DISTINGUISHED LECTURE SERIES 2024/25

BIG AI
FOR SMALL DEVICES

23 OCT 2024 (WED)
2:30–3:30 PM (HKT)

Mr. & Mrs. Lee Siu Lun Lecture Theatre (WLB205),
The Wing Lung Bank Building for Business Studies,
Shaw Campus

PROF. YIRAN CHEN

Department of Electrical and Computer Engineering
Duke University

Abstract:

As artificial intelligence (AI) transforms industries, state-of-the-art models have exploded in size and capability. However, deploying these models on resource-constrained edge devices remains a significant challenge. Smartphones, wearables, and IoT sensors face stringent limitations on compute, memory, power, and communication, creating a gap between demanding AI models and edge hardware capabilities that hinders the deployment of intelligence. In this talk, we will re-examine techniques to bridge this gap and embed big AI on small devices. We will begin by discussing how the properties of various hardware platforms impact the design strategies of efficient deep neural network (DNN) models, such as quantization and pruning. Next, we will discuss techniques aimed at reducing the inference and training costs of distributed collaborative edge AI systems. Finally, we will delve into the underlying design philosophies and their evolution toward efficient, scalable, robust, and secure edge computing systems.

Enquiries: +852 3411 5935 / comp@comp.hkbu.edu.hk

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

CS DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

Speaker's Biography

Event Registration