Sizhe (Steven) Liu

CONTACT INFORMATION

✉ sliu0727@usc.edu
in linkedin.com/in/sliu0727
⌂ steven51516.github.io/
G Google Scholar
○ github.com/Steven51516

213-590-0762
Los Angeles, CA
United States, 90089
Department of Computer Science
University of Southern California

EDUCATION

University of Southern California — Aug 2021 - May 2025
Bachelor of Science in Computer Science — GPA: 3.94/4.00
Advisors and Mentors: USC: Prof. Liang Chen, Prof. Yue Zhao, and Prof. Tea Jashashvili. I collaborate with Prof. Tianfan Fu at RPI, and Chair Prof. Stan Z. Li at Westlake.

RESEARCH

AI for Drug Discovery, Large Language Models, Graph Neural Networks, ML Systems

PUBLICATION & PREPRINT

Sizhe Liu, Jun Xia, Lecheng Zhang, Yuchen Liu, Yue Liu, Wenjie Du, Zhangyang Gao, Bozhen Hu, Cheng Tan, Hongxin Xiang, Stan Z. Li: FlexMol: A Flexible Toolkit for Benchmarking Molecular Relational Learning. Neural Information Processing Systems (**NeurIPS** 2024) Track on Datasets and Benchmarks.

Sizhe Liu, Liang Chen: Deciphering Single-Cell Gene Expression Variability and Its Role in Drug Response. **Human Molecular Genetics**, 2024.

Sizhe Liu, Yuchen Liu, Haofeng Xu, Jun Xia, Stan Z. Li: SP-DTI: Subpocket-Informed Transformer for Drug-Target Interaction Prediction. **Bioinformatics**. (Under Review)

Jun Xia, **Sizhe Liu**, Jingbo Zhou, Shaorong Chen, Hongxin Xiang, Zicheng Liu, Yue Liu, Stan Z. Li: Bridging the Gap between Database Search and *De Novo* Peptide Sequencing with SearchNovo. International Conference on Learning Representation (**ICLR**), 2025. (Submitted)

Tea Jashashvili, **Sizhe Liu**, Kristian J. Carlson: Multi-Bone Micro-CT Scanning with a Leap Towards Reproducibility and Efficiency Using Automated Tools. American Association of Biological Anthropologists(**AABA**) Conference, 2024.

Jingbo Zhou, Shaorong Chen, Jun Xia, **Sizhe Liu**, Tianze Ling, Wenjie Du, Yue Liu, Jianwei Yin, Stan Z. Li: NovoBench: Benchmarking Deep Learning-based *De Novo* Sequencing Methods in Proteomics. Neural Information Processing Systems (**NeurIPS** 2024) Track on Datasets and Benchmarks.

SELECTED PROJECTS

A Flexible Toolkit for Benchmarking Molecular Relational Learning
Project Lead | Supervisor: Prof. Stan Z. Li — Jul 2023 - May 2024

- Proposed a Molecular Relational Learning(MRL) toolkit that supports the dynamic construction of over 70, 000 distinct model architectures.
- Implemented a robust suite of preset model components, including 16 drug encoders, 13 protein sequence encoders, 9 protein structure encoders, and 7 interaction layers.
- Interfaced with the Therapeutics Data Commons (TDC) project to support various applications, including binary classification, regression, and multi-class classification tasks, with 21 preset metrics for automatic training and testing.

Predicting Cellular Response to Genetic Perturbation

Project Lead | Supervisor: Prof. Liang Chen   Aug 2024 - Present

- Proposed integrating LLM-derived literature embeddings with knowledge graph-based embeddings for predicting unseen genetic perturbations.

- Designed a pathway-based linear attention mechanism for interpretable and scalable cross-gene effect modeling.

LLM-Based Agents for Drug Discovery

Project Lead | Supervisors:

Prof. Yue Zhao, Prof. Tianfan Fu, Prof. Jieyu Zhao   Jul 2024 - Present

- Developed a framework to enhance LLM agents for machine learning research tasks requiring domain knowledge, with a focus on drug discovery.

- Implemented key components including domain knowledge identification, API search, tool construction, unit testing, and reflective analysis.

- Demonstrated the framework's effectiveness in advancing research areas such as AD-MET and DTI prediction.

Integrating Database Search and De Novo Peptide Sequencing

Core Developer | Supervisor: Prof. Stan Z. Li   Jul 2024 - Sep 2024

- Introduced a unified framework that combines the advantages of database search and de novo peptide sequencing to improve peptide sequencing.

- Implemented a retriever module for retrieving peptide spectrum matches (PSMs) and a fusion module for guiding target peptide sequence generation.

- Improved robustness to noisy reference PSMs through a re-designed loss function with pseudo references.

Role of Single-Cell Gene Expression Variability in Drug Response

Project Lead | Supervisor: Prof. Liang Chen   Sep 2022 - Oct 2023

- Employed local coefficient of variation (LCV) algorithm to quantify gene expression variability in scRNA-seq data.

- Analyzed LCV among pharmacogenes in a wide array of cell types across eight different human tissues.

- Established key correlations between pharmacogenetic expression variability at both cross-cell and cross-individual levels and drug efficacy.

Subpocket-Informed Transformer for Drug-Target Interaction Prediction

Project Lead | Supervisor: Prof. Stan Z. Li   Oct 2023 - May 2024

- Designed to enhance drug-target interaction(DTI) prediction in unseen drug/protein and cross-domain settings.

- Utilized Cavity Identification and Analysis Routine(CAVIAR) for detailed subpocket analysis, providing a finer resolution of binding sites compared to traditional pocket-based methods.

- Introduced an embedding approach for both drugs and proteins using pre-trained language models while preserving valuable geometrical information.

INTERNSHIP/
TRAININGS

Shanghai Fullhan Microelectonics Co.,Ltd.

Intern | ML Engineer   Jul 2023 - Aug 2023

- Designed and implemented an image/video upscaling algorithm by integrating edge-directed methods with Nvidia Image Scaling, enhancing boundary transition smoothness.

- Achieved top-tier performance in PSNR and SSIM metrics, demonstrating robustness in real-world video processing scenarios.

USC Viterbi School of Engineering

Teaching Assistant   Jan 2023 - May 2023
- Assisted in teaching, prepared course materials and held office hours for CSCI170. (Discrete Methods in Computer Science)

SERVICE | **ICLR2025, IEEE TNNLS** | Reviewer

SELECTED
COURSEWORK

Computer Science/Mathematics
- Algorithms and Theory of Computing • Introduction to Artificial Intelligence • Data Structures and Object Oriented Design • Machine Learning • Computer Systems • Mathematical Statistics • Probability Theory • Linear Algebra • Calculus III

Biology
- Organismal Biology and Evolution • Cell Biology and Physiology • Molecular Biology • Structural Bioinformatics: From Atoms to Cells

Note: Received "A" in all listed courses.

SKILLS

Programming: C++/C, Python, Java, SQL, Matlab, R, CSS/HTML
Frameworks & Tools: PyTorch, Numpy, Pandas, OpenCV, Scikit-Learn, Matplotlib, LaTeX, PyG, DGL, Git, LangChain, LangGraph, AutoGen
Languages: Chinese and English - Bilingual, German - Elementary