Caden Roberts

408-599-9701 | cawrober@ucsc.edu | https://www.linkedin.com/in/cwro | https://github.com/cadenroberts

AI Infrastructure Engineer specializing in distributed ML systems, deterministic LLM evaluation frameworks, and production-grade automation platforms across fintech and scientific computing environments.

Experience

Paystand												**Jan 2026 – Present**
Machine Learning Engineer									*Santa Cruz, CA*
- Designed and deployed LLM-driven automation infrastructure orchestrating crawl, retrieval, inference, adversarial evaluation, and publishing stages for enterprise marketing and finance workflows.
- Architected a deterministic consensus-gated evaluation system leveraging 50 grading agents and 88-criterion validation checks (4,400 automated constraints per output), enforcing $\geq 90\%$ supermajority approval prior to release.
- Built a financial variance automation platform integrating Oracle NetSuite and Abacum data pipelines with structured LLM-generated narratives, automating $>70\%$ of monthly reporting and reducing cycle time by $\sim 40\%$.
- Implemented cost-aware model routing and latency optimization strategies across multiple LLM providers, balancing reasoning depth, throughput, and token expenditure under production constraints.

BioMedAI Lab, UCSC										**Jul 2025 – Present**
Machine Learning Researcher								*Santa Cruz, CA*
- Engineered GPU-accelerated ML benchmarking pipelines for coarse-grained DNA–protein modeling using OpenMM and WESTPA across NERSC A100 infrastructure.
- Designed distributed experiment orchestration workflows enabling scalable parameter sweeps across 10 molecular systems under multi-GPU workloads.
- Optimized sampling efficiency and compute utilization by benchmarking GNN-based architectures across enhanced MD and CG simulation regimes.

Baskin Engineering, UCSC									**Jan 2025 – Mar 2025**
Research Assistant										*Santa Cruz, CA*
- Implemented custom Linux scheduling policies (FIFO, LIFO, Round-Robin) via sched_ext to benchmark kernel-level performance tradeoffs in resource-constrained environments.
- Evaluated scheduler behavior within Ecovisor, an AI-assisted cloud optimization framework targeting energy-efficient compute allocation for renewable workloads.

Selected Systems Projects

CliniRepGen | *Python, PostgreSQL, RAG, LLM Evaluation Systems*				**2025**
- Architected a reusable regulatory reporting platform (ingestion $\rightarrow$ retrieval $\rightarrow$ LLM synthesis $\rightarrow$ structured post-processing) generating CONSORT/ICH-E3 clinical trial reports from heterogeneous biomedical datasets.
- Designed deterministic retrieval pipelines with traceable source attribution and reproducibility guarantees for audit-ready regulatory documentation.
- Implemented iterative critique-loop validation enforcing checklist coverage thresholds before report finalization, reducing unsupported claims to zero.

ClinImCL | *PyTorch, MONAI, Google Cloud A100*						**2025**
- Built a distributed self-supervised contrastive learning pipeline for 3D longitudinal MRI analysis across 3,000+ scans using GPU-accelerated training workflows.
- Benchmarked embedding stability and representation consistency via UMAP and cosine similarity to validate downstream disease progression modeling.

OllamaBot | *Swift, Local LLM Inference, Multi-Model Orchestration*			**2025**
- Designed a multi-model agent orchestration runtime coordinating 4 specialized LLMs under structured tool-execution with priority-based context budgeting and frame-coalesced streaming.
- Implemented deterministic tool routing, parallel execution pipelines, and checkpoint-based state recovery for autonomous local inference workflows.

Technical Skills

ML Infrastructure: Distributed training, GPU workflows (A100), CUDA, Slurm, LLM evaluation pipelines, RAG architectures, model routing

Systems & Platform: Docker, Linux, sched_ext, orchestration pipelines, cost-aware inference, structured logging, checkpoint recovery

ML Frameworks: PyTorch, MONAI, OpenMM, WESTPA, contrastive learning, GNN benchmarking

Languages: Python, C/C++, SQL, JavaScript, Swift, Bash

Data & Integration: PostgreSQL, REST APIs, OpenAPI, Oracle NetSuite, Abacum

Education

University of California, Santa Cruz							**Sep 2025 – Mar 2026**
M.S., Computer Science and Engineering GPA: 3.97						*Santa Cruz, CA*

University of California, Santa Cruz							**Aug 2023 – Jun 2025**
B.S., Computer Engineering									*Santa Cruz, CA*