

THOMAS KRENDL GILBERT

2 W Loop Rd, New York, NY 10044 • (510) 388-4841 • tg299@cornell.edu
www.thomaskrendlgilbert.com

PROFESSIONAL POSITIONS

AI Ethics Lead, daios Sept. 2021–Present
Postdoctoral Fellow, Digital Life Initiative, Cornell Tech Sept. 2021–Present

LEADERSHIP & COMMUNITY BUILDING

Industry and Public Policy Workshops

Co-Organizer Dec. 2021–Present

- Inaugural “PERLS” workshop at NeurIPS 2021, the world’s premiere machine learning conference
- “Building Accountable and Transparent RL” at RLDM 2022
- “AI and Humanity” at the 2022 Simons Institute Summer Cluster
- “Designing Technology and Policy Simultaneously” at CHI 2023

Political Economy of Reinforcement Learning Systems (PERLS)

Founder Dec. 2020–Present

- Awarded and served as the inaugural Law and Society Fellow at the Simons Institute (\$16,470) in support of fall 2020 program on “Theory of Reinforcement Learning”
- Created and convened an international reading group on the social impacts of reinforcement learning systems across North America and Europe (130+ active participants)
- Assembled the first ever syllabus on concepts and case studies in the political economy of reinforcement learning, based on regular group meetings with leading AI researchers and policy experts

Ethics of Sociotechnical Systems

Research Lead Aug. 2017–Present

- Formed a community of UC Berkeley doctoral students across computer science, social science, and the Humanities (“GEESE”) to examine the social and political implications of advanced AI
- Published a whitepaper on machine learning’s societal risks with support from the AI Security Initiative at the Center for Long-Term Cybersecurity (\$45,000)
- Received a 2023 Mozilla Technology Fund Award: Auditing Tools for AI Systems (\$50,000) to further develop “Reward Reports”, a form of dynamic documentation for AI capabilities

RECENT PEER-REVIEWED FIRST AUTHOR PAPERS

- "Fleets on the streets: how number, affiliation and purpose of shared-lane automated vehicle convoys influence public perception and blame." *Transportation research part F: traffic psychology and behaviour*, February 2023.
- “Reward Reports for Reinforcement Learning”. Proceedings of the Responsible Decision Making in Dynamic Environments workshop, ICML 2022.
- "Sociotechnical Specification for the Broader Impacts of Autonomous Vehicles." Proceedings of the Fresh Perspectives on the Future of Autonomous Driving workshop, ICRA 2022.
- “Hard Choices in Artificial Intelligence.” *Artificial Intelligence*, November 2021.

SELECTED EDUCATION

PhD in Machine Ethics and Epistemology, University of California-Berkeley Aug. 2021
MA in Sociology, University of California-Berkeley Aug. 2015
MPhil in Political Thought and Intellectual History, University of Cambridge June 2013
BA in Sociology and Philosophy, Northwestern University June 2011