Finn Sheng

Raleigh, NC (919) 619-6896
zsheng2@ncsu.edu github.com/robcholz linkedin.com/in/finn-sheng robcholz.com

Education

North Carolina State University, Raleigh, NC - BS IN COMPUTER ENGINEERING Aug 2024 - May 2028
- Minor in Business Entrepreneurship

Experience

GPU Power Research Assistant, NC State University Dec 2025 - Present
- Proposed a hybrid datacenter power forecasting method that decomposes load into base power, predicted by a time-series model, and periodic components, captured via statistical period estimation, enabling offline training with online inference.
- Evaluated asynchronous checkpoint scheduling's impact on training overhead and power fluctuation under multi-GPU execution; reproducing GoCkpt and exploring efficient async cross-GPU checkpoint synchronization to reduce training pause time.

AI Intern, Espressif Systems | https://github.com/espressif/esp-claw Jun 2026 - Aug 2026

ESP-Claw — Espressif's official open-source AI agent framework
- Designed layered memory with automatic context compaction, improving context recall and cutting context drift across long, multi-turn agent sessions.
- Improved observability and long-horizon task stability by introducing a semantic finite-state machine into the agent core/agent loop.
- Migrated the agent core from C to Rust with an async runtime, sharply reducing memory consumption while improving concurrent resource usage and memory safety.
- Re-architected the core into a clean four-layer architecture (iteration loop → base agent → special agent → orchestrator) for maintainability and component extensibility.
- Built an ESP-IDF (HAL) simulation layer and 40+ end-to-end tests; surfaced and fixed 12 memory-leak bugs.

Research Intern, Shanghai Jiao Tong University May 2025 - Jul 2025
- Benchmarked LSTM models and time-series foundation models (TimesFM) for multivariate forecasting under latency/memory constraints; deployed and profiled on Jetson Orin Nano for edge feasibility.

Selected Engineering Projects

VibeBox |Rust/Dev Tool | https://github.com/robcholz/vibebox (~200 stars) Feb 2026

Sandboxed runtime for secure coding agents
- Open-source per-project micro-VM sandbox for running coding agents safely on Apple Silicon macOS (Apple Virtualization Framework); built as a daily coding-agent user who wanted to give agents a real shell without exposing the host.
- Designed a security model with a hard guest-VM isolation boundary, repo-first mount allowlists, and a reusable session lifecycle (supervisor/socket sessions, multi-terminal attach, <5s warm re-entry, cleanup); shipped across 11 releases.

lc3-toolchain |Rust/CI Tool | https://crates.io/crates/lc3-toolchain Mar 2025

Developer tooling for the LC-3 ecosystem, published on crates.io (8K+ downloads).
- Built a **formatter and linter** for LC-3 assembly, helping developers enforce consistent style, catch common mistakes.
- Added validation passes and correctness checks to prevent regressions, ensuring stable behavior across multiple toolchain versions.

Technical Focus

Edge AI: efficient deployment
Languages: Rust · C / C++ · Python · TypeScript
Agent Systems: Agent Loop · context management & compaction · tool use · long-term memory · multi-agent orchestration · context engineering

Leadership & Activities

2nd Place - StarkHacks Hackathon, Purdue University | https://github.com/robcholz/SentinelGlove April 2026
- **Led architecture and integration of a real-time edge-ML system** on Espressif hardware (sensor capture → on-device inference → live telemetry), choosing low-latency local inference over cloud to meet industrial deployment constraints.
- Debugged firmware-level cross-core synchronization and runtime integration under a 36-hour deadline; demoed to Espressif engineers and placed 2nd in the Espressif IoT Track at StarkHacks — the world's largest hardware hackathon (750+ builders, 80 universities, 10 countries).

2nd Place - VenturePack Competition, North Carolina State University April 2025
- Secured **$8K prize** in NC State's largest entrepreneurship competition among **20+ teams**.

Publications & Patents

Patent: ZL 2022 2 3316373.9 and Z2017 2 1306730.X May 2023
- **Commercialized oReminder (A post-operative care wearable device) versions** (v1/v2) sold at scale. (2018, 2023).