



Governing Artificial Intelligence: UPHOLDING HUMAN RIGHTS & DIGNITY

Mark Latonero

Data&Society

Governing Artificial Intelligence: UPHOLDING HUMAN RIGHTS & DIGNITY

Mark Latonero

EXECUTIVE SUMMARY

Can international human rights help guide and govern artificial intelligence (AI)? Currently, much of society is uncertain about the real human impacts of AI systems. Amid hopes that AI can bring forth “global good” there is evidence that some AI systems are already violating fundamental rights and freedoms. As stakeholders look for a North Star to guide AI development, we can rely on human rights to help chart the course ahead. International human rights are a powerful tool for identifying, preventing, and redressing an important class of risks and harms. A human rights-based frame could provide those developing AI with the aspirational, normative, and legal guidance to uphold human dignity and the inherent worth of every individual regardless of country or jurisdiction. Simply put:

In order for AI to benefit the common good, at the very least its design and deployment should avoid harms to fundamental human values. International human rights provide a robust and global formulation of those values.

This report is intended as a resource for anyone working in the field of AI and governance. It is also intended for those in the human rights field, outlining why they should be concerned about the present-day impacts of AI. What follows translates between

these fields by reframing the societal impact of AI systems through the lens of human rights. As a starting point, we focus on five initial examples of human rights areas – nondiscrimination, equality, political participation, privacy, and freedom of expression – and demonstrate how each one is implicated in a number of recent controversies generated as a result of AI-related systems. Despite these well-publicized examples of rights harms, some progress is already underway. Anticipating negative impacts to persons with disabilities, for example, can lead designers to build AI systems that protect and promote their rights.

This primer provides a snapshot of stakeholder engagement at the intersection of AI and human rights. While some companies in the private sector have scrambled to react in the face of criticism, others are proactively assessing the human rights impact of their AI products. In addition, the sectors of government, intergovernmental organizations, civil society, and academia have had their own nascent developments. There may be some momentum for adopting a human rights approach for AI among large tech companies and civil society organizations. To date, there are only a few, albeit significant, number of examples at the United Nations (UN), in government, and academia that bring human rights to the center of AI governance debates.

Human rights cannot address all the present and unforeseen concerns pertaining to AI. Near-term work in this area should focus on how a human rights approach could be practically implemented through policy, practice, and organizational change. Further to this goal, this report offers some initial recommendations:

- Technology companies should find effective channels of communication with local civil society groups and researchers, particularly in geographic areas where human rights concerns are high, in order to identify and respond to risks related to AI deployments.
- Technology companies and researchers should conduct Human Rights Impact Assessments (HRIAs) through the life cycle of their AI systems. Researchers should reevaluate HRIA methodology for AI, particularly in light of new developments in algorithmic impact assessments. Toolkits should be developed to assess specific industry needs.
- Governments should acknowledge their human rights obligations and incorporate a duty to protect fundamental rights in national AI policies, guidelines, and possible regulations. Governments can play a more active role in multilateral institutions, like the UN, to advocate for AI development that respects human rights.
- Since human rights principles were not written as technical specifications, human rights lawyers, policy makers, social scientists, computer scientists, and engineers should work together to operationalize human rights into business models, workflows, and product design.

- Academics should further examine the value, limitations, and interactions between human rights law and human dignity approaches, humanitarian law, and ethics in relation to emerging AI technologies. Human rights and legal scholars should work with other stakeholders on the tradeoffs between rights when faced with specific AI risks and harms. Social science researchers should empirically investigate the on-the-ground impact of AI on human rights.
- UN human rights investigators and special rapporteurs should continue researching and publicizing the human rights impacts resulting from AI systems. UN officials and participating governments should evaluate whether existing UN mechanisms for international rights monitoring, accountability, and redress are adequate to respond to AI and other rapidly emerging technologies. UN leadership should also assume a central role in international technology debates by promoting shared global values based on fundamental rights and human dignity.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
INTRODUCTION	5
BRIDGING AI AND HUMAN RIGHTS.....	7
A HUMAN RIGHTS FRAME FOR AI RISKS AND HARMS.....	10
Nondiscrimination and Equality.....	10
Political Participation.....	12
Privacy.....	13
Freedom of Expression.....	14
The Disability Rights Approach and Accessible Design	15
STAKEHOLDER OVERVIEW.....	17
Business.....	18
Civil Society.....	20
Governments	20
United Nations.....	21
Intergovernmental Organizations	22
Academia	23
CONCLUSION.....	24
ENDNOTES.....	26
ACKNOWLEDGMENTS	36

Author: Mark Latonero; Research Lead, Data & Society; PhD Annenberg School for
Communication, University of Southern California.



