NJIT
New Jersey Institute of Technology

Department of Data Science & Institute for Data Science

Faculty Candidate Seminar Talk

Performance Engineering for Scalable AI

by

Xuhao Chen

MIT CSAIL

Hosted by Shuai Zhang

DATE: Tuesday, March 5, 2024
COFFEE: 2:15pm-2:30pm
TIME: 2:30pm-3:30pm
LOCATION: GITC 4402 (4th floor lecture hall)
Zoom: https://njit-edu.zoom.us/j/91325381832?pwd=bGxsUkZDdmhZMDNCSDVWaDU3dngxZz09
Passcode: 852345

Abstract:

AI applications are computationally expensive and hard to scale, which poses great challenges in computer system design. In this talk, I will discuss my approach called cross-stack performance engineering, to address this challenge. This approach involves performance optimization techniques and automation methods across different layers of the system stack. I will describe my experiences in building software and hardware systems for machine learning and data mining on graph data. I will also showcase promising results, to demonstrate that this cross-stack approach is effective to make AI scalable.

Bio:

Xuhao Chen is a Research Scientist at MIT CSAIL. Dr. Chen is broadly interested in parallel systems and architectures, with a focus on AI and big-data applications. His recent work aims to make AI scalable by designing efficient algorithms, software systems and hardware accelerators. His work has been published in OSDI, ISCA, MICRO, VLDB, ICS, etc.