Arian Amani *Machine Learning Scientist*

ArianAmaani@gmail.com Rome, Italy linkedin.com/in/arianamani
ArianAmani ArianAmani.com

Profile

Machine Learning Scientist working at the intersection of **computational biology** and **drug discovery**, developing **deep generative and foundation models** for molecules and cells. Experienced with **molecule generation** and **single-cell perturbation modeling (VAEs, Diffusions, Transformers, GNNs, FlowMatchings)**, and **representation learning** for biological and chemical systems. Passionate about building AI methods that accelerate target discovery and therapeutic design.

Professional Experience

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

- Develop **deep learning and generative models** for **drug discovery**, leveraging **transformer and flow matching-based architectures** for molecular representation and generation.
- Deploy and scale deep learning pipelines on **Google Cloud (GCP)** using **PyTorch Lightning** and **Dockerized** workflows.
- Design **multi-modal ML pipelines** integrating molecular structure and biological assay for **target mechanism prediction**.
- Maintain scalable ML pipelines using **PyTorch, Lightning**, **RDKit**, and **HuggingFace Transformers**.
- Gained experience using computational chemistry software including commercial and open-source molecular docking and design tools (e.g., **BioSolveIt**, AutoDock **Vina, Boltz-2**).

Wellcome Sanger Institute 11/2022 – present
Data Scientist Hinxton, UK (Remote)

- Co–first author of **CellDISECT**, a **deep generative model** for **disentangled cellular representations** and **in silico perturbation analysis**, developed to study **perturbation effects** across single-cell populations.
- Conducting research at Lotfollahi Group alongside >10 PhD students and Postdocs
- **Fine-Tuned** 40 million parameter Transformer **Foundation Models**: LoRA, P-Tuning
- Contributed to projects such as **CPA:** Compositional Perturbation Autoencoder (**~60 Commits** and maintaining the repository)
- More than **60** reviewed and **merged Pull Requests** | **500** reviewed **commits** in 2024

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

- Delivered >95% accuracy solutions for tasks with limited data (15 images per class)
- Worked on 5 diverse projects meeting client requirements
- Led 2 individual projects, enhancing development pipelines for data augmentation

Publications & Blogs

Shortest-Path Flow Matching with Mixture-Conditioned Bases for OOD Generalization to Unseen Conditions 2026
arxiv
Developed SP-FM, a novel conditional flow-matching framework that achieves robust out-of-distribution generalization by learning condition-dependent base distributions.

Integrating multi-covariate disentanglement with counterfactual analysis on synthetic data enables cell type discovery and counterfactual predictions (Co-First Author) 2025
bioRxiv

- Developed CellDISECT as a first author, a novel causal generative model for single-cell analysis that disentangles covariate effects and enables counterfactual predictions.
- Achieved flexible fairness through expert models, capturing both covariate-specific information and new biological insights.
- Enhanced cell type discovery and biological interpretation using multi-covariate disentanglement and advanced counterfactual analysis.github
- https://github.com/Lotfollahi-lab/CellDISECT

Leveraging Machine Learning to Predict Cellular Behavior in Drug Treatments 2024
- Wrote a Medium article, reviewing the current state of ML in Drug Discovery

A Deep Learning Road Map And Where To Start 2022
- Shared experiences: The Deep Learning Road Map That I Took

Skills

Key Skills
- Generative Machine Learning
- Deep Learning
- Drug Discovery
- Single-Cell Genomics
- Drug Discovery

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

Programming & Engineering
- Python, C++
- Google Cloud Platform (GCP)
- Git, Linux, Docker
- Probability and Statistics
- Linear Algebra

Education

Bachelor's degree, Applied Computer Science & Artificial Intelligence 09/2023 – 06/2026
Sapienza University of Rome Rome, Italy

Bachelor's degree, Computer Science 09/2020 – 06/2023
Amirkabir University of Technology Tehran, Iran
GPA: 17.39/20, Completed 65 credits out of 134 before transferring to Rome
Teaching Assistant: ML for Bioinformatics (Masters) | Introduction to ML | C++ Programming

Certificates

Deep Learning Specialization
Coursera

Upwork Skill Certification - Machine Learning
Certified freelancer with proficiency in applied machine learning

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