

RIT

Center for
Human-aware
Artificial
Intelligence



CHAI Seminar Series

Refreshments will be served

DATE: **Monday, March 25, 2024, 12:00-1:00 PM**

SPEAKER: **Ali Baheri PhD**, Assistant Professor, Department of Mechanical Engineering, and Head, Safe AI Lab, RIT

TITLE: **Evolving AI Decision-Making: From Safe Reinforcement Learning to Intelligent Systems with Language Models**

IN PERSON: **Golisano Hall (070), Room CYB-1710/1720**

ABSTRACT: Reinforcement learning (RL) has been a driving force in advancing decision-making across various domains, despite the inherent challenge of ensuring safety in real-world deployments. Traditional safe RL methodologies, which focus on embedding pre-defined safety constraints within the policy learning process, face limitations in dynamic environments where constraints are not easily defined or adaptable. In the first part of the talk, we propose an approach that concurrently learns a safe RL control policy and identifies unknown safety constraint parameters. This method has proven effective in diverse environmental conditions, accurately learning safety constraint parameters and demonstrating proficiency in adhering to environmental safety without comprehensive prior knowledge. The second part of the talk will explore the synergy between recent advances in large language models (LLMs) and RL, highlighting how the integration of LLMs with RL can enhance informed decision-making by leveraging the predictive and generative capabilities of LLMs to inform and refine RL policies. This fusion aims to create more intelligent, adaptable, and safe decision-making systems.



BIO: Ali Baheri is Assistant Professor of Mechanical Engineering, where he leads the Safe AI Lab. His lab focuses on research at the intersection of autonomy, controls, and machine learning, aiming to advance safe, certified, and efficient autonomy for intelligent systems. Before joining RIT, he was a visiting scholar at Stanford University, an Assistant Professor (research track) at West Virginia University, and a postdoctoral research fellow at the University of Michigan—Ann Arbor. He received his Ph.D. from the University of North Carolina at Charlotte in 2018. His work, funded by NSF, FAA, and NASA, leverages insights from control theory, optimization, and machine learning to contribute to the progress of autonomy, ensuring safety and efficiency in the application of intelligent systems.

NOTE: To schedule interpreter and/or services for this event, please use <https://myaccess.rit.edu>.

<https://www.rit.edu/chai/>