Ethan Cai

ethan_cai@berkeley.edu | Portfolio | LinkedIn | GitHub

Education

University of California, Berkeley Expected May 2028
B.S. Electrical Engineering and Computer Sciences (EECS)
Prior coursework: SDSU Calculus III, Linear Algebra (4.0); UCSD Extended Studies: Deep Neural Networks (4.0).

Technical Skills

Languages: Python, TypeScript, JavaScript, Java, C++, C#, HTML/CSS
Frameworks & Tools: PyTorch, scikit-learn, pandas, NumPy, CleanRL, React, Node.js, Firebase, Unity, Azure, GitHub Actions
Foundations: Machine Learning, Reinforcement Learning (PPO, curriculum learning), Large Language Models (LLMs), LLM Evaluation, Web Scraping, Statistical Analysis, Data Structures & Algorithms

Publication

Cai, E.; Kouznetsova, V.L.; Tsigelny, I.F. Metabolomics-Based Machine Learning Diagnostics of Post-Acute Sequelae of SARS-CoV-2 Infection. *Metabolites* 2025, 15(12), 801. doi.org/10.3390/metabo15120801

Research Experience

UCSF Sohn Lab (UC Berkeley URAP) September 2026 - Present
Research Apprentice
- Post-training an open-weight LLM with PPO for clinical history summarization: adapting and tuning a reliability-monitoring framework into an LLM-as-a-judge whose scores are the reward, and validating the final results against UCSF radiologists' assessments.

University of California, Santa Cruz May 2025 - June 2026
Research Intern
- Implemented prioritized level replay on CleanRL's PPO in Scenic/MetaDrive highway simulations, with a disk-backed replay buffer and rank-based resampling; replay-trained agents beat baseline PPO in 21 of 30 seeds (sign test, $p < 0.05$).

University of California, San Diego / San Diego Supercomputer Center May 2024 - December 2025
Research Intern
- Built an assay-agnostic Long COVID diagnostic from metabolomic profiles; a multi-layer perceptron classifier reached 0.899 AUC-ROC and separated Long COVID from four of five clinically similar conditions (ME/CFS, Lyme disease, POTS, IBS).

Industry Experience

RapDevPro January 2025 - May 2025
Machine Learning Engineer
- Built a golf-score prediction model for FanSlices, a fantasy sports betting platform, and an internal HTML/CSS/JS dashboard to test, compare, and visualize it.

Projects

Vigil — AI Compliance Scanner | Python, TypeScript October 2026
- Won the grand prize (Bright Data track) at the Talk to Users Hackathon (SF Tech Week, 20 selected teams) with a 12-hour build of an LLM scanner that flags misleading SEC Marketing Rule claims in a firm's marketing.
- Led the project: defined the problem and pitch, ran live interviews with target users and turned their feedback into product changes, and ramped up on SEC marketing compliance mid-event.
- Took over and rebuilt from scratch a two-pass, cross-referenced inference pipeline and a harness testing it against a one-prompt LLM wrapper (F1 0.97 vs. 0.57, 4 vs. 111 wrong flags, 125 cases); added 24 cases from 12 SEC orders fining advisers.

Human Benchmark 3D — VR Memory Benchmark | Unity, C#, Firebase May 2025 - May 2026
- Built and published a VR block-tapping memory test on SideQuest (1,800+ downloads), with a Firebase backend that logs performance and benchmarks players by age bracket. Founded and led a 20+ member VR Club.

Grant Test Auto — AI Grant Office | TypeScript, React, SpacetimeDB, Claude, Azure October 2026
- Designed and built an AI grant office (Cal Hacks 13.0) automating a lab's tedious checks against obscure funding rules (no supplies in a grant's last 90 days) and conflicting official records; built its database backend, React front end, evaluation and GitHub Actions CI/CD to Azure; caught 38 of 38 freezes on 37 NIH grants and 28 of 28 audited unallowable items.

Honors & Awards

Talk to Users Hackathon, Grand Prize (Bright Data Track, 2026) · USACO Gold (2024) · AIME Qualifier, 3 years (2023-2025) · Games for Change Winner (2024) · National Merit Scholar (2026)