Department of Computer Science
香港城市大學
City University of Hong Kong

COMPUTER SCIENCE COLLOQUIUM

Exploring Trustworthy AI: Research on Explainability and Fairness

SPEAKER **Dr. Changwu HUANG**

Research Associate Professor
Department of Computer Science and Engineering (CSE) Southern University of Science and Technology (SUSTech) Shenzhen, China

DATE 21 May, 2025 (Wed)
TIME 10:00 AM - 12:00 PM
VENUE G7315, 7th Floor Green Zone, Yeung Kin Man Academic Building, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong

ABSTRACT

Artificial Intelligence (AI) is profoundly reshaping various domains of our society and exerting far-reaching impacts on economic development, social governance, and international political landscapes. However, the rapid advancement and widespread application of AI technologies have also introduced numerous ethical challenges that have garnered global attention. Consequently, strengthening AI ethics governance and promoting the development of Trustworthy AI have become an international consensus. This talk first elucidates the developmental background of Trustworthy AI. Then, the talk focuses on two core dimensions of trustworthy AI: explainability and fairness. In terms of AI explainability, this talk presents a research roadmap for explainable artificial intelligence (XAI) and proposes a multi-objective feature attribution explanation method. Regarding AI fairness, the talk develops an explainable fairness-aware feature selection method and uses evolutionary constrained learning approach to improve distributive fairness in machine learning (ML). Furthermore, leveraging XAI technique, this talk investigates the issue of procedural fairness in ML, encompassing its quantitative metrics, enhancement strategies, and the intricate relationship between procedural fairness and distributive fairness.

BIOGRAPHY

Dr. Changwu HUANG is currently a Research Associate Professor in the Department of Computer Science and Engineering (CSE) at Southern University of Science and Technology (SUSTech), Shenzhen, China. He received his Bachelor's degree from Southwest Jiaotong University in 2010, Master's degree from Beijing Jiaotong University in 2013, and Ph.D. degree from National Institute for Applied Sciences of Rouen (INSA Rouen Normandie) in 2018. Upon completing his doctoral study, he joined Prof. Xin Yao's research group at SUSTech in March 2018, and progressed from a Postdoctoral Fellow (03/2018-09/2020) to a Research Assistant Professor (10/2020-09/2023), and then a Research Associate Professor (10/2023-Present). His research interests mainly include Artificial Intelligence Ethics (AI Ethics), Trustworthy Artificial Intelligence (Trustworthy AI), Evolutionary Computation (EC), and their practical applications.

All are welcome!

In case of questions, please contact Prof Zhichao Lu at zhichao.lu@cityu.edu.hk, or visit the CS Departmental Seminar Web at https://www.cs.cityu.edu.hk/events/cs-seminars/recent-cs-colloquiums.