Human Rights and AI: Impacts, Risks, and Opportunities
Fall 2024
Malika Saada Saar
Senior Fellow at the Watson Institute

Welcome to an interdisciplinary study group on human rights and AI, facilitated by the Watson Institute's Center for Human Rights and Humanitarian Studies and the Data Science Institute's Center for Technological Responsibility, Reimagination, and Redesign. We will meet Thursdays from 2:30-4:00 on *9/26, 10/10, 10/24, 11/7, and 12/5*. In addition, I will be available for office hours on the Fridays after our sessions (9/27; 10/11; 10/25, 11/8, and 12/6).

<u>**Study Group Description:**</u>
The Human Rights and AI study group will explore the complex and rapidly evolving landscape of AI and its profound impacts on human rights, at the intersection of technology, public policy, and international law. We will critically analyze both the harms and opportunities that AI poses to fundamental rights. Will these technologies further entrench hierarchies of racial, gendered, and economic privilege? Will AI undermine democratic institutions and the integrity of the electoral process, in the US and globally? Is there such a thing as AI for liberation and justice, if the technology is shaped, designed and deployed from a rights respecting framework? We will wrestle with these questions and examine if and how human rights principles, impact assessments, and due diligence best practices can be applied to ensure the responsible and equitable governance of AI.

Our study group will include guest speakers at the frontlines of these very questions, from the tech, academic, government, and human rights sectors.

<u>**Course Objectives**</u>:
1. Understand the core principles of human rights and how AI can both advance and undermine them.
2. Analyze how AI-driven systems can perpetuate and reify existing structural inequalities and infringe on fundamental rights and freedoms.
3. Evaluate existing international human rights laws, principles, and best practices to interrogate their applicability in AI governance.
4. Develop a critical understanding of the role of technology companies, governments, and civil society in shaping the future of AI and its impacts on human rights.
5. Explore actionable strategies for promoting responsible AI development and deployment that prioritize human rights and equity.

<u>**Study Session 1**</u>
September 26

Introduction to a Human Rights Framework for the Responsible Use and Governance of AI
What exactly is AI? Is it just math? Why should we consider a human rights approach to AI? What are human rights frameworks, and can international human rights laws, treaties, and principles intended to constrain the power of the state from violating fundamental human rights, be applied to tech companies and at the rapid pace development of AI technologies?

- Introduction to Human Rights: Key principles, international legal frameworks, and the significance of a rights-based approach to AI.
- Mapping out the implications, opportunities, and limits of human rights frameworks for the design and deployment of AI.

Study Session 2

October 10

AI Impacts on Human Rights: Algorithmic Bias and Discrimination

Bias and discrimination run through the training, design, and deployment of AI features. Can the harms to fundamental rights of equal protection and nondiscrimination be mitigated or outputs de-risked for bias?

- AI Bias and Discrimination: Examining how biases in AI algorithms can perpetuate and amplify existing societal inequalities, with specific attention to harms of AI bias in areas such as criminal justice, employment, healthcare, and education.
  - Combatting Algorithmic Bias: Exploring due diligence strategies for mitigating bias in AI systems, red teaming, and adversarial testing, mapping out taxonomies of sociotechnical harms.

Study Session 3

October 26

AI Impacts on Human Rights: Surveillance, Deep Fakes, and Algorithmic Wars and the Rights to Privacy, Freedom of Expression, and Assembly

AI-driven technologies like facial recognition, predictive policing, and other AI systems for surveillance intensify the power of the state to control its citizens and potentially undermine democratic processes**.**

- Human Rights at Risk: Analyzing the impacts of AI surveillance on the rights to privacy, freedom of movement, and freedom of expression.
- Deep Fakes and Algorithmic Wars: How use of AI technologies in elections and armed conflicts threatens fundamental rights and freedoms.
- Use of due diligence and human rights impact assessments to identify and mitigate human rights infringements.

