專題演講

講 者： 陳尚澤 教授（National Taiwan University）

題 目： AI-infused Security: Robust Defense by Bridging Theory and Practice

摘 要：

While Artificial Intelligence (AI) has tremendous potential as a defense against real-world cybersecurity threats, understanding the capabilities and robustness of AI remains a fundamental challenge, especially in adversarial environments. In this talk, I address two interrelated problems that are essential to the successful deployment of AI in security settings. (1) Discovering real-world vulnerabilities of deep neural networks and the countermeasures to mitigate such threats. I will present ShapeShifter, the first targeted physical adversarial attack that fools state-of-the-art object detectors, and SHIELD, a real-time defense that removes adversarial noise by stochastic data compression. (2) Developing theoretically-principled methods for choosing machine models to defend against unknown future attacks. I will introduce a novel game theory concept called “diversified strategy” to help make the optimal decision with limited risk, and then show how to use this concept to design efficient learning algorithms with strong theoretical guarantees for distributed and noisy data. Finally, I will share my vision on making AI more robust under different threat models, and research directions on deploying AI in security-critical and high-stakes problems.