Responsibility in AI Systems & Experiences (RAISE) at the University of Washington presents:

Sandhya Saisubramanian

AI in the Open World: Leveraging Human Guidance to Mitigate Undesirable Effects of Incomplete Specification

Friday January 21, 2022, 9-10am PT

Join: https://washington.zoom.us/j/94636255672

Despite recent advances in AI, ensuring reliable and fair operation of deployed systems remains a challenge. AI systems fielded in the open world often suffer from incomplete specification, which causes concerns around reliability and fairness. In the first part of the talk, I will present techniques that allow autonomous systems to improve reliability by mitigating their negative side effects using different forms of feedback. In the second part of the talk, I will discuss how to leverage human demonstrations to avoid biases when clustering with incompletely specified fairness metrics.

Sandhya Saisubramanian is an Assistant Professor in the School of Electrical Engineering and Computer Science at Oregon State University where she directs the Laboratory for Adaptive Intelligent Systems. She completed her Ph.D. from the University of Massachusetts Amherst. Her research interests are in designing AI systems that are safe, reliable, and unbiased. Her current research focuses on reliable decision-making in autonomous systems that operate in the open world. She is a recipient of IJCAI 2020 Distinguished Paper Award.

RAISE is a UW-wide group of students and faculty interested in the broad space of responsible AI, trustworthy machine learning, human-centered computing and data science. As part of this group, our mission is to engage in scholarly, educational, and outreach activities that lead to foundational research in these areas. https://www.raise.uw.edu.

UNIVERSITY of WASHINGTON

RAISE
RESPONSIBILITY IN AI SYSTEMS & EXPERIENCES