

Lewis Lui

(909) 323-9915 | lui1@uci.edu | github.com/LewisLui777/ | linkedin.com/in/lewislui | lewislui.com

EDUCATION

University of California, Irvine

Master of Computer Science | Expected Dec 2026 | **GPA: 4.0/4.0**

Focus: Artificial Intelligence, Machine Learning, Data Science, NLP, Computer Vision

University of California, Berkeley

B.S. in Electrical Engineering and Computer Science (**EECS Honors**) | May 2025 | **GPA: 3.9/4.0**

Honors: Dean's List, College of Engineering

Relevant Coursework: Data Structures & Algorithms, Operating Systems, Distributed Systems

WORK EXPERIENCE

Tsavorite Scalable Intelligence

Machine Learning/Software Engineer | Milpitas, CA | June 2025 – August 2025

- Designed and built Python-based REST APIs on AWS to deploy Hugging Face AI/ML models, supporting enterprise customer integrations and contributing to **\$100M+** deals.
- Debugged and optimized backend C/C++ Linux libraries to improve CPU efficiency, achieving up to **18x** faster token generation than industry benchmarks.
- Supported x86 architectures and optimized FPGA-based inference pipelines, delivering up to **7x** inference speedups over standard hardware platforms.

Surge AI (DataAnnotation Tech)

Software Developer | Independent Contractor | Jan 2025 – June 2025

- Evaluated and debugged Python, C, and Java outputs from LLMs, improving correctness by **17%**.
- Performed large-scale QA across **10+** AI/ML models, validating execution and reasoning.
- Validated JavaScript training datasets for enterprise AI systems, reducing hallucinations by **23%**.

Haddee.com

Software Engineering Intern | Berkeley, CA | Oct 2024 – Jan 2025

- Used Pandas/NumPy/PyTorch to make a personalized AI chatbot, increasing user retention by **9%**.
- Designed AI inference algorithms and data structures to improve client-college matches by **14%**.
- Analyzed databases of **2200+** college programs with MySQL to train our transformer model.

PROJECTS

Deep Q-Network Pong Agent

- Engineered a Deep Q-Network (DQN) agent to learn optimal policies from raw pixel inputs.
- Implemented Convolutional Neural Networks and Replay buffers for training over **10,000** episodes.
- Optimized hyperparameters like learning rate/discount factor to achieve **95%-win** rate vs humans.

Procedural World Generation Engine

- Built a generation system using Object-Oriented Design (OOD) to render **5,000+** environments.
- Designed a UI with JavaScript and optimized the backend render pipeline, reducing runtime by **32%**.
- Implemented multithreading/parallel processing to handle states and reduce latency by **47%**.

Large-Scale Ngram Analytics

- Processed Google Ngram dataset of **500B+** words to find usage trends and optimize text searches.
- Developed custom data structures (Custom HashMaps, Red-Black Trees) to query in **O(log n)** time.

TECHNICAL SKILLS

Languages: Python, Java, C/C++, C#, SQL, JavaScript, TypeScript, Rust, Go, HTML/CSS

ML/Backend: PyTorch, TensorFlow, NumPy, Pandas, Flask, Spring, React

Systems/Cloud: Linux, Git, Docker, Kubernetes, AWS, GCP, Azure, MongoDB, CI/CD