Jay Yeung

650-293-7697 | jayyeung@berkeley.edu | linkedin.com/in/jayyeung | https://github.com/JayYeung | jayyeung.vercel.app

Education

University of California, Berkeley May 2027
Bachelor's Degree Computer Science, Double in Data Science *GPA: 4.0*

- **Relevant Coursework:** Machine Learning (A+), Digital Design and Integrated Circuits (A+), Probability & Random Processes (A), Discrete Mathematics & Probability Theory (A+), Operating Systems (A), Compilers (A), Efficient Algorithms (A+), Computer Security (A+), Optimization Models in Engineering (A+), Advanced Matrix Computations (A), Computational Geometry (A)
- **Relevant Teaching:** Discrete Mathematics and Probability Theory (CS70) Teaching Assistant (UGSI 2)
- **Awards:** AIME Qualifier (2021), LeetCode Guardian (2300 rating, top 0.4%) my profile

Experience

Hudson River Trading New York, NY
Algo Dev Intern *May 2026 – Aug 2026*

- Developed a high-frequency trading strategy for B3 markets achieving a Sharpe ratio of 14
- Regularized mid-frequency factor models, improving factor return stability and reducing return volatility

Amazon Annapurna Labs Berkeley, CA
Technical Project Manager *Sep 2025 – Jan 2026*

- Migrated Group Relative Policy Optimization (GRPO) Library on AWS Trainium chips within Hugging Face's Optimum Neuron stack, enabling reinforcement learning for Trainium (GitHub)(Report)

Meta Menlo Park, CA
Software Engineer Intern *May 2025 – Aug 2025*

- Saved 300k/yr on database usage and compute for Messenger ML Infrastructure Team

Berkeley AI Research (BAIR), LLM Interpretability Berkeley, CA
Undergraduate Researcher at Berkeley Speech Group *Jan 2025 – May 2025*

- Built layer-wise inference analysis pipeline using TunedLens to decode logits and track token rank trajectories
- Submitted arXiv:2510.18871 to ACL 2026 and released a fully reproducible pipeline (GitHub)

Cadence Design Systems San Jose, CA
Hardware Machine Learning Intern *May 2024 – Aug 2024*

- Developed Hypergraph Neural Network (HGNN) and Graph Attention Network (GAT) for analog model classification
- Optimized simulations for Spectre (EDA tool) under the mixed analog-digital signal (AMS) team

Projects

Apple x Berkeley EECS151 ASIC Design Competition

- Placed #1 in Apple's ASIC Design Contest among 40 teams

TreeHacks 2025

- Won Tesla's Autonomy Track Grand Prize (1st place), and Tesla's Excellence Prize (3rd place). $\sim$ 8k in prizes.

Datathon 5.0 C-Light at Berkeley

- Won a Berkeley Datathon using Generative AI for eye retinal scans
- Used GANS, Auto-encoders, Latent Space Sampling, and Embedding Interpolation to generate synthetic videos

Efficient Deep Learning Model Compression using Blocked Sparse Matrices (Paper)

- Wrote a custom forward using the experimental Triton `bsr_dense_addmm`; worked around PyTorch/cuSPARSELt gaps

Technical Skills

Programming Languages: Python, JavaScript, HTML, CSS, Java, C++, Perl, TypeScript

Frameworks: TensorFlow, React, Next.js, OpenGL, Flask, Django, Bootstrap, Express.js

Libraries: Git, NumPy, Pandas, SciPy, XGBoost, Matplotlib, PyTorch, Sk-Learn, Scrapy, LangChain, OpenCV, BeautifulSoup, MediaPipe, PyAutoGUI, pickle, tabula, NLTK, RegEx, pyimgui, seaborn