Liyang Xie

xieliyang@utexas.edu | (510)944-8886 | linkedin.com/in/liyangxie

Research Interests

AI4Science • Diffusion & Flow-matching models • Distillation • Protein Design • Spatial Transcriptomics

Education

University of Texas, Austin (Aug. 2022 ~ Present)
Ph.D. in Statistics. GPA 4.00/4.00
University of California, Berkeley (Aug. 2018 ~ Dec. 2021)
B.A. in Computer Science; B.A. in Applied Mathematics with concentration in Statistics. GPA: 3.98/4.00

Publications

- **Liyang Xie**, Haoran Zhang, Zhendong Wang, Wesley Tansey, Mingyuan Zhou. "Distilled Protein Backbone Generation". *Under review.*
- Leonid Hanin, **Liyang Xie**, Rainer K. Sachs. "Mathematical Properties of Incremental Effect Additivity and Other Synergy Theories." *Mathematical Methods in the Applied Sciences.* https://doi.org/10.1002/mma.9387
- Edward Greg Huang, Ren-Yi Wang, **Liyang Xie**, Polly Chang, Gracie Yao, Borong Zhang, Dae Woong Ham, Yimin Lin, Eleanor A. Blakely, and Rainer K. Sachs. "Simulating galactic cosmic ray effects: synergy modeling of murine tumor prevalence after exposure to two one-ion beams in rapid sequence." *Life Sciences in Space Research* 25 (2020): 107-118. https://doi.org/10.1016/j.lssr.2020.01.001

Research Projects

Protein Diffusion Distillation, UT Austin, Sept. 2024 ~ present
- Explored **diffusion** and **flow-matching** models for protein backbone generation tasks (e.g. RFDiffusion, Proteina) and protein folding tasks (e.g. AlphaFold 3, SimpleFold)
- Developed **distillation** strategy to accelerate inference of **protein backbone generation** models
- Rebuilt the self-consistency evaluation pipeline to confirm over 94% designability and effective sampling time of less than 0.3s per designable sample

ST Deconvolution with Sparsity, MD Anderson Cancer Center, May. 2024 ~ present
- Proposed novel deconvolution method for sparse spatial transcriptomic and microRNA data to estimate tumor transcript proportions, validated with simulations

ML Explainability with Shapley Values, UT Austin, Sept. 2023 ~ Dec. 2023
- Designed decomposition method using **Shapley theories** to increase interpretability of ML models
- Reconstructed the causal structure based on the decomposition

Predict TmS with Computer Vision, MD Anderson Cancer Center, Jun. 2023 ~ May. 2024
- Developed **deep learning** models under the **multiple instance learning** (MIL) framework to predict the TmS biomarker from H&E whole slide images
- Benchmarked penalized regression techniques (Lasso/Elastic-Net, SCAD, MCP)

Mathematical Properties of Synergy Theories, UC Berkeley, Mar. 2021 ~ Aug. 2021
- Provided visualization and proofs using numerical differentiation and integration to illustrate properties of different synergy models
- Co-authored paper "Mathematical Properties of Incremental Effect Additivity and Other Synergy Theories" submitted to the Physica D: Nonlinear Phenomena

IEA Theory to Model Effect of Radiation, UC Berkeley, Sept. 2019 ~ Dec.2019
- Applied mathematical and statistical analysis to study the effects of Galactic Cosmic Rays radiation on murine Harderian gland tumorigenesis

- Fitted and evaluated the Incremental Energy Additivity (IEA) Theory using data from Bevalac in LBNL and NASA Space Radiation Laboratory
- Co-authored paper "Simulating galactic cosmic ray effects: Synergy modeling of murine tumor prevalence after exposure to two one-ion beams in rapid sequence"

Professional Experience

Software Engineer, Percassist, Jan. 2021 ~ Aug. 2022, May. 2024 ~ Aug. 2024
- Designed real-time ECG QRS detection algorithm in Matlab and C++ with milliseconds of delay
- Developed embedded control system features (STM32 MCUs) to for medical device control (pump, fans, temperature and pressure sensors, etc.)
- Improved C# UI application for system control during animal and clinical studies
- Built unit test and integration test pipeline for the algorithm and the control system
- Analyzed animal study and clinical study data for quality control

Software Engineer Intern, Future International Information Co., Ltd., Feb. 2018 ~ Apr. 2018
- Maintained network infrastructure (50+ switches, 1000+ machines); performed troubleshooting and client support

Skills

- **Machine Learning and Statistics:** Deep learning (PyTorch, TensorFlow), Diffusion & flow models, Computer vision, Causal inference, Multiple Instance Learning, Transfer learning, Penalized regression
- **Programming and Tools:** Python, R, C++, Matlab, Java, SQL, Git, LaTeX, Unix, Kubernetes