

DEPARTMENT OF COMPUTER SCIENCE

SEMINAR

2026 SERIES

Agentic AI System Security Through the Lens of Real-World Interaction

DATE & TIME

21 JAN 2026 (WED) 10:45 - 11:45 AM

ONLINE VIA ZOOM



DR. XINFENG LI

Postdoctoral Fellow
Nanyang Technological University

ABSTRACT

In the large language model (LLM) era, recent breakthroughs are transforming AI systems from passive information processors into autonomous agents that act in the real world. Industry leaders are rapidly deploying LLM agents to automate complex workflows, ranging from software development and cloud infrastructure management to scientific discovery. However, this acceleration in capability has far outpaced a systematic understanding of agent safety, security, and privacy in open-world settings. As a result, agentic AI introduces a dual threat: agents are simultaneously vulnerable to adversarial attacks and powerful enough to be weaponized for malicious or deceptive purposes at scale. In this talk, I argue that these risks arise from fundamental mismatches at the interfaces between agents and the real world. I structure these challenges into three gaps. First, the perception gap, where subtle physical perturbations undermine sensory trust and propagate failures downstream. Second, the cognition gap, where utility-driven objectives conflict with human-centric values such as safety and privacy. Third, the execution gap, where general-purpose agents interact insecurely with specialized digital systems, enabling large-scale misuse. I will present a systematic framework for securing agent-world interaction, combining principled security analysis with lightweight, deployable defenses across all three stages. Together, my work aims to advance trustworthy agentic systems that are responsible and robust for real-world deployment.



**SPEAKER'S
BIOGRAPHY**



REGISTER NOW