

Tornike Onoprishvili

+41 76 267 53 34 Ticino, Switzerland ✉ tornikeonoprishvili@gmail.com 🌐 [tornikeo](#) in [Tornike Onoprishvili](#) 🌐 [tornikeo](#)

WORK EXPERIENCE

Machine Learning Consultant

Apr 2024 — Present

[Pangea Bio](#)

Remote

Designed **SpectruMS**: a cost-efficient foundation model learning the language of tandem mass spectrometry.

Authored a novel, proprietary 90GB MS/MS pretraining dataset:

- Set up an EC2 instance to scrape 800TB of raw MS/MS data from [GNPS](#) to S3.
- Implemented an efficient curation and de-duplication scripts with Python and Pandas.
- Used AWS Lambda to parse curated raw 100TB data into 90GB parquet files in 5 hours, with 900 processes.

Designed SpectruMS, a proprietary 1B multimodal BART-style model. Specifically:

- Designed a unified chemical vocabulary with tokens for both chemical SMILES and MS/MS data.
- Created a scalable unsupervised pretraining approach to leverage the two unlabeled datasets.
- Set up a distributed pretraining of SpectruMS on Google Cloud TPUs.
- Implemented a supervised GPU-based fine-tuning of SpectruMS using proprietary annotated MS/MS data on a GPU.

Created **SimMS**: GPU-Accelerated Cosine Similarity for Tandem Mass Spectrometry:

- Created a custom CUDA C++ kernel for up to 1000x faster cosine similarity search.
- Re-implemented an existing AWS Batch job, reducing costs per typical invocation by 20x.
- Wrote a paper describing the method (currently in review) [1].

Machine Learning Consultant

Jun 2021 — Present

[Scalexa](#)

Remote

Optimized **VIMAGE** AI back-end on Google Cloud:

- Ported GPU-dependent image processing steps to Vertex AI Endpoint for scalability.
- Accelerated image processing pipeline by 10x via model quantization and low-resolution masked in-painting.
- Provided documentation and training to the VIMAGE team and created a technical [write-up](#) on the project.

Created **AtomicArch**: a frame-differencing computer vision tool for automated nano-structure assembly in Python.

- Leveraged [Roboflow](#) platform to create a baseline video object segmentation (VOS) dataset.
- Explored state-of-the-art VOS models for the nano-structure tracking – due to lack of sufficient data opt for classical computer vision approaches.
- Designed a classical frame-differencing computer vision tool for accurate tracking of nano-structures, using OpenCV and Scikit-Image.

Lecturer

Jun 2020 — Aug 2021

Free University of Tbilisi

Tbilisi, Georgia

- Designed curriculum for an undergraduate course “Introduction to Electrical and Computer Engineering”
- Delivered lectures in computer architecture (nand2tetris) as a graduate student.

Research Assistant

Jun 2017 — Aug 2020

Free University of Tbilisi

Tbilisi, Georgia

- Developed [FDTD simulation software](#) for optical nonlinear photonic crystals in MatLab [2].
- Developed software for optical logical gate simulations in MatLab [3].

OPEN-SOURCE CONTRIBUTIONS

Author, SimMS (github.com/PangeAI/SimMS)

Jun 2024 — Present

- Primary author and maintainer

Contributor, MatchMS (github.com/matchms/matchms)

Jul 2024 — Present

- Performance optimization, profiling and maintenance.

Contributor, DreaMS (github.com/pluskal-lab/DreaMS)

Jun 2024 — Present

- Maintenance and code review

COMPETITIONS

Neural Wave

Nov 2024

Participated in [Neural Wave hackathon](#) in Switzerland, Lugano. Led a team of five in designing a retrieval augmented generator (RAG) chat-bot for Swisscom customer support bot and [winning](#) a 2nd place prize of CHF 2500. More specifically:

- Proposed scraping responses from an existing Swisscom chat-bot to form the sample response dataset for automated chat-bot testing.
- Proposed and implemented an LLM judge that automatically compared our chat-bot responses to the baseline.
- Designed an automated chat-bot test suite that allowed the team to rapidly test the effects of prompt and code modifications in a principled manner.

Published the [code and methods](#) as an open-source package (see [demo](#)). Participated in talks with Swisscom employees on adopting the new chat-bot for use in production.

SKILLS

Programming Languages: Python, CUDA C++, Bash

Technologies: Git, UNIX, Docker, Google Cloud Platform, AWS, Slurm

Python Libraries: PyTorch, Flax, Pandas, HuggingFace Ecosystem, NUMBA, CuPy

EDUCATION

Università della Svizzera italiana <i>Master of Science, Data Science (double degree)</i>	Lugano, Switzerland <i>Aug 2024 — Present</i>
Lappeenranta-Lahti University of Technology <i>Master of Science, Data-centric Engineering</i>	Lappeenranta, Finland <i>Aug 2023 — Present</i>
Free University of Tbilisi <i>Bachelor of Engineering, Electrical Engineering</i>	Tbilisi, Georgia <i>Aug 2015 — May 2019</i>

Bibliography

[1] T. Onoprishvili *et al.*, “SimMS: A GPU-Accelerated Cosine Similarity implementation for Tandem Mass Spectrometry,” *bioRxiv*, 2024, doi: [10.1101/2024.07.24.605006](#).

[2] V. Jandieri, R. Khomeriki, T. Onoprishvili, D. H. Werner, J. Berakdar, and D. Erni, “Functional all-optical logic gates for true time-domain signal processing in nonlinear photonic crystal waveguides,” *Opt. Express*, vol. 28, no. 12, pp. 18317–18331, Jun. 2020, doi: [10.1364/OE.395015](#).

[3] V. Jandieri, T. Onoprishvili, R. Khomeriki, D. Erni, and J. Pistora, “Digital signal processing in coupled photonic crystal waveguides and its application to an all-optical AND logic gate,” *Optical and Quantum Electronics*, vol. 51, no. 4, p. 121, 2019, doi: [10.1007/s11082-019-1833-9](#).