

Xiangyu (Leo) Shi

xiangyu.shi@gmail.com | (858) 568-0271 | [linkedin.com/in/xleoshi/](https://www.linkedin.com/in/xleoshi/) | github.com/xiangyshi

EDUCATION

Carnegie Mellon University, Silicon Valley

Master of Science, Software Engineering

September 2025 – December 2026

- **1st Place**, CMU-SV Hackathon

University of California San Diego

Bachelor of Science, Mathematics - Computer Science

September 2021 - March 2025

- Overall GPA: 3.9, Provost Honors
-

PROFESSIONAL EXPERIENCE

J. Craig Venter Institute

June 2025 – August 2025

Data Science & Machine Learning Engineer

- Developed ML-driven applications for protein structure and stability prediction and integrated them into a researcher-facing web platform (**Express.js** + HTML/JS/CSS) deployed with **Docker** and **AWS EC2**.
- Automated data ingestion and preprocessing pipelines through **Python**, **Jupyter** and **Bash** scripting, supporting large-scale proteome translation of **thousands of viral sequences** (~20k amino acids each).

Yunming Technology

January 2025 – June 2025

Software Engineer

- Engineered a modular **Retrieval-Augmented Generation (RAG)** chatbot system for **CRM** workflows, integrating query reconstruction, semantic search, knowledge filtering, and SSE-based streaming using **Python**, **Flask**, **React.js**, **TypeScript**, **Taro**, and **LangChain**.
- Improved knowledge query **accuracy by ~50%**, reducing noise and enhancing CRM agent responses.

Human Internet

September 2024 – June 2025

Full Stack Engineer

- Enhanced the humanID developer console by refactoring backend APIs (**Django/REST**, **MySQL**).
- Redesigned humanID's **Single Sign-On (SSO)** system with **JWT**-based session management, async exception handling, international phone syntax authentication, **Dockerized** for reliable multi-region deployment and **reducing login errors by ~80%**.

STEMz Learning

July 2024 – Present

Full Stack Engineer

- **Led a team of 6 developers** to extend core systems with **Express.js** and **MongoDB**, scaling the frontend-only prototype to a service with **30+ API endpoints** and multi-classroom data support.
- Designed and deployed the teacher-platform web application, supporting roughly **~200 active student/teacher accounts** with announcement, chatroom, and direct message features.

Saier Lab, UC San Diego

July 2024 - Present

Research Assistant

- Developed a protein family architecture processing tool using **Python** and CDD/Pfam, implementing algorithms (e.g., union-find) to detect conserved domain gaps in large protein datasets.
-

TECHNICAL SKILLS

Languages: Python, JavaScript (ES6+), TypeScript, Java, C/C++, F#, Bash, SQL

Frameworks & Libraries: React.js, Next.js, Node.js, Express.js, Flask, Django, TailwindCSS, LangChain, Socket.IO

Databases & Data Tools: MySQL, PostgreSQL, MongoDB, Redis, PyTorch, NumPy, Pandas, SciPy

Cloud, DevOps & Other Tools: AWS (EC2, S3, Lambda), Azure, Docker, Git/GitHub Actions, Vercel, Render, Linux/Unix, UML, REST APIs, GraphQL, SSE, OpenAI API, Agile, Scrum, CI/CD, Pgvector, Nginx