

HUI SUN

✉ hsun26@ncsu.edu ☎ 352-328-6856 🌐 <https://www.linkedin.com/in/huisun2022/>

EDUCATION

Ph.D. in Computer Science

North Carolina State University, GPA: 4.3/4.0 Advisor: Dr. Wesley K. G. Assunção
Research Interests: LLMs, Software Engineering, HCI, AI4AI, AI4SE, Agent

Raleigh, NC

Aug. 2023 - Present

M.S. in Computer Science

University of Florida, GPA: 3.93/4.0

Gainesville, FL

Aug. 2020 - Dec. 2022

B.E. in Software Engineering

Xidian University, GPA: 3.6/4.0

Xi'an, CN

Aug. 2015 - Dec. 2019

WORK EXPERIENCE

Google Inc.

Mountain View, CA

Core-ML Ph.D. SWE Intern || PyTorch/XLA, TPU, Agent || **Host: Ann Yan**

May. 2026 - Aug. 2026

- Designed and built a fully automated local agent framework that deploys and migrates models on remote Google Cloud TPU VMs — a multi-tiered MCP routing loop drives local LLM reasoning to handle environment setup, GPU-to-TPU code migration (PyTorch/XLA), and self-healing recovery, with no manual intervention required.
- Engineered a fully automated runtime self-healing pipeline for the above system: real-time telemetry tracks memory usage, crash signatures, and CPU fallback events across distributed workloads, automatically triggering an LLM-driven debugging and code-correction loop to recover failed runs without manual intervention.

PAX Technology Inc.

Jacksonville, FL

Software Engineering - Full Time || Java, Vue.js || **Mentor: Yurii Maidaniuk**

Jan. 2023 - Aug. 2023

- Contributed to the development and deployment of a return merchandise authorization system that manages repairs and tracks equipment for over 3,400 client companies in North America, from initial documentation to final deployment.
- Participated in developing a warehouse management system to manage 1.4 million devices.
- Participated in the development of the counting tax function of Key Injection purchase system using Spring Boot, which generated approximately \$840,000 in annual revenue for our company.
- Participated in the development of the CyberLab Function, a government-approved robotics kit used by global testing teams to evaluate financially sensitive features, such as PIN entry functions and other processes involving merchant-sensitive information.

IBM

Shanghai, CN

Software Engineering - Intern || Java, Struts2, Spring, WebSphere

Mar. 2021 - Jun. 2021

- Developed and deployed a monitoring function for the Development Bank of Singapore's bank card system, enhancing security across all branches in Asia.
- Used IBM WebSphere to control and deploy different versions of the Bank Card System, which helps manage \$134.5 billion in deposits for DBS Bank.
- Responsible for writing a product requirement document and defining the format of raw data to support IBM's Singapore and Hong Kong team in developing Blacklist and Whitelist functions for Bank Card System.

SELECTED PROJECTS

Evaluating Language Models on Cross-Language Code Functional Equivalence

Raleigh, NC

Aug. 2025 - May. 2026

Google Cloud, FUNCAP, CNPq funded project

- **A new dataset, PolyHuman** — the first dataset for evaluating *cross-language code functional equivalence* on human-written code (CPP / Java / Python, 5,035 problems from CodeContests, 15 sub-tasks), addressing a gap left by prior benchmarks that focus on a single language or synthetically transformed code.

- **LLMs are far from reliable** — a systematic evaluation of 9 open- and closed-source models shows accuracy drops from 0.998 on synthetic code to 0.84 on human-written code; prediction bias is driven by the *model*, not the language; as problem difficulty rises, models increasingly misclassify non-equivalent code as equivalent; and similarity signals act as a double-edged sword behind the scenes.
- **A three-tier failure hierarchy** — manual analysis of 81 misjudged cases reveals three root causes: *Knowledge Failures* (21%), *Abstraction-Level Reasoning Failures* (53% — premature conclusions, semantic misinterpretation, and implementation-detail tracing errors), and a small share of dataset-labeling issues (19%); Claude Opus excels at deep reasoning while Gemini is stronger on fine-grained knowledge, motivating complementary multi-model strategies over single-model use.

Understanding Reviewer Concerns on Test Code in Pull Requests

Raleigh, NC

NSF funded project in the WISER lab

Dec. 2024 – Mar. 2025

- We are the first to analyze the GitHub Action’s influence on test code review, leveraging 213,511 pull requests and 600,905 review comments from six large-scale GitHub Projects.
- Demonstrated platform-level differences: GitHub reviews are less frequent but more balanced between test and production code compared to Gerrit.
- Discovered that adoption of GitHub Actions significantly reduced reviewer and developer attention to test code: lower review probability, comment density, and test code churn.
- Highlighted risks of over-reliance on automation, showing how it may increase technical debt by discouraging defect detection in test code, and delivered actionable implications for developers, reviewers, and tool designers, and released a public replication package to ensure transparency and reuse.

Software-hardware Co-design Differential Privacy

Raleigh, NC

NSF funded project in the Network Lab

Aug. 2023 – May 2024

- We are the first to leverage the tendency of hardware transistors to degrade under voltage drop as a source of noise for Local Differential Privacy (LDP). Since this hardware-induced noise is **uncontrollable**, it offers a more robust alternative to traditional software-based noise injection.
- We propose **SRAM-DP**, the first hardware-software co-design for achieving differential privacy on semi-structured data. Experiments on a custom SRAM design, using the WHO-released disease dataset, show that our method preserves individual privacy while retaining meaningful data distribution analysis, validating its applicability to real-world privacy protection.
- We show that the naturally occurring hardware noise satisfies LDP requirements, yielding a fully randomized noise injection process that strengthens privacy guarantees. This work was published at ISVLSI 2024.

SKILLS

Skills: Agent building, RAG, XLA, Java, Python, C++, R, Matlab, SQL, J2EE, ASP.NET, AI4SE, LLMs, LaTeX, Git

Frameworks and Libraries: Spring Boot, MyBatis, PyTorch, scikit-learn, Ollama, Pandas, HuggingFace

PUBLICATIONS

- [1] [Hui Sun](#), Kyle Mooney, Mario Renteria-Pinon, Tingxiang Ji, Hritom Das, Na Gong and Jianqing Liu, “Low-power and Computing-free Privacy Design for IoT Systems”, IEEE Computer Society Annual Symposium on VLSI, 2024. [Published]
- [2] Jinming Xing, Guoheng Sun, [Hui Sun](#), Linchao Pan, Shakir Mahmood, Xuanhao Luo, Muhammad Shahzad, “NetSight: Graph Attention Based Traffic Forecasting in Computer Networks”. [Under Double-Blind Review]
- [3] [Hui Sun](#), Yanan Wu, Wesley K. G. Assunção, Kathryn T. Stolee, “Test Code Review in the Era of GitHub Actions: A Replication Study”. [Under Double-Blind Review]
- [4] [Hui Sun](#), Anderson Uchôa, Rohit Gheyi, Wesley K. G. Assunção, “Evaluating Language Models on Cross-Language Code Functional Equivalence”. International Symposium on Empirical Software Engineering and Measurement, ESEM 2026 [Published], Google Cloud Research Award Supported.
- [5] Anderson Uchôa, Jhefferson Souza, Emanuel Ávila, Gabriel Silva, Kevin Lira, Nelson Araújo, Arthur Cordeir, [Hui Sun](#), Rohit Gheyi, Wesley K. G. Assunção, “Evaluating Language Models on Cross-Language Code Functional Equivalence”. [Under Double-Blind Review]