DEPARTMENT OF
COMPUTER SCIENCE

TEXAS TECH
Whitacre College of Engineering

On the Trustworthiness of ML for Security-Related Tasks

Michael Reiter, Ph.D.
Duke University
Tuesday, November 18, 2025
3:30 p.m.
Zoom

Abstract: In this talk, we will discuss our results over the past decade in evaluating the trustworthiness of ML models in performing security-related tasks. First, by casting authorization decisions as the task of classifying requests into those that should be permitted or denied, we summarize our results showing that ML classifiers tend to be susceptible to being fooled by adversaries in two important authorization contexts: facial authentication and malware detection. Second, we consider the ability of ML models to keep secrets, and put their weaknesses in doing so to a constructive use: enabling a data owner to audit an ML model for use of her data in training. In particular, we will detail our progress on developing general techniques for black-box, data-use auditing of ML models that enable the data owner to bound her false-detection rate.

Bio: Michael Reiter is a James B. Duke Distinguished Professor in the Departments of Computer Science and Electrical & Computer Engineering at Duke University. His research interests include all aspects of computer security and fault-tolerant distributed systems, including blockchains. He is a Fellow of the ACM and the IEEE, and the recipient of: the Outstanding Contributions Award from the ACM Special Interest Group on Security, Audit and Control; the ACM CODASPY Lasting Research Award; and Test of Time Awards from ACM CCS (x2) and from Intel for specific research contributions. More information can be found at https://reitermk.github.io.