Safe and Interactive Autonomy: A Journey Starting from Formal Methods

(Keynote)

Dorsa Sadigh

Computer Science Department and Electrical Engineering Department
Stanford University, Stanford, CA, USA
https://dorsa.fyi/

Abstract

Todays society is rapidly advancing towards autonomous systems that interact and collaborate with humans, e.g., semiautonomous vehicles interacting with drivers and pedestrians, medical robots used in collaboration with doctors, or service robots interacting with their users in smart homes. With the emergence of autonomous systems in our every day lives, we need to design algorithms and tools that enable safe and seamless interactions with people.

In this talk, I will start with my journey in providing safety for human-robot systems by discussing a spectrum of views on safe autonomous systems including a formal methods perspective for synthesizing provably correct controllers, a robust control approach, and more recent advances in safe learning and verification. I will then discuss one of the main challenges of safety of human-robot systems, i.e., studying how robots influence humans actions in one-on-one or group settings. This is usually overlooked by assuming humans act as external disturbances just like moving obstacles, or assuming that automation can always help societies without actually considering how humans can be impacted. I will talk about our recent work in building computational models of human behavior from expert demonstrations and preferences in interaction with autonomous systems and challenges it introduces for safety and robustness verification.