Yale University
Computer Science Talk

Trustworthy Deep Learning: methods, systems and theory

Host: Abhishek Bhattacharjee

Matthew Mirman

Thursday – March 10, 2022
10:30 a.m.

Zoom Presentation

Abstract: Deep learning models are quickly becoming an integral part of a plethora of high stakes applications, including autonomous driving and health care. As the discovery of vulnerabilities and flaws in these models has become frequent, so has the interest in ensuring their safety, robustness and reliability. My research addresses this need by introducing new core methods and systems that can establish desirable mathematical guarantees of deep learning models.

In the first part of my talk I will describe how we leverage abstract interpretation to scale verification to orders of magnitude larger deep neural networks than prior work, at the same time demonstrating the correctness of significantly more properties. I will then show how these techniques can be extended to ensure, for the first time, formal guarantees of probabilistic semantic specifications using generative models.

In the second part, I will show how to fuse abstract interpretation with the training phase so as to improve a model's amenability to certification, allowing us to guarantee orders of magnitude more properties than possible with prior work. Finally, I will discuss exciting theoretical advances which address fundamental questions on the very existence of certified deep learning.

Bio: Matthew Mirman is a final-year PhD student at ETH Zürich, supervised by Martin Vechev. His main research interests sit at the intersection of programming languages, machine learning, and theory with applications to creating safe and reliable artificial intelligence systems. Prior to ETH, he completed his B.Sc. and M.Sc. at Carnegie-Mellon University supervised by Frank Pfenning.