

Jack Yang

jack-yang.com

EDUCATION

University of California, Berkeley

Berkeley, CA

BA, Computer Science and Applied Math (Machine Learning and Optimization); Minor, German

May 2027

- **Coursework:** Efficient Algorithms & Intractable Problems, Numerical Analysis, Machine Learning, Abstract Algebra, Linear Algebra, Optimization Models in Engineering, OS, Real + Complex Analysis, Robotics

WORK EXPERIENCE

SDE / ML Intern

May 2025 - Aug 2025

AWS Bedrock: Capacity Optimization Team

New York, NY

- Architected a high-throughput Capacity Optimization platform; automated fleet-scale resource allocation for 60+ engineers using IaC, achieving 90%+ test coverage
- Engineered a low-latency AI reasoning engine utilizing RLHF and dynamic Chain-of-Thought (CoT) prompting; reduced P99 resolution latency by **94%** (15m to 1m) for mission-critical capacity issues

Software Engineering Intern

June – August 2024

LiveX AI

Palo Alto, CA

- Engineered a high-performance ranking engine and feedback profiling pipeline; optimized vector search/retrieval to recover **\$120K ARR** by mitigating churn

Investment Research Intern

June 2023 – Aug 2023

Ocean Arête Limited

Hong Kong S.A.R.

- Synthesized quantitative unit economics for GenAI infrastructure at a \$1B Macro Fund; performed sensitivity analysis on hardware/compute costs across 3 asset classes
- Presented investment theses to CEO on long-term impact of LLM adoption on enterprise SaaS productivity.

LEADERSHIP EXPERIENCE

President of Risk Management / SVSH Prevention

2024 - 2025

UC Berkeley CalGreeks

Berkeley, CA

- Managed 10+ weekly events for 500+ attendees across 35 fraternities, implementing emergency response protocols and leading risk workshops for 70+ delegates

Team Captain / Jiu Jitsu Instructor

2024 - 2025

UC Berkeley

Berkeley, CA

- Founded and lead a 20-member competitive team (3rd overall at USA Grappling CSUN); lead weekly lessons for 50+ members with a 95% retention rate.

RESEARCH

UCSF Memory and Aging Center: Co-pilot Project

Present

- Developing an LLM-based system to integrate and standardize patient data (neuroimaging, medical history) to assist in neurodegenerative disease diagnosis
- Quantifying model robustness via adversarial perturbations; implementing gradient-based attacks to identify feature sensitivity and mitigate data biases

TECHNICAL SKILLS

Programming Languages: C/C++, Python (FastAPI, Sklearn), Java, SQL, TypeScript

AI & Machine Learning: LLMs, Model Context Protocol, Prompt engineering, RLHF, Heuristic design

Cloud & DevOps: Figma, IaC; AWS: Lambda, EventBridge, IAM, CloudWatch, Cloudshell, Bedrock, S3

Software: Git, Unix CLI, CI/CD Pipelines, VIM, Observability & Telemetry (CloudWatch, Prometheus)