Xingjun Ma
Assistant Professor
School of Information Technology
Deakin University, Australia
Email: danxjma@gmail.com
Homepage: http://xingjunma.com

Dr. Xingjun (Daniel) Ma is an assistant professor in the School of Information Technology at Deakin University and an honorary fellow at The University of Melbourne. He obtained his Ph.D. degree in machine learning from The University of Melbourne where he also worked as a postdoctoral research fellow for one year and a half. His research interests include adversarial machine learning, weakly supervised learning, AI security, and data privacy. He has published 20+ works at top-tier conferences such as ICML, ICLR, CVPR, ICCV, ECCV, AAAI, and IJCAI. These works have made substantial impacts in the machine learning community with either theoretical contributions or new SOTA results. His work on “unlearnable examples” in 2021 was recently featured by MIT Technology Review. He also serves as a PC/SPC member or reviewer for a number of leading machine learning conferences and journals.

Title: Adversarial, Backdoor and Unlearnable

Abstract:

In this talk, I will introduce my three ICLR2021 works on 1) backdoor defense, 2) adversarial defense, and 3) data protection, respectively. The first work explores a neural attention distillation approach to erase backdoors from deep neural networks (DNNs). The second work reveals the magnitude and frequency characteristics of adversarially robust activations at the intermediate

layers of DNNs and introduces a channel-wise activation suppressing (CAS) technique to robustify DNNs. The third work proposes a type of error-minimizing noise to fool DNNs to believe there is nothing to learn from the training data so as to achieve an effect of “unlearnable” for the purpose of data protection. I will share some insights into unlearnable examples, their current limitations and the opportunities.