Application: Digital Economics and AI Tutorial

Contact Information

Name: Rubi Hudson

Legal Name: Jeremy Rubinoff

Association: University of Toronto, Department of Economics

Email: jeremy.rubinoff@gmail.utoronto.ca

Phone: 647-853-5329

Address: 100 Spadina Rd, Unit 2002, Toronto, ON, Canada, M5R 2T7

Course of Graduate Study Pursued

I have completed the first two years of a PhD in Economics at the University of Toronto. In the second year, I focused on field courses in microeconomics theory. During this time, I also collaborated with research scientists at Anthropic, a major AI firm, and PhD students in Computer Science on modeling the risks posed by advanced AI systems.

Description of Research Interests

My research interests are primarily focused on assessing the risks posed by AI systems as they become increasingly powerful and widespread, along with developing tools to mitigate these risks. In the last year, I have been focusing on modeling cutting edge systems as optimizing for predictive accuracy, and how doing so introduces an incentive for using predictions to manipulate the outcomes being predicted. This focus has resulted in collaborations with AI experts on two papers (1, 2).

I am continuing this research direction in my second year paper, supervised by Gabriel Carroll, which studies how jointly scoring multiple agents can get around previously established impossibility results for eliciting predictions from a single agent. I do not yet have a dissertation or committee, and may either continue with this direction or switch to studying a different threat model of AI.

I also organize the Mechanism Design for AI Safety reading group, and I am leading a research project with Economics and Computer Science PhD students from top schools to develop an agenda on how mechanism design tools can be applied to AI systems.