Xiangyu (Leo) Shi

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

EDUCATION

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

- Overall GPA: 3.9, Provost Honors
- Technical Proficiencies: Python, JavaScript, Java
- Frameworks & Libraries: Flask, Django, React.js, OpenAI API, LangChain
- Tools & Technologies: NumPy, Pandas, MySQL, PostgresSQL, MongoDB, Docker

PROFESSIONAL EXPERIENCE

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

- Contributed to the development and deployment of an AI-powered **Retrieval-Augmented Generation (RAG)** chatbot system tailored for companies with CRM service needs.
- Implemented features such as **query reconstruction**, **live search**, **tool-augmented LLM reasoning**, and **Server-Sent Events (SSE)** protocol integration.
- Designed database schemas for efficient **Object Storage Service** (**OSS) document management.**

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

- **Optimized** bi-clustering algorithms for single nucleotide polymorphism (SNP) mutation analysis, reducing average runtime from **16 hours** to **10 minutes**.
- Conducted performance profiling and algorithm tuning, streamlining genetic data analysis and boosting computational efficiency.

Foundation for a Human Internet **September 2024 - Present**
Full Stack Engineer

- Contributed to the development and optimization of humanID developer console utilizing **Django** and **MySQL**, improving usability and functionality.
- Improved humanID's **Single Sign-On (SSO)** system by integrating JWT-based session management and exception handling across asynchronous login flows.
- Optimized phone verification with international phone number APIs to support global authentication.

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

- Independently developed a protein family architecture processing tool using **Python** and CDD/Pfam for **Transporter Classification Database (TCDB),** a database containing over 23000 protein strands.
- Implemented algorithms to identify conserved domain gaps in protein sequences, using union-find methods to discover potential novel domains.

STEMz Learning **July 2024 - Present**
Full Stack Engineer

- Developed a dynamic quiz feature using **React.js** to enhance user engagement and learning outcomes.
- Deployed an **Express API** on **Vercel**, facilitating cross-platform communication.
- Managed end-to-end integration of user and classroom system with a **MongoDB** database.