SCHOOL OF COMPUTING & DATA SCIENCE
The University of Hong Kong

AI & Data Science

HKU AI&DS Dept Seminar on

Do Generalist Robots Need Specialist Models?

Date & Time:
14 Oct 2025 (Tue), 16:00-17:00

Venue :
CB308

Abstract

Large Vision-Language Models (VLMs) have demonstrated impressive generalization in the digital realm, but translating this into reliable robot manipulation and navigation remains a fundamental challenge. This talk explores a hybrid path forward: augmenting generalist "brains" with specialist "nervous systems." I will first present two foundation model efforts: SeeDo, which leverages VLMs to interpret long-horizon human videos and generate executable task plans, and INT-ACT, an evaluation suite that diagnoses a critical intention-to-execution gap in current Vision-Language-Action (VLA) systems. This gap reveals a key generalization boundary: robust task understanding does not guarantee robust physical control. To bridge this divide, I will introduce specialist models that provide two missing ingredients: fine-grained physical understanding and acquiring data for learning at scale. EgoPAT3Dv2 grounds robot action by learning 3D human intention forecasting from real-world egocentric videos. To address the data-scaling challenge, RAP employs a real-to-sim-to-real paradigm, while CityWalker explores web-scale video to learn robust, specialized skills. I will conclude by drawing analogies from the only known generalist agents—ourselves—to offer my answer to the question posed in the title.

About Speaker

Prof. Chen Feng

Chen Feng is an Institute Associate Professor at New York University, Director of the AI4CE Lab, and Founding Co-Director of the NYU Center for Robotics and Embodied Intelligence. His research focuses on active and collaborative robot perception and robot learning to address multidisciplinary, use-inspired challenges in construction, manufacturing, and transportation. He is dedicated to developing novel algorithms and systems that enable intelligent agents to understand and interact with dynamic, unstructured environments. Prior to NYU, he worked as a research scientist in the Computer Vision Group at Mitsubishi Electric Research Laboratories (MERL) in Cambridge, Massachusetts, where he developed patented algorithms for localization, mapping, and 3D deep learning in autonomous vehicles and robotics. Chen earned his doctoral and master's degrees from the University of Michigan between 2010 and 2015, and his bachelor's degree in 2010 from Wuhan University. As an active contributor to the AI and robotics communities, he has published over 90 papers in top conferences and journals such as CVPR, ICCV, RA-L, ICRA, and IROS, and has served as an area chair and associate editor. In 2023, he was awarded the NSF CAREER Award. More information about his research can be found at https://ai4ce.github.io.