Daniel Kasenberg

PhD Candidate, Tufts University

(857) 800 3131
dmk@cs.tufts.edu
http://dkasenberg.github.io/
Advisor: Matthias Scheutz

Education

2015–2021 (expected) **PhD, Computer Science and Cognitive Science**, *Tufts University*, Medford, MA, Advisor: Matthias Scheutz.
Qualifying exam fields: Artificial Intelligence; Statistical Pattern Recognition; Template Metaprogramming

2015–2018 **MS, Computer Science**, *Tufts University*, Medford, MA, Advisor: Matthias Scheutz.
Relevant courses: Moral Psychology; Ethics for AI, Robotics, and Human Robot Interaction; Information Theory; Computational Models in Cognitive Science

2007–2014 **BMath, Applied Mathematics and Computer Science**, *University of Waterloo*, Waterloo, ON, Canada, Excellent standing; Dean's honour list.
Relevant courses: Introduction to Artificial Intelligence; Machine Learning: Statistical and Computational Foundations; Calculus of Variations; Algorithms; Introduction to Computational Mathematics; Logic and Computation

Research Interests

My ultimate goal is to develop **morally competent artificial agents**: agents that can represent, learn, reason about, and follow human moral and social norms, and explain their behavior with respect to these decisions. My current work lies at the intersection of **reinforcement learning**, **logic**, and **machine ethics**: I develop algorithms that allow agents to maximally satisfy multiple (potentially conflicting) goals, to infer others' goals from behavior, and to explain their behavior (in natural language) with respect to these goals, where such goals are represented in **temporal logic**.

Research Experience

2018-2018 **Research Intern**, *DeepMind*.
Developing exploration algorithms for reinforcement learning agents

2015-2018 **Graduate Research Assistant**, *Tufts University*, Human-Robot Interaction Lab.
Applying moral and social norms in stochastic domains

2014-2014 **Undergraduate Research Assistant**, *University of Waterloo*, HCI Lab.
Extracting procedural information from online software tutorials using machine learning

Publications

Conference Papers

[C1] Daniel Kasenberg, Antonio Roque, Ravenna Thielstrom, Meia Chita-Tegmark, and Matthias Scheutz. "Generating Justifications for Norm-related Agent Decisions". In: *Proceedings of the 12th International Conference on Natural Language Generation*. 2019.

[C2] Daniel Kasenberg, Vasanth Sarathy, Thomas Arnold, Matthias Scheutz, and Tom Williams. "Quasi-Dilemmas for Artificial Moral Agents". In: *Proceedings of the International Conference on Robot Ethics and Standards*. 2018.

[C3] Daniel Kasenberg and Matthias Scheutz. "Inverse Norm Conflict Resolution". In: *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. 2018.

[C4] Daniel Kasenberg, Thomas Arnold, and Matthias Scheutz. "Norms, Rewards, and the Intentional Stance: Comparing Machine Learning Approaches to Ethical Training". In: *Proceedings of the First AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. 2018.

[C5] Daniel Kasenberg and Matthias Scheutz. "Norm Conflict Resolution in Stochastic Domains". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

[C6] Daniel Kasenberg and Matthias Scheutz. "Interpretable Apprenticeship Learning with Temporal Logic Specifications". In: *Proceedings of the 56th IEEE Conference on Decision and Control (CDC)*. 2017.

Workshop Papers

[W1] Daniel Kasenberg. "Inferring and Obeying Norms in Temporal Logic". In: *Human Robot Interaction Pioneers Workshop (HRI 2018)*. In press.

[W2] Daniel Kasenberg. "Learning and Obeying Conflicting Norms in Stochastic Domains". In: *Proceedings of the Student Program, 1st AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society*. 2018.

[W3] Thomas Arnold, Daniel Kasenberg, and Matthias Scheutz. "Value Alignment or Misalignment - What Will Keep Systems Accountable?" In: *Proceedings of the 3rd International Workshop on AI, Ethics and Society*. 2017.

Honors and Awards

Doctoral consortia, HRI 2018, AIES 2018

Teaching Experience

Teaching Assistant

2016 **Introduction to Machine Learning**, *Tufts University*, taught by Roni Khardon.

Volunteer Experience

2019–2019 **Organizer**, *RLDM Workshop on Moral Decision-Making (MoDeM)*.

2018–2019 **Web and Publicity Chair**, *HRI Pioneers Workshop 2019*.

2017–2018 **Secretary**, *Tufts Graduate Student Council (GSC)*.

2016–2017 **First-year representative**, *Tufts Computer Science Graduate Student Organization*.