JOSEPH LIU

joseph@liu.us | https://joseph.liu.us | +1 (650) 276-8035

EDUCATION

University of Southern California (USC), Los Angeles, CA **08/2022 - 05/2025 (Expected)**
Bachelor of Science in Computer Science (**GPA**: 4.0/4.0) | W.V.T. Rusch Engineering Honors Program

- **Courses**: Artificial Intelligence, Machine Learning, LLMs in Natural Language Processing, Probability, Statistics

Santa Clara University (SCU), Santa Clara, CA **09/2021 - 05/2022**
Bachelor of Science in Computer Science (**GPA**: 3.94/4.0)

- **Courses**: Algorithms, Linear Algebra, Differential Equations

PUBLICATIONS

[1] **Liu, J.**, Cui, X., Nam, Y., & Swabha, S. (2025). Judging Text Simplicity with Large Language Models. (in preparation)

[2] Chen, J., Zhu, X., Wang, Y., Liu, T., Chen, X., Chen, Y., Leong, C., Ke, Y., **Liu, J.**, Yuan, Y., McAuley, J., & Li, L. (2025). Symbolic Representation for Any-to-Any Generative Tasks. *CVPR*. (under review)

[3] Chen, X.*, Yuan, Y.*, **Liu, J.***, Leong, C., Zhu, X., & Chen, J. (2024). Generative Models in Protein Engineering: A Comprehensive Survey. *NeurIPS 2024 Workshop FM4Science*. (Poster)

[4] Smith, R., Patel, A., Soraisam, M.D., Guhathakurta, P., Tadepalli, P., Zhu, S., **Liu, J.**, et al. (2024). Variable Stars in M31 Stellar Clusters from the Panchromatic Hubble Andromeda Treasury. *The Astrophysical Journal*, 974(2), p.292.

[5] He, K.*, Li, M.*, & **Liu, J.*** (2023). Enhancing Debugging Skills of LLMs with Prompt Engineering. *Technical report*.

[6] **Liu, J.** (2023). Predicting Game Popularity from Steam Descriptions. *Technical report*.

[7] Patel, A., Mukherjee, S., Soraisam, M., Guhathakurta, P., **Liu, J.**, & Tadepalli, P. (2022). Variable Stars in M31 Stellar Clusters using the Panchromatic Hubble Andromeda Treasury. *Bulletin of the AAS*, 54(6).

RESEARCH EXPERIENCE

Learning Heuristics for Multi-Agent Pathfinding, IDM Lab, USC, Los Angeles, CA **05/2024 - Present**
Mentored by Yimin Tang, advised by Prof. Sven Koenig

- **Trainable Heuristic Environment**: Developed an RL environment to train heuristics for multi-robot path planning, leveraging 4D representations to capture spatial-temporal relationships between robot paths and environmental constraints.
- **Two-Phase Training Strategy**: Crafted a two-phase training strategy, initially replicating traditional heuristics and subsequently enhancing search efficiency with a node expansion reward system.
- **Search Efficiency Assessment Tool**: Implementing a quantitative evaluation system based on node expansion metrics, enabling direct measurement of search efficiency improvements for the learned heuristic function.

LLM-based Text Simplification Evaluation System [1], DILL Lab, USC, Los Angeles, CA **01/2024 - Present**
Mentored by Xinyue Cui and Yoonsoo Nam, advised by Prof. Swabha Swayamdipta

- **Text Simplification Metrics**: Designed a novel reference-free metric for text simplification by introducing LLM judges, eliminating the need for specialized training data.
- **Model Architecture Design**: Developed an efficient evaluation framework utilizing pre-trained models such as Llama 3 without fine-tuning, enabling broad domain coverage and robust simplification assessment.
- **Evaluation**: Demonstrated superior performance in evaluating simplifications, achieving 51.2% correlation with human judgment and outperforming traditional metrics, such as FKGL and SARI, and trained metrics such as LENS.

Symbolic Representation for Any-to-Any Generative Tasks [2] **09/2024 - 12/2024**

- **Symbolic Any-to-any Paradigm**: Introduced a symbolic language with functions, parameters, and topologies, enabling flexible representation of any-to-any generative tasks (e.g., image-to-video, image-to-3D, image merging, etc.).
- **Training-free Inference:** Developed a training-free inference engine that transforms natural language task descriptions into executable symbolic flows, allowing seamless task execution as a program.

