Xiangyu (Leo) Shi

xiangy.shi@gmail.com | (858) 568-0271 | linkedin.com/in/xleoshi/ | github.com/xiangyshi

EDUCATION

Carnegie Mellon University, Silicon Valley
Master of Science, Software Engineering **September 2025 - December 2026**

University of California San Diego
Bachelor of Science, Mathematics - Computer Science **September 2021 - March 2025**

- Overall GPA: 3.9, Provost Honors

RESEARCH EXPERIENCE

Saier Lab, UC San Diego **July 2024 – Present**
Research Assistant

- Designed and implemented a **data processing pipeline with a full command-line interface** to analyze protein family architectures in the Transporter Classification Database (**>23000** protein systems).
- Devised algorithms to parse **CDD and Pfam** output, detect gaps between conserved domains via adjacent domain pair analysis, and identify optimal domain pairs for classification.
- Created **family-level visualization tools** to generate domain architecture plots, highlighting both general and characteristic domains, as well as architecture summary plots for comparative analysis.
- **Co-author** on manuscript *in preparation* describing a novel transporter classification algorithm and computational tool.

J Craig. Venter Institute **June 2025 – August 2025**
Viroinformatics Intern

- Developed **deep learning** models to predict antibody titer values and binding properties of **COVID-19** variant proteins, supporting rational **vaccine design**.
- Automated large-scale data transformation and processing pipelines with **Python** and **Bash**.
- Deployed ML workflows on **Docker + AWS EC2**, ensuring scalability for collaborative research use.
- Presented research findings in an institute-wide seminar at JCVI, demonstrating ability to communicate complex computational biology results to interdisciplinary audiences.

Ehlers Lab, The Scripps Research Institute **September 2024 – April 2025**
Research Assistant

- **Optimized bi-clustering algorithms** for SNP epistasis analysis (GWES), reducing average runtime from **16 hours to 10 minutes**, enabling large-scale genetic studies previously infeasible.
- Performed **performance profiling and algorithmic tuning** to streamline genetic data analysis pipelines, improving efficiency and reproducibility in genomic research workflows.
- **Co-author (second author)** on manuscript *in preparation* on GWES data analysis, contributing to algorithm design, implementation, and experimental validation.

TECHNICAL EXPERIENCE

Yunming Technology **January 2025 – June 2025**
Software Engineer

- Engineered an **AI-powered** Retrieval-Augmented Generation (**RAG**) chatbot system for enterprise CRM research applications.
- Implemented query reconstruction, live search, knowledge base filtering, and Server-Sent Events (SSE) integration for **scalable deployment**.