

# Osama Yousuf

✉ osamayousuf@gwu.edu    ☎ (202) 568 9380    🔗 osama-yousuf.github.io    in osama-yousuf    🌐 osama-usuf

## Professional Summary

---

- 5+ years of experience in deep learning models, architectures, and frameworks (excelling in PyTorch).
- 4+ years of experience in advanced memory device modeling, characterization, and mixed-signal prototyping.
- Proficient in academic writing with numerous first-author publications in peer-reviewed journals/ conferences on neuromorphic computing (Nature, IEEE, ACM, Frontiers).
- **Research interests:** Neuromorphic computing, low-rank training, robust algorithms for noisy hardware neural networks, neural network quantization, hardware accelerators for Artificial Intelligence (AI).

## Education

---

### The George Washington University (GWU)

*PhD in Computer Engineering*

*Washington, DC*

*Jan 2021 – May 2025*

- Focus: Machine Learning and Intelligent Systems | CGPA: 4.00/4.00, [Transcript](#) 🔗.
- **Dissertation:** Modeling, Prototyping, and Algorithm Design for *Memristive Neural Networks*.

### Habib University

*Bachelors in Computer Science & Mathematics*

*Karachi, Pakistan*

*Aug 2016 – July 2020*

- Awards: Gold Medalist (Class Rank: 1), President's Honor List, Dean's Honor List | CGPA: 3.92/4.00, [Transcript](#) 🔗.
- **Final Year Project:** [Personal Outfit Recommendation System](#) for eastern clothing (transfer learning).

## Work Experience

---

### Graduate Research Assistant

*The George Washington University*

*Washington, DC*

*Jan 2021 – May 2025*

Advised by [Gina C. Adam](#) at the [Adaptive Devices And Microsystems \(ADAM\) Lab](#).

- **Robust Inference:** Developed a noise-tolerant inference scheme for analog neural networks, improving performance by  $\geq 150\%$  compared to prior methods under similar fault conditions (patent pending).
- **Low-rank Training:** Developed low-rank training algorithms for hardware neural networks that enable in-situ training from scratch. Unlike low-rank adaptation (LoRA) in LLMs, this method directly approximates backpropagated gradients in low-rank subspaces. Hardware networks achieve near-software performance ( $< 1\%$ ) with substantial memory savings.
- **NeuroTorch:** Led the development of *NeuroTorch*, a deep learning library tailored for studying in-memory accelerators for deep neural networks. Fully compatible with PyTorch, it supports diverse quantization schemes, fault-tolerance algorithms, and analog device models. NeuroTorch layers leverage fast pyBind11 bindings to offload backpropagation and inference calculations to custom C++ acceleration cores.
- **Data-driven Modeling:** Introduced a data-driven algorithm to model various network layers (fully connected, recurrent, convolutional, transformer) as memristive crossbars, achieving greater experimental realism compared to existing methods.
- **Quality Metrics:** Conceived the concept of *modeling bias* and several quantitative metrics for measuring deviations between analog neural network models against experimental hardware realizations.

### Research Intern (Machine Learning)

*Western Digital Corporation*

*San Jose, CA*

*Oct 2024 – Dec 2024*

Advised by [Mike Grobis](#), [Martin Lueker-Boden](#), and [Michael Tran](#) from *Western Digital Research*.

- **Accelerating Large Language Models (LLMs):** Developed a framework for benchmarking large language models (LLMs) under a variety of noise sources in analog AI accelerators. It extends PyTorch and integrates with *transformers* (by HuggingFace) and *lm-eval* (by EleutherAI) for efficient language model evaluation. All operations are CUDA-compatible; pre-trained models from the Hub can be tested under various noise sources on natural language processing benchmarks (PIQA, HellaSwag, Big-Bench-Hard, etc.) with a minimal worst-case overhead of  $1.1\times$  in simulation time.
- **LLM Robustness:** Studied noise limitations for various state-of-the-art LLMs and derived specifications for an analog LLM accelerator. For example, the accelerator must have device-to-device variability  $\sigma \leq 15\%$  for attaining near-software performance ( $< 3\%$  of baseline).
- **Device Characterization:** Performed electrical device characterization experiments on in-house fabricated vehicles consisting of Magnetic Random Access Memory (MRAM) crossbar arrays. Wrote Python code for controlling a probe card setup based on PXIe hardware and implemented device testing and data visualization routines for a voltage-assisted write scheme for MRAM devices. This scheme could enable ultra low-power inference for large-scale deep learning workloads, such as LLM inference.

