Date: 29th May, 2024 [Wednesday]

Time: 12 pm to 1 pm

Venue: RM 101, Rajeev Motwani Building, Department of Computer Science and Engineering

Speaker: Prof. Siddharth Garg (https://engineering.nyu.edu/faculty/siddharth-garg)

Title: Foundation Models: The Good, the Backdoors and the Ugly

Abstract:

Foundation models, massive neural networks trained at great expense, will form the backbone of a rapidly growing AI/ML ecosystem. Indeed, foundation models have shown impressive capabilities in generalising to a broad range of tasks. I will start with a note of optimism (the good) and describe our recent success in tailoring large language models (LLMs) for chip design. With few exceptions, however, these models are black-boxed and only accessible via cloud APIs, with no visibility into training data and training scripts. As users, we are expected to trust foundation models with our sensitive data and to trust their responses. I argue that there is good reason to be skeptical of blackbox foundation models. I will highlight four key concerns, and in some cases, mitigations, from the work in my group. These are: (1) data breaches and privacy-preserving model inference; (2) security bugs in LLM generated code; (3) malicious backdoors; and (4) demographic bias.

Bio:

Siddharth Garg is currently the Institute Associate Professor of ECE at NYU Tandon, where he leads the EnSuRe Research group (https://wp.nyu.edu/ensure_group/). Prior to that he was in Assistant Professor also in ECE from 2014-2020, and an Assistant Professor of ECE at the Unversity of Waterloo from 2010-2014. His research interests are in machine learning, cyber-security and computer hardware design.

He received his Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University in 2009, and a B.Tech. degree in Electrical Engineering from the Indian Institute of Technology Madras. In 2016, Siddharth was listed in Popular Science Magazine's annual list of "Brilliant 10" researchers. Siddharth has received the NSF CAREER Award (2015), and paper awards at the IEEE Symposium on Security and Privacy (S&P) 2016, USENIX Security Symposium 2013, at the Semiconductor Research Consortium TECHCON in 2010, and the International Symposium on Quality in Electronic Design (ISQED) in 2009. Siddharth also received the Angel G. Jordan Award from ECE department of Carnegie Mellon University for outstanding thesis contributions and service to the community. He serves on the technical program committee of several top conferences in the area of computer engineering and computer hardware, and has served as a reviewer for several IEEE and ACM journals.