

Noah Lee

513-609-6390 | noah.lee@berkeley.edu | noahlee.org | [LinkedIn](#) | [GitHub](#) | Berkeley, CA

EDUCATION

University of California, Berkeley

Berkeley, CA

B.A. Data Science, Minor in Mechanical Engineering | GPA: 4.0/4.0

Expected May 2028

- **Coursework:** Computational Structures in Data Science, Linear Algebra & Differential Equations
- **Activities:** CalSol Solar Vehicle Team – race strategy and energy simulation; Open Project – Game Dev

EXPERIENCE

Software Engineering Intern

Oct 2026 – Present

Sidekick (YC S26)

Engineering Analytics Intern

May 2026 – Aug 2026

Barnes Aerospace

West Chester, OH

- Reduced indirect time waste **11%** and increased weekly productivity **13%** by deploying a Python constrained-scheduling application with greedy optimization and an operator UI for **30 employees, 20 machines**, and **3 shifts**; scheduled **80 part types** from **600K–1M records** under training, staffing, and supply constraints.
- Surfaced **\$50K** in held inventory (**18% of inventory**) by building a Power BI dashboard from SQL extracts and Excel Power Query, classifying order status and delivering recurring self-service reporting to supervisors and OEM management.
- Supported a **\$10K increase in average daily shipment value** by translating dashboard findings into production priorities that reduced unnecessary production and contributed to a projected **17% gain** toward monthly targets.

Undergraduate Research Apprentice

Jan 2026 – May 2026

UC Berkeley Space Sciences Laboratory, Prof. Ali Rahmati

Berkeley, CA

- Classified **3 solar energetic particle events** across **6 years** of NASA MAVEN data (2019–2025) by analyzing particle flux in Python and IDL and correlating count spikes across ion, electron, and X-ray sensor channels.
- Modeled Scorpius X-1 X-ray emission by fitting regression curves to observed flux and quantifying pulse periodicity, creating a quantitative baseline for studying behavior in similar sources.

PROJECTS

Lesion Atlas | Python, JavaScript, Cloudflare Workers, FastAPI, IndexedDB | [Live Demo](#) | [GitHub](#)

2026

- Built and deployed an AI skincare web application on Cloudflare, integrating a Workers AI LLM chatbot with a privacy-first browser workspace that stores user photos, scale calibration, and lesion annotations locally in IndexedDB.
- Developed a personalized CNN training, evaluation, and inference pipeline in Python using chronological train/validation/test windows with **3-day leakage gaps**, validation-selected thresholds, and lesion-center matching within **2 mm**.
- Hardened the ML and deployment stack with profile isolation, dataset/hash validation, reproducible model/evaluation artifacts, FastAPI integration, and **58 automated tests** covering temporal splits, training, evaluation, inference, and request boundaries.

MLB Game Outcome Prediction System | Python, scikit-learn, pandas | [Live Demo](#) | [GitHub](#)

2026

- Built an end-to-end MLB prediction pipeline on **11 seasons** (approximately **25K games**), engineering point-in-time temporal features and training logistic regression and HistGradientBoosting classifiers.
- Deployed calibrated game and team-win probabilities with Platt scaling and validated point-in-time correctness using reference-implementation, future-deletion, and shuffled-label tests.

Point-in-Time Sector Return Study | Python, pandas, pytest

2026

- Identified look-ahead bias in recession-based sector analysis: after accounting for NBER announcement lags, **4 of 9 sectors** reversed recession-versus-expansion return gaps; validated results with a **331-offset circular-shift permutation test** ($p = 0.232$) and **278 automated tests** catching all 11 targeted mutations.

PUBLICATION

Lee, N., Banks, A., & Zheng, V. (2024). *Orbit Determination of the Near-Earth Asteroid 2012 FN62. Summer Science Program.*

- Tracked NEA with Morehead Observatory, calculated orbital elements, fit trajectory with Monte Carlo simulation in Python.

TECHNICAL SKILLS

Languages & Libraries: Python, SQL, JavaScript/TypeScript, HTML/CSS, IDL; pandas, NumPy, scikit-learn, Matplotlib, OpenCV, Streamlit

ML & Statistics: Regression, Pytorch, classification, gradient boosting, model calibration, computer vision, self-supervised learning, time-series analysis, statistical inference, hypothesis testing, Monte Carlo simulation

Data & Platforms: Data cleaning, data pipelines, exploratory data analysis (EDA), Power BI, Excel (Power Query), FastAPI, SQLite, IndexedDB, Cloudflare Workers, Workers AI

Software & AI: Git/GitHub, pytest (unit/mutation testing), LLM application development, Agentic AI, workflow automation, Arduino