

Date: January 29

Speaker: Shikhar Kumar, PhD student. Department of Industrial Engineering & Management, BGU

Title: Understandable robots

Abstract: For the widespread deployment of autonomous robots in everyday tasks robot's actions, decisions and intentions must be understood by users. Inspired by recent works in the field of explainable AI the focus of this research is on contributing to the development of understandable robots by enhancing the robot with explanations. The explanation is based on three questions i.e., *what* needs to be explained, *when* should it be explained and *why* an action is being taken by the robot. These three questions provide a basis for defining different levels of explanation (corresponding to levels of automation) to increase understandability.

In the current presentation we focus on a model for the explanations which includes verbosity, a pattern in which explanations are being communicated to the user (*what*), and justification for a particular option (*why*). This relates to a fundamental factor affecting robot understandability, the amount of discrepancy between the robot's and human's state-of-minds. We provide valuable insights of this discrepancy based on a user study and developing an algorithm that addresses and manages it. A user study which involved 36 participants who executed a collaborative task based on the robot's instructions verified that the lesser discrepancy between robot and human, the more efficient and successful the interactions are, despite vague or short robot instructions. Based on this, we developed a general algorithm based on Hidden Markov Models, for the robot to continuously assess the discrepancy during an interaction and in turn adapt its behavior with the aim of decreasing discrepancy.

Bio: Shikhar Kumar is a final year PhD student at Ben-Gurion university of the Negev. Previously, he was an Early-Stage researcher in the SOCRATES EU ITN Project at BGU. He holds a master's degree from the Indian Institute of Technology (Indian School of Mines), Dhanbad. His research interest is in Human-robot interaction.