EXPLAINABLE AI PLANNING IN URBAN TRAFFIC

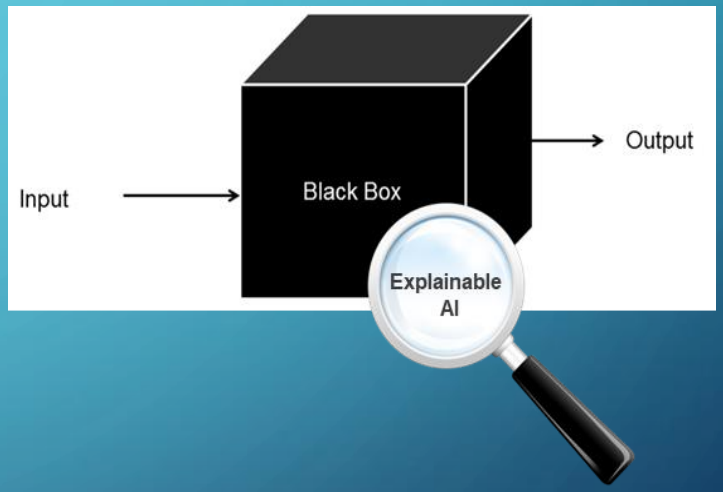
MARLON PEREIRA – MARLON.P@EDU.PUCRS.BR

AUTOMATED PLANNING – GRADUATE PROGRAM IN COMPUTER SCIENCE, POLYTECHNICAL SCHOOL

PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL (PUCRS)

OVERVIEW

- AI is a powerful tool but often viewed as a black box due to the opacity of the decision process to outside observers
- There is an arising need in that direction, both for collaborative interactions between AI and humans and to establish trust with end-users in general



RECENT DEVELOPMENTS

- Explainable AI Planning (XAIP) Workshop at (ICAPS)
  - Derived from the XAI Program (DARPA) – focused on Machine Learning
- Plan Explicability [Zhang et al., 2017] – human's interpretation of plans
- Model reconciliation [Chakraborti et al., 2017] – exploring human and AI model differences

THE DECISION MAKING PROBLEM

- *Why did you do that? And why didn't you do something else?*
- *Why is what you're proposing is better than the alternatives?*
- *Why can't you do that?*
- *Why is there a reason to replan at this point (or not)?*

SCENARIO – URBAN TRAFFIC

GIVEN A SET OF N AGENTS WITH DISTINCT COORDINATES AND OBJECTIVES, WHAT IS THE MOST OPTIMIZED SET OF STRUCTURES THAT SATISFY THE OVERALL GOAL?

OBJECTIVE

- Develop a model capable of purposing a solution to resource allocation and offer insights to why such answer is the best one
  - PDDL+ modelling – expansion of the PDDL language, which introduces processes and events
  - Model-Based Explanation – focus on the Reconciliation of human and AI models

TASK SCHEDULE

	PDDL+	BUILD THE MODEL	GENERATE INSIGHTS	TESTING/ TUNNING	EVALUATION	REPORT + PRESENTATION
WEEK 1	X					
WEEK 2	X	X				
WEEK 3		X	X			
WEEK 4		X	X	X		
WEEK 5				X	X	
WEEK 6						X

THANK YOU!