YooJung Choi

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING | Computer Science

"Tractable Probabilistic Reasoning for Trustworthy AI"

📅 Tuesday, March 29, 2022

🕐 10:45 AM – 12:00 PM

▐▐▐ https://wse.zoom.us/j/98323092626

ABSTRACT

Automated decision-making systems are increasingly being deployed in areas with personal and societal impacts, leading to growing interest and need for AI and ML systems that are robust, explainable, fair, and so on. It is important to note that these guarantees only hold with respect to a certain model of the world, with inherent uncertainties. In this talk, I will present how probabilistic modeling and reasoning, by incorporating a distribution, offer a principled way to handle different kinds of uncertainties when learning and deploying trustworthy AI systems. For example, when learning classifiers, the labels in the training data may be biased; I will show that probabilistic circuits, a family of tractable probabilistic models, can be effective in enforcing and auditing fairness properties by explicitly modeling a latent unbiased label. In addition, I will also discuss recent breakthroughs in tractable inference of more complex queries including information-theoretic quantities, to demonstrate the full potential of probabilistic reasoning. Finally, I will conclude with my future work towards a framework to more flexibly reason about and enforce trustworthy AI/ML system behaviors.

BIOGRAPHY

YooJung Choi is a Ph.D. candidate in Computer Science at the University of California, Los Angeles, advised by Guy Van den Broeck. Her research is broadly in the areas of artificial intelligence and machine learning, with focus on probabilistic modeling and reasoning for automated decision-making. Click here for more information.

HOW TO REACH US

✉ Contactus@cs.jhu.edu

☎ 410-516-8775

🌐 cs.jhu.edu

Johns Hopkins University
Department of Computer Science
3400 N. Charles St | Malone 160
Baltimore, MD 21218