Proposal Defense
Doctor of Philosophy in Computer Science

Interpretable and Human-Aligned AI for Multi-Objective Traffic Signal Control
by **Marcelo Andrade Rodrigues d'Almeida**

Date: December 1, 2025
Time: 11:00 a.m. – 1:00 p.m.
Place: Sennott Square 6329 - Board Room,
3810 Forbes Ave, Pittsburgh, PA 15213

Committee:

- Dr. Daniel Mosse, Professor, Department of Computer Science, School of Computing and Information
- Dr. Xiaowei Jia, Associate Professor, Department of Computer Science, School of Computing and Information
- Dr. Ryan Shi, Assistant Professor, Department of Computer Science, School of Computing and Information
- Dr. Na Du, Assistant Professor, Department of Informatics and Networked Systems, School of Computing and Information

Abstract:
Traffic Signal Control (TSC) is a Cyber-Physical System problem that has traditionally relied on optimizing signal timings. Traffic Engineers (TEs) require such systems to be mostly automated, taking their input to address challenges, such as prioritizing different road users with conflicting objectives and adjusting day-to-day traffic operations. These systems must also be transparent, enabling TEs to analyze and audit traffic operation logic, and be robust, handling traffic incidents and sensor issues automatically. To address the above challenges, the project devises (a) a novel interpretable reinforcement learning (RL) TSC system handling preferences for many road users (multimodal) and many objectives (≥ 5), following TEs' recommendations, (b) identify which features are necessary for fault-tolerant traffic control and how to identify them automatically, and (c) a post hoc explainability analysis to further help TEs' understanding of control policies. We test the proposed methods in a rich multi-modal traffic simulation.