BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK | Computer Science

THE DEPARTMENT OF COMPUTER SCIENCE PRESENTs

# INVITED SPEAKER SERIES

**Dr. Eugene Santos Jr.**
**Dartmouth College**

**Friday, October, 8, 2021 at 12 p.m.**
https://binghamton.zoom.us/j/92872321997

## Inferring and Understanding Team Behavior Through Learning Rewards and Reward Structures: Interference and Human-Machine Teams

**Abstract:** This talk focuses on understanding how a team (machine-machine, human-machine, multiples) performs or will perform and why they performed in the way they did. For the latter, the ability to make any sort of explanation has obvious implications including identifying future improvements for the team. I will present a methodology which considers both the team's and the team members' reward functions. These reward functions and structures can be inferred via inverse reinforcement learning (IRL) using their observed behaviors. Reward functions reflect underlying agent goals and preferences and can be used to capture unique agent behavior. We map these rewards to high-level behavioral attributes (behA) that are related to a team’s performance metrics. These behA can provide insights that will then help to explain a team’s performance. Central to understanding team behavior is interference which impacts team member interactions. Interference occurs when the goals of one agent affect the goals of the other agents. When positive, it boosts a team's performance, and when negative, it degrades the team’s performance. Interference is likely to arise due to many differences including communication mechanisms, roles, capabilities, adaptiveness, and responsibility such as between human and machine members. I will describe two recent experiments: one focused on machine-machine teams, and the other, our initial results on human-machine teams, both within a popular computer game.

**Bio:** Dr. Eugene Santos, Jr. received his Ph.D. in Computer Science from Brown University. He is currently Professor of Engineering at the Thayer School of Engineering and Adjunct Professor of Computer Science at Dartmouth College. Dr. Santos’ work on artificial intelligence intersects the areas of information, cognition, human factors, and mathematics. His current focus is on computational intent, dynamic human behavior, and decision-making with an emphasis on learning nonlinear and emergent behaviors and explainable AI. Dr. Santos has applied his work with the goal of better understanding how we, both as individuals and our society, can best leverage knowledge through AI to improve our world for social good. These application areas include computational social systems – group to individual decision-making, opinion and belief change, socio-cultural attitudes and factors, and social resilience; user and team modeling – inferring user intentions and needs, surgical errors and team intention gaps, and effective use of text and data analytics; and, cybersecurity – insider threat and deception detection, misinformation vs. disinformation, and adversarial intent and course of action analysis. He was appointed to the State of Vermont Taskforce on AI (first state-level taskforce in nation), serves on the Board of Directors for Sigma Xi, and is a 2019 Public Voices fellow of the OpEd Project. He is a Fellow of the AAAS and IEEE.