INTRODUCTION

Can international human rights help guide and govern artificial intelligence (AI)? According to the global ethics initiative of the Institute of Electrical and Electronics Engineers (IEEE), the largest organization of technical professionals, the answer is clear. The IEEE's 2017 report on ethically aligned design for AI lists as its first principle that AI design should not infringe upon international human rights.¹ Yet some AI systems are already infringing on such rights. For instance, in March 2018, human rights investigators from the United Nations (UN) found that Facebook – and its algorithmically driven news feed – exacerbated the circulation of hate speech and incitement to violence in Myanmar.² During a US Congressional hearing in April 2018, Senator Patrick Leahy questioned CEO Mark Zuckerberg about the failure of Facebook's AI for content detection in the face of possible genocide against Myanmar's Rohingya ethnic minority. While Zuckerberg initially told senators that more advanced AI tools would help solve the problem, he later conceded to investors that Facebook's AI systems will be unable to detect “hate” in local contexts with reasonable accuracy anytime soon.³

Just a month after Zuckerberg's hearing, the UN's International Telecommunications Union (ITU) hosted its second annual AI for Global Good summit in Geneva.⁴ For many involved in the summit, AI is not just a source of potential risks, it can bring a better future of worldwide benefits. Between these hopes and fears lies an increased sense of uncertainty. As stakeholders look for a North Star to guide AI development, we can rely on human rights to help chart the course ahead. Simply put:

In order for AI to benefit the common good, at the very least its design and deployment should avoid harms to fundamental human values. International human rights provide a robust and global formulation of those values.

In bridging AI and human rights, what's at stake is human dignity.* As an international framework, human rights law is intended to establish global principles (“norms”) and mechanisms of accountability for the treatment of individuals. As such, a rights-based

* The definition of human dignity is contested and its normative value is debated in an extensive literature that is outside the scope of this report. For the present purposes, the term human dignity gestures towards its usage in Western moral philosophy, such as Kant's notions of dignity linked to human autonomy and agency, while acknowledging that dignity has been linked to traditions such as Eastern philosophy as well. This report's usage of human dignity also evokes the United Nation's charter, the Universal Declaration on Human Rights, and the major rights treaties, which link fundamental human rights, the dignity and worth of the human person, and the equal rights of men and women. The interactions between humans and AI may further challenge or refine the concept of human dignity, which is an important topic for future work.

approach provides actors developing AI with the aspirational and normative guidance to uphold human dignity and the inherent worth of every individual, regardless of country or jurisdiction. Implementing human rights can help identify and anticipate some of AI's worst social harms and guide those developing technical and policy safeguards to promote positive uses. Those working on AI accountability can activate the international system of human rights practice – including binding treaties, UN investigations, and advocacy initiatives – to monitor social impacts and establish processes of redress. Importantly, advocates can use human rights to focus attention on power relationships and inequalities that impact vulnerable or marginalized groups around the globe.

IMPLEMENTING HUMAN RIGHTS CAN HELP IDENTIFY AND ANTICIPATE SOME OF AI'S WORST SOCIAL HARMS AND GUIDE THOSE DEVELOPING TECHNICAL AND POLICY SAFEGUARDS TO PROMOTE POSITIVE USES.

