

Rubi Hudson

PhD Student in Economics
University of Toronto

Legal Name: Jeremy Rubinoff
rubi.hudson@mail.utoronto.ca
1-647-853-5329

Research Fields	Mechanism Design, Artificial Intelligence	
Education	Ph.D., University of Toronto Economics	2021-Present
	MRes, Universitat Pompeu Fabra Economics	2021
	M.A., Barcelona School of Economics Economics (Advanced Track)	2020
	BMath, University of Waterloo Statistics (Major), Combinatorics and Optimization (Minor), Economics (Minor)	2017
Publications	Joint Scoring Rules: Zero-Sum Competition Avoids Performative Prediction <i>Rubi Hudson</i> AAAI 2025	
	Incentivizing honest performative predictions with proper scoring rules Caspar Oesterheld*, Johannes Treutlein*, Emery Cooper, <i>Rubi Hudson</i> UAI 2023	
Preprints	Conditioning Predictive Models: Risks and Strategies Evan Hubinger, Adam S. Jermyn, Johannes Treutlein, <i>Rubi Hudson</i> , Kate Woolverton arXiv Preprint 2023	
Under Review	Superalignment Anti-Literature Review Michael Cohen*, <i>Rubi Hudson</i> *, Yoshua Bengio	
	Corrigibility Transformation: Constructing Goals That Accept Updates <i>Rubi Hudson</i>	
Awards and Scholarships	Open Philanthropy Career Development Funding	2023
	Open Philanthropy Early Career Funding for Individuals Interested in Improving the Long-Term Future	2022
	University of Toronto Doctoral Recruitment Award	2021
Service	Mechanism Design for AI Safety Founder and Lead Organizer	2022-Present
	Supervised Program for Alignment Research Mentor	2023-2024

Professional Experience	Manifund Independently Funded AI Safety Researcher	2023-2024
	Stanford Existential Risks Initiative Machine Learning Alignment Theory Scholar	2022-2023
	University of Toronto Research Assistant (Gabriel Carroll)	2022
	University of Toronto Teaching Assistant (Principles of Fair Decisions, Microeconomic Theory)	2021-2022
	Universitat Pompeu Fabra Research Assistant (Massimo Motta and Antonio Penta, Ruben Durante)	2020
	Crack the Ice Founder	2015-2019
	Preyra Solutions Group Senior Analyst	2018-2019
Invited Talks	Cancer Care Ontario Analyst (Funding Unit)	2017-2018
	AI Futures Fellowship Scheduled Programming	2024
	Condor Camp MasterClass	2024
	University of Oxford 13th Oxford Workshop on Global Priorities Research	2023