## Research Associate

National Institute of Standards and Technology (NIST)

Gaithersburg, MD  
Jan 2022 – Aug 2024

Hosted by [Matthew W. Daniels](#) and [Jabez J. McClelland](#) at the [Alternative Computing Group](#).

- **Daffodil:** Contributed to *Daffodil* - a mixed-signal hardware prototyping system for benchmarking resistive memory-based deep neural networks. Led the Linux kernel bring-up using *Vivado & PetaLinux tools*, edited off-the-shelf drivers for SPI devices using C++, wrote low-level drivers for interfacing between the kernel and programmable logic (PL) fabric using *ctypes*, and implemented core abstraction modules using *Python* that enable execution of neural network layers (training and inference) natively on the platform.
- **Resource-efficient Training:** Assisted on the testing of *QSArray* - a specialized hardware architecture based on systolic arrays for accelerating deep neural network training. Developed supporting code that enables high-level PyTorch networks (transformers, multi-layer perceptrons, etc.) to delegate gradient calculations to acceleration cores or Verilated modules in C++. The system produces low-rank approximations of network gradients in a streaming, pipelined fashion over mini-batches, and has crucial applications in distributed training and federated learning.

## Software Engineer (Frontend)

Stellic

Karachi, Pakistan  
Jul 2020 – Dec 2020

Stellic is an ed-tech platform headquartered in San Mateo, CA, focused on improving student graduation rates.

- **Feature Development:** Developed and implemented key product features, including comprehensive time zone support, enabling seamless collaboration between students and advisors during the pandemic. This innovation distinguished Stellic from competitors, secured multiple call-for-proposal (CFP) wins, and drove adoption by six new colleges across the U.S., directly boosting revenue.
- **Frontend Engineering:** Designed and developed user interfaces to enhance user experience for students and advisors.
- **Collaboration:** Worked closely with design and backend teams to ensure seamless integration and functionality.

## Software Engineer (Analyst)

Afiniti

Karachi, Pakistan  
Mar 2020 – Jul 2020

Afiniti is a Washington, DC-based platform that uses AI to optimize contact center pairings.

- **Production Support:** Ensured smooth production operations for global clients, maintaining system uptime and reliability.
- **Incident Management:** Researched, documented, and reported incidents to relevant teams to facilitate quick resolutions.
- **Database Health:** Monitored and maintained the health of system databases for product deployments.
- **Process Automation:** Automated manual workflows with Python and Bash, reducing real-time problem identification time from 5+ minutes to < 30 seconds.

## Research Intern (Artificial Intelligence)

Texas A&M University

College Station, TX  
May 2019 – Aug 2019

Hosted by [Guni Sharon](#) at the  [\$\pi^\*\$  AI & Optimization Lab](#).

- **Sim-to-Real RL Framework:** Developed a Reinforcement Learning (RL) framework for the CARLA simulator to enable sim-to-real policy transfer for autonomous vehicles, including the DeepRacer robot. The framework was adopted by lab members to validate novel RL algorithms for autonomous driving.
- **Deep Q-Learning for Autonomous Control:** Trained a Deep Q-learning agent to control a vehicle using video inputs as the sole sensory data. Demonstrated the effectiveness of vision-based RL in learning driving policies through environment interactions.
- **Policy Optimization and Real-World Alignment:** Applied reward shaping, action constraints, and curriculum learning to enhance policy performance and ensure applicability to real-world driving scenarios.

## Teaching Experience

---

### Graduate Teaching Assistant

The George Washington University

Washington, DC  
Jan 2024 – May 2024

- Course taught: *ECE 2140: Design of Logic Systems*.
- Trained a class of 15+ students on breadboard circuit prototyping, Verilog module and testbench design, waveform verification, and FPGA prototyping using Xilinx Basys 3 boards and Vivado.

