Ahmad Hossein Yazdani

✉ ahmadyazdani@vt.edu
ayazdani1997.github.io/
Github LinkedIn

Work Experience

August,2020 – ***Research Assistant at Distributed System and Storage Lab*, Virginia Tech**.

Advisor : Dr. Ali Butt, *Professor, Department of Computer Science*, Virginia Tech

- **LLMStore:** Designed caching and gradient-based activation ranking to reduce tensor offload latency in distributed LLM training; identified 300GB+ (LLaMA-70B) offload traffic via trace analysis.
- Implemented gradient compression in the gRPC communication layer for cross-cloud/HPC federated learning, reducing cross-site communication overhead.
- Co-developed data access pattern (FAST '23) improving cache hit ratio by up to 4.5× over LRU for distributed deep learning workloads.

Spring 2026 ***Research Intern, Oak Ridge National Laboratory***.

- Selected for ORNL research internship (2026) to explore AI-driven optimization of memory- and storage-intensive HPC workflows.

Summer 2024 ***Student Assistant at NERSC, Lawrence Berkeley National Laboratory (LBNL), internship***.

- Examined Drishti, an HPC I/O recommendation tool; Detected many false positive warnings, and derived insights to improve the accuracy of the the Drishti I/O recommendation tool.

Summer 2023 ***Student Assistant at Lawrence Berkeley National Laboratory (LBNL), internship***.

- Identified I/O variability in HPC workloads (E3SM, LAMMPS) under interference; extended analysis to large-scale LLM training (LLMStore).

Summer 2021 ***Internship at Oak Ridge National Laboratory, Analytics & AI Methods at Scale Group***.

- Predicted the I/O pattern of the next job given the features from the past submissions of the same user with an accuracy of nearly 90% for HPC jobs.

Computer skills

Languages Python, C, C++, pthread, CUDA

ML & Systems PyTorch, DeepSpeed, LLM, Computer Vision, Federated Learning, Distributed systems, File Systems, Lustre, GPFS, Slurm, MapReduce, Redis, MPI

Conference & Workshop publications

[ICDCN'25] **Yazdani, Ahmad Hossein** et al. User-based i/o profiling for leadership scale hpc workloads. ICDCN '25, New York, NY, USA, Jan. 2025. Association for Computing Machinery. URL `https://doi.org/10.1145/3700838.3700865`.

[FAST'23] Redwan Ibne Seraj Khan and **Yazdani, Ahmad Hossein** et al. Shade: Enable fundamental cacheability for distributed deep learning training. Santa Clara, CA, US, Feb. 2023. USENIX Association. URL `https://www.usenix.org/conference/fast23/presentation/khan`.

Education

2020-2026 **PhD, Computer Science**, *Virginia Tech*, Blacksburg, VA, US.
Advised by Dr Ali Butt

2025 **Masters of Computer Science**, *Virginia Tech*, Blacksburg, VA, US.
Advised by Dr Ali Butt

2015–2020 : **Bachelor of Computer Software Engineering**, *University of Tehran*, Tehran, Iran.