PURDUE UNIVERSITY | Elmore Family School of Electrical and Computer Engineering

Faculty Candidate Seminar – Software Engineering

Kexin Pei

PhD Candidate in Computer Science
Columbia University

Wednesday, January 25, 2023
Presentation: 10:30 A.M. – 11:30 A.M.
Reception: 11:30 A.M. – 12:00 P.M.
MSEE 239

Analyzing and Securing Software with Robust and Generalizable Learning

Abstract: Software is powering every aspect of our society, but it remains plagued with errors and prone to critical failures and security breaches. Program analysis has been a predominant technique for building trustworthy software. However, traditional approaches rely on hand-curated rules tailored for specific analysis tasks and thus require significant manual effort to tune for different applications. While recent machine learning-based approaches have shown some early promise, they, too, tend to learn spurious features and overfit to specific tasks without understanding the underlying program semantics.

In this talk, I will describe my research on building machine learning (ML) models toward learning program semantics so they can remain robust against transformations in program syntax and generalize to various program analysis tasks and security applications. The corresponding research tools, such as XDA, Trex, StateFormer, and NeuDep, have outperformed commercial tools and prior arts by up to 117x in speed and by 35% in precision and have helped identify security vulnerabilities in real-world firmware that run on billions of devices. To ensure the developed ML models are robust and generalizable, I will briefly describe my research on building testing and verification frameworks for checking the safety properties of deep learning systems. The corresponding research tools, such as DeepXplore, DeepTest, ReluVal, and Neurify, have been adopted and followed up by the industry (e.g., in TensorFuzz built by Google), been covered in media such as Scientific American, IEEE Spectrum, Newsweek, and TechRadar, and inspired over thousands of follow-up projects.

Bio: Kexin Pei is a Ph.D. candidate in Computer Science at Columbia University, advised by Suman Jana and Junfeng Yang. His research lies at the intersection of security, software engineering, and machine learning, with a focus on building machine-learning tools that utilize program structure and behavior to analyze and secure software. His research has received the Best Paper Award in SOSP, a Distinguished Artifact Award, been featured in CACM Research Highlight, and won CSAW Applied Research Competition Runner-Up. He was part of the learning for code team when he interned at Google Brain, building program analysis tools based on large language models.