

Rohan Sawai

sawairohan90@gmail.com | +1 540-824-9544 | linkedin.com/in/rohan-sawai | github.com/rohansawai | Washington DC

Education

Master of Science in Computer Engineering, Machine Learning Specialization

Jan 2025 – Dec 2026

Virginia Tech, Blacksburg, VA

GPA: 3.81/4.00

Relevant Coursework: Big Data Systems, Compiler Optimization, Trustworthy ML, Applications of ML

Bachelor of Technology in Electronics and Communications

Aug 2019 – May 2023

Veermata Jijabai Technological Institute (VJTI Mumbai)

Mumbai, India

Experience

Amazon Web Services

May 2026 – Aug 2026

Software Development Engineer Intern | Java, AWS Cloud Infra

Herndon, VA

- Designed and built a serverless Java pipeline to detect and remediate CVEs that a deployment target inherits transitively from other software on its host, closing a security blind spot across 60M+ targets.
- Implemented fault-tolerant backend services on Lambda, SNS, SQS, S3, and DynamoDB with thread pool management and optimistic locking, processing 10K+ events per batch; replaced week-long manual remediation campaigns run by Security Engineers with one-click remediation completing in under 30 minutes, cutting manual effort by 98.6%.
- Owned observability and release quality across the services: built a CloudWatch monitoring stack that cut incident detection time by 70%, drove 95%+ test coverage (JUnit, Mockito), and managed CI/CD end-to-end.

Center for Geospatial Information Technology at Virginia Tech

Feb 2025 – Present

Graduate Research Assistant | Django, JavaScript, Leaflet, OpenLayers, PostgreSQL

Blacksburg, VA

- Built reusable analytics widgets for a state-wide crash-analysis platform, now used across 10+ dashboard views by utilizing common object oriented design patterns. Designed them to be configurable, so a single widget handles 20-30 configurations instead of a new component per view.
- Delivered UI components and new analytics features, letting users combine map selections — county, radius, freeform drawn regions — with 20+ filters in a single query.
- Automated CI/CD pipelines for test setup & traffic replay services, replacing ad-hoc manual deployment, with repeatable, pipeline-driven releases and saving 2-4 hours of manual work per release.

Citibank

Jul 2023 – Jan 2025

Software Development Engineer | React, Redux State Management, Java, AWS

Pune, India

- Migrated 10+ legacy Citi Velocity components to React 18 over two quarters, covering high-volume dashboards, analyst research pages, and live-streaming modules; cutting page load and interaction latency on the slowest views from 4+ seconds to under 400ms.
- Applied code splitting and memoization to cut redundant renders, and throttled WebSocket updates to eliminate render thrashing on high-frequency streaming views.
- Built and extended Java REST APIs backing the migrated dashboards and maintained Jenkins CI/CD pipelines for deployments to AWS, alongside primary frontend ownership.
- Provided technical guidance to 6 engineers on architecture, implementation, and deployment; mentored 2 SDE Interns through code reviews and debugging.

Projects and Publications

CDC Stream — Real-time Database Change Capture & Replay

[github.com/rohansawai/proto-sstream]

PostgreSQL WAL, Redis Streams, Apache Kafka, Protobuf, gRPC, Django

- Built WAL-based CDC pipeline with **dual delivery**: Redis Streams (low-latency) + Kafka (durable archival); used PostgreSQL LSN as global ordering/deduplication key across both paths, mirroring production-grade CDC systems (Debezium, AWS DMS).
- Built gRPC-based replay service with time-range-scoped historical replay and Django admin dashboard for job management.

LLM Reasoning via Supervised Fine-Tuning

[github.com/reds-lab/Project-Reasoning-SFT-LLM]

Qwen2.5-3B, vLLM, LLaMA-Factory, DeepSpeed ZeRO-3, FlashAttention-2, SLURM, lighteval

- Fine-tuned Qwen2.5-3B-Instruct on a curated 15K subset of AceReason-1.1-SFT to improve mathematical and scientific reasoning, training with DeepSpeed ZeRO-3 and FlashAttention-2 on SLURM-scheduled H200 nodes.
- Placed 3rd of 14 teams, averaging 0.518 across AIME25, MATH, GPQA, and Olympiad benchmarks against a 0.446 base-model baseline — beating random-subset selection (0.492) through a targeted data-selection strategy.

DARN: Research on adapting Computer Vision Models for Satellite Imagery

[openaccess.thecvf.com/CVPR2026]

Co-author | PyTorch, AWS SageMaker, MLflow, Computer Vision | Research at a YC W26 geospatial AI startup

- Co-authored paper on adaptation of foundation Computer Vision models to satellite imagery (SAR), accepted at CVPR 2026.

Skills

Languages: Java, Python, C++, JavaScript | **Cloud and Data Infra:** AWS, Kafka, Docker

CI/CD & DevOps: Jenkins, Git | **Backend & Protocols:** gRPC, REST API, NodeJs, React, TailwindCSS, Django, Protobuf

AI/ML: PyTorch, vLLM, Hugging Face | **Databases:** PostgreSQL, SQL, DynamoDB, Redis