Varmin Singh

singh.varmindip@gmail.com | (559) 903-2599 | Berkeley, CA | linkedin.com/in/varmin | github.com/Varminn | Website

EDUCATION

University of California, Berkeley **Berkeley, CA**
Bachelor of Arts: **Data Science** **08/2022 – 05/2026**
Master of Science: **Statistics** **08/2026 - 05/2027**
Relevant Coursework: Applied Machine Learning (Graduate), Big Data & Development (Graduate), Data Engineering/Databases, Computer Architecture, Probability for Data Science, Data Structures, Artificial Intelligence, Designing Deep Neural Networks, Data Mining and Analytics, Operating System & System Programming

Experience

Boeing **05/2025 - Pres.**
Data Science Intern **Long Beach, CA**
- Engineered and deployed an end-to-end Retrieval-Augmented Generation solution that enables 400+ program managers to perform complex, natural language queries across 1M+ rows of project schedule data.
- Architected a reusable Python ETL framework to process, standardize, and index data from over 2,500 project schedules, creating a scalable foundation for the primary internal LLM and future AI initiatives.
- Refactored legacy pipelines to run on a new CI/CD framework using GitLab Runner and Kubernetes, transitioning from local machines to a centralized, automated system for real-time database updates.

Penserra **06/2024 – 05/2025**
Data Engineer Intern **Orinda, CA**
- Developed a cloud-based NLP pipeline using AWS and Gemini API to ingest streaming transaction data from financial documents into PostgreSQL, ensuring near real-time compliance checks for trades and 98% accuracy.
- Integrated a Python-based data pipeline to support the launch of the new Qualified Contingent Transactions (QCT) line of business to reduce manual data reconciliation time by over 35%, ensuring compliance with SEC regulations.
- Designed and implemented a data governance framework to ensure data quality and integrity for financial compliance logs, centralizing data accessibility and creating a single source of truth for the operations team.

University of California, Berkeley **08/2024 - 12/2024**
Academic Intern, Computer Architecture **Berkeley, CA**
- Mentored 100+ students during office hours, explaining complex technical concepts and C++ implementations.
- Collaborated and led technical workshops with live-coding demonstrations on RISC-V assembly, pipelining, parallel processing, multi-threading, and CPU design to improve student understanding and project success.

University of California, Berkeley **01/2024 – 05/2024**
Research Assistant **Berkeley, CA**
- Conducted research on AWS components for real-time anomaly detection aimed at predicting cloud outages, leading to a 30% improvement in detection accuracy compared to baseline models.
- Developed automated reporting systems that analyzed data related to AWS infrastructure for faster issue resolution.

Projects

Panorama - News Across the Political Spectrum (SF Hacks 2025 Winner)
- Developed a full-stack, multi-partisan news search engine using Python (FastAPI), React, and MongoDB to aggregate articles across the political spectrum (left, center, right), providing balanced topic coverage.
-Leveraged Perplexity API for live news aggregation and OpenAI API for AI-driven article summarization and Q&A; implemented user authentication, search history, and bookmarking features via MongoDB.

Tops Formatter (Penserra)
- Implemented a full-stack web application for automated end-of-day trade recaps, ensuring 100% compliance with internal audit standards while accurately calculating partner broker commissions.
- Engineered a real-time data pipeline from TOPS SEQ to Azure SQL Server with secure authentication, saving the operations team 15+ hours weekly and providing traders with instant activity updates.

Skills and Interests

Languages: Python, SQL, Java, C, C++,
Tools: Pandas, AWS, Docker, Node.js, NumPy, Django, CI/CD, PostgreSQL, Scikit-Learn, PyTorch, OpenCV