北京大学前沿计算研究中心
Center on Frontiers of Computing Studies, Peking University

静园5号院
前 沿 讲 座

On the Safety and Generalizability of Trustworthy Intelligent Autonomy

Dr. Ding Zhao

Department of Mechanical Engineering
Carnegie Mellon University

董豪 助理教授
2022年12月23日 星期五 11:00–12:00
在线讲座

Abstract

As machine learning has started to shift towards deployment in the physical world, its rapid development is coupled with as much risk as benefits. In this talk, I will discuss two key aspects of developing trustworthy autonomy: generalizability and safety. I will introduce our recent progress in both mathematical foundations and real-world applications. I will also share the lessons we learned and our outlook for future research directions.

Biography

Ding Zhao is currently an assistant professor in the Department of Mechanical Engineering at Carnegie Mellon University, with affiliations at the Computer Science Department, Robotics Institute, and CyLab on Security and Privacy. Directing the CMU Safe AI Lab, his research focuses on the theoretical and practical aspects of safely deploying AI to safety-critical applications, including self-driving, assistant robots, healthcare diagnosis, and cybersecurity. He is the recipient of the National Science Foundation CAREER Award, CMU George Tallman Ladd Research Award, MIT Technology Review 35 under 35 Award in China, Ford University Collaboration Award, Carnegie-Bosch Research Award, Struminger Teaching Award, and many industrial fellowship awards. He worked with leading industrial partners, including Google, Apple, Ford, Uber, IBM, Adobe, Bosch, Toyota, and Rolls-Royce. He is a visiting researcher in the Robotics Team at Google Brain.

http: //cfcs.pku.edu.cn/