Robotics Seminar Series

Value Alignment and Safety via Interactive and Explainable Human-Robot Learning

Dr. Matthew Gombolay
Associate Professor of Interactive Computing
Georgia Institute of Technology

Value alignment and safety in generative Artificial Intelligence (AI) are key problems receiving intense focus writ large due to the emergent success of large models in many intelligence benchmarks. The threat of misaligned models becomes even more stark as we look to embody agents physically for human-robot interaction. In this talk, I will present human-centered approaches to address these critical challenges in generative AI through interactive and explainable robot learning. First, I will define challenges for safety and alignment in AI/Robotics and show that addressing the problem is more difficult than it appears at first glance. Second, I will share our recent advances to align and verify the properties of learned robot behavior through interactive machine learning and explainable artificial intelligence. Finally, I will present a roadmap for future work to bring together critical stakeholders for transdisciplinary research for beneficent AI.

Dr. Matthew Gombolay is an Associate Professor of Interactive Computing at the Georgia Institute of Technology. He was named the Anne and Alan Taetle Early-career Assistant Professor in 2018. He received a B.S. in Mechanical Engineering from Johns Hopkins University in 2011, an S.M. in Aeronautics and Astronautics from MIT in 2013, and a Ph.D. in Autonomous Systems from MIT in 2017. Between defending his dissertation and joining the faculty at Georgia Tech, Dr. Gombolay served as technical staff at MIT Lincoln Laboratory, transitioning his research to the U.S. Navy and earning an R&D 100 Award. His publication record includes best paper awards and nominations from the American Institute for Aeronautics and Astronautics, the ACM/IEEE Conference on Human-Robot Interaction, the Conference on Robot Learning, and Robotics: Science and Systems. Dr. Gombolay was selected as a DARPA Riser and received the Early Career Award from the National Fire Control Symposium, a NASA Early Career Fellowship, and the NSF CAREER award.

FRIDAY
Jan 10
10-11AM

Rogers Hall 230
FREE
Refreshments Served

OSU Robotics
robotics.oregonstate.edu

Oregon State University