Le Li

Winchester, MA | lile.moziya@gmail.com | (607) 280-5283 | Homepage: moziya.github.io/moziya
GitHub: /moziya | LinkedIn: /leli-Comp | Google Scholar: /citations?user=fmtx_vUAAAAJ

SUMMARY

ML engineer-leaning scientist with ~10 years of experience in designing, developing, and deploying deep learning and statistical models in computational biology. Strong track record in building scalable ML pipelines, web server deployment, time-series forecasting, and large-scale data workflows. Passionate about bridging research and production, ensuring models are reliable, maintainable, and impact-oriented.

TECHNICAL SKILLS

- **Core Languages**: Python, C++, Bash, SQL
- **ML Frameworks**: PyTorch, TensorFlow, Scikit-learn
- **MLOps & Tools**: Docker, Git (GitHub Actions), MLflow, FastAPI, AWS, HPC clusters
- **Expertise**: Time-series forecasting, Multimodal learning, Graph neural networks, Anomaly detection, Deep learning, Predictive modeling
- **Other**: Data wrangling & visualization, Statistical modeling, Optimization, API & Web server development

EXPERIENCE

- **Senior Computational Scientist** (ML focused) – Prologue Medicines, MA (01/2025 – 06/2025)
  - Developed and deployed the company's first production-level ML pipeline, reducing experimental costs by 15%.
  - Built automated data pipelines for ingestion, training, evaluation, and retraining workflows, ensuring reproducibility.
  - Containerized workflows with Docker and integrated into CI/CD pipelines for continuous integration and deployment.
  - Deployed ML workflows on AWS (EC2/S3) and HPC clusters for scalability.
  - Optimized model performance and reduced inference latency for large biological datasets.
- **Postdoctoral Associate** (ML & Computational Biology) – Cornell University, NY (02/2019 – 12/2024)
  - Led the 3D viral-human protein interactome project: built a large-scale structural database, deployed a web server, and developed visualization tools for AlphaFold-based interaction models.
  - Designed multimodal deep learning models for drug-target interaction, scaling to tens of millions of pairs; improved AUROC and enabled discovery of novel interactions.
  - Engineered genomic anomaly detection pipelines (OMSV, COMSV) with 20× speed-up and improved recall, deployed on HPC clusters for population-scale studies.
  - Collaborated cross-functionally with engineers, statisticians, and clinicians; ensured reproducibility through version control (Git) and unit testing of ML modules.

PROJECTS (Selected)

- 3D Viral-Human Protein Interactome: Full pipeline from data curation and AlphaFold structure prediction to interface analysis and deployment of interactive web server.
- Opioid Relapse Forecasting System: LSTM-based multivariate forecasting with preprocessing, uncertainty metrics, and API support for clinical integration.
- Drug-Target Interaction Prediction: Combined sequential and graph data; trained models on millions of pairs; engineered scalable pipelines for inference and discovery.

EDUCATION

Ph.D., Computer Science – The Chinese University of Hong Kong
M.Sc. & B.Sc., Computer Science – South China University of Technology