

Sourabh Sahoo

Senior Product Engineer (AI / ML)

San Francisco, CA | sourabhsahoo10@gmail.com | [linkedin.com/in/sourabhsahoo](https://www.linkedin.com/in/sourabhsahoo)

SUMMARY

Product engineer building full-stack, AI-native products from concept through production. Shipped developer tools, retrieval systems and user-facing workflows at an early-stage startup, working directly with customers and owning technical and product decisions. Helped grow the startup from pre-revenue to \$500K+ ARR.

CORE SKILLS

Languages: Python, TypeScript, Java, SQL

AI / LLM: RAG & semantic retrieval, agents & tool use, LLM evals, Fine-tuning (SFT, LoRA, RLHF), MCP servers, PyTorch, Hugging Face, CNNs & computer vision, embeddings & vector search

Systems & Infra: APIs & dashboards, React, AWS, GCP, Airflow, Chroma, CI/CD, MLOps

EXPERIENCE

Senior Product Engineer (AI / ML) | MacroX

2025 – Present | San Francisco

- Translated recurring client feedback into MacroX’s first **MCP server** over proprietary data, automating manual analyst workflows and becoming the default interface for internal analysts and clients to query data and run analysis
- Built a **hybrid retrieval and reasoning system** over S&P 500 earnings calls since 2004, combining structured SQL, reranking, and temporal/regime comparison; increased financial Q&A accuracy to 90%, substantially outperforming a vanilla RAG baseline
- Fine-tuned **open-weight LLMs** on expert-labeled data for hawkish/dovish classification and central bank reports, optimizing agreement with domain experts on a subjective task. Shipped a LoRA-adapted model after benchmarking macro F1, recall, grounding, latency, and cost
- Architected and built MacroX’s API platform from the ground up, exposing 1,000+ data endpoints with sub-second retrieval; designed the production backend, caching, and invalidation mechanisms for fast, reliable access to all serving clients

Data Scientist | MacroX

2023 – 2025 | San Francisco

- Built a real-time city-level housing supply-demand indicator using diverse alternative data, leading Zillow and Redfin by one month and increasing out-of-sample Case-Shiller prediction R^2 from 0.04 to 0.18
- Built and maintained ML and data pipelines (Airflow, AWS/GCP) with ongoing model monitoring and maintenance
- Designed and deployed real-time alternative-data indicators for consumer financial stress, delivering signals earlier than the University of Michigan Consumer Sentiment Index and improving prediction accuracy by 30%
- Built and continuously evolved MacroX’s primary client dashboard, owning React and Dash pages, backend integrations, and end-to-end feature delivery for all customers

Senior Development Architect | Standard Chartered

2018 – 2021 | India

- Built ML models on credit-card data to surface customer insights; contributed signal-processing and model-architecture work
- Promoted within first year; team recognized with the company's “Best Newcomers” award

PROJECTS

LLM Memory Manager | Independent Project

- Developing a model-agnostic memory layer for LLMs - tackling what to retain, retrieve, and expire over long horizons, where context windows alone fall short
- Designing a local-first architecture with on-device storage and a drop-in interface compatible with multiple LLM providers

EDUCATION

M.S. in Robotics: University of California San Diego

2021 – 2023

Focus: convolutional neural networks & deep neural networks

B.Tech. – Aerospace (Honors): Indian Institute of Technology Kharagpur

2014 – 2018