

Abdullah Naveed

Ph.D. Candidate in Computer Engineering | University of Florida

✉ abdullahnaveed102@gmail.com | 🌐 abdullahnaveed102.github.io | 🔗 linkedin.com/in/abdullahnaveed102
🐙 github.com/Abdullahnaveed102 | 📄 scholar.google.com/citations?user=LL4EX88AAAAJ

Research Interests

Dependable artificial intelligence; large language model reliability; resilience for high-performance computing; hardware–software co-design; scientific data compression; machine learning systems.

Education

University of Florida

Ph.D. in Computer Engineering, GPA: 3.93/4.00

Advisor: Prof. Guanpeng Li.

Expected May 2027

Gainesville, FL

University of Iowa

Master of Computer Science, GPA: 3.99/4.00

Aug 2025

Iowa City, IA

Lahore University of Management Sciences

B.S. in Computer Science, GPA: 3.61/4.00

May 2023

Lahore, Pakistan

Research Experience

Argonne National Laboratory

Visiting Graduate Researcher

Jun 2025 – Present

Lemont, IL

- Designed an end-to-end hardware–software co-design framework for FPGA/CPU scientific data compression at the Advanced Photon Source, achieving up to **19.8× cycle speedup** over matched software and **46–65% lower end-to-end workflow time**.
- Developed a modular Chisel/RTL compression library for constraint-driven composition of reusable components into system-optimized pipelines, supporting LZ4, Snappy, StreamVByte, SZ3, SZp, SZx, zfp, and extensible new designs.

University of Florida

Graduate Research Assistant, Dependable Systems Lab

Aug 2025 – Present

Gainesville, FL

- Developed **TRACE**, a runtime detector for silent data corruption during large language model inference, achieving **95.3% average detection coverage** across 15 LLMs and 3 datasets with **0.56% runtime overhead**; paper accepted at SC26.
- Characterize the resilience of large language model inference, pretraining, and parameter-efficient fine-tuning to transient hardware faults, including models fine-tuned with **LoRA, QLoRA, LoRA-FA, and DoRA**, across fault intensities, parameter types, model architectures, and workloads.

University of Iowa

Graduate Research Assistant

Aug 2023 – Aug 2025

Iowa City, IA

- Developed an LLVM-based silent-error detector for scientific and HPC applications using selective monitoring of informative program locations, achieving **86.19% average detection coverage at 4.90% runtime overhead** across 15 benchmarks.
- Compared compiler- and runtime-based error-detection techniques across scientific workloads, analyzing their behavior under different fault-propagation patterns and quantifying trade-offs between detection coverage and runtime overhead.

Publications

- Abdullah Naveed, Mahina Sheikh, and Guanpeng Li.
“**Low-Overhead Silent Error Detection for Large Language Model Inference at Scale.**”
International Conference for High Performance Computing, Networking, Storage and Analysis (SC '26), 2026.
Accepted; to appear.
- Darren Ng, Abdullah Naveed, Duo Zhang, Guanpeng Li, Sheng Di, and Xiaoyi Lu.
“**Z-SPDK: Accelerating Lossy Compressed I/O on NVMe-SSDs.**”
International Conference for High Performance Computing, Networking, Storage and Analysis (SC '26), 2026.
Accepted; to appear.

3. Muhammad Hassan, Talha Tahir, Muhammad Farrukh, **Abdullah Naveed**, Anas Naeem, Fareed Zaffar, Fahad Shaon, Ashish Gehani, and Sazzadur Rahaman.
“Evaluating Container Debloaters.”
2023 IEEE Secure Development Conference (SecDev), pp. 88–98, 2023. doi:10.1109/SecDev56634.2023.00023.

Manuscripts Under Review

1. LLM Pretraining Reliability

Large-scale characterization of transient soft-error effects and resilience during Transformer pretraining.
Manuscript under review, 2026.

2. LLM Fault Resilience

Large-scale study of large language model behavior under increasing fault intensity, multi-bit corruption, and emerging high-error computing environments.
Manuscript under review, 2026.

3. HPC Reliability

Compiler-based, low-overhead detection of silent errors in scientific and HPC applications.
Manuscript under review, 2026.

Teaching Experience

University of Iowa

Aug 2023 – Dec 2023

Teaching Assistant, Database Systems

Iowa City, IA

- Guided more than 30 students in applying relational database concepts to MySQL projects, including schema design, query formulation, debugging, and implementation decisions.

Lahore University of Management Sciences

Sep 2022 – Jan 2023

Teaching Assistant, Data Structures

Lahore, Pakistan

- Led tutorials for more than 100 students on data structures and algorithms, reinforcing core concepts through worked examples and guiding students in applying them to programming and problem-solving assignments.

Professional Service

Conference Reviewer / Subreviewer: SC; DSN; ICS; AAAI; HPDC.

Research Community Contribution: Contributing team member to the ProdeBench benchmarking project, supporting implementation and experimental evaluation.

Selected Technical Projects

Multimodal Fake News Classification

Developed a multimodal classification pipeline that integrates textual and visual representations to identify misleading content using complementary evidence across language and image modalities.

Technologies: Python, NLP, Computer Vision, Multimodal Learning

Cozmo’s Star Wars Odyssey

Built an interactive robotics experience combining visual character recognition with LLM-driven narrative generation and question answering to enable context-aware interaction with a Cozmo robot.

Technologies: Computer Vision, Large Language Models, Robotics

COVID SOP Violation Detection

Developed a real-time multi-camera vision pipeline for mask detection and crowd monitoring, integrating YOLOv5-based detection, top-view projection, and spatial heatmaps for density analysis.

Technologies: Python, PyTorch, OpenCV, YOLOv5

LUMS Mobile App

Developed a cross-platform student portal for deadlines, GPA calculation, and communication with real-time cloud-backed data management.

Technologies: React Native, JavaScript, Firebase

Technical Skills

Programming: Python, C, C++, Bash, SQL, Scala, JavaScript, Git.

Machine Learning / AI: PyTorch, Hugging Face Transformers, TensorFlow/Keras, scikit-learn, OpenCV.

HPC / Parallel Computing: MPI, OpenMP, Slurm, large-scale CPU/GPU experimentation, fault injection, resilience analysis.

Hardware / Systems: Chisel, Verilog/RTL, Chipyard, RISC-V, Verilator, Vivado, LLVM, Linux, Docker.

Scientific Computing / Data Analysis: NumPy, Pandas, Matplotlib, Jupyter, MATLAB.