Study Session 4

November 7

Human Rights as a Baseline for AI Governance: The Role of Technology Companies, Governments, and Civil Society

How are tech companies, government bodies and civil society seeking to integrate human rights principles into AI design, deployment, and governance? Is it an undeveloped approach, a "human rights washing" or a meaningful attempt to advance internationally agreed norms, laws, and processes to ensure that AI is rights respecting?

- The Responsibilities of Tech Companies: Examining the ethical and legal obligations of the tech sector to apply human rights frameworks and best practices in the AI lifecycle. UN Guiding Principles on Business; Human Rights, US State Department Risk Management Profile for AI and Human Rights, and the Council of Europe Framework Convention on AI, Human Rights, Democracy and Rule of Law
- US and EU Government Regulation of AI: Analyzing existing and proposed AI regulations their effectiveness in promoting human rights to govern AI design and deployment and achieve accountability. Blueprint for an AI Bill of Rights, EU AI Act
- The Role of Civil Society: Exploring the work of human rights organizations, advocates, and researchers in shaping the responsible development and deployment of AI.

Study Session 5
December 5
AI for Humanity: Bending the Moral Arc of AI
Exploring existing efforts to harness AI to address human rights and humanitarian challenges in the areas of:

- Humanitarian Logistics: AI can help optimize the logistics of humanitarian aid delivery, ensuring that supplies are delivered to the right places at the right time.
- Refugee and Displacement: AI can be used to analyze data on refugee flows, identify vulnerable populations, and improve the delivery of aid and services to refugees. It can also be combined with remote sensing and satellite imagery to document war crimes in hard-to-reach settings.
- Optimizing Renewable Energy: AI is used to optimize the performance of solar and wind farms, improving energy efficiency, and reducing reliance on fossil fuels.
- Improving Access to Healthcare: AI-powered clinical decision support tools can help clinicians in low-resource settings better diagnose and manage patients with a variety of illnesses.
- Making Education More Accessible: AI can provide accessibility features for students with disabilities, making education more accessible to diverse learners.

Course Leadership
Malika Saada Saar is a highly accomplished human rights lawyer with extensive experience in civil and human rights law, tech policy development, multi-stakeholder engagement, and the responsible governance and use of AI. As Google's Global Head of Human Rights at YouTube, she led a team responsible for integrating human rights principles across Trust & Safety, Government Affairs and Public Policy, Legal, and Product teams.

Prior to joining YouTube, Malika served as Google's Senior Counsel on Civil and Human Rights, developing civil and human rights frameworks to guide multi stakeholder engagements and operationalizing safeguards to ensure supply chains were not complicit with modern day slavery. In this position, she also convened state and federal lawmakers on Google initiatives to promote democracy, human rights, and protect vulnerable communities, and architected partnerships with UN Women and the UN High Commission on Refugees, and other key global partners.

Earlier in her career, Malika was founder and Executive Director of the Human Rights Project for Girls (Rights4Girls), a legal and advocacy organization dedicated to combating child sex trafficking, child detention, and gender-based violence. As a human rights lawyer and advocate, Malika led efforts to shut down online sex ads that served as the leading sites for child sex trafficking, ended the federal practice of shackling pregnant mothers behind bars in U.S. prisons, and successfully advocated for millions in federal funding for treatment services for at-risk families.

Newsweek and the Daily Beast have named Malika as one of "150 Women Who Shake the World." The Obama White House selected Malika to serve on the Presidential Advisory Council on HIV/AIDS, and she is presently on the Board of Directors for Brown University's Watson Institute for International and Public Affairs and the Peabody's Board of Directors. She has also been honored by Spelman College's National Service Award, Brown University's highest alumni honor, the Roger Williams Award, and the Georgetown Law Center's esteemed Robert F. Drinan Award for Public Service.

Malika holds a B.A. from Brown University, M.A. in Education from Stanford University, and a J.D. from Georgetown University Law Center.