

# NATHAN GIN

+1 (650) 759-3494 | nathanbsgin@gmail.com | San Mateo & Irvine, CA

---

## EDUCATION

### University of California, Irvine

Master of Science (M.S.) in Statistics | GPA 3.85

December 2026

Bachelor of Science (B.S.) in Data Science | GPA 3.6

December 2023

## WORK EXPERIENCE

### Tarsus Pharmaceuticals — IT Intern, ServiceNow Enterprise RAG Chatbot

Jun 2025 – Sep 2025

- Designed an enterprise Retrieval-Augmented Generation (RAG) chatbot integrating ServiceNow, Databricks, vector search, and LLMs, improving response time from days to seconds; authored end-to-end architecture documentation.
- Built secure OAuth 2.0 ingestion pipelines with full/incremental sync to Delta tables; implemented metadata logging.
- Implemented embedding pipelines with incremental updates, ID generation, session memory, and similarity scoring.
- Configured Azure API Management for routing, authentication, CORS, and request transforms; supported a single-page HTML/CSS/JS UI with Okta Authentication.

### Outlier AI (Remote) - Data Science AI Trainer

January 2024 to March 2025

- **Model Optimization:** Optimized and evaluated prompt responses for quality and consistency in large language models
- **Peer Review:** Reviewed output for technical correctness and contributed to improving model truthfulness

### Field AI (Mission Viejo, CA) - Data Labeler

May to October 2024

- **Autonomous Vehicles:** Labeling datasets related to autonomous vehicle navigation in diverse environments.
- **Data Precision:** Ensured high-quality data annotations to train machine learning models effectively.

### Capstone Project (Irvine, CA)

Collaborative Research with UCI Stem Cell Research Center

January to June 2023

- **Spinal Cord Injury Analysis:** Evaluated the effectiveness of stem cell treatments for SCI using statistical models and hypothesis testing on murine data. Proposed longitudinal analysis to track recovery trends between treatment groups.
- **Machine Learning Automation:** Utilized DeepLabCut to automate movement tracking in ladder beam task videos. Trained a deep neural network, achieving a final error rate of 0.096%, which improved scoring accuracy and efficiency.

### Cepheid (Sunnyvale, CA)

Global Access Digital Research & Analysis Intern - Web Strategy

June to September 2022

- **Dashboard Development:** Developed interactive dashboards using Google Data Studio to visualize quality control metrics, enabling leadership to identify trends and improve decision-making processes.
- **Search Optimization:** Utilized Google Data Studio, Google Analytics, and Conductor to enhance search performance, resulting in a 7% increase in impressions and a 15% improvement in position for website visibility.
- **Performance Monitoring:** Monitored search trends for Cepheid's products and analyzed data from the past year to implement improvements, including content mapping and optimizing international and U.S. websites.

Global Access Digital Research & Analysis Intern - Marketing Communications

June to September 2021

- **Media Sentiment Analysis:** Evaluated sentiment across various media platforms, identifying a 12% rise
- **Data Management:** Tracked and cleaned sales data using Smartsheet and Salesforce, ensuring actionable insights

## PROJECTS

**Stroke Prediction (R, Python):** Built logistic regression and decision tree classifiers to analyze RACE score, age, and stroke status; evaluated predictive performance across models.

**Machine Learning (Python, scikit-learn):** Predicted property prices with random forests, neural nets, logistic regression, and KNN; cleaned inconsistent data and improved model accuracy using MAPE.

**SQL (PostgreSQL):** Designed ETL pipelines for IMDB datasets with advanced queries for trend analysis and relational insight

**AI Game Agents:** Built Checkers AI in Python with Monte Carlo Tree Search (top 10% in class) and Othello AI in C++ with heuristic search (top 15%).

**Mobile App Development:** Co-designed a food & movie recommendation app with Figma prototypes and Python backend.

**Data Visualization:** Developed dashboards in Tableau and Google Data Studio; currently completing Google Data Analytics Professional Certificate and Tableau Specialization.

**Biostatistics Application:** Analyzed longitudinal beta-carotene serum data from a Phase II clinical trial (n=46) with linear mixed effects models, identifying dose-dependent pharmacokinetics.