

Samira Jahangiri

Samira.Jahangiri@ucf.edu — Website — GitHub

AI/ML Engineer — Data Scientist — Bioinformatics Engineer

About Me

PhD candidate with 8+ years of interdisciplinary experience in machine learning (transformers, GNNs), big data analytics, and large-scale genomic/proteomic analysis. Proven ability to bridge biotech and AI through projects in bioinformatics, NLP for biomedical text, and federated learning. Strong statistical foundation with applied expertise in Python/R and cloud computing, with proven ability to adapt domain-specific knowledge to new problem spaces. For detailed project portfolios: samirajahangiri.github.io

Technical Skills

- **Programming:** Python (PyTorch, TensorFlow, scikit-learn), R (tidyverse, Bioconductor), SQL, Bash, Julia Lang, \LaTeX
- **AI/ML:** Transformer models (text embedding), GNNs (3D pose estimation), RNN/CNNs (genomic prediction), federated learning
- **Data Engineering:** AWS (S3, EC2), NGS pipelines (RNA-seq/WGS), Docker, PowerBI, Azure
- **Geometric ML:** Matrix algebra (SVD, eigendecomposition), manifold learning, graph embeddings

Professional Experience

Bioinformatics Specialist *Howard Hughes Medical Institute* Mar 2022 – Jul 2023

- Led NGS data analysis for neurological disease studies, creating end-to-end pipelines using SQL and tree-based models to prioritize causal variants from RNA-seq and WGS data

Genomics Data Engineer *Shriners Hospital for Children* Jul 2021 – Feb 2022

- Migrated petabyte-scale WGS data between AWS clouds (S3 → Glacier), reducing storage costs significantly
- Automated ETL workflows for pediatric precision medicine research (cerebral palsy, scoliosis)

Data Scientist *USF Omics Hub* Aug 2018 – Dec 2021

- Built R Shiny dashboards for microbiome/RNA-seq analysis (16S, metagenomics) used by Florida DOH
- Created statistical power calculator for microbiome studies (PERMANOVA effect size estimation)

Bioinformatics Consultant *Pars Silico Bioinformatics (Tehran)* August 2015 – June 2018

- Enhanced an in-house molecular docking algorithm by integrating Cytoscape network analysis with DrugBank data to support pharmaceutical R&D and compound repurposing workflows.

Research Experience

PhD Candidate in Big Data Analytics *University of Central Florida* 2023 – Present

- Developing complex-valued NLP models for biomedical text embedding, capturing semantic relationships via phase angles
- Applied GNNs/LSTMs to 3D pose estimation in autism research (privacy-preserving federated learning)

Graduate Research Assistant *University of South Florida* 2018 – 2021

- Developed AI models (RNNs/CNNs) to identify epigenetic drug-resistance markers in malaria

Graduate Research Assistant *University of Tehran* 2015 – 2018

- Integrated compound screening data with structural features to prioritize candidate molecules for antimicrobial targeting.

Education

- **PhD in Big Data Analytics** / Statistics track *University of Central Florida* 2022 – Present
- **MS in Public Health** (Genomics/Biostatistics) *University of South Florida* 2018 – 2021
- **MS in Biotechnology** *Azad University* 2014 – 2016

Publications

- Full list: [Google Scholar](#)