

Mihir Srivastava

mihir-s.com | (408) 431-4134 | m_srivastava@ucsb.edu | [linkedin.com/in/mihir-srivastava-19a342240](https://www.linkedin.com/in/mihir-srivastava-19a342240) | <https://github.com/mihir-s-05>

EDUCATION

University of California, Santa Barbara (UCSB) | Combined B.S./M.S. Program

B.S. Computer Engineering, expected 2027 | M.S. Computer Science, expected 2028

- Data Structures & Algorithms, Object Oriented Design, Deep Learning, Machine Learning, Synchronous Digital Systems, Signal Processing, Operating Systems, Computer Architecture, Embedded Devices, Networking, Parallel Computing, Programming Languages

EXPERIENCE

Everpure - Software Engineering Intern / Contractor

June '26 - Present

- Leading open-source LLM post-training for Jira/Jenkins triage, using supervised fine-tuning (SFT) & reinforcement learning (RL) with Dr. GRPO/CISPO & an ECHO auxiliary loss; running training & base-model evaluations on Baseten.
- Curated ~9 months of production-agent trajectories using successful-trace filtering & LLM-guided iterative SFT (iSFT) repair; built historical-ticket RL environments with time-filtered Jira access & model-specific tool-call formatting.
- Built an infrastructure metrics MCP server for an internal AI triage agent to distinguish infrastructure issues from FlashBlade simulator code issues, reducing false positives by 89% & increasing accuracy by 45%.
- Built a custom agent harness replacing the Claude Agent SDK, reducing token costs while supporting native adapters for MCP tools & triage-specific functionality.

Cisco - Software Engineering Intern

June '25 - Sept '25

- Extracted ASA LiNA's legacy LDAP subsystem into a standalone Linux authentication agent; replaced a customized fork with upstream OpenLDAP, remediating 45+ CVEs & achieving 5x faster authentication.

Spillbox - Software Engineering Intern

July '24 - Oct '24

- Engineered a globally distributed predictive file caching system, processing 100K+ access logs with 95%+ prediction accuracy.

Logistics Director - SBHacks

Oct '23 - Present

- Organized UCSB's largest hackathon for 300+ attendees with workshops from sponsors such as Anthropic, Google, & GitHub, raising over \$30k 3 times.

OpenROAD Project - Software Engineering Intern

March '22 - Oct '22

- Redesigned OpenROAD's Jenkins CI interface with version tracking, performance comparisons, & test-monitoring workflows for regression analysis.

RESEARCH & PROJECTS

Modular LLM Learning & Removal | Manuscript submitted to ICLR 2027

July '26 - Sept '26

- Built a low-rank residual module testbed across 75 runs (135M-1.7B); showed that removing earlier modules can damage retained-source accuracy & evaluated module-level dropout to reduce that damage.

Parameter Golf - OpenAI Competition | PyTorch, Distributed Training, Quantization

March '26 - April '26

- Developed transformer language models in PyTorch, implementing a Titans-inspired neural-memory layer & LoRA test-time training; [submitted](#) a 14.5 MB quantized code-and-model artifact achieving 1.1844 bits/byte on FineWeb.
- Prototyped JEPA & VAE architectures; trained a compressed byte-level Universal Transformer using multi-GPU distributed training on 8 H100s, EMA, & QAT, [submitting](#) a 14.71 MB artifact achieving 1.4457 bits/byte.

UCSB NLP Group - Persona-Based Language Model Debate

Sept '25 - Present

- Researching persona-based LLM debate by designing a task-adaptive persona generator & evaluation pipeline to study how reasoning-style diversity influences debate outcomes, convergence, and failure modes on HLE and game theory-style problems.

Agent Mailbox | Go, PostgreSQL, MCP, pgx, mTLS

Feb '26

- Built a secure PostgreSQL-backed MCP message bus for Codex & Claude Code agents, supporting direct and broadcast messaging, priority long polling, acknowledgements, idempotent delivery, keyset pagination, & event-driven notifications via LISTEN/NOTIFY.

Policy Radar | Python, FastAPI, React, TypeScript, OpenAI API, ChromaDB

Dec '25 - Jan '26

- Built a full-stack policy research agent spanning nine government data sources, with SSE streaming, persistent histories, & session-scoped PDF retrieval; secured document fetching with SSRF protection, authentication, & rate limiting.

LATTICE | Python, OpenAI, Groq, Gemini, LM Studio, FastAPI, SearXNG

Aug '25 - Feb '26

- Built a multi-agent orchestration engine that routes and coordinates specialized agents across sequential & parallel workflows, using agent meetings for collaborative planning with decisions & artifacts accessible through semantic search & RAG.

Decompiling Higher Order Functions in Digital Circuits – Early Research Scholars Program (ERSP)

Sept '24 - June '25

- Developed PyRTL & Egglog [analyses](#) to recover higher-order functions from netlists through fold detection for reducers, adders, & trees.

FacEmotion | PyTorch, OpenCV, Gemini, Streamlit | Data Science SB

Jan '24 - May '24

- Won 1st place among 53 teams for a CNN-based emotion-classification application trained on 35K+ images, achieving 87% accuracy & integrating Gemini-powered personalized recommendations.

TECHNICAL SKILLS

Languages: Python, Java, C/C++, Rust, JavaScript/TypeScript, Go, SystemVerilog

Frameworks: PyTorch, FastAPI, React, React Native, OpenCV, scikit-learn, Pandas, NumPy, PyRTL, vLLM, Hugging Face Transformers

Tools/Platforms: Git, GitHub, Docker, Linux, PostgreSQL, Jenkins, Arduino, STM32, ModelSim, Quartus II, VS Code