

Efficient AI Through and For Intelligent Programming Systems



Xipeng Shen

Professor

North Carolina State University

Wednesday, April 23, 2025

10:30 AM • MSEE 112

<https://purdue-edu.zoom.us/j/91315463036>

Abstract

Fueled by deep neural networks, recent years have witnessed tremendous advances in Artificial Intelligence. This talk explores the reciprocal relationship between modern AI—especially Large Language Models (LLMs)—and the development of next-generation programming systems and high-performance computing.

The first half of the talk will examine how AI can fundamentally reshape software development. We will explore how intelligent programming systems can harness the power of LLMs to enable software that can learn, adapt, and improve over time—much like humans do. The second half will shift focus to the other side of the relationship: how innovations in programming systems can drive more efficient and effective AI. In particular, I will present reuse-centric optimizations that eliminate over 99% of computations in deep neural networks without compromising prediction accuracy. The talk will conclude with a broader overview of related research efforts and a forward-looking vision for AI-powered programming systems that are both intelligent and efficient.

Bio

Xipeng Shen is a Professor in the Computer Science Department at North Carolina State University. His research focuses on intelligent computing and programming systems, emphasizing interdisciplinary challenges and cross-cutting solutions. His work has significantly impacted the development of machine learning infrastructures and programming systems for multicore processors and accelerators. He has received several prestigious honors, including the DOE Early Career Award, NSF CAREER Award, Google Faculty Research Award, IBM CAS Faculty Fellow Award, and the NCSU University Faculty Scholars Award. He is a Distinguished Member of the ACM and a Senior Member of IEEE. Professor Shen has chaired numerous major international conferences and served on several steering committees of major conferences. He is also a co-founder of CoCoPIE LLC, a startup aimed at enhancing AI deployment efficiency. Beyond academia, he has consulted for leading tech companies such as Intel, Meta, Microsoft, and Cisco.

Host Milind Kulkarni, milind@purdue.edu, 765-494-3539



Elmore Family School of Electrical
and Computer Engineering