PURDUE UNIVERSITY | Elmore Family School of Electrical and Computer Engineering

Faculty Candidate Seminar - Optimization for ML

Haiying (Helen) Shen

Dr. Haiying Shen Associate Professor

University of Virginia

Thursday, February 26, 2026

Presentation: 11:15 A.M.– 12:15 P.M.

Q&A: 12:15 P.M. - 12:45 P.M.

SCPI 102 (Student Center, Purdue Indianapolis)

Resource Optimization for ML Inference Serving

Abstract: Dr. Shen's research focuses on job scheduling and resource management for Machine Learning (ML) and Large Language Model (LLM) systems. With the rapid growth of deep learning models, minimizing monetary costs and maximizing the goodput of inference-serving systems – across cloud, edge servers, and edge devices– have become critical challenges. Addressing these challenges requires efficient task scheduling both within and across nodes, as well as optimized resource management to ensure high resource utilizations, and adherence to Service Level Objectives (SLOs). However, current approaches often fall short of fully addressing these challenges. In this talk, Dr. Shen will present our novel methods for bridging these gaps and enabling efficient execution of ML and LLM workloads. Dr. Shen will also briefly outline ongoing and future research plans aimed at advancing LLM systems.

Bio: Dr. Haiying Shen is an Associate Professor in the Department of Computer Science at the University of Virginia. During her 2024 sabbatical, she served as a Consulting Researcher at Microsoft in Redmond, WA, where she focused on LLM systems. Her research area is distributed systems, with a focus on ML/LLM systems, cloud computing, edge computing, and cyber-physical systems (CPS). Dr. Shen has made significant contributions to her field, with an H-index of 52 and over 380 publications in top conferences and journals such as SIGCOMM, OSDI, EuroSys, SoCC, ASPLOS, CoNext, Infocom, IEEE/ACM Transactions on Networking (TON), IEEE Transactions on Parallel and Distributed Systems (TPDS), and IEEE Transactions on Mobile Computing (TMC).

Zoom

https://purdue-edu.zoom.us/j/91028490864?pwd=8COKwBeQo1cJmnMaAksaKZ5k31rZgA.1

Meeting ID: 910 2849 0864

Passcode: 592791

Hosted by Lingxi Li ~ lingxili@purdue.edu