Evan Lee

(925)-791-9632 exl2813@berkeley.edu evxlee evxlee.github.io 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, Data Structures, Foundations of Data Science, Structure and Interpretation of Computer Programs, Linear Algebra

Experience

LG Nova Feb 2026 – May 2026

Machine Learning Engineer (Contract) *Berkeley, CA*

- Engineered a RAG pipeline over LG Nova's startup portfolio using Amazon Bedrock Titan v2 and pgvector to enable semantic search for an internal portfolio chatbot
- Implemented incremental embedding logic to re-process only changed records, reducing Bedrock API costs as portfolio data scales
- Built end-to-end integration tests and a Precision@5 evaluation framework to validate retrieval quality across the full ingest → embed → search pipeline

FactGrid Cuneiform Feb 2026 – Present

NLP Researcher - advised under Dr. Adam Anderson *Berkeley, CA*

- Fine-tuning a Byte-T5 model on ancient Akkadian cuneiform transliterations to automatically detect and correct OCR errors for FactGrid's linked cuneiform language database

UCSF Tech Lab Jan 2026 – Present

Data Science Researcher - advised under Faris Raza *San Francisco, CA*

- Conducting ML research on CGM data from the AI-READI multimodal diabetes dataset to identify predictive patterns in blood glucose dynamics

Ergonosis Jan 2026 – Mar 2026

Data Engineer Intern *Berkeley, CA*

- Built a Python ETL pipeline integrating Plaid and Microsoft Graph APIs to extract and normalize bank transactions and enterprise email data into structured JSON for downstream analytics

FibuVerse Dec 2025 – Feb 2026

Machine Learning Engineer Intern *Berkeley, CA*

- Developed a natural-language query processing system using semantic embeddings to classify and route user questions to predefined SQL queries for cardio workout metrics
- Reduced LLM usage costs by 78% and improved query routing speed by 2.7x by replacing a pure-LLM pipeline with a hierarchical template-matching system

Projects

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

- Reached 2nd place in Phase 1 Finance Track (sponsored by OpenAI); built and deployed infrastructure for a Schema Merging Evaluation agent, automating benchmark assessments via CI/CD pipelines

Equity-Focused College Matcher (EquiMatch) | *Python, Tableau* Hackathon, Nov 2025

- Built a college matching pipeline encoding 6,000+ institutions by equity metrics, ranking by Euclidean distance
- Integrated a LLM chatbot (Llama 3.1) to support natural-language queries over recommended colleges; project awarded 3rd place at UC Berkeley's 7th Annual Datathon for Social Good

Personalized Song Recommender (LyricCal) | *Python* Sept 2025 - Dec 2025

- Engineered a data pipeline ingesting raw lyrical and acoustic data from Spotify and Genius APIs, preprocessed features via lemmatization and stopword removal
- Utilized a Transformer model (DistilRoBERTa) to classify song moods from lyrics and trained a neural network autoencoder on acoustic features to produce song embeddings for cosine similarity-based top-K recommendations

Technical Skills

- **Languages:** Python, SQL, R, Java
- **Data Science/Machine Learning:** Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn, OpenCV, PyTorch, TensorFlow
- **Tools:** Git, Docker, Hugging Face, Tableau