CS VIRTUAL SEMINAR SERIES

Eric Wong

JOHNS HOPKINS WHITING SCHOOL of ENGINEERING | Computer Science

"Building the Reliability Stack for Machine Learning"

Tuesday, March 22, 2022

10:45 AM - 12:00 PM

https://wse.zoom.us/j/97595802965

ABSTRACT

Currently, machine learning (ML) systems have impressive performance but can behave in unexpected ways. These systems latch onto unintuitive patterns and are easily compromised, a source of grave concern for deployed ML in settings such as healthcare, security, and autonomous driving. In this talk, I will discuss how we can redesign the core ML pipeline to create reliable systems. First, I will show how to train provably robust models, which enables formal robustness guarantees for complex deep networks. Next, I will demonstrate how to make ML models more debuggable. This amplifies our ability to diagnose failure modes, such as hidden biases or spurious correlations. To conclude, I will discuss how we can build upon this "reliability stack" to enable broader robustness requirements, and develop new primitives that make ML debuggable by design.

BIOGRAPHY

Eric Wong is a postdoctoral researcher in the Computer Science and Artificial Intelligence Laboratory at Massachusetts Institute of Technology. His research focuses on the foundations for reliable systems: methods that allow us to diagnose, create, and verify robust systems. He is a 2020 Siebel Scholar and received an honorable mention for his thesis on the robustness of deep networks to adversarial examples at Carnegie Mellon University.

HOW TO REACH US

Contactus@cs.jhu.edu

410-516-8775

cs.jhu.edu

Johns Hopkins University
Department of Computer Science
3400 N. Charles St | Malone 160
Baltimore, MD 21218