### Undergraduate Teaching Assistant

Habib University

Karachi, Pakistan  
Aug 2017 – May 2018, Jan 2020 – Jun 2020

- Courses taught: *Digital Logic & Design, Discrete Mathematics, & Nature of Computation*.
- Designed assessment materials, evaluated homework assignments, and provided peer tutoring services to support learning and academic success.

## Skills

---

**Languages** Python, C++, Verilog, SQL, JavaScript  
**Frameworks** PyTorch, PyTorch Lightning, Scikit-Learn, Huggingface, Streamlit, Gradio, wandb, Gym, Numpy, Pandas  
**Visualization** Matplotlib, Altair, Seaborn, Wolfram Mathematica, MATLAB  
**Web Dev.** Django, Flask, HTML/CSS, REST APIs, Angular, WebGL  
**Prototyping** Vivado, PYNQ, cocotb, hls4ml, Uboot, Synopsys Design Vision, Waveforms, PSpice  
**Equipment** Oscilloscopes, function generators, multi-meters, Analog Discovery 2, FPGA programming (Basys3, VCU118, Zynq)  
**Misc.** Jupyter Notebook, pyBind11, ctypes, Sphinx, Git, L<sup>A</sup>T<sub>E</sub>X

## Publications

---

- Robust Hardware-Aware Neural Networks for FeFET-based Accelerators** 2024  
**O. Yousuf**, A. L. Glasmann, A. L. Mazzoni, S. Najmaei, G. C. Adam.  
Under review at *IEEE Transactions on Nanotechnology*.
- Layer Ensemble Averaging for Fault Tolerance in Memristive Neural Networks** 2024  
**O. Yousuf**, B. Hoskins, K. Ramu, M. Fream, W. A. Borders, A. Madhavan, M. W. Daniels, et al.  
Accepted at *Nature Communications: Neuromorphic Computing*, [10.48550/arXiv.2404.15621](https://arxiv.org/abs/10.48550/arXiv.2404.15621) [↗](#).
- Neural Network Modeling Bias for Hafnia-based FeFETs** 2023  
**O. Yousuf**, I. Hossen, A. L. Glasmann, S. Najmaei, G. C. Adam.  
*International Symposium on Nanoscale Architectures (NANOARCH)*, [10.1145/3611315](https://doi.org/10.1145/3611315) [↗](#), [Slides ↗](#).
- Device Modeling Bias in ReRAM-based Neural Network Simulations** 2023  
**O. Yousuf**, I. Hossen, M. W. Daniels, M. Lueker-Boden, A. Dienstfrey, G. C. Adam.  
*IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, [10.1109/JETCAS.2023.3238295](https://doi.org/10.1109/JETCAS.2023.3238295) [↗](#).
- Gradient Decomposition Methods for Training Networks with Non-Ideal Synapses** 2021  
J. Zhao, S. Huang, **O. Yousuf**, Y. Gao, B. Hoskins, G. C. Adam.  
*Frontiers in Neuroscience: Neuromorphic Computing*, [10.3389/fnins.2021.749811](https://doi.org/10.3389/fnins.2021.749811) [↗](#).
- A System for Validating Resistive Neural Network Prototypes** 2021  
B. Hoskins, W. Ma, M. Fream, **O. Yousuf**, Y. Gao, B. Hoskins, G. C. Adam.  
*International Conference on Neuromorphic Systems (ICONS)*, [10.1145/3477145.3477260](https://doi.org/10.1145/3477145.3477260) [↗](#).

