Jay (Zhuosheng) Liu

zsliu@ucdavis.edu | jlk666.github.io | linkedin.com/in/zhuosheng-liu | 🎓 Google Scholar

2026 *graduate seeking internship/full-time opportunities in DS/MLE*

Experience

Capital One: *Data Science PhD Internship* *June 2025 - Aug 2025*

- Working with the **AI Foundation Customer Core** team on **Graph Representation Learning** across multiple modalities of customer data to predict and **identify targeted prospective users** and **prevent potential fraud suspects.**
- Investigating different **graph embedding methods** (Cube, Node2Vec, Graph Autoencoder, Variational Graph Autoencoder) to predict future shared accounts among **2 million** customers, achieving **73%** boost of AUC score

Metagenomi Inc: *Machine Learning Internship* *June 2024 - Sept 2024*

- Conducted data preprocessing and feature engineering using **highly noisy and high-dimensional data** from inhouse PostgreSQL database to improve the accuracy of predictive models, therefore help to **reduce operational cost by approximately 40%**.
- Assisted in the development and implementation of machine learning models **(KNN, Random Forest Regressor achieving 88.0% cosine similarity)** to predict the indel profiles of inhouse Crispr nuclease
- Assisted in prototyping AI solutions by implementing and optimizing deep learning frameworks**(CNN, LSTM achieving 99.5% accuracy of prediction the most common gene editing outcomes)** to enhance the understanding of key features in CRISPR nuclease activity, ultimately guiding in-house CRISPR screening.

Information Fusion and Mining Lab, *Graduate Research Assistant co-advised by Dr. Jiawei Zhang* *May 2023 - Now*

- Drove advanced solutions and compared different **parameter efficient-fine-tuning** methods for **vision language model**.
- Systematical evaluation of different **GenAI text-to-video and image-to-video models**. Priprint available at **Here**.

Education

3.97/4.0 **PhD in Microbiology**, *University of California, Davis* | California, USA *2020-26*

3.91/4.0 **Master's degree in Computer Science**, *University of California, Davis* | California, USA *2023-25*

3.84/4.0 **Master's degree in Biotechnology**, *Columbia University* | New York, USA *2019-20*

Research Focus: Exploratory Data Science | Applied ML/AI |AI4Science

Courses: ECS 271 **Advanced Machine Learning (A+)** | ECS **174 Computer Vision (A)** | ECS 122A/222A **Algorithm Analysis and Design (A)** | ECS 124 Bioinformatics Theory and Practice (A) | STA 141 Advanced Statistical Computing A-B-C (A) | ECS 201A Computer Architecture (A)|ECS 289 G **Advanced Deep Learning (A+)|ECS 289 L Advanced Artificial Intelligence (A+)|ECS 230 Applied Linear Algebra (A)**

Skills

Programming	**Python**(NumPy, SciPy, pandas, Matlotlib, Jupyter), **C/C++** , **Java**, **Spark/Pyspark**, HTML, CSS, Perl, R, SQL, Scala, Git, Bash Scripting, LaTeX
AI|LLM	Prompt Tuning, PEFT, RAG, Langchain, LangGraph Hugging Face Model cards
Software	Linux, AWS, Tensorflow, Pytorch, Keras, OpenCV, Snakemake, scikit-learn, Docker, Kubernetes, DBeaver

Selected Projects

Parameter-Efficient Fine-Tuning for Vision-Language Models *June 2023 - Now*

Preliminary report available at ***Here***

- Conducted an extensive study on the application of **Parameter-Efficient Fine-Tuning** (PEFT) methods to vision-language models, particularly focusing on the **CLIP** .
- Conducted comprehensive empirical analysis across various datasets, such as VTAB-1K, to understand the adaptability and effectiveness of PEFT methods in different scenarios.
- Investigated the effectiveness of **prompt engineering** and **adapter technique such as CoOP, VPT and UPT** in improving model performance
- Implemented and optimized various hyperparameters through **systematic grid search** methods to ascertain the most effective settings for the models.

Utilizing the Efficient Segment-Anything Model (ESAM) for Enhanced Biomedical Image Segmentation *June 2023 - June 2024*

Preliminary report available at ***Here***

- Implemented **ESAM** for complex biomedical image segmentation using box and point prompts, significantly improving model performance through fine-tuning.
- Developed a Python-based framework for extracting and analyzing segmentation results with **GPT-4** for insightful medical recommendations.
- Conducted **statistical evaluation** using Intersection over Union (IoU) and Dice Similarity Coefficient (DSC) metrics on over 1500 medical images, enhancing diagnostic accuracy.

Virulence prediction of pathogenic foodborne pathogens using ML/DL based on Pangenome *June 2023 - Now*

Manuscript in preparation

- Constructed pangenome (100,000 unique gene feature) using **fully integrated and scalable bioinformatics pipeline** (automation achieved using Snakemake)
- Visualized high-dimensional pangenomic gene feature using **dimensional reduction PCA**
- Trained, optimized, validated and statistically compared fundamental **unsupervised and supervised machine learning model** performance, including **Kmean, GMM, KNN, SVM, RF, Naive Bayes**
- Trained and validated **deep neural network (MLP and CNN**) and achieved 93% virulence prediction accuracy
- Obtained valuable biological understandings by applying an **interpretable CNN with Grad-CAM**

Publications

2025	**Pathogenic potential prediction of Vibrio parahaemolyticus by using pangenome data with high performance machine learning algorithms**, First-author	*Under review on PNAS*
2025	**A Survey of AI-Generated Video Evaluation**, Co-author	*Arxiv*
2025	**AIGVE-Tool: AI-Generated Video Evaluation Toolkit with Multifaceted Benchmark**, Co-author	*VISAPP*
2024	**Recent advances in understanding the fitness and survival mechanisms of Vibrio parahaemolyticus**, First-author	*IJFP*

Peer Review Activity

2025	**Frontiers in Microbiology**, Reviewer	**Journal**
2025	**Conference on Neural Information Processing Systems (NeurIPS)**, Reviewer	**Conference**
2025	**ACM Transactions on Computing for Healthcare**, Reviewer	**Journal**
2025	**Journal of Medical Imaging**, Reviewer	**Journal**
2025	**IEEE Journal of Biomedical and Health Informatics**, Reviewer	**Journal**
2025	**IEEE Transactions on Neural Networks and Learning Systems**, Reviewer	**Journal**
2025	**IEEE Data Descriptions**, Reviewer	**Journal**
2025	**IEEE Transactions on Multimedia**, Reviewer	**Journal**
2025	**Frontiers in Marine Science**, Reviewer	**Journal**
2024	**International Conference on Learning Representations (ICLR)**, Reviewer	**Conference**
2023	**IFT FIRST Technical Research Paper**, Reviewer	**Conference**