JOSEPH LIU

+1 (650) 276-8035 | joseph@liu.us | github.com/MajikalExplosions | linkedin.com/in/joseph-liu-pub

EDUCATION

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

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

RESEARCH EXPERIENCE

IDM Lab, USC, Los Angeles, CA **05/2024 - Present**
- Implementing custom Gymnasium reinforcement learning environment for model training
- Designing and optimizing reward functions to enhance model performance
- Developed predictive models to optimize node exploration in ECBS multi-agent pathfinding algorithm
- Mentored by Yimin Tang, advised by Prof. Sven Koenig

Data, Interpretability, Language, and Learning (DILL) Lab, USC, Los Angeles, CA **01/2024 - Present**
- Developing novel evaluation methods for text simplification tasks
- Implementing Large Language Model (LLM)-driven metrics for model assessment
- Designed and implemented perplexity-based methods, and evaluated performance on human ratings
- Mentored by Xinyue Cui and Yoonsoo Nam, advised by Prof. Swabha Swayamdipta

Computation and Data Driven Discovery Group, USC, Los Angeles, CA **08/2023 - 12/2023**
- Worked on physics-informed machine learning techniques to model wildfire spread using diffusion and GAN models
- Mentored by Bryan Shaddy, advised by Prof. Assad Oberai

Interaction Lab, USC, Los Angeles, CA **04/2023 - 08/2023**
- Created an autonomous lab tour using Amazon Astro Robot for Prof. Maja Mataric

Researcher (Variable Stars in Andromeda Galaxy), UC Santa Cruz **06/2020 - 08/2021**
- Built a webpage and a graphical Python application to help efficiently visualize, categorize, and compare 1000+ stars
- Discovered and identified missing portions of the PHAT dataset around brick edges with computational geometry
- Optimized database queries by a factor of over 100, by reducing request count using batching, slicing data, and joins
- Parallelized cubic-time array operations in difference imaging data pipelines using NumPy matrix operations
- Mentored and supervised team of high school students, providing technical guidance and reference materials to Python libraries and previous work during and outside of weekly meetings
- Co-authored a paper in 2024 and a poster in 2022

INDUSTRY EXPERIENCE

Data Science Intern, Stellantis N.V., Auburn Hills, MI (Remote) **05/2023 - 08/2023**
- Refactored and automated machine learning sales prediction pipeline, decreasing interruptions by 86%, runtime by 30%, and cost by 25%
- Increased data quality by identifying and fixing multiple functional bugs that affected 60% of sales dataset
- Investigated potential features for better model performance and candidates for further feature engineering
- Presented findings and work to an audience of 80, including multiple directors and VPs
- Received return offer for Summer 2024

Machine Learning Intern, iKala Interactive Media Inc., Taipei, Taiwan **06/2022 - 08/2022**
- Researched state-of-the-art methodologies in Computer Vision (CV) and Natural Language Processing (NLP) for video analysis, presenting weekly findings to intern team
- Designed and implemented a Transformer-based model for multimodal (video and audio) embedding generation with PyTorch, achieving 60% precision on AudioSet dataset

TEACHING EXPERIENCE

Teaching Assistant, USC, Los Angeles, CA **05/2024 - 07/2024**

- Teaching Assistant for CSCI-201: Principles of Software Development
- 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
- As a freshman, graded homework and exams for a course primarily taken by upperclassmen

PROJECTS

Medical Case Report Search with AI **09/2024 - Present**
Developing website to search case report database using intelligent symptom and complication matching

Generative Models in Protein Engineering **08/2024 - Present**
Investigated protein representations and generation approaches, and current challenges and opportunities in generative protein models

Enhancing Debugging Skills of LLMs with Prompt Engineering **09/2023 - 01/2024**
Investigated the effects of prompt engineering techniques like chain-of-thought on debugging skills of general LLMs and analyzed failure cases to identify causes

Earnings Call Analysis with Machine Learning **11/2021 - 06/2022**
Automated scraping of earnings call transcripts from multiple online sources, analyzed sentiment using AWS SageMaker, built pipeline to extract audio features, and computed embeddings for machine learning using open-source libraries

Using CNNs to Identify Exoplanet Candidates **09/2021 - 03/2022**
Conducted independent research into feasibility of using CNNs on photometric data to identify candidate exoplanets; collected and augmented data to train 3DCNN model for binary classification with PyTorch, achieving 61% accuracy

PUBLICATIONS

Chen, X*., Yiwen, Y.*, **Liu, J**.*, Leong, C., Zhu, X., & Chen, J. (2024). Generative Models in Protein Engineering: A Comprehensive Survey. *NeurIPS 2024 Workshop Foundation Models for Science.* **(oral)**

Smith, R., Patel, A., Soraisam, M.D., Guhathakurta, P., Tadepalli, P., Zhu, S., **Liu, J.**, Girardi, L., Johnson, L.C., Mukherjee, S., Olsen, K.A. (2024). Variable Stars in M31 Stellar Clusters from the Panchromatic Hubble Andromeda Treasury. *The Astrophysical Journal*, 974(2), p.292.

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).

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)
- Northeastern University Honors Scholarship: 2021 ($22,000 renewable yearly as a student and up to $88,000)

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