

Employment

GSK

2024 – present **Senior AI/ML Engineer**, Causal Machine Learning

2023 – 2024 **AI/ML Engineer**, Causal Machine Learning

Dana Farber Cancer Institute & Harvard Medical School

2022 – 2023 **Instructor** (faculty), Medical Oncology, Division of Population Sciences

2018 – 2022 **Research Fellow**

Broad Institute of MIT and Harvard

2019 – 2023 **Postdoctoral Scholar** (affiliate), Medical & Population Genetics

Education

2014 – 2018 **University of Oxford**

- **DPhil (PhD), Theoretical Physics**
- Weak integrability breaking and full counting statistics
- Supervisor: Dr. Fabian Essler

2013 – 2014 **Rutgers, The State University of New Jersey**

- **MSc, Physics**, GPA: 4.0
- Interaction and external field quantum quenches in the Lieb-Liniger and Gaudin-Yang model
- Supervisor: Dr. Natan Andrei

2010 – 2013 **Julius-Maximilians University Würzburg**

- **BSc, Physics** with distinction (4.0 equivalent)

Research Experience

(Senior) AI/ML Engineer
GSK
Apr 2023 – present
I lead research in causal machine learning and generative models for biomarker discovery, target identification, and drug response prediction. My work leverages large-scale multi-omic datasets, genetic data, medical imaging, electronic health records, and clinical trial data to uncover causal biological mechanisms. I bring extensive expertise in bulk, single-cell, and spatial RNA sequencing, as well as perturbation data analysis, combining state-of-the-art machine learning with classical statistical methods.

Research Fellow
Dana-Farber Cancer Institute
Oct 2018 – Mar 2023
I developed advanced statistical and machine learning methods to investigate genetic mechanisms underlying complex diseases, with emphasis on population and statistical genetics, disease progression modeling, response prediction, and time-to-event analysis. I bring extensive expertise in real-world clinical, genetic, and EHR data, as well as multi-modal omics data. Through collaborations with pharmaceutical and machine learning companies, I've applied cutting-edge techniques to clinical trial and real-world data.

PhD Theoretical Physics
University of Oxford
Oct 2014 – July 2018
My PhD research focused on the intersection of theoretical condensed matter physics, non-equilibrium statistical mechanics, and quantum computing. I developed expertise in advanced analytical techniques including quantum field theory, quantum statistical mechanics, and exact algebraic methods, alongside computational approaches for many-body physics such as exact diagonalization and tensor product methods.

Skills

Modern ML methods • NODEs • Flow matching/Diffusion models • VAEs • Transformers
Statistical methods • causal inference • TMLE • survival analysis • state-space models • Bayesian statistics
Programming languages • Python • R • Julia • C/C++ • SQL • bash
ML frameworks • pytorch • jax • lightning • pyro • equinox • diffraction • numpyro • scikit-learn
Data science frameworks • pandas • numpy • scipy • luigi • prefect • hydra • data.table • tidyverse • survival
Software development • git • docker • gcloud • CI/CD
Computational biology • scanpy • seurat • scvi • plink • samtools • hstlib • WDL
Data experience • EHR • RNAseq • scRNAseq • perturb seq • spatial tx • DNAseq
Physics methods • quantum mechanics • statistical mechanics • tensor product methods

Scholarships and Honors

- 2025 Bronze Recognition Award, GSK
- 2024 Individual Translational Medicines and Vaccines Award 2024, GSK
- 2024 Bronze Recognition Award, GSK
- 2021 – 2023 DFCI Trustee Science Committee Fellowship
- 2021 Top Reviewer Award at ML4H 2021
- 2020 ASHG Reviewer's Choice
- 2020 ESHG Young Investigator Award Candidate
- 2014 – 2018 University of Oxford Clarendon Scholarship (top 1.8% of admitted graduate students)
- 2014 – 2017 Santander Graduate Award
- 2013 – 2014 Fellowship from the DAAD (German Academic Exchange Service)
- 2012 – 2014 Fellowship of the Studienstiftung des deutschen Volkes
- 2010 Exceptional prize of the German Physical Society (DPG) at the German finals "Jugend forscht" (biggest youth science competition in Europe) 2010, physics section
- 2010 Finalist of the German selection for the International Physics Olympiad 2010 (11th place).

