

Richard Shan

(704) 297-1039 | richardmshan@gmail.com | github.com/richard-shan | linkedin.com/in/richshan | richardshan.com

EDUCATION

Yale University

B.S. Computer Science

New Haven, CT

Aug 2026 - May 2030

North Carolina School of Science and Mathematics

GPA: 5.0/4.0 (Rank 1/354), SAT: 1580/1600

Durham, NC

Aug 2024 - May 2026

EXPERIENCE

Generative AI Engineer

The MITRE Corporation

Jun 2024 - Present

McLean, VA

- Built MERIT, a mechanistic interpretability framework that analyzes DeepSeek-R1 Distill Llama-8B latent activations across 2,000 mathematical problems of varying difficulty and domain
- Discovered feature fracturing: general reasoning features specialize into domain-specific sub-experts (geometry, number theory) under increased task complexity; demonstrated causal feature steering between reasoning modalities
- Trained Sparse Autoencoders for feature extraction with automated interpretation via judge-LLM domain classification and feature family identification
- Presented findings on LLM offensive cybersecurity capabilities to Congressional and DoD staff

AI Research Intern

Stanford University School of Medicine, AIMI Center

Aug 2025 - Jun 2026

Stanford, CA

- Designed a RAG pipeline for clinical discharge anomaly detection; conducted red-teaming and adversarial evaluation of LLMs in clinical deployment contexts
- Led development of a chest X-ray pneumonia detection model with NLP augmentations on clinical summaries; won 2nd place at Stanford-internal hackathon
- Built an Automated Bone Fracture Detection System with the Integrative Biomedical Imaging Informatics lab

Computer Science Teaching Assistant

North Carolina School of Science and Mathematics

Nov 2025 - May 2026

Durham, NC

- TA for Machine Learning, Data Structures & Algorithms, and AP Computer Science courses at the #1 ranked public high school in the US
- Built internal tooling and projects for the CS department; supported students through office hours, code review, and assignment development

Technical Director

SMathHacks Hackathon

Oct 2025 - Mar 2026

Durham, NC

- Directed NC's premiere high school hackathon to record-breaking \$80,000+ in sponsorship and 300+ hackers, both all-time records in event history
- Led technical operations (keynotes, livestreams, infrastructure) and coordinated outreach with sponsors including HRT, Google, and others

PROJECTS

MERIT: Mechanistic Explainability of Reasoning Integrity and Transparency | *Python, PyTorch*

2026

- Created 6 metrics quantifying LLM reasoning quality by logical structure, self-corrections, and coherence, moving evaluation beyond binary correctness
- Trained Sparse Autoencoders on model activations; used correlation and causal intervention analysis across difficulty levels to show features fracture into domain-specific specialists
- Identified distinct feature-governed reasoning modalities (concise calculation vs. verbose explanation) and steered between them via targeted activation manipulation

SparseForge: SAE Training on Qwen2.5-7B <i>Python, PyTorch, CUDA</i>	2025
- Trained a 205M-parameter Sparse Autoencoder (8× expansion, 28,672 learned features) on 10M activation vectors from Qwen2.5-7B’s layer-20 residual stream, extracted from SlimPajama - Achieved 90% explained variance with ~4,000 active features per input using 20-100× fewer training tokens than typical published work - Solved catastrophic feature death through linear L1 warmup scheduling, positive encoder bias initialization, soft decoder normalization via penalty term, and tied encoder-decoder initialization	
Brailiant <i>C++, ESP32, Raspberry Pi, Custom PCB, CAD</i>	2023 - Present
- Built a refreshable Braille display at \$77 build cost (market price: \$3,500-\$15,000) using solenoid-driven 3×2 tactile cells with real-time multilingual text-to-Braille translation - Designed custom PCB with ATTiny1614 microcontroller and IRF520 MOSFET array driving solenoids; ESP32CAM captures text via GPT-4o OCR, Raspberry Pi maps characters to Braille patterns over WebSocket - Raised \$10,000+ in pre-seed funding; AWS Startups Showcase Featured Startup; invited to present at Heureka Finland to audience of international scientists	
VeinVigil <i>C++, ESP32, React, Node.js, Socket.io</i>	2026
- Built a continuous IV catheter monitoring system that detects early infiltration via multi-signal fusion (temperature, pressure, humidity, rate-of-change) and computes a real-time risk score mapped to clinical INS Infiltration Scale grades - ESP32 polls sensors every 2 seconds; Node.js backend enriches readings with computed metrics (dew point, rolling median) and broadcasts via WebSocket to React dashboard with live trend charts	
ARES-M: Atmospheric Remote Environmental System for Mars <i>C++, Next.js, PostgreSQL</i>	2025
- Built a distributed environmental sensing platform using ESP32S3 sensor pods in a peer-to-peer mesh network, monitoring temperature, humidity, seismic activity, UV/IR, and soil moisture - Wrote full-stack telemetry dashboard in Next.js with Prisma ORM, Chart.js, and PostgreSQL optimized for time-series queries; embedded firmware transmits JSON payloads over HTTP every 5 seconds - Won 1st Place (Hardware) at SMathHacks 2025	

PUBLICATIONS

96 citations | h-index: 7

- “Certifying Generative AI: RAG Chatbots in High-Stakes Environments,” *IEEE Computer*, 2024. **Cover feature.**
- “Language AI at a Crossroads,” *IEEE Computer*, 2024. **Cover feature, No. 1 trending article.**
- “AI That Learns, Thinks, and Acts,” *IEEE Computer*, 2025
- “Enterprise LLMOps,” *IEEE Cloud Summit*, 2024

HONORS & AWARDS

ISEF Grand Award Winner <i>3rd Place in Robotics and Intelligent Machines (ROBO037)</i>	May 2026
ISEF Special Award Winner <i>1 of 4 winners out of 1700+ projects (ROBO037)</i>	May 2026
Lockheed Martin CodeQuest, World Champion <i>1st out of 2000+ programmers</i>	Feb 2026
USACO Gold <i>United States of America Computing Olympiad, top 5% worldwide</i>	Jan 2024
National Merit Scholar <i>Top 0.4% nationally (1/6000 of 1.4M students)</i>	Feb 2026
HackDuke Sponsor Award <i>VeinVigil: real-time IV infiltration detection (1/54 projects)</i>	Mar 2026
SMathHacks Hackathon 1st Place Hardware and Overall Winner <i>1/30+</i>	Mar 2025

TECHNICAL SKILLS

Languages: Python, C/C++, Java, JavaScript/TypeScript, SQL, L ^A T _E X
ML/AI: PyTorch, Hugging Face Transformers, Sparse Autoencoders, scikit-learn, OpenCV, LangChain
Systems & Tools: Git, Docker, AWS, Linux, CUDA, PostgreSQL, Next.js, React, ESP32/Embedded
Areas: Mechanistic Interpretability, AI Alignment, Computer Vision, NLP, RAG, Embedded Systems