

RACHEL FREEDMAN

rachel.freedman@berkeley.edu | <https://rachelfreedman.github.io/>

Education

2019-pres.	Artificial Intelligence PhD, UC Berkeley <ul style="list-style-type: none">• I work with the Center for Human-Compatible Artificial Intelligence (CHAI) on techniques to ensure that advanced artificial agents act in ways that align with human values. My advisor is Prof. Stuart Russell.	3.91 GPA
2013-2017	Computer Science/Psychology BA, Duke University <ul style="list-style-type: none">• Graduated Phi Beta Kappa and <i>magna cum laude</i> with High Distinction in Computer Science• Designed interdepartmental major entitled "Artificial Intelligence Systems," combining computer science, psychology, neuroscience and philosophy to explore interdisciplinary perspectives on AI	3.93 GPA
2015-2016	CS/Philosophy Registered Visiting Student, Oxford University <ul style="list-style-type: none">• Completed 9 courses in computer science and philosophy• Co-founded existential risk research discussion society	3.93 GPA
2011-2017	Part-Time Student, UNC Chapel Hill	4.00 GPA

Publications

Rachel Freedman, Rohin Shah, and Anca Dragan. "Choice Set Misspecification in Reward Inference". In *Workshop on Artificial Intelligence Safety at IJCAI 2020*.

Best Paper Award at AISafety 2020.

Rohin Shah, Pedro Freire, Neel Alex, **Rachel Freedman**, Dmitrii Krasheninnikov, Lawrence Chan, Michael Dennis, Pieter Abbeel, Anca Dragan and Stuart Russell. "Benefits of Assistance over Reward Learning". In *Cooperative AI Workshop at NeurIPS 2020*.

Best Paper Award at CoopAI 2020.

Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John Dickerson, and Vincent Conitzer. "Adapting a Kidney Exchange Algorithm to Align with Human Values". *Artificial Intelligence*, v. 283, 2020.

Also presented at *AAAI 2018*, *AIES 2018*, *MD4SG 2018*, and the *Participatory ML workshop at ICML 2020*.

Outstanding Student Paper Honorable Mention at AAAI-18.

Rachel Freedman. "Aligning with Heterogeneous Preferences for Kidney Exchange". In *Workshop on Artificial Intelligence Safety at IJCAI 2020*.

Honors

2020	Best Paper Award , <i>CoopAI Workshop at NeurIPS 2020</i>
2020	Best Paper Award , <i>AI Safety Workshop at IJCAI-PRICAI 2020</i>
2019	EECS Excellence Award , <i>UC Berkeley</i>
2019	EECS Departmental Fellowship , <i>UC Berkeley</i>
2019	Gates Cambridge Finalist , <i>Gates Cambridge Trust</i>
2018	Outstanding Student Paper Honorable Mention , <i>AAAI 2018</i>
2017	Rhodes Scholarship Finalist , <i>Rhodes Trust</i>

2017	Phi Beta Kappa , <i>Duke University</i>
2017	magna cum laude , <i>Duke University</i>
2017	High Distinction in Computer Science , <i>Duke University</i>
2013-2017	Robertson Scholarship , <i>Robertson Scholarship Leadership Program</i> <ul style="list-style-type: none"> • The Robertson Scholars Leadership Program "invests in young leaders who strive to make transformational contributions to society," awarding four years of full tuition, stipend, and project funding.
2013-2016	Thomas J. Watson Memorial Scholarship , <i>IBM</i>
2013	President's Volunteer Service Gold Award , <i>US Government</i>

Work and Research

2019-pres.	CHAI PhD Student , <i>UC Berkeley (Berkeley, CA)</i> <ul style="list-style-type: none"> • Researching methodologies to allow reinforcement learning agents to robustly infer reward functions by integrating human feedback from different modalities • Advised by Prof. Stuart Russell, also working with Prof. Anca Dragan
2017-2019	Software Engineer , <i>Galatea Associates (London, UK)</i> <ul style="list-style-type: none"> • Designed and developed complex position-keeping and regulatory compliance solutions for financial institutions as part of a small international consulting team
2016-2017	Moral AI Researcher , <i>Duke University (Durham, US)</i> <ul style="list-style-type: none"> • Conducted a year-long independent research project incorporating computer science and psychology methodologies in the field of artificial morality • Adapted a kidney exchange algorithm to align with human values by modeling underlying frameworks in human responses to moral dilemmas • Awarded High Distinction for my thesis research; presented and published this work at AAAI 2018, where it won Outstanding Student Paper Honorable Mention • Published in the <i>AI Journal</i>, poster presented at AIES 2018 and MD4SG 2018.
2015 summer	Software Engineering Intern , <i>Microsoft (Seattle, US)</i> <ul style="list-style-type: none"> • Collaborated with team of interns to build a Skype data visualization service
2014 summer	Independent Educational Psychology Researcher , <i>(Cleveland MS, US)</i> <ul style="list-style-type: none"> • Designed, administered and evaluated a psychological survey characterizing the impact of the Mississippi Summer Arts Institute on student learning strategies
2013-2014	Psychology and Neuroscience Research Assistant , <i>Duke University (Durham NC, US)</i> <ul style="list-style-type: none"> • Evaluated impact of technology-based intervention on at-risk students' numerical cognition, and contributed to a manuscript submitted for publication

Service and Leadership

2020	Program Committee Member , <i>IJCAI Workshop on Artificial Intelligence Safety</i>
2020	AI Ambassador and Mentor , <i>Effective Altruism Global (Berkeley, CA)</i> <ul style="list-style-type: none"> • Advised 18 aspiring AI safety researchers on resources and career paths at the EAG SF 2020 and the EA Student Summit's Fellowship programmes
2015-2016	Co-founder and President , <i>Oxford Existential Risk Society (Oxford, UK)</i> <ul style="list-style-type: none"> • Co-founded society that grew to 25+ students, academics, and community members who met weekly at the Oxford Future of Humanity Institute to discuss academic papers, philosophical arguments, and technologies • Recruited speakers from Machine Intelligence Research Institute, Oxford Future of Humanity Institute, and UT Austin

- Advised students founding similar groups at Princeton and Cambridge

2014 summer **Publicist**, *Mississippi Summer Arts Institute (Cleveland, MS)*

- Volunteered 40-60 hours/week to writing press releases and managing social media to support creative opportunities for rural students ages 5-18

2011-2013 **Tutor**, *John Avery Boys and Girls Club (Durham, US)*

- Tutored at-risk elementary school students weekly
- Designed and implemented a self-confidence intervention based on my social psychological coursework