

EVAN LEE

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

EDUCATION

University of California, Berkeley

Expected Graduation: May 2029

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

Relevant Coursework: Principles and Techniques of Data Science, Discrete Mathematics and Probability Theory, Data Structures, Linear Algebra, Introduction to Business Analytics, Human Contexts and Ethics of Data Science

EXPERIENCE

Ithildin (Berkeley SkyDeck) ☞

Jun 2026 – Aug 2026

AI/ML Engineer Intern

Berkeley, CA

- Designed and built the real-time transcription and speaker diarization engine powering the deposition platform, pairing Deepgram ASR with diart
- Engineered a benchmarking framework scoring candidate ASR and diarization models against certified court-reporter transcripts, weighting speaker accuracy above word accuracy to select the production stack

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 using BioBERT, SciFive, and DistilBERT; automated prompt optimization with DSPy and GEPA to raise accuracy by 9% on 60 held-out papers
- Building a multi-agent system that automates patient cohort selection for clinical studies, incorporating a steering layer that applies clinician corrections mid-run

UCSF Tech Lab

Mar 2026 – Present

Data Science Researcher - advised under Faris Raza

San Francisco, CA

- Engineered a multimodal data pipeline over the NIH AI-READI diabetes dataset, integrating CGM, ECG, wearable, and clinical signals into a unified patient-level feature table
- Ran cross-modal statistical analysis on glycemic control, applying non-parametric testing, multiple-comparison correction, and regression adjustment for confounders

LG NOVA

Feb 2026 – May 2026

Machine Learning Engineer (Contract)

Santa Clara, CA

- Engineered a RAG pipeline over the startup portfolio using Amazon Bedrock Titan v2 embeddings and pgvector, enabling semantic search for an internal analyst chatbot
- Built integration tests spanning the full ingest → embed → search pipeline and a Precision@5 framework scoring retrieval quality, catching result regressions before they reached analysts

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 Akkadian cuneiform, training on 18,579 scholar-verified transliteration lines across 2,005 tablets
- Decreased CER by 75% on held-out tablets, improving chrF++ by 18.5 points and BLEU by 28.2 points

PROJECTS

Data Matchmaker Agent ☞ | *AgentX – AgentBeats Competition (Berkeley RDI)*

Dec 2025 – Jan 2026

- Placed 2nd in the OpenAI-sponsored Finance Agent track; deployed an LLM evaluator agent that auto-scores candidate agents on multi-source data integration tasks

EquiMatch ☞ | *Python, scikit-learn, FastAPI*

Nov 2025

- Placed 3rd of 50+ teams at UC Berkeley's 7th Annual Datathon for Social Good; built a college recommender pipeline for underserved students ranking 6,000+ institutions by RBF-normalized L2 similarity over KNN-imputed equity metrics
- Integrated an LLM chatbot (Llama 3.1) to support natural-language queries over recommended colleges

TECHNICAL SKILLS

- Languages:** Python, SQL, R, Java
- Data Science/ML:** Pandas, NumPy, scikit-learn, XGBoost, PyTorch, TensorFlow/Keras, Matplotlib, Seaborn, OpenCV, Tableau, Hugging Face, DSPy, RAG, Hypothesis Testing, Regression, Clustering
- Tools/Systems:** AWS (Bedrock), Azure, PostgreSQL, pgvector, Docker, Git