Title: Interactive Learning and Control in the Era of Large Models
Presenter: Dorsa Sadigh

Abstract:
In this talk, I will discuss the problem of interactive learning by discussing how we can actively learn objective functions from human feedback capturing their preferences. I will then talk about how the value alignment and reward design problem can have solutions beyond active preference-based learning by tapping into the rich context available from large language models. In the second section of the talk, I will more generally talk about the role of large pretrained models in today's robotics and control systems. Specifically, I will present two viewpoints: 1) pretraining large models for downstream robotics tasks, and 2) finding creative ways of tapping into the rich context of large models to enable more aligned embodied AI agents. For pretraining, I will introduce Voltron, a language-informed visual representation learning approach that leverages language to ground pretrained visual representations for robotics. For leveraging large models, I will talk about a few vignettes about how we can leverage LLMs and VLMs to learn human preferences, allow for grounded social reasoning, or enable teaching humans using corrective feedback. Finally, I will conclude the talk by discussing some preliminary results on how large models can be effective pattern machines that can identify patterns in a token invariant fashion and enable pattern transformation, extrapolation, and even show some evidence of pattern optimization for solving control problems.