Those working on AI commercially might wonder why they should care about human rights. Increasingly, stakeholders are holding the private sector responsible for upholding rights.⁵ In 2011, the UN released a landmark document – *The Guiding Principles on Business and Human Rights* – that calls on industry to respect, protect, and provide remedies for human rights.⁶ These principles can provide AI executives and developers alike with a template for conducting due diligence on human rights impacts. They provide guidelines for how businesses should assume a higher duty of care when developing and deploying their products.⁷

Although a milestone in the field of business and human rights, the *UN Guiding Principles* reflects but a starting point for the application of human rights in the tech sector. Those working directly on AI need regulation, or “hard” laws, along with technical standards, social norms, and market incentives, to effectively incorporate a respect for human rights into their business models, policies, and practices. At the same time, those working on human rights need to be actively engaged in AI governance and monitoring. When necessary, they should be ready to invoke a human rights framework to challenge how AI is developed and deployed by business or government. Civil society and AI developers should work together to help assess risk areas and anticipate the needs of vulnerable groups. Only when stakeholders are working across silos to safeguard against harms can AI systems avoid human rights abuses and advance the enjoyment of human rights.

This report is intended as a resource for anyone working in the field of AI and governance. Anywhere that AI is being researched, developed, or deployed, a human rights frame can identify, anticipate, and minimize an important class of risks and harms. This

work is also intended for those in the human rights field, outlining why they should be concerned about the present and near-term impacts of AI. What follows translates between these fields by reframing the societal impact of AI systems through the lens of human rights.

For those seeking to govern AI – from governments looking to craft regulation to companies looking to self-regulate – this document offers a perspective based on established human rights accountability and norms. The field of human rights has limitations and will certainly not address all the ethical issues arising from AI. Yet it offers a strong value proposition: an approach to AI governance that upholds human dignity based on international human rights law.

The first part of this report, “Bridging AI and Human Rights,” connects the entry points between AI and human rights for governance discussions. Next, “A Human Rights Frame for AI Risks and Harms” reviews a number of current AI risks and harms from a human rights perspective, describing how such rights can be applied. Part three, “Stakeholder Overview,” catalogues the current state of play among stakeholders active in this space, with examples of progress and challenges. Finally, the conclusion discusses the limitations and presents several recommendations for incorporating a human rights approach for AI governance.



BRIDGING AI AND HUMAN RIGHTS

Human rights have only appeared at the periphery of our prominent AI debates.⁸ Both AI and human rights are highly technical fields; to fully digest either would require far more of an exegesis than can be attempted in this report. Instead, we shall draw on basic entry points from both fields to inform AI governance discussions.

Discussions about AI can be fragmented; some people speak of AI colloquially in the popular press or in tech marketing materials, while others speak of concrete methods in scientific proceedings.⁹ Moreover, the nuances of terminology and the speed at which the field is moving can make cross-disciplinary discussions difficult to

have. When considering social and policy implications, it is useful to think of “AI” as a catchphrase for a cluster of technologies embedded in social systems. This includes machine learning, natural language processing, computer vision, neural networks, deep learning, big data analytics, predictive models, algorithms, and robotics—all of which are intrinsically situated in the social contexts where they are developed and deployed.

While some areas of AI remain only theoretical, others, such as machine learning, are already having an impact in real-world contexts.¹⁰ Machine learning systems process large amounts of historical training data and “learn” from these examples to detect patterns that can be useful in decision-making.¹¹ All machine learning algorithms contain some level of statistical bias, which produces incorrect decisions some of the time.¹² However, if the historical data are incomplete or are not representative of a specified population, these biases can scale quickly and inexplicably across AI systems. Such systems can further entrench discriminatory outcomes in people’s lives. How far should we as a society allow machine learning systems to influence human decision-making or even make decisions on their own? These concerns are at the heart of AI debates.¹³

While these questions have yet to be answered, the fact is that today, automated systems are making predictions about human behavior and producing decisions and recommendations that are impacting people in everyday life. These systems are increasingly becoming embedded in a number of social contexts, from policing and judicial sentencing to medicine and finance. We do not know the unintentional impacts or unforeseen consequences of current or future AI systems. As this uncertainty has brought urgent calls to govern AI, we can now turn to the value of human rights.

The field of human rights can be complex for nonexperts. For the purposes of this report, we anchor international human rights law in the drafting and implementation of the Universal Declaration on Human Rights (UDHR) by the United Nations in 1948. The UDHR’s aspirational language established that human rights were grounded in a respect for all individuals that derived from our equal status as bearers of inherent human dignity. This was a response to the “disregard and contempt for human rights,”¹⁴ which precipitated two world wars and the Holocaust. Human dignity and fundamental rights are not tied to country citizenship, legal regime, or socioeconomic position. These rights are universal in the sense that they apply to everyone, everywhere, which provides a *frame* for discussing global AI impact and governance.