Talks and Poster Presentations

- 2023 Invited talk: IARC Collider bias and Mendelian randomization (CMR) working group
- 2022 Poster presentation: Time series for Health, NeurIPS workshop
- 2022 Invited talk: NCI Immuno-Oncology Translational Network Bioinformatics and Computational Biology Working Group
- 2022 Invited talk: American Conference on Pharmacometrics, Denver Colorado
- 2022 Poster presentation: American Society for Human Genetics Conference
- 2022 Invited talk: Invitae research seminar
- 2021 Invited talk: ML4H seminar series, Broad Institute of MIT and Harvard
- 2021 Oral presentation: AAAI Symposium 2021 on survival prediction
- 2021 Invited talk: Probably Genetic Research Forum
- 2021 Invited talk: Modeling & Simulation Forum, Genentech
- 2020 Poster Presentation: Machine Learning for Health (ML4H), NeurIPS workshop
- 2020 Oral, poster presentation: Learning Meaningful Representations of Life, NeurIPS workshop
- 2020 Invited talk: UCLA/UChicago joint journal club on statistical genetics
- 2020 Poster presentation: American Society for Human Genetics Conference
- 2020 Oral presentation: European Society for Human Genetics Conference
- 2019 Poster presentation: Learning Meaningful Representations of Life, NeurIPS workshop
- 2019 Poster presentation: Harvard PQG, Quantitative Challenges in Cancer Immunology and Immunotherapy
- 2019 Poster presentation: American Society for Human Genetics Conference
- 2018 Invited talk: Conference "Quantum Paths", Erwin Schrödinger Institute, Vienna
- 2018 Invited talk: Rudolf Peierls Centre for Theoretical Physics, University of Oxford
- 2017 Invited talk: Brookhaven National Laboratory

Selected Publications

- "A comprehensive analysis of clinical and polygenic germline influences on somatic mutational burden", Kodi Taraszka, **Stefan Groha**, et al, The American Journal of Human Genetics, 2024
- "Discovery of disease-associated cellular states using ResidPCA in single-cell RNA and ATAC sequencing data", Shaye Carver, Kodi Taraszka, **Stefan Groha**, Alexander Gusev, bioRxiv, 2024
- "SurvivAE: Variational Autoencoders for Clustering Time Series", **Stefan Groha**, Alexander Gusev, Sebastian Schmon, Proceedings of Learning from Time Series for Health, 2022
- "Common germline variants associated with immunotherapy-related adverse events", **Stefan Groha** et al, Nature Medicine, 28, pages 2584–2591 (2022)

- “SurvLatent ODE: A Neural ODE based time-to-event model with competing risks for longitudinal data improves cancer-associated Venous Thromboembolism (VTE) prediction”, Intae Moon, **Stefan Groha**, Alexander Gusev, Machine Learning for Healthcare Conference, 800-827 (2022)
- “Constructing germline research cohorts from the discarded reads of clinical tumor sequences”, Alexander Gusev, **Stefan Groha**, et al, Genome medicine 13, 1-14 (2022)
- “Automated identification of immune related adverse events in oncology patients using machine learning.”, Wenxin Xu, Alexander Gusev, **Stefan Groha**, et al, Journal of Clinical Oncology 39, 1551-1551 (2021)
- “Clinical inflection point detection on the basis of EHR data to identify clinical trial-ready patients with cancer”, Kenneth Kehl, **Stefan Groha**, et al, JCO Clinical Cancer Informatics 5, 622-630 (2021)
- “Topological Data Analysis of Copy Number Alterations in Cancer”, **Stefan Groha**[†], Caroline Weis[†], Alexander Gusev, Bastian Rieck, Learning Meaningful Representations of Life Workshop. Neural Information Processing Systems (NeurIPS) (2020), arxiv:2011.11070
- “A General Framework for Survival Analysis and Multi-State Modelling”, **Stefan Groha**[†], Sebastian Schmon[†], Alexander Gusev, arxiv:2006.04893
- “Full Counting Statistics in the Transverse Field Ising Model after a Quantum Quench”, **Stefan Groha**, Fabian H. L. Essler, Pasquale Calabrese, SciPost Phys. 4, 043 (2018)
- “Full counting statistics in the spin-1/2 Heisenberg XXZ chain”, Mario Collura[†], Fabian H. L. Essler[†], **Stefan Groha**[†], J. Phys. A Math. Theor. 50 414002 (2017)
- “Spinon decay in the spin-1/2 Heisenberg chain with weak next nearest neighbour exchange”, **Stefan Groha**, Fabian H. L. Essler, J. Phys. A Math. Theor. 50 334002 (2017)
- “Thermalization and light cones in a model with weak integrability breaking”, Bruno Bertini[†], Fabian H. L. Essler[†], **Stefan Groha**[†], and Neil J. Robinson[†], Phys. Rev. B 94, 245117 (2016)
- “Prethermalization and thermalization in models with weak integrability breaking”, Bruno Bertini[†], Fabian H. L. Essler[†], **Stefan Groha**[†], and Neil J. Robinson[†], Phys. Rev. Lett. 115, 180601 (2015)

Professional Service

- Program Chair Assistant (Organising Committee) for NeurIPS 2024
- Reviewer for NeurIPS 2022, 2024, 2025
- Reviewer for ICLR 2022, 2025
- Reviewer for AISTATS 2025
- Reviewer for ML4H 2021, 2022, 2023, 2024
- Reviewer for workshop on Spurious Correlations, Invariance, and Stability 2022, 2023
- Reviewer for workshop on Time Series for Health 2022, 2024
- Reviewer for workshop on Causal Representation Learning 2023
- Reviewer for AAAI Symposium 2021 on survival prediction
- Reviewer for eLife
- Reviewer for Journal of the American Statistical Association
- Reviewer for Journal of Statistical Mechanics: Theory and Experiment