Generative Models in Protein Engineering [3] **08/2024 - 12/2024**

- **Protein Model Classification**: Systematically categorized protein generative models through a multi-dimensional framework, encompassing inference methodologies (diffusion-based/autoregressive) and modeling targets (sequence/structure), establishing a structured overview of this emerging field's technical landscape.
- **Protein Diffusion Model Comparison**: Established a comparison framework for protein diffusion models across two fundamental dimensions: the mathematical representation level and the structural invariance level, revealing how modeling choices affect protein structure design.

- **Future Directions in Protein Modeling**: Identified critical challenges and future opportunities in protein generative models, emphasizing the transition from data limitations to large-scale datasets and hybrid modeling approaches.

Enhancing Debugging Skills of LLMs with Prompt Engineering [5] **09/2023 - 01/2024**
Advised by Prof. Swabha Swayamdipta

- **Debugging Prompt Engineering**: Integrated prompt engineering into LLMs to boost performance in debugging tasks through few-shot learning and chain-of-thought prompting.
- **Multidimensional Evaluation Metrics**: Developed and implemented a comprehensive set of evaluation metrics, including CodeBLEU, CodeROUGE, and CodeF1, to quantitatively assess LLM debugging performance.
- **Real-World Error Dataset Construction**: Constructed a dataset by integrating Java and Leetcode to replicate real-world programming bugs for dynamic analysis.

Wildfire Spread Prediction, Computation and Data Driven Discovery Group, USC, Los Angeles, CA **08/2023 - 12/2023**
Mentored by Bryan Shaddy, advised by Prof. Assad Oberai

- Worked on physics-informed machine learning techniques to model wildfire spread using diffusion and GAN models.

Variable Stars in Andromeda Galaxy [4][7], UC Santa Cruz, Santa Cruz, CA **06/2020 - 08/2021**
Mentored by Sagnick Mukherjee, advised by Prof. Puragra Guhathakurta

- **Data Cleaning and Collection:** Organized, filtered, and cleaned datapoints of millions of stars, including work in database query optimization, parallelization, and computational geometry.
- **Variable Star Census and Classification**: Established a catalog of 86 luminous variables in M31 clusters, with comprehensive characterization of their evolutionary phases and initial masses based on theoretical isochrones.

TEACHING EXPERIENCE

Teaching Assistant, University of Southern California, Los Angeles, CA **05/2024 - 07/2024**

- Teaching Assistant for CSCI-201: Principles of Software Development for *Prof. Victor Adamchik*
- Helped the professor prepare the computer lab exercises and coached students in the lab for their coding assignments.

Grader, Santa Clara University, Santa Clara, CA **03/2022 - 06/2022**

- Grader for CSCI 163: Theory of Algorithms for *Prof. Nicholas Tran*
- As a freshman, graded homework and exams for a course primarily taken by sophomores and juniors.

INDUSTRY EXPERIENCE

Data Science Intern, Stellantis N.V., Auburn Hills, MI (Remote) **05/2023 - 08/2023**

- **Pipeline Optimization**: Led end-to-end optimization of ML sales prediction pipeline, achieving 86% reduction in interruptions, 30% faster runtime, and 25% cost savings while improving data quality by fixing critical bugs affecting 60% of the dataset.
- **Research Leadership**: Spearheaded feature engineering initiatives and performance optimization research, presenting findings to 80+ stakeholders including directors and VPs.
- **Performance Recognition**: Demonstrated exceptional performance resulting in return offer for Summer 2024.

Machine Learning Intern, iKala Interactive Media Inc., Taipei, Taiwan **06/2022 - 08/2022**

- **Video Analysis Research**: Researched state-of-the-art methodologies in Computer Vision (CV) and Natural Language Processing (NLP) for video analysis.
- **Audio-Video Embedding**: Designed and implemented a Transformer-based model for multimodal (video and audio) embedding generation with PyTorch, achieving 60% precision on AudioSet dataset.

AWARDS

- USC Provost's Undergrad Research Fellowship: Fall 2024 ($1,000)
- USC Center for Undergraduate Research in Viterbi Engineering Fellowship: Fall 2023; Spring, Summer 2024 ($5,500)
- USC Viterbi Dean's List: Spring, Fall 2023; Spring 2024
- SCU Dean's Scholarship: 2021-2022 ($8,100)

SKILLS

Languages: Python, Java, C++, C#, SQL, JavaScript, x86-64 Assembly
Frameworks/Tools: PyTorch, Pandas, NumPy, Git, AWS
Environments: Unix/Linux, Windows
Areas of Expertise: Machine Learning, Natural Language Processing (NLP), Large Language Models (LLMs), Data Structures & Algorithms