Over the last 70 years, human rights proponents have developed the principles of the UDHR into a body of international human rights law that includes nine major human rights treaties; regional rights instruments in the Americas, Africa, and Europe; incorporation in state constitutions and national laws; and customary and case law.¹⁵ Yet because of a divergence in political ideologies and claims to sovereignty, governments enforce international human rights law to wildly varying degrees.¹⁶ Thus, a human rights *framework* has emerged to monitor, promote, and protect human rights. This involves the further development of international human rights law and the interaction

of a diverse network of actors in the UN system, nation-states, international organizations, NGOs, civil society, the private sector, academia, and advocates at the local or individual level.

Those looking for first principles to ground AI governance can use the language of human rights. For example, one of the most hotly debated topics in AI is discriminatory algorithms and systems. This includes empirical research on facial recognition systems that cannot “see” people, particularly women, with darker skin due to a lack of adequate training data or to faulty models and therefore reproduce culturally engrained biases against people of color.¹⁷ Human rights principles of nondiscrimination have been propagated through a multitude of UN treaties, national laws, UN commentary, academic interpretation, and other policies and guidelines. This body of work offers not only a distinct value commitment but also a global perspective on how to identify the impact of discrimination. Equality and nondiscrimination are foundational to practically every major human rights instrument that has emerged from debate and input from representatives from the world’s nations. The development of human rights has its own controversies and politics, but over the last 70 years, international human rights have come to represent shared global values.

Those working on technology policy are faced with the difficult task of deciding what standards, values, or norms to apply in different social contexts. They need to balance the tradeoffs of developing or deploying technologies. They need to understand the potential misuses and abuses, unintended consequences, biases in sociotechnical systems, and even the costs of not deploying a tool when it may help someone in need. Human rights provide those working on AI with a basis for understanding why governing systems – from technical standards to policy – should address values like nondiscrimination in the first place. This is important for tech companies whose products will be used across national borders where laws and values vary.

While it is outside the present scope of this report, an area that demands more foundational work concerns pathways for human rights accountability and remedy when AI harms become manifest. A purely legal, regulatory, or compliance framework would lag behind the velocity of change associated with emerging AI technologies. Thus, other components of the human rights framework, such as UN special rapporteurs, independent investigators, and monitors from civil society are crucial for calling attention to AI risks and harms. As scholars Christiaan Van Veen of New York University (NYU) and Corinne Cath of Oxford Internet Institute state, “Human rights, as a language and legal framework, is itself a source of power because human rights carry significant moral legitimacy and the reputational cost of being perceived as a human rights violator can be very high.”¹⁸ Thus, human rights can provide the link between an AI system’s negative social impact on even one individual in places like Myanmar and the most powerful companies in Silicon Valley.



A HUMAN RIGHTS FRAME FOR AI RISKS AND HARMS

This section reframes a number of well-publicized controversies generated by AI systems. By taking up a human rights lens, we can see how these classes of risks and harms fall within the purview of human rights. We will focus on rights found in the UDHR and the most significant human rights treaties: The International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR), which have been ratified by roughly 170 countries. Together, these three documents make up the International Bill of Rights, which illustrates that human rights are “indivisible, interdependent, and interrelated.”¹⁹

This section provides the foundation for viewing these challenges (and future ones) as not just local problems impacted by individual technologies but as concerns addressable through a framework of universal rights. Intended as a starting point rather than an exhaustive analysis, this section will touch upon five areas of human rights: nondiscrimination, equality, political participation, privacy, and freedom of expression. In addition, we will examine how the rights of persons with disabilities can help us anticipate harms to the human dignity of vulnerable groups and, in doing so, allow us to develop AI technologies to advance human rights.

