

Daniel Lee

408-826-2633 | danlee4293@gmail.com | linkedin.com/in/lee4293 | github.com/lee4293

EDUCATION

University of California, Davis
Master of Science in Computer Science
Purdue University
Bachelor of Science in Data Science

Davis, CA
September 2026 – June 2028
West Lafayette, IN
August 2022 – May 2026

WORK EXPERIENCE

Software Engineer Intern

July 2024 – October 2024

Daedalus Labs

San Francisco, CA

- Built a **Node.js** ingestion pipeline on **Microsoft Azure** to standardize fragmented blockchain webhook payloads, persisting **10,000+** Bitcoin Rune transactions in **Azure SQL** for reliable transaction monitoring
- Developed a near-real-time monitoring system for Bitcoin transactions and Nostr protocol events, delivering alerts through Discord webhooks and a Chrome extension
- Evaluated Fractal Bitcoin's transaction model using protocol documentation and transaction-level data, translating findings into monitoring and ingestion requirements during **Agile** sprint reviews

Data Analyst Intern

August 2023 – May 2024

NSF I-GUIDE

West Lafayette, IN

- Integrated flood-inundation rasters from **345+** dams with census-block socioeconomic data using **Python** to assess population exposure and vulnerability across flood-prone regions
- Containerized and distributed geospatial pipelines using **Docker**, **Apache Spark**, and **Apache Sedona**, standardizing local and **HPC** environments while scaling raster-vector operations and large-area spatial joins
- Profiled raster-heavy Python workloads using timing and memory diagnostics, isolating computational and data-access bottlenecks to target high-cost pipeline stages

RESEARCH AND PUBLICATIONS

Undergraduate Student Research

August 2022 – May 2023

The Data Mine (Purdue University)

West Lafayette, IN

- Investigated rising pediatric melatonin ingestion incidents with the **CDC PROTECT initiative**, analyzing product labels and online marketplace descriptions to identify potential sources of consumer misunderstanding and accidental misuse
- Filtered and standardized labeling data for **347+** melatonin products from the NIH Dietary Supplement Label Database, creating a reference schema for comparing ingredient, dosage, warning, and marketing language across products
- Built **NLP** pipelines with **Hugging Face Transformers** to perform topic modeling and sentiment analysis on **3,500+** Amazon reviews and **410+** Reddit posts, surfacing recurring consumer concerns, misconceptions, and usage patterns to inform the broader labeling analysis

PROJECTS

Nuveri.app | TypeScript, Next.js, React, Anthropic API, Supabase

August 2026 – Present

- Built a full-stack research platform with **Next.js**, **TypeScript**, and **Supabase/PostgreSQL**, aggregating scholarly data to let users search, evaluate, and trace scientific claims back to primary research
- Engineered a **TypeScript** and **Anthropic** analysis pipeline combining LLM inference with deterministic validation to produce source-grounded research assessments while reducing unsupported conclusions

LEO Satellite Telemetry Anomaly Detection | Python, scikit-learn, Pandas

January 2026 – May 2026

- Built a reproducible **Python** and **scikit-learn** anomaly-detection pipeline, engineering 28 statistical and channel features from **2,123 telemetry segments** and benchmarking supervised models to identify satellite outages, interference, and link failures
- Extended detection to unlabeled orbital data by engineering 18 features from **9,919 Starlink records** and ensembling Isolation Forest with One-Class SVM to flag potential position anomalies and unexpected orbital behavior without ground-truth labels

TECHNICAL SKILLS

Languages: Python, Java, JavaScript, C++, SQL

Frameworks/Libraries: React, Next.js, Node.js, scikit-learn, Pandas, Apache Spark, Apache Sedona, Hugging Face Transformers, Tailwind CSS

Databases/Cloud/Tools: PostgreSQL, Azure SQL, Microsoft Azure, Docker, Git, Jupyter