

## RESEARCH SEMINAR

# Edge Computing Systems for Real-time AI-Driven Applications



### Dr Neuwen Ling

Postdoctoral Associate  
Efficient Computing Lab  
Yale University  
USA

**Date** : 17 May 2024 (Fri)  
**Time** : 3:00 pm - 4:00 pm  
**Venue** : HJ303

## Abstract

Edge AI is pivotal for enabling low-latency and resource-efficient solutions across a variety of critical applications like autonomous driving. Developing robust and optimized Edge AI systems is essential to realizing these advancements. In this talk, I will highlight the opportunities and challenges for deploying Edge AI systems and introduce several edge computing systems I have designed for real-time AI-driven applications. I will begin with my research on supporting concurrent DL tasks on a single resource-constrained edge device, focusing on addressing resource contention among these tasks and optimizing the utilization of heterogeneous CPU-GPU resources. Following this, I will touch on my research on cooperative Edge AI, showcasing a family of cooperative edge systems that support distributed real-time AI applications. This includes their implementation in a smart roadside infrastructure system on a university campus. I will conclude by outlining future directions, including my vision for developing edge systems to support LLM-powered autonomous agents.

## About the Speaker

Dr Neuwen Ling is currently a Postdoctoral Associate in the Efficient Computing Lab at Yale, working with Prof. Lin Zhong. She completed her Ph.D. at the Chinese University of Hong Kong, where she was advised by Prof. Guoliang Xing. Her research falls in the intersection of Edge Computing, Machine Learning and Real-time System, with the goal of developing edge computing systems for real-time AI-driven applications, such as autonomous driving. Dr. Ling has published papers on ACM/IEEE flagship conferences, including MobiCom, SenSys, IPSN, MobiSys, and IoTDI. She received one Best Paper Award Finalist and one Best Poster Award from the prestigious international conference SenSys. Besides, she has organized the first FMSys workshop at CPS-IoT Week 2024 and served as a reviewer for top ACM/IEEE journals and conferences like TMC, IMWUT/UbiComp, TOSN, and INFOCOM. She is also on the technical program committee for CHASE 2023 and IoTDI 2023 Poster & Demo.