UNIVERSITY OF CHICAGO
DEPARTMENT OF COMPUTER SCIENCE
DISTINGUISHED LECTURE SERIES
PRESENTS:

"Moral Artificial Intelligence: How to Learn Objectives from People"

Vincent Conitzer
Duke University

Abstract:
Algorithms in machine learning (ML) and artificial intelligence (AI) generally require us to specify an objective function, which formalizes what it is that we want our algorithm to optimize. While these algorithms were confined to the laboratories in which they were developed, the exact objective function often did not matter much. But now that they are being broadly deployed in the world, we find that simplistic objectives often have unintended side effects. For one, AI systems increasingly need to make decisions with a moral component. E.g., should a self-driving car prioritize the safety of its passengers over that of others, and to what extent? I will briefly discuss some general approaches to such problems.

I will then go into detail on the application of these techniques to kidney exchanges (no prior familiarity required). A kidney exchange allows patients who are in need of a kidney transplant, and who have willing but incompatible donors, to exchange donors. Some real kidney exchanges use algorithms to determine an optimal matching. Should such an algorithm take features such as the patient's age into account, and to what extent? What would be the consequences for a reasonably large exchange?

Finally, if time permits, I will discuss in more detail the problem that, generally, not everyone will agree on what the morally preferred option is, and how this can be addressed using techniques from computational social choice and computational learning theory.

Bio
Vincent Conitzer is the Kimberly J. Jenkins University Professor of New Technologies and Professor of Computer Science, Professor of Economics, and Professor of Philosophy at Duke University. He received Ph.D. (2006) and M.S. (2003) degrees in Computer Science from Carnegie Mellon University, and an A.B. (2001) degree in Applied Mathematics from Harvard University. Conitzer works on artificial intelligence (AI). Much of his work has focused on AI and game theory, for example designing algorithms for the optimal strategic placement of defensive resources. More recently, he has started to work on AI and ethics: how should we determine the objectives that AI systems pursue, when these objectives have complex effects on various stakeholders? Conitzer has received the Social Choice and Welfare Prize, a Presidential Early Career Award for Scientists and Engineers (PECASE), the IJCAI Computers and Thought Award, an NSF CAREER award, the inaugural Victor Lesser dissertation award, an honorable mention for the ACM dissertation award, and several awards for papers and service at the AAAI and AAMAS conferences. He has also been named a AAAI Fellow, a Guggenheim Fellow, a Kavli Fellow, a Bass Fellow, a Sloan Fellow, and one of AI's Ten to Watch. He is program co-chair of the 2019 Conference on AI, Ethics, and Society (AIES'19) and the 2020 AAAI conference.

Thursday, May 23, 2019
3:30 pm
Crerar 298
Host: Rebecca Willett