Arian Amani Machine Learning Scientist

ArianAmaani@gmail.com +39 328 361 8263 Rome, Italy
linkedin.com/in/arianamani ArianAmani ArianAmani.com

Profile

Machine Learning Scientist with 4 years of research experience building **conditional generative models** for **single-cell perturbation biology** and **molecular drug design**. Co-first author of **CellDISECT** (bioRxiv 2025, under review at Nature Methods) and co-author of **SP-FM** (arXiv 2026), contributing novel methods in **causal VAEs**, **flow matching**, and **out-of-distribution generalization** to unseen biological conditions. Core technical stack: PyTorch, Transformers, VAEs, diffusion models, flow matching, **scRNA-seq** analysis (Scanpy/scVI), and **cheminformatics** (SMILES, RDKit, AutoDock Vina/Boltz-2/BioSolveIt). Experienced deploying ML pipelines at scale on **GCP**.
Research goal: develop generalizable AI systems that bridge **cellular perturbation modeling** and **de novo molecular generation** for **target-driven therapeutic design**.

Professional Experience

AI VIVO
Machine Learning Scientist
12/2024 – present
Cambridge, UK
(Remote)

- Developed **deep learning and generative models** for **drug discovery**, leveraging **transformer and flow matching-based architectures** for molecular representation and generation using **SMILES**-encoded inputs and **ADMET** property constraints.
- Deployed and scaled deep learning pipelines on **Google Cloud (GCP)** using **PyTorch Lightning** and **Dockerized** workflows, supporting concurrent experiment iterations across multiple molecular generative model configurations.
- Built **multi-modal ML pipelines** fusing **SMILES**-encoded molecular structures with HTS biological assay readouts for **mechanism-of-action classification** across >1 billion compound-target pairs (PyTorch, RDKit, HuggingFace Transformers).
- Applied **structure-based drug design** workflows using **AutoDock Vina** and **Boltz-2** across >10,000 target proteins, integrating predicted binding affinities as reward signals in **generative model** training loops.

Wellcome Sanger Institute
Machine Learning Researcher
11/2022 – present
Hinxton, UK
(Remote)

- Co-First Author of **CellDISECT**, a **causal VAE** framework for **disentangled single-cell representation** and **in silico perturbation prediction**, trained on 2M+ **scRNA-seq** cells across 100+ counterfactual conditions; submitted to peer-reviewed journal (under review at Nature Methods).
- Drive collaborative research outputs across a >10-member PhD/Postdoc team at Lotfollahi Lab, co-authoring 2 preprints and maintaining core open-source tools used by the broader **single-cell genomics** community.
- Fine-tuned 300M-parameter single-cell **Transformer Foundation Models** (scFoundation, Geneformer, scGPT) using **LoRA** and **P-Tuning** on 1500+ perturbation conditions, improving downstream perturbation prediction via **self-supervised** pre-training.
- Took on the role of **lead maintainer** for **CPA** (Compositional Perturbation Autoencoder; ~84K+ downloads),
- Merged 70+ PRs and reviewed 500+ commits in 2025, maintaining widely-adopted **single-cell perturbation modeling** libraries.

Virasad
Computer Vision Engineer
01/2022 – 05/2022
Tehran, Iran

- Delivered >95% accuracy computer vision solutions across 5 client projects with limited training data ($\leq$15 images/class), leading 2 end-to-end pipelines including data augmentation and model deployment

Publications

Shortest-Path Flow Matching with Mixture-Conditioned Bases for OOD Generalization to Unseen Perturbations 2026
arXiv:2601.11827 [cs.LG]

Proposed SP-FM, a conditional **flow matching** framework with learnable condition-dependent base distributions that achieves robust **out-of-distribution generalization** to unseen biological perturbations. Introduces shortest-path interpolation with mixture-conditioned bases, outperforming standard conditional flow matching baselines on held-out perturbation generalization benchmarks.

Integrating multi-covariate disentanglement with counterfactual analysis on synthetic data enables cell type discovery and counterfactual predictions 2025
bioRxiv 2025.06.03.657578 **Co-First Author** *(Under review at Nature Methods)*
Developed CellDISECT, a **causal VAE** with expert-module architecture enabling **covariate-aware disentanglement** across single-cell populations, separating batch effects from biologically meaningful latent structure. Enables **counterfactual in silico perturbation prediction** and unsupervised cell type discovery across >1000 perturbation conditions and 50+ cell types from **scRNA-seq** data.

Workshops & Presentations

University of Tehran *RSG Iran*
- **ML in Single-Cell Genomics Workshop** (Instructor) – Oct 2024

Amirkabir University of Technology *MCS Talk*
- **Introduction to Adversarial Machine Learning** (Speaker) – May 2023

Science Communication

Why Perturbation Prediction Needs Much Better Metrics 2026
Medium — Essay on evaluation metrics for single-cell perturbation prediction, mean-predictor baselines, and implications for benchmarking in computational biology.

Leveraging Machine Learning to Predict Cellular Behavior in Drug Treatments 2024
Medium — Review of the current state of ML in drug discovery for a general scientific audience.

A Deep Learning Road Map And Where To Start 2022
Medium — Practical guide to entering deep learning, drawing from personal experience.

Skills

ML & Research: Generative Models, VAEs, Diffusion Models, Flow Matching, Transformers, Representation Learning, Perturbation Modeling, scRNA-seq Analysis, Causal Inference, Multi-omics, ADMET Property Prediction, Cheminformatics (SMILES, RDKit), Self-Supervised Learning, Drug Discovery, Single-Cell Genomics, Benchmarking, Out-of-Distribution Generalization

Libraries & Frameworks: PyTorch, PyTorch Lightning, HuggingFace Transformers, Scanpy, scVI, RDKit, Scikit-Learn, Matplotlib, BioSolveIt, AutoDock Vina, Boltz-2

Engineering: Python, C++, Google Cloud Platform (GCP), Git, Linux, Docker

Education

BSc, Applied Computer Science & Artificial Intelligence 09/2023 – 06/2026
Sapienza University of Rome Rome, Italy
GPA: 27.06/30

Bachelor's degree, Computer Science 09/2020 – 06/2023
Amirkabir University of Technology Tehran, Iran
GPA: 17.39/20 (~3.7/4.0 equivalent). Completed 65 credits out of 134 before transferring to Sapienza University of Rome in 2023.

Languages

English
- Full professional proficiency
- IELTS Overall 8.0/9.0

Italian
- Elementary proficiency

Persian
- Native or bilingual proficiency

Teaching Experience

Sharif University of Technology
- Machine Learning for Bioinformatics (Graduate Course) – Spring 2023
- Introduction to Machine Learning – Fall 2022

Amirkabir University of Technology
- Introduction to Image Processing and Neural Networks – Fall 2022
- Advanced Programming with C++ – Spring 2022