

# RACHEL FREEDMAN

---

rachel.freedman@berkeley.edu | [github.com/RachelFreedman](https://github.com/RachelFreedman) | <https://rachelfreedman.github.io/>

## Education

---

2019-2025 (expected)	<b>Artificial Intelligence PhD, UC Berkeley</b> I work with the Center for Human-Compatible Artificial Intelligence (CHAI) on AI safety via reinforcement learning, reward modeling, and assistance games. My advisor is Prof. Stuart Russell.	<b>3.93 GPA</b>
2013-2017	Computer Science/Psychology BA, <b>Duke University</b> <ul style="list-style-type: none"><li>• Graduated Phi Beta Kappa and <i>magna cum laude</i></li><li>• Earned High Distinction in Computer Science for my research thesis (now published in the journal <i>Artificial Intelligence</i>)</li><li>• Designed original interdepartmental major entitled "Artificial Intelligence Systems," combining computer science, neuroscience and philosophy to explore interdisciplinary perspectives on AI</li></ul>	<b>3.93 GPA</b>
2015-2016	CS/Philosophy Registered Visiting Student, <b>Oxford University</b> Co-founded "Stop the Clock", an existential risk research discussion society that held weekly meetings at the Future of Humanity Institute	<b>3.93 GPA</b>
2011-2017	Part-Time Student, <b>UNC Chapel Hill</b>	<b>4.00 GPA</b>

## 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 IJCAI 2020 AISafety Workshop.

Rohin Shah, Pedro Freire, Neel Alex, **Rachel Freedman**, Dmitrii Krashennnikov, 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 NeurIPS 2020 CoopAI Workshop.

**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*.

## Work and Research

---

2019-pres.	<b>CHAI PhD Researcher UC Berkeley (Berkeley, CA)</b> <ul style="list-style-type: none"><li>• Research topics include reinforcement learning, active learning, and reward modeling in support of robustly beneficial and safe AI</li><li>• Pivoted from interdisciplinary undergraduate degree to technical graduate program via extensive independent study of foundational mathematics, reinforcement learning, deep learning and advanced robotics alongside coursework and research</li><li>• Contributed technical feedback to reports for the United Nations and the World Economic Forum and Brian Christian's book <i>The Alignment Problem</i></li><li>• Advised by Prof. Stuart Russell, also worked with Prof. Anca Dragan</li></ul>
------------	---

- 2017-2019 **Software Engineer and Consultant** *Galatea Associates (London, UK)*
- Designed and developed complex position-keeping and regulatory compliance solutions for financial institutions
  - Collaborated with international team of consultants, engineers, contractors and stakeholders to manage ever-changing business problems and constraints
  - Learned new tools and languages rapidly as required
- 2016-2017 **Moral AI Researcher** *Duke University (Durham, US)*
- Conducted a year-long independent research project incorporating computer science and psychology methodologies into 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 *Artificial Intelligence*, poster presented at AIES 2018 and MD4SG 2018.
- 2015 summer **Software Engineering Intern** *Microsoft (Seattle, US)*
- Collaborated with a team of interns to build a data processing and visualization service that delivered actionable insights for the Skype for Business team
- 2013-2014 **Psychology and Neuroscience Research Assistant** *Duke University (Durham NC, US)*
- Evaluated the impact of technology-based intervention on at-risk students' numerical cognition, and contributed to a manuscript submitted for publication

## Honors

---

- Conferences & Workshops
- Invited Talk** *Institute for Advanced Study WAM Program (2022)*
  - Best Paper Award** *CoopAI Workshop at NeurIPS (2020)*
  - Best Paper Award** *AI Safety Workshop at IJCAI-PRICAI (2020)*
  - Outstanding Student Paper Honorable Mention** *AAAI (2018)*
- Graduate
- Rising Star in AI Ethics** *Women in AI Ethics (2021)*
  - EECS Excellence Award** *UC Berkeley (2019)*
  - EECS Departmental Fellowship** *UC Berkeley (2019)*
- Undergraduate
- Phi Beta Kappa** *Duke University (2017)*
  - magna cum laude** *Duke University (2017)*
  - High Distinction in Computer Science** *Duke University (2017)*
  - Robertson Scholarship** *Robertson Scholarship Leadership Program (2013-2017)*
  - Thomas J. Watson Memorial Scholarship** *IBM (2013-2016)*
- Other
- Rhodes Scholarship Finalist** *Rhodes Trust (2017)*
  - Gates Cambridge Finalist** *Gates Cambridge Trust (2019)*
  - President's Volunteer Service Gold Award** *US Government (2013)*

## Service and Leadership

---

- 2022-pres **AI Safety Advising Contractor** *80000 Hours (remote)*
- Advise promising early-career individuals on high-impact careers in AI safety
- 2020 **Program Committee Member** *IJCAI Workshop on Artificial Intelligence Safety*
- 2015-2016 **Co-founder and President** *Oxford Existential Risk Society (Oxford, UK)*
- Co-founded the society, which grew to comprise over twenty five students, academics, and community members who met weekly at the Oxford Future of Humanity Institute to discuss academic research on
  - Recruited speakers from Machine Intelligence Research Institute, Oxford Future of Humanity Institute, and UT Austin
  - Advised students founding similar groups at Princeton and Cambridge

## Skills and Hobbies

---

Programming	Java, python, git; extensive practice learning new technologies quickly and proactively via engineering and consulting work
Service	Tutored and mentored at-risk elementary school students regularly for two years in North Carolina (for which I earned the President's Volunteer Service Gold Award), volunteered full-time for several months to support creative opportunities for students in rural Mississippi, mentor young people seeking a career in AI (particularly those from underrepresented backgrounds) on an ongoing basis
Martial Arts	Earned World Taekwondo Federation-certified master (4th degree) black belt in Taekwondo after a decade of training. Trained, taught and competed for 12 years.
Creative Writing	Served as writer-in-residence at Italian artist residency "Arte Studio Ginestrelle." Stories and plays published at Duke, UNC and Oxford University student publications, and performed at Duke New Works Festival in 2014.