

EVAN LEE

📞 (925)-791-9632 ✉ exl2813@berkeley.edu 💻 [evxlee](https://evxlee.github.io) 🏠 evxlee.github.io 🌐 [EvxLee](https://EvxLee.com)

Data Science and Economics student at UC Berkeley building ML systems behind products used by litigators, clinicians, and venture analysts. Currently running customer discovery and outbound at an early-stage GPU infrastructure startup.

EDUCATION

University of California, Berkeley

Expected Graduation: May 2029

Bachelor's in Data Science, Bachelor's in Economics, Minor in Computer Science

EXPERIENCE

Git Machine 📁

Jun 2026 – Present

Forward Deployed Engineer Intern

San Francisco, CA

Git Machine builds software that recovers wasted GPU compute so teams run more workloads on hardware they already own.

- Ran 10+ discovery calls with engineers and technical leads at target companies, diagnosing GPU serving bottlenecks and moving qualified prospects to founder conversations
- Built the outbound motion from zero, created a qualifying test that screened prospects on GPU workload ownership, then ran campaigns to 100+ prospects at 50+ companies across multimodal AI, robotics, and biotech industries
- Tested which pain points landed across discovery calls, reframing the pitch to lead with the capacity and throughput gains engineers raised over cost savings

Ithildin (Berkeley SkyDeck) 📁

Jun 2026 – Aug 2026

AI/ML Engineer Intern

Berkeley, CA

Startup building a deposition platform that transcribes testimony and flags inconsistencies for litigators and court reporting firms.

- Designed the transcription and diarization engine in two passes to cover litigator use cases: a real-time transcript to follow testimony in the room, and a full-recording pass to review and correct the final transcript
- Benchmarked candidate transcription and speaker-labeling models against official deposition transcripts to select the production stack, weighting speaker accuracy above word accuracy
- Replaced voiceprint enrollment with one-click speaker naming, eliminating upfront setup for users and proving more reliable in speaker labeling

UCSF Yang Lab

May 2026 – Present

NLP Researcher - advised under Dr. Hanxue Gu

San Francisco, CA

- Developed an NLP pipeline extracting 6 clinical-trial metadata fields from biomedical articles; automated prompt optimization using DSPy to raise accuracy by 9% on 60 held-out papers
- Building a multi-agent system that finds eligible patients for clinical studies, letting clinicians correct selections mid-run

LG NOVA

Feb 2026 – May 2026

Machine Learning Engineer (Contract)

Santa Clara, CA

LG Corp's venture arm; Atlas is its platform that consolidates Monday.com and PitchBook data into one portfolio view.

- Automated the GCED team's quarterly cycle report, turning a day of manually sorting startups by vertical into an on-demand AI-generated summary inside Atlas
- Built the semantic search behind Atlas's portfolio chatbot, helping internal teams answer portfolio questions in one query instead of navigating multiple tools by hand

FactGrid Cuneiform

Jan 2026 – Apr 2026

NLP Researcher - advised under Dr. Adam Anderson

Berkeley, CA

- Fine-tuned a ByT5 model to correct OCR errors in scanned Akkadian transliterations, cutting character error rate by 75% so Assyriologists retrieve the right words when querying FactGrid's open cuneiform database

PROJECTS

Data Matchmaker Benchmark AI Agent 📁 | *AgentX – AgentBeats Competition (Berkeley RDI)* Dec 2025 – Jan 2026

- Placed 2nd in the OpenAI-sponsored Finance Agent track; deployed an evaluator agent that scores other AI agents' accuracy on data integration tasks

Equity-Focused College Matcher (EquiMatch) 📁

Nov 2025

- Placed 3rd of 50+ teams at UC Berkeley's 7th Annual Datathon for Social Good; built a college recommender app for underserved students, ranking 6,000+ institutions by L2 distance over equity metrics
- Added a chatbot that let users learn about their recommended colleges and understand the reasoning behind their results

TECHNICAL SKILLS

- **Languages:** Python, SQL, R, Java
- **Tools & Frameworks:** PyTorch, Hugging Face, RAG, scikit-learn, XGBoost, OpenCV, Pandas, Tableau, PostgreSQL, Docker, AWS, Git