Darius A. Faroughy, Ph.D

Jersey City, NJ 07302, USA • (848) 313-3093 • darius.faroughy@rutgers.edu • GitHub • Google Scholar • LinkedIn

Citizenships: *USA | Slovenia* • ***Languages:*** *English, Spanish, French*

I build generative AI for scientific discovery. Trained as a theoretical particle physicist, I now research at the intersection of generative modeling and the natural sciences, with a focus on flow-matching, diffusion, foundation models, and agentic systems for complex scientific data. Recent contributions include introducing the first multimodal generative model for collider data, developing the first foundation model trained on real data from the Large Hadron Collider (LHC), and demonstrating neural scaling laws for jet generation. Beyond physics, I led the development of a generative foundation model for pharmacokinetics time-series. My latest research explores generative AI in three directions:

- Flow-matching models for large-scale structure of the Universe and other cosmological data.
- LLM-driven evolutionary search for *de novo* design of synthesizable molecules.
- Autonomous agents for scientific reasoning and discovery.

Technical Skills

- **ML:** *Generative AI, Transformers, Flow-Matching, Diffusion, Neural Operators, Time-series, LLMs, Agentic AI*
- **Languages & Frameworks:** *Python, PyTorch, Lightning, HuggingFace, scikit-learn, NumPy, C++.*
- **Computing:** *Bash, Git, Distributed GPU, SLURM, Comet.*

Experience

Postdoctoral Researcher (2022 – Present) — *Rutgers University, New High Energy Theory Center, NJ, USA*

- **My Chemical Harness**. Built the agentic harness and ran the molecular-design experiments for *My Chemical Harness*, an LLM-guided evolutionary framework over executable synthetic routes. On a soluble epoxide hydrolase proxy task, our agentic algorithm reached state-of-the-art performance across multiple molecular-design metrics, including sEH score, synthetic accessibility, and AiZynthFinder success rate, surpassing dedicated synthesis-aware generative baselines such as NVIDIA's ReaSyn.
- **Foundation model for Pharmacokinetics**. Led a multidisciplinary collaboration with mathematics and pharmacology groups to develop a generative foundation model for pharmacokinetic time-series, enabling zero-shot population synthesis and individual drug-response forecasting. Our method outperforms prior generative approaches..
- **Collider-Bench**. Led the design and release of an agentic-AI benchmark grounded in real scientific tasks, evaluating autonomous reproduction of experimental particle physics analyses at the LHC.
- **Multimodal flow-matching for particle clouds.** Built the first multimodal generative model for point-cloud data from hadron colliders, jointly modeling continuous kinematics with flow-matching and discrete particle attributes (charge, flavor) via Markov jump processes. The resulting transformer-based model achieves current state-of-the-art.
- **Foundation model for Particle Physics — Aspen Open Jets**. Led the development of the first foundation model trained on real Large Hadron Collider (LHC) data, alongside *AspenOpenJets* the first large-scale real-world LHC dataset for ML.
- **Neural scaling laws for jet generation**. Co-led the first systematic study of neural scaling laws for generative models trained on real LHC data, establishing predictable performance gains with model size.

Postdoctoral Researcher (2019 – 2022) — *University of Zürich, Physik-Institut, Zürich, Switzerland*

- **NLP and Bayesian models for anomaly detection.** Applied natural-language-processing techniques to tokenized jet substructure and Bayesian probabilistic modeling to anomaly detection at colliders. Contributed to the LHC Olympics 2020 community anomaly-detection challenge.
- **Effective field theories.** Classified flavor operators in the Standard Model Effective Field Theory under various flavor symmetries, and developed methods combining low- and high-energy constraints on flavorful new physics.
- **Flavor physics.** Investigated the interplay between low-energy flavor observables and high-momentum LHC searches, applied to leptoquark and Pati-Salam-like UV-complete models proposed as solutions to the B-physics anomalies.
- **HighPT.** Led the development of *HighPT*, an open-source Mathematica tool for global analyses of high-momentum Drell-Yan scattering within the Standard Model Effective Field Theory and beyond. Our code is widely adopted by the community for new-physics searches at the LHC.

Teaching Assistant (2019 – 2022) — *ETH Zürich & University of Zürich, Switzerland*

- Teaching assistant for 6 semesters for master's-level courses in Quantum Field Theory I & II, Particle Physics, and the Standard Model of Particle Physics.