NONDISCRIMINATION AND EQUALITY

As mentioned above, bias and discrimination have become central topics for those concerned with the governance and social impact of AI systems.²⁰ A number of high profile studies have demonstrated that, as in the case of detecting skin color, certain AI systems are inherently discriminatory. Alarming reports have detailed how discriminatory algorithms are already deployed in the justice system, wherein judges use these tools for sentencing that purport to predict the likelihood a criminal defendant will reoffend.²¹

In *Automating Inequality*, Virginia Eubanks details how government actors implement automated and surveillance technologies that harm marginalized groups.²² Eubanks studied automated systems in the US that discriminated against the poor’s receipt of

government assistance. *Automating Inequality* includes a discussion of the Allegheny Family Screening Tool (AFST), a predictive risk model deployed by the County Office of Children, Youth, and Families to forecast child abuse and neglect. While the AFST is only one step in a process that includes human decision-makers, Eubanks argues that it makes workers in the agency question their own judgment and “is already subtly changing how some intake screeners do their jobs.” Moreover, the system can override its human co-workers to automatically trigger investigations into reports. The model has inherent flaws: it only contains information about families who use public services, making it more effective at targeting poor residents.²³ Such discriminatory effects can lead to harms in other human rights areas, such as education, housing, family, and work.

Some governments are already using algorithmic systems to classify people based on problematic categories. For example, there are reports that the government of China is deploying systems to categorize people by social characteristics.²⁴ This Social Credit System is being developed to collect data on Chinese citizens and score them according to their social trustworthiness, as defined by the government. The system has punitive functions, such as shaming “debtors” by displaying their faces on large screens in public spaces or blacklisting them from booking trains or flights.²⁵

Historically, we have seen how governments’ use of national systems of social sorting along predetermined physical categories can lead to discrimination against marginalized groups. In South Africa, a classification system built on databases that sorted citizens by pseudoscientific racial taxonomies was deployed to implement the racist and violent policies of the apartheid regime. This well-documented case serves as an important cautionary tale for any widespread deployment of AI social scoring systems.²⁶ Without safeguards, even AI systems built for mundane bureaucratic functions can be repurposed to enact discriminatory policies of control.

The importance of equality and nondiscrimination has filtered down through the ratification of treaties to provide the basis for post-war constitutions, state law, and judicial interpretation.²⁷ For example, the South African constitution, adopted in 1996, directly accounted for the discriminatory policies of the past. The constitution establishes equality, human dignity, and human rights as its legal foundations and core values.

There have been attempts to frame discrimination in machine learning algorithms as a human rights issue, as in a recent World Economic Forum (WEF) report that raised both concerns and possible solutions for biased decision-making.²⁸ The report calls for human rights to move to the center of AI discussions:

even when there is no intention for discrimination, ML [machine learning] systems for which success is strictly measured in terms of efficiency and profit may end up achieving these at the expense of a company’s responsibility to respect human rights.²⁹

The report challenges companies to prioritize compliance with human rights standards and to perform rights-based due diligence. Among the recommendations is a call for companies to actively include a diversity of input and norms in systems design. Companies are also encouraged to provide mechanisms of access and redress that make developers responsible for discriminatory outputs.

In May 2018, Amnesty International and Access Now led the drafting of *The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems*.³⁰ The document grounds the current attention on AI bias in binding international legal principles. The Toronto Declaration outlines the responsibilities of both states and private sector actors in respect to the use of machine learning systems, including mitigating discriminatory effects, transparency, and provision of effective remedies to those harmed. It remains to be seen how influential the declaration will become, as organizers are currently in the process of seeking endorsements, particularly from AI companies. Even so, it represents a significant effort to translate fundamental human rights for the AI space.

POLITICAL PARTICIPATION

A report from the Brookings Institution states that “advancements in artificial intelligence and cyber capabilities will open opportunities for malicious actors to undermine democracies more covertly and effectively than what we have seen so far.”³¹ Russian disinformation campaigns through automated bots on social media have been highlighted by researchers as attempts to interfere with the 2016 American presidential election.³² Because AI is being designed to mimic human behavior in online conversations, detecting those online bots that are weaponized to spread disinformation in political discourse could become more difficult. Bots have many useful purposes, including helping search engines find content. Yet those designed for malicious purposes, such as spreading disinformation, have been identified on platforms like Twitter,³³ which undercuts the possibility of an informed citizenry as needed for meaningful democratic elections.

VIEWED THROUGH THIS HUMAN RIGHTS LENS, THE
CO-OPTED USE OF AN AUTOMATED SYSTEM BY
A BAD FAITH ACTOR CREATES A HUMAN RIGHTS
LIABILITY THAT DEMANDS REDRESS.

