Yixuan Li

1088 Xueyuan Ave., Shenzhen, Guangdong, China
shellyli1105@gmail.com ShellyLeee shellyleee.github.io

Education

Southern University of Science and Technology (SUSTech) **Aug. 2022 – Jun. 2026 (Expected)**
Bachelor of Science in Data Science *Shenzhen, China*

- **GPA:** 3.79/4.00 (3.92/4.00 in Junior Year)
- **Mathematics Courses:** Calculus, Linear Algebra, Mathematical Analysis, Discrete Mathematics and Its Applications, Probability Theory, Mathematical Statistics, Statistical Linear Models, Operational Research and Optimization
- **CS Courses:** Introduction to Computer Programming, Data Structures and Algorithm Analysis, Principles of Database Systems, Distributed Storage and Parallel Computing, Advanced Natural Language Processing, Artificial Intelligence

University of California, San Diego (UCSD) **Mar. 2025 – Jun. 2025**
University and Professional Studies (UPS) Program *San Diego, CA, United States*

- **GPA:** 4.00/4.00
- **Courses:** Machine Learning Algorithms, AI Search and Reasoning, Topics of Data Science: Trustworthy Machine Learning

Research Interests

- **Trustworthy Machine Learning**: Uncertainty Estimation & Calibration, Hallucination Detection, Model Interpretability, LLM Alignment, Generative AI
- **AI for Science (Biomedical)**: Multimodal Learning, Graph Neural Networks (GNNs), Computational Biology

Research Experience

SUSTech Machine Learning Lab | *Advisor: Prof. Hongxin Wei* | SUSTech **Shenzhen, China**

Project: Multimodal Learning for Spatial Transcriptomics-to-Proteomics Prediction
Main Developer *Oct. 2025 – Present*

- Built an end-to-end multimodal framework for the *STP Open Challenge* to predict spatial protein expression from transcriptomic measurements in glioma tissue.
- Extracted high-fidelity morphological representations from H&E patches using *pathology-specific foundation models* as encoders, enhanced with targeted *data augmentation* for robust generalization.
- Developed a topology-aware *multimodal Graph Attention Network* over spatial neighborhood graphs, integrating morphological and transcriptomic embeddings via *adaptive gated fusion* and regularizing training with protein co-expression priors embedded in the loss.
- Achieved **0.75 on Spearman@Top10** and currently **holding 1st position** on the *public leaderboard*, demonstrating a reliable and scalable multimodal strategy for spatial proteomics prediction.

Project: Benchmarking Uncertainty Quantification Methods in Large Language Models
Independent Researcher *Jun. 2025 – Oct. 2025*

- Reproduced and extended a *semantic uncertainty evaluation pipeline* for LLMs, integrating a new method and adapting the codebase to handle model-format inconsistencies, enabling systematic comparisons across QA datasets (AUROC, AUARC, PRR).
- Conducted an extensive literature review on *uncertainty quantification in LLMs*, connecting with *hallucination detection* and *calibration*, and summarized insights in internal slide decks and short blog-style write-ups.

Jianqing Shi's Lab | *Advisor: Prof. Jianqing Shi* | SUSTech **Shenzhen, China**

Project: Statistical Analysis of Mental Health Impact among Menopausal Women
Research Assistant *Sep. 2025 – Nov. 2025*

- Analyzed large-scale survey data on menopausal women in Malaysia to examine the relationship between menopausal stages and mental health outcomes.
- Applied statistical methods including *Kruskal–Wallis test*, *Fisher's exact test*, *Chi-square test*, *ANCOVA*, *Cochran-Mantel-Haenszel test*, *Paired t-tests*, and *multiple imputation* for hypothesis testing and subgroup analysis.
- Identified stage-adjusted psychological risk factors and enhanced inference robustness in women's health research.
- Results reported internally due to data confidentiality.