Education

- **Ph.D.** in Theoretical Particle Physics *(2014 – 2019)* — *Jozef Stefan Institute, Ljubljana, Slovenia*
- **Postgraduate Diploma** in High Energy Physics *(2011 – 2013)* — *ICTP, Trieste, Italy*
- **B.Sc.** in Physics *(2005 – 2010)* — *Universidad Simon Bolivar, Caracas, Venezuela*

Selected Papers

Applied ML

- ***My Chemical Harness: Evolutionary Molecular Design over Synthetic Pathways with Large Language Model Agents***
  C. Ojeda, D. A. Faroughy, M. Karimi, P. Zarrintaj, M. M.i Seyedebrahimi, and M. Carballo-Pacheco
  arXiv preprint: 2606.11256.
- ***Neural scaling laws for jet generation*** *
  O. Amran, D. A. Faroughy, T. Gerdes, A. Hallin, G. Kasieczka, M. Krämer, H. Reyes, D. Shih.
  arXiv preprint: 2605.28940.
- ***Collider-Bench: Benchmarking AI Agents with Particle Physics Analysis Reproduction***
  D. A. Faroughy, S. Palacios Schweitzer, I. Pang, S. Mishra-Sharma, D. Shih.
  NeurIPS 2026 (submitted) – arXiv preprint: 2605.13950.
- ***Prior-Fitted Functional Flows: In-Context Generative Models for Pharmacokinetics***
  C. Ojeda, N. Hartung, W. Huisinga, T. Jahn, P. Kavwele, M. Klose, P. Kumar, R. J. Sánchez, D. A. Faroughy
  UAI 2026 – arXiv preprint: 2604.17670.
- ***Multimodal Generative Flows for LHC Jets***
  D. A. Faroughy, M. Opper, C. Ojeda.
  NeurIPS 2025, ML4PS – arXiv preprint: 2509.01736.
- ***Aspen Open Jets: Unlocking LHC Data for Foundation Models in Particle Physics*** *
  O. Amram, L. Anzalone, J. Birk, D. A. Faroughy, A. Hallin, G. Kasieczka, M. Krämer, I. Pang, D. Shih
  Mach. Learn. Sci. Tech. 6 (2025) 3, 030601 – arXiv preprint: 2412.10504.
- ***EPiC-ly Fast Particle Cloud Generation with Flow-Matching and Diffusion*** *
  E. Buhmann, C. Ewen, D. A. Faroughy, T. Golling, G. Kasieczka, M. Leigh, J. A. Raine, D. Sengupta, D. Shih
  arXiv preprint: 2310.00049.
- ***The Machine Learning Landscape of Top Taggers***
  G. Kasieczka, T. Plehn, A. Butter, K. Cranmer, B. M. Dillon, D. A. Faroughy, et al.

SciPost Physics, 7, 014 (2019) – arXiv preprint: 1902.09914

Physics:

- ***Surfing to the Fundamental Limit of Jet Tagging***
  I. Pang, D. A. Faroughy, D. Shih, R. Das, G. Kasieczka
  arXiv preprint:2511.15779.
- ***On the Effective Field Theory validity for Drell–Yan tails at the LHC*** *
  L. Allwicher, D. A. Faroughy, M. Martines, O. Sumensari, F. Wilsch
  European Physical Journal C, 85 (2025) – arXiv preprint: 2412.14162.
- ***Drell-Yan Tails Beyond the Standard Model*** *
  L. Allwicher, D. A. Faroughy, F. Jaffredo, O. Sumensari, F. Wilsch
  Journal of High Energy Physics, (2023)064 – arXiv preprint: 2207.10714
- ***HighPT: A tool for high-pT Drell-Yan tails beyond the standard model*** *
  L. Allwicher, D. A. Faroughy, F. Jaffredo, O. Sumensari, F. Wilsch
  Computational Physics Communications, 289 (2023) 108749 – arXiv preprint: 2207.10756
- ***Collider signatures of coannihilating dark matter in light of the B-physics anomalies*** *
  M. J. Baker, D. A. Faroughy, S. Trifinopoulos
  Journal of High Energy Physics, 11 (2021) 084 – arXiv preprint: 2109.08689
- ***Flavour symmetries in the Standard Model Effective Field Theory*** *
  D. A. Faroughy, G. Isidori, F. Wilsch, K. Yamamoto
  Journal of High Energy Physics, 08 (2020) 166 – arXiv preprint: 2005.05366
- ***Uncovering new physics patterns in collider events with Bayesian probabilistic models***
  D. A. Faroughy, ICHEP 2020 – arXiv preprint: 2012.08579.
- ***Uncovering latent jet substructure*** *
  B. Dillon, D. A. Faroughy, J. F. Kamenik
  Physical Review D, 100 (2019) 5, 056002 – arXiv preprint: 1904.04200
- ***Closing the window on single leptoquark solutions to the B-physics anomalies*** *
  A. Angelescu, D. Bečirević, D. A. Faroughy, O. Sumensari
  Journal of High Energy Physics, 10 (2018) 183 – arXiv preprint: 1808.08179.

(*) authors listed alphabetically