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

Deploying Intelligent Autonomy at a Large Scale - Generalizability, Safety, Embodiment

Prof. Ding Zhao

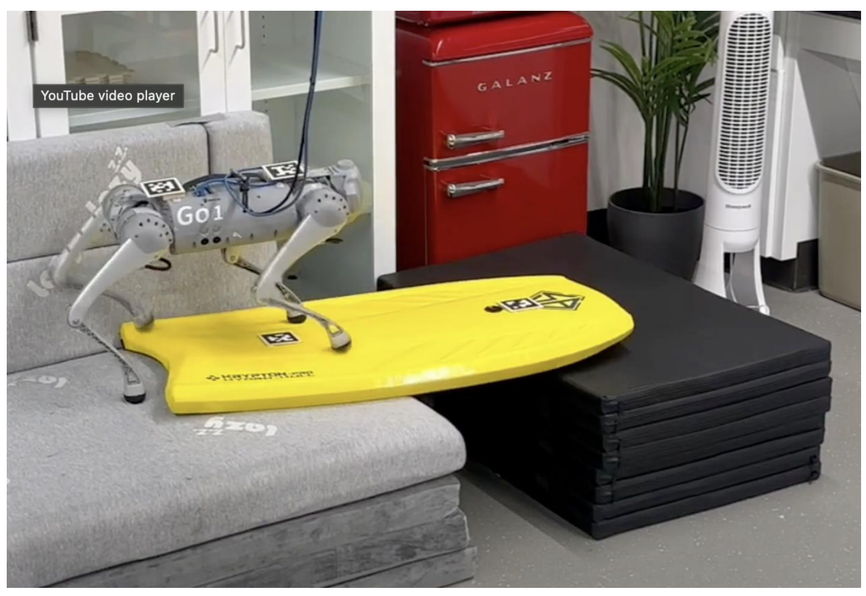
Department of Mechanical Engineering
Carnegie Mellon University

Host: 董豪 助理教授
2023年12月6日 星期三 15:00
静园五院204室

Abstract

As AI becomes more integrated into physical autonomy, it presents a dual spectrum of opportunities and risks. In this talk, I will introduce our efforts in deploying trustworthy intelligent autonomy at a large scale for vital civil usage such as self-driving cars, assistant robots, and autonomous surgery. During the deployment and transition, training data often exhibit significant imbalance, multi-modal complexity, and nonstationarity. I will initiate the discussion by analyzing 'long-tailed' problems with rare events and their connection to safety evaluation and safe reinforcement learning. I will then discuss how modeling multi-modal uncertainties as ‘tasks’ may enhance generalizability by learning across domains. To facilitate task delineation with high-dimensional inputs in vision and language, we have developed prompt-transformer-based structures for efficient adaptation and mitigation of catastrophic forgetting.

In cases involving unknown-unknown tasks with severely limited data, we explore the potential of leveraging external knowledge from legislative sources, causal reasoning, and large language models. Lastly, we will expand intelligence development into the realm of system-level design space with meta physical robot morphologies, which may achieve generalizability and safety more effectively than relying solely on software solutions.



Biography

Ding Zhao is the Dean's Early Career Fellow Associate Professor of Mechanical Engineering at Carnegie Mellon University. He directs the CMU Safe AI Lab, where his research focuses on large scale deployment of intelligent autonomy, encompassing generalizability, safety, physical embodiment, as well as considerations of privacy, equity, and sustainability. His work spans self-driving cars, assistant robots, autonomous surgical robots, and co-designing smart cities/buildings/infrastructure with autonomy. Ding Zhao has received numerous awards, including IEEE George N. Saridis Best Transactions Paper Award, National Science Foundation CAREER Award, MIT Technology Review 35 under 35 Award in China, Struminger Teaching Award, George Tallman Ladd Research Award, Ford University Collaboration Award, Qualcomm Innovation Award, Carnegie-Bosch Research Award, and many other industrial awards.