Internship Experience

Machine Learning for Clinical Decision Support in Orthodontics **Jun. 2024 – Aug. 2024**

Clinical Data Science Intern — Mentor: Dr. Ying Shan *Peking University Shenzhen Hospital*

- Performed data preprocessing on clinical craniofacial imaging datasets, dental records, and treatment histories, which included cleaning, de-duplication, and standardization.
- Developed *linear and neural network models* based on craniofacial measurement variables to predict patient-specific tooth extraction plans. Achieved strong predictive performance, contributing to internal clinical adoption.

Personal Projects & Competitions

Concept Bottleneck in Generative Models: Reproduction Study | *Core Member* **Apr. 2025 – Jun. 2025**

- Reproduced the CVPR paper on Post-hoc Concept Bottleneck frameworks: *Concept Bottleneck Autoencoder (CB-AE)*, *Concept Controller (CC)* using *StyleGAN*, replicating key quantitative and qualitative results from the original study.
- Conducted additional experiments on *Concept Extra-Interpolation, Entanglement, and Leakage* to evaluate model interpretability, robustness, and reliability under concept-level perturbations.
- Delivered a comprehensive reproduction package including *GitHub code*, *presentation slides*, and a public *technical blog*, earning **43.4/40** on the final project (**Top 1** in a UCSD class enrolling both undergraduate and graduate students).

Kaggle Competition: Fine-Tuning Gemma-2B for Chinese QA Dataset | *Team Leader* **Oct. 2024 – Jan. 2025**

- Led a team to fine-tune *Gemma-2B* on the *GPT-4 Chinese QA dataset*. Built a complete fine-tuning pipeline covering data preprocessing, LoRA integration, and evaluation. Published notebook can be viewed *here*.
- Integrated *Retrieval-Augmented Generation (RAG)* for data augmentation and *LoRA-based PEFT* for parameter-efficient fine-tuning, resulting in improved dataset diversity, contextual grounding, and instruction-following performance.
- Evaluated model performance using *BLEU, ROUGE, METEOR, and BERTScore* metrics, and performed *cross-lingual robustness analysis* showing stable results across four languages.

Reinforcement Learning on FrozenLake Game Setting | *Core Member* **Oct. 2024 – Jan. 2025**

- Implemented and evaluated the *Double DQN* and *Dueling DQN* algorithms on the FrozenLake environment, performing hyperparameter tuning and robustness analysis to achieve up to **98%** success rate in optimal path learning.
- Compiled a comprehensive report and structured a well-documented *GitHub code repository* to ensure reproducibility.

Awards

- **First Class of the Merit Student Scholarship, SUSTech** (Top 5%, two-time recipient) **2023 – 2025**
- **Third Class of the Merit Student Scholarship, SUSTech** (Top 15%) **2022 – 2023**
- **Outstanding Student Award, SUSTech** (three-time recipient) **2022 – 2025**
- **Second Prize** China Undergraduate Mathematical Contest in Modeling (CUMCM), Guangdong Division **Oct. 2024**
- **Honorable Mention** Mathematical Contest In Modeling (MCM) **May 2024**

Leadership & Activities

Community Involvement **Oct. 2023 – Present**

Volunteer *SUSTech*

- Taught online academic enrichment classes for students in rural areas, supporting long-term educational outreach efforts.
- Contributed over **179 hours** of volunteer service.
- Awarded **Outstanding Volunteer** honors at both the university and School of Science levels.

Shuren College Media Center **Sep. 2023 – Aug. 2024**

President *SUSTech*

- Led a 15-member team producing digital communications and organizing college events across official media channels.
- Received the **Excellent Student Leader Award** and the **Outstanding College Organization Award**.

Skills

- **Programming:** Python, R, SQL, Java, JavaScript, HTML
- **Development Tools:** PyTorch, Git, Jupyter Notebook, LaTeX, Markdown, Hadoop
- **Languages:** English (TOEFL 109: R28/L27/S27/W27), Mandarin (Native), Cantonese (Fluent)