Matt Ebrahim

Senior Data Scientist — Generative AI — Drug Discovery — Biotech
m.ebrahimkhani1993@gmail.com — (631) 275-5369
linkedin.com/in/matt-ebrahim github.com/matt-ebrahim Google Scholar Portfolio

Research & Industry Experience

Senior Data Scientist, Formation Bio – *Jun 2025 – Present*

- Fine-tuned domain-specific LLMs (BioMistral, BioMegatron, SapBERT, etc.) for ontology mapping framed as a named-entity-recognition task.
- Expanded the SPOKE biomedical knowledge graph and developed GNN models for indication–asset mapping to accelerate portfolio decisions.

AI Scientist II, 1910 Genetics – *2023 – May 2025*

- Led design of generative-AI models (DDPMs, GNNs, Transformers) for molecular generation and property optimization.
- Built E(3)-equivariant autoregressive models and SMILES-based RNNs with RL fine-tuning for CNS-targeted drug discovery.
- Integrated QM-derived features into graph models for improved ADMET prediction; benchmarked against Graphormer, AttentiveFP, and QM cross-attention embeddings.
- Curated datasets (CrossDocked, ChEMBL, MOSES) and deployed pipelines on AzureML.

Lecturer, Northeastern University – *Spring 2025*

- Designed and taught CSYE 7374: *Applied Deep Learning & Generative Models in Healthcare.*
- Covered diffusion models, transformers, GNNs, protein-ligand modelling, and molecular generation.
- Course site: CSYE 7374-2025

Postdoctoral Researcher, Northwestern University – *2022 – 2023*

- Developed DL pipelines to estimate aortic blood flow from wearable SCG signals.
- Built CycleGAN + CNN-MLP systems for non-invasive cardiovascular diagnostics.

Ph.D. Research, Stony Brook University – *2017 – 2022*

- Developed signal- and image-processing ML models for THz imaging and burn diagnostics.
- Published 8+ first-author papers and co-authored 30+ peer-reviewed publications.

Technical Skills

Languages: Python, C++, MATLAB, Bash
AI/ML Frameworks: PyTorch, TensorFlow, Scikit-learn, Optuna, PyCaret
Generative Modelling: Diffusion (DDPM), RNN, Transformer (GPT, T5), VAE, GAN
Molecular Modelling: RDKit, DeepChem, AutoDock Vina, Open Babel, ESMFold, AlphaFold
Graph ML: Graphormer, GCN, GAT, MPNN, DGL, PyG
Cloud/Deployment: Azure ML, AWS, Docker, SLURM, Git, VS Code
Domains: Drug discovery, ADMET modelling, protein-ligand design, medical imaging, spectroscopy

Education

Ph.D., Biomedical Engineering – Stony Brook University, 2022
Postdoc, AI for Medical Imaging – Northwestern University, 2023
B.Sc., Electrical Engineering – Amirkabir University of Technology, 2016

Select Publications & Patents

- Multimodal Graph-Attention Networks with QM-Guided Cross-Attention for ADMET Prediction — first author, submitted to NeurIPS 2025.
- Deep Learning for Aortic Flow Estimation from SCG — *Annals of Biomedical Engineering*, 2023.
- US Patent (2025): *Personalized Chest Acceleration . . . Using Deep Learning.* freepatentsonline.com/y2025/0040894.html
- Deep Learning for Triage of *in vivo* burn injuries - *Biomed. Opt. Express*, 2022.
- MolGPT + Reinforcement Learning — github.com/matt-ebrahim