## Talks & Presentations

---

- Accelerating Large Language Models with Magnetic RAM Crosspoint Arrays** Dec 2024  
**O. Yousuf**  
Invited Talk at *Frontier Seminar Series, Western Digital Corporation, San Jose, CA*.  
[Abstract ↗](#) (Slides are confidential)
- Layer Ensemble Averaging for Improving Memristive Neural Network Performance** Oct 2024  
**O. Yousuf**, B. Hoskins, K. Ramu, M. Fream, W. A. Borders, A. Madhavan, M. W. Daniels, et al.  
Poster Presentation at *Innovation Bazaar, Western Digital, San Jose, CA*.  
[Poster ↗](#)
- Daffodil: A Robust Prototyping System for Resistive Device Characterization and Neural Network Implementation** Nov 2023  
**O. Yousuf**, B. Hoskins, K. Ramu, M. Fream, W. A. Borders, A. Madhavan, M. W. Daniels, et al.  
Poster Presentation at *NIST Prep Symposium, Gaithersburg, MD*.  
[Poster ↗](#)
- Investigating Bias in the Modeling of Resistive RAM Devices** Dec 2022  
**O. Yousuf**, I. Hossen, M. W. Daniels, M. Lueker-Boden, A. Dienstfrey, G. C. Adam.  
Poster Presentation at *International Conference on Memristive Materials, Devices & Systems (MEMRISYS), Cambridge, MA*.  
[Poster ↗](#)
- Towards a Hardware-Aware Decomposition Method for ReRAM Neural Networks** Jul 2022  
**O. Yousuf**, M. W. Daniels, A. Dienstfrey, G. C. Adam.  
Oral Presentation at *International Conference on Neuromorphic Systems (ICONS), Knoxville, TN*.  
[Recording ↗](#), [Abstract ↗](#), [Slides ↗](#)

## Streaming Gradient Tracking using Non-negative Matrix Factorization

Apr 2022

O. Yousuf, M. W. Daniels, A. Dienstfrey, G. C. Adam.

Poster Presentation at *GW Research Showcase, George Washington University, Washington, DC.*

Poster [🔗](#)

## Algorithmic improvements and optimizations for training memristive neural networks

Oct 2021

O. Yousuf, J. Zhao, M. W. Daniels, B. Hoskins, W. Ma, et al.

Poster Presentation at *Innovation Bazaar, Western Digital, San Jose, CA.*

Slides [🔗](#)

## Quasisystolic Arrays for Pipelined and Resource-Efficient Neural Network Training

May 2021

M. W. Daniels, B. Hoskins, A. Madhavan, O. Yousuf, G. C. Adam et al.

Best Poster Presentation at *Sigma Xi NIST, Gaithersburg, MD.*

Poster [🔗](#)

## Honors & Awards

---

- o Awarded funding for a research internship at Western Digital by the National Science Foundation (NSF) *Aug 2024*  
[INTERN](#) program.
- o [Best Poster Award](#) at *Sigma Xi NIST: AI, Machine Learning, Engineering, Nanotechnology, and Math.* *May 2021*
- o Received [Academic Spotlight](#) for publishing at *Tezhib* - Habib University's student research journal. *Nov 2020*
- o Awarded the [Dean's Gold Medal](#) for being the top candidate in Computer Science among the graduating class. *Jul 2020*
- o Awarded the [Research Abroad Scholarship](#) at Texas A&M University (sole recipient). *May 2019*
- o Awarded the prestigious [Study Abroad Scholarship](#) by Habib University (awarded to top 1% students) for attending *Summer 2018* at Stanford University, CA as part of the International Honors Program. *May 2018*
- o Recipient of the [Top of the TOPS program](#) award by Habib University. *Aug 2016*
- o Awarded the full-ride [TOPS scholarship](#) by Habib University. *May 2016*

## Academic Services

---

Reviewed a total of **2** papers on deep learning and neuromorphic computing for the following venues:

- o IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS).

*2024*

## Extra-Curriculars

---

Full list of activities available at: [Meta Curricular Transcript](#) [🔗](#)

- o Served in the student orchestra as a Violin player. *Nov 2019 - Jul 2020*
- o Volunteered at *Fixit School* as a Computer Science Instructor. *Dec 2019 - Mar 2020*
- o Represented Pakistan at Harvard University's HPAIR Conference. *Jan 2019 - Feb 2019*
- o Served as the CEO of the *Physics & Astronomy Club* at Habib University. *Jan 2018 - May 2018*
- o Served as an executive member of the *Tachi* (gaming) club at Habib University. *Sep 2016 - Dec 2016*