Farshid Varno

Intro

Seasoned AI/Hardware Co-Design Engineer with 15+ years of experience in hardware-aware software development, hardware design, and AI research. I leverage deep expertise in systems programming, computer architecture, and system simulation to design and optimize solutions for large-scale machine learning infrastructure. My track record includes patented AI accelerator architectures, impactful research in distributed machine learning, development of FPGA-based telecom and encryption modules, and optimizing hardware-software integration for AI systems. Passionate about bridging AI, software, and hardware for next-gen computing challenges, I am actively deepening my expertise in low-level systems programming and compiler technologies by learning Rust, MLIR, and LLVM, to contribute to high-performance ML infrastructure.

Experience

Research Scientist (AI/HW co-design), Rain

August 2023–present, San Francisco, USA

- CIM: LUT-based approximations, Online Softmax, Quantization.
- Multi-level simulation: performance, behavioral, cycle-accurate (PyTorch, SystemC, QEMU).
- Custom RISC-V instructions.

Research Scientist, Imagia

May 2018–March 2022 (internship period included), Montreal, Canada

- Federated Learning, Hypothesis Transfer Learning, Meta Learning, Few-Shot Learning.
- AI experimentation orchestration.

Research Assistant, Institute for Big Data Analytics

May 2017–May 2018, Halifax, Canada

- CUDA programming, OpenMP, AIS data, Deep Learning research.

FPGA Engineer, Kara Telephone Co.

Jun 2013–Jun 2014, Tehran, Iran

- FPGA-based switches for PBX systems (lead engineer), communication protocols.

RTL Designer Intern, SarvNet Telecommunication Inc.

March 2012–Sep 2012, Isfahan, Iran

- FPGA-based encryption modules, resource sharing algorithms for AES in STEM4.

Education

- **Ph.D., Computer Science** | Dalhousie University (2016–2023), CGPA: 4.19
- **M.Sc., Computer Architecture** | University of Isfahan (2012–2015), CGPA: 4.02
- **B.Sc., Computer Engineering** | Guilan University (2008–2012)

Skills

Programming Languages: Python, C++, CUDA

Systems Programming & Compiler Technologies:

- **Proficient In:** RISC-V Extensions, PyTorch Dynamo Integration, Low-Level Performance Optimization
- **Actively Learning:** Rust, MLIR, LLVM, Triton

AI & Machine Learning: Distributed Machine Learning (Federated Learning), Transfer Learning, Quantization & Compression, On-Device Training, LLM & Attention Architecture Performance Modeling

ML Infrastructure & HPC: System Simulation (SystemC, QEMU), AI Performance Modeling

Hardware Design & Verification: VHDL, Verilog, SystemC

Developer & MLOps Tools: Git, GitHub Actions, Bazel, Polyaxon, MLflow

Selected Achievements

- **Patents**: Lead inventor on patented AI accelerator and Transfer Learning advancements.
- **Publications**: Published Federated Learning and Transfer Learning research at ECCV and ICLR.
- **Awards**: Scotia Scholar Award ($45k), Best Graduate Research Award, Mitacs Accelerate Award ($56k).

This resume is compiled using LaTeX and is also available in HTML format at https://farshid.varnio.com/resume/farshid/index.html.