香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

35 YEARS OF EXCELLENCE

DEPARTMENT OF COMPUTER SCIENCE

SEMINAR

2023 SERIES

When Attribution Methods Meet Robustness of Deep Networks

DATE & TIME

24 JUL 2023 (MON) 10:00 - 11:00 AM

VENUE

Mr. and Mrs. Lee Siu Lun Lecture Theatre (WLB 205),
Shaw Campus

DR. ADAMS WAI-KIN KONG

Associate Professor
School of Computer Science and Engineering
Nanyang Technological University

ABSTRACT

In recent years, deep networks have demonstrated their powerful capabilities, often outperforming human experts, such as AlphaGO. Many have been deployed to real applications, creating great business value. Because of their great impact to human beings and our society, many have concerned about their potential risks. In this talk, the speaker will first give some background information about adversarial attacks, including black-box and white-box attacks and attribution methods, including integrated gradients. Then, he will discuss how to use attribution methods, in particular integrated gradients, which was originally developed for explaining deep methods, to study risks of transferable attacks. Similar to deep network output, attributions can be manipulated by adversarial attacks. In the second half of this talk, the speaker will discuss how to protect attributions against adversarial attacks.

SPEAKER'S BIOGRAPHY

REGISTER NOW

Enquiries: 3411-2385 **Email:** comp@comp.hkbu.edu.hk **Website:** https://bit.ly/bucs-events