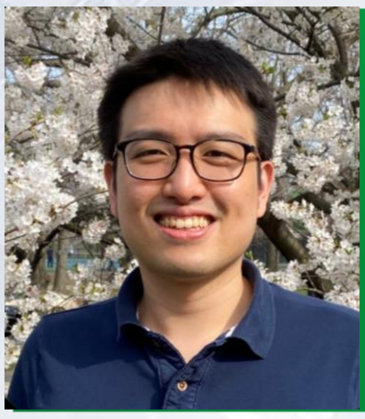
香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

ONLINE SEMINAR
2022 SERIES

Department of Computer Science



Dr. Pin-Yu Chen

Principal Research Scientist
IBM Thomas J. Watson Research Center, USA.

Register Now

Date: 23 August 2022 (Tuesday)
Time: 9:00am – 10:00am
Registration: http://bit.ly/bucs-ereg

(*Zoom details will only be provided to registrants)

AI Model Inspector: Towards Holistic Adversarial Robustness for Deep Learning

ABSTRACT

In this talk, I will share my research journey toward building an AI model inspector for evaluating, improving, and exploiting adversarial robustness for deep learning. I will start by providing an overview of research topics concerning adversarial robustness and machine learning, including attacks, defenses, verification, and novel applications. For each topic, I will summarize my key research findings, such as (i) practical optimization-based attacks and their applications to explainability and scientific discovery; (ii) Plug-and-play defenses for model repairing and patching; (iii) attack-agnostic robustness assessment; and (iv) data-efficient transfer learning via model reprogramming. Finally, I will conclude my talk with my vision of preparing deep learning for the real world and the research methodology of learning with an adversary. More information about my research can be found at: http://www.pinyuchen.com

BIOGRAPHY

Dr. Pin-Yu Chen is a principal research scientist at IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. He is also the chief scientist of RPI-IBM AI Research Collaboration and PI of ongoing MIT-IBM Watson AI Lab projects. Dr. Chen received his Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor, USA, in 2016. Dr. Chen's recent research focuses on adversarial machine learning and robustness of neural networks. His long-term research vision is to build trustworthy machine learning systems. At IBM Research, he received the honor of IBM Master Inventor and several research accomplishment awards, including an IBM Master Inventor and IBM Corporate Technical Award in 2021. His research works contribute to IBM open-source libraries including Adversarial Robustness Toolbox (ART 360) and AI Explainability 360 (AIX 360). He has published more than 50 papers related to trustworthy machine learning at major AI and machine learning conferences, given tutorials at AAAI'22, IJCAI'21, CVPR('20,'21), ECCV'20, ICASSP'20, KDD'19, and Big Data'18, and organized several workshops for adversarial machine learning. He received the IEEE GLOBECOM 2010 GOLD Best Paper Award.

ENQUIRY

Tel: 3411-2385 Email: comp@comp.hkbu.edu.hk Website: https://www.comp.hkbu.edu.hk/v1/?page=events