

Jason Trinh

jasontrinh@berkeley.edu | [linkedin.com/in/jason-trinh](https://www.linkedin.com/in/jason-trinh) | jaizunt.github.io | Berkeley, CA

EDUCATION

University of California, Berkeley

Aug. 2024 – May 2028

B.S. Electrical Engineering & Computer Science (EECS) | GPA: 4.0/4.0

Berkeley, CA

- **Coursework:** Scalable AI, Deep Learning, Reinforcement Learning, Probability & Random Processes, Convex Optimization, Data Structures, Computer Architecture, Signals & Systems
- **Awards:** Dean's List, AIME Qualifier, National Merit Finalist

TECHNICAL SKILLS

Languages: Python, Java, C, C++, SQL, RISC-V Assembly, GDScript

Frameworks: PyTorch, TensorFlow, vLLM, NeMo, Stable-Baselines3, Scikit-learn, pandas, NumPy, Matplotlib, cvxpy

ML/AI: Foundation Models, LLMs, Pre-training, SFT, RL Post-training, Model Parallelism, Distributed Training

Tools: Linux, Git, CI/CD, REST APIs, Backend Development

EXPERIENCE

AI Research Intern

June 2026 – Aug. 2026

NASA Ames Research Center – Space Life Sciences Training Program

Mountain View, CA

- Developing a **biomedical foundation model** on RNA sequencing data for space life sciences research
- Representation Learning via **sparse autoencoders** to uncover gene expression patterns in spaceflight conditions
- Applied **gene-gene covariance analysis** and prior biological knowledge to enhance model interpretability through constrained **latent space disentanglement** and **feature attribution**

Network Engineer Intern

June 2025 – Aug. 2025

Verizon

Richardson, TX

- Built **supervised machine learning** models in **Python** to detect mislabeled network location codes
- Identified **~100,000 mislabeled sites** with **99.99% accuracy**, improving data integrity at scale

Firmware Developer

Jan. 2025 – Present

Formula Electric at Berkeley

Berkeley, CA

- Designed a **battery state machine** to monitor accumulator charge and circuit output in real time
- Developed **PCU firmware** in **C** for **ADC** voltage processing, plausibility logic, and **CAN** torque commands

Machine Learning Research Intern

June 2024 – Aug. 2024

University of California, San Francisco

San Francisco, CA

- Classified **craniectomy outcome** in stroke patients with **supervised machine learning** on **EMR data**
- Trained **random forest** and **decision tree** classifiers with **scikit-learn** for clinical prediction
- Applied **recursive feature elimination** and **class rebalancing** to improve precision and recall

PROJECTS

EchoKV: Adaptive KV Cache Quantization 🌀 | *NeurIPS 2026 Submission*

May 2026

- Built **closed-loop adaptive per-layer KV cache quantization** for LLM serving on **vLLM**
- Reduced **KV-cache memory** by **59–75%** on **Llama-3.1-70B** while preserving FP16 accuracy
- Designed **canary sensor** and **EMA controller** for live **INT2/INT4/INT8** per-layer bit-width selection

Block Blast: Reinforcement Learning Agent

Jan. 2026

- Trained an autonomous **deep reinforcement learning** agent in **PyTorch** using **PPO**
- Applied **model-based reinforcement learning** with a learned **world model** for next-state dynamics
- Developed a real-time **computer vision pipeline** to extract board state from gameplay for model training

AI Multiverse

Oct. 2025

- Built a **real-time 3D multiplayer voice chat** game in **Godot** with interactive **LLM-powered AI characters**
- Integrated live **speech-to-text** via **Python WebSocket** for low-latency audio transcription
- Designed a conversation manager with **AI turn-taking**, smart interruption, and **text-to-speech** playback

RISC-V CPU

July 2025

- Engineered a **5-stage pipelined CPU simulator** supporting **RISC-V** instructions (add, lw, beq, jal, etc.)
- Designed a complete **instruction fetch–writeback datapath** with simulated **memory** and **register file**
- Achieved **instruction-level parallelism** through **data forwarding**, **hazard detection**, and **stall logic**