Vittorio Zampinetti PhD

Stockholm, Sweden

data scientist · probabilistic ml · data engineer

zampinetti@gmail.com · +39 333 839 0600 · linkedin.com/in/vzamp · github.com/vzampinetti · vitto.zampi.net

01 profile

Data scientist with a PhD in Applied Mathematics and Computer Science, five years of research experience in machine learning, statistical modelling and data engineering. Experienced in taking poorly defined problems from first principles through method design, implementation, evaluation and open-source release. During my PhD, I independently developed probabilistic inference methods, designed evaluation frameworks where no established benchmarks existed, and built reproducible data pipelines in Python. Four first-author research papers. Looking to apply this combination of analytical rigour and hands-on engineering to building impactful data products.

02 experience

PhD Researcher Politecnico di Torino & KTH, Turin / Stockholm 2021–2026

- Designed, implemented and released an open-source probabilistic inference method for reconstructing tumour evolution from single-cell DNA data.
- Designed evaluation metrics and benchmarks for unexplored problem settings, enabling systematic assessment of competing methods.
- Built Python data pipelines from raw whole-genome sequencing output to versioned, analysis-ready datasets, implementing validation, quality control, unit tests and HPC automation.
- Taught MSc statistics in R for two academic years, plus guest phylogenetics lectures in 2023 and 2024.
- Managed the journal club meetings and admin of the lab's resources documentation.

Visiting Researcher Memorial Sloan Kettering Cancer Center, New York Apr–Aug 2023

- Worked with Sohrab Shah's Computational Oncology Lab, applying our methods to tumor whole-genome sequencing data in a leading computational oncology research environment.

03 education

PhD, Applied Mathematics and Computer Science Politecnico di Torino & KTH 2021–2026

Joint doctoral programme between two major technical universities. Thesis: *Single-cell tumor phylogenetics*.

MSc, Data Science and Engineering Politecnico di Torino & Politecnico di Milano 2019–2021

Maximum grade with honours. Selected for Alta Scuola Politecnica (ASP), the joint Milano-Torino multidisciplinary honours programme for 150 MSc students per year. One-year Erasmus exchange at KTH.

BSc, Computer Engineering Università di Firenze 2016–2019

Maximum grade with honours. Thesis: *Automated brain cell measurement in 3D images*.

04 publications

- Scalable and robust phylogenetic tree reconstruction from copy-number data with Sparse Rooted Neighbor Joining. bioRxiv 2026.
- Cellmates: a method for single-cell phylogeny reconstruction with proper evolutionary distances from copy numbers. pre-print 2026.
- Sampling random spanning arborescences in graphs with low conductance. Statistics and Probability Letters, 2025.
- VICTree - A variational inference method for clonal tree reconstruction. RECOMB 2024.

05 skills

Programming and data Python · R · SQL · C++ · Java · Bash
Machine learning and statistics Probabilistic modelling · statistical modelling · PyTorch · JAX
Software and infrastructure Data pipelines · Snakemake · Docker · Linux · HPC · unit testing
AI-assisted development Claude Code · Codex · OpenCode · Herdr
Languages Italian (native) · English (fluent) · Swedish (A2)

Beyond work: Custom PC building, technical tutorials and eight years of capoeira.