risk assessment and incident management processes

Teaching Assistant – ETH Zurich Feb 2023 – Jun 2023

- Data Modelling and Databases, Prof. Ce Zhang
- Prepared course material and conducted exercise sessions for the entire student cohort (200+ students)

Teaching Assistant – ABZ at ETH Zurich Feb 2022 – Mar 2022

- Taught multiple classes of middle school students about cryptography and self-verifying codes

SKILLS

Languages English (proficient), German (native), French (intermediate), Chinese (intermediate)

Programming Python, C++, Bash, JavaScript, Haskell, Elixir, Go, SQL

Machine Learning PyTorch, torchrun, Megatron-LM, HuggingFace Transformers, Accelerate, vLLM, scikit-learn, NumPy, RAG (retrieval augmented generation), embeddings

Infrastructure FastAPI, Weights & Biases, Runpod, Slurm, Docker, Kubernetes, Git, Linux/Unix