DEPARTMENT OF COMPUTER SCIENCE
TEXAS TECH
Whitacre College of Engineering

Unveiling Privacy Risks in AI: Data, Models, and Systems

Shengwei An
Purdue University
Wednesday, February 19, 2025
10:00 a.m.
IMSE 121

Abstract: Artificial Intelligence (AI) has become deeply integrated into diverse systems, transforming industries and reshaping our daily lives. However, this widespread adoption also introduces critical privacy risks across the training data, AI models, and AI-powered systems. This talk will explore privacy challenges through these three aspects. First, I will introduce the first high-fidelity attack that exposes the privacy vulnerabilities of training data in pre-trained models and commercial AI services. Next, I will present a novel physical impersonating attack that highlights the privacy risks inherent in AI-based authentication systems. Additionally, I will discuss the first data-free framework designed to eliminate trigger-based model watermarks in diffusion models that aim to protect their intellectual property. Finally, I will conclude with a forward-looking perspective on addressing privacy risks in emerging generative AI techniques, such as Large Language Models and Stable Diffusion Models.

Bio: Shengwei An is a Ph.D. candidate in the Department of Computer Science at Purdue University, advised by Prof. Xiangyu Zhang. His research focuses on AI security and privacy, with an emphasis on designing state-of-the-art tools to investigate and mitigate privacy vulnerabilities in real-world AI systems. His work has been published in top-tier conferences, including S&P, USENIX Security, NDSS, and AAAI. He is the recipient of the Ross Fellowship from Purdue University and the Best Paper Award in the ECCV 2022 AROW Workshop.
More information can be found on his website: https://www.cs.purdue.edu/homes/an93/