Mark Zuckerberg’s written submission to the US Congress in 2018 stated how bad-faith actors, in this instance operatives of the Russian government, have been able to manipulate the political process through Facebook.³⁴ Further studies have demonstrated that bots have and continue to be used to manipulate the media in countries across the world to interfere with the outcomes of democratic elections.³⁵

The rights around political participation are referenced, for example, in the right to self-determination and the right to equal participation in political and public affairs in the ICCPR. Viewed through this human rights lens, the co-opted use of an automated system by a bad-faith actor creates a human rights liability that demands redress. Yet finding the right remedy is one of the most contentious areas in platform technology today. Platforms are more likely to remove bots because they violate their terms of service rather than to protect users' right to political participation. Further exploration would be needed to see how human rights principles could inform contractual disputes or litigation in this context.

PRIVACY

Privacy has long been a major concern for a broad field that includes government, business, academia, and civil society organizations. For example, there has been a surge in interest from developers and engineers to follow privacy-by-design³⁶ principles, which demonstrate how norms can be incorporated at the systems-design level. Conducting a privacy impact assessment for technological deployments is an established tool for privacy compliance. Yet we already see tensions around the human right to privacy and AI development.

For instance, Stanford University researchers trained a deep neural network to “predict” the sexual orientation of their subjects, without obtaining consent, using a set of images collected from online dating websites.³⁷ Beyond various methodological shortcomings, the research demonstrated how a disregard for privacy rights increases the risks of algorithmic surveillance, where data that is collected and analyzed threatens to reveal personal information about users. This can put individuals and groups at risk, particularly those living under regimes that would use such information to repress and discriminate.³⁸

IF AI DEVELOPERS TREAT PRIVACY NOT AS AN ETHICAL PREFERENCE BUT AS A FUNDAMENTAL HUMAN RIGHT, IT WOULD STRENGTHEN THE PRIVACY CONSIDERATIONS THAT ALREADY EXIST IN INDUSTRY NORMS AND TECHNICAL STANDARDS.

Another example is Amazon's AI-powered facial recognition software, which was made widely available in 2016. In July 2018, American Civil Liberties Union (ACLU) researchers ran an experiment matching pictures of all 535 members of Congress to a database of 25,000 public images of arrested individuals. Researchers found that the software not only produced 28 false matches but was also racially biased. Since Amazon has sold this software to police departments, the ACLU expressed concern about further use of facial recognition for government surveillance, which is pervasive, opaque, and unregulated.³⁹

If AI developers treat privacy as a fundamental human right rather than an ethical preference, the privacy considerations that already exist in industry norms and technical standards would be stronger.⁴⁰ The right to privacy is found in Article 12 of the Universal Declaration, Article 17 of the ICCPR, and in a number of other human rights documents, national constitutions, and national laws.⁴¹

International human rights law and principles around privacy can help AI developers analyze, identify, and respond to emerging risks. The AI capabilities demonstrated by the Stanford study give a glimpse into how AI can threaten privacy: both through the rampant collection of data and the capacity for de-anonymizing subjects. These concerns have recently been documented in a report by human rights organizations Article 19 and Privacy International, which notes that “AI-driven consumer products ... are frequently equipped with sensors that generate and collect vast amounts of data without the knowledge or consent of those in its proximity.”⁴² The report states that AI can be used to infer sensitive facts from relatively mundane data, learning about people’s emotional states, health, politics, and others from data like location histories and social media interactions.⁴³ Protecting the right to privacy is key to the enjoyment of a number of related rights, such as freedoms of expression, association, political participation, and information.

FREEDOM OF EXPRESSION

The right to freedom of expression is particularly important in an environment wherein social media platforms use algorithms that decide whose voices we hear. In 2014, researchers from Cornell collaborated with Facebook to undertake an emotional-contagion study, examining how emotions spread through the social network. The researchers manipulated the experience of almost 700,000 Facebook users by using a sentiment analysis tool to identify if friends posted negative comments or posts. Those negative posts were then removed from users’ newsfeeds to test whether algorithmically skewing the feed to display positive posts would keep users on the site longer.⁴⁴ This study demonstrates how platforms can make decisions based on users’ expressions that cause the world to appear in certain ways, strengthening one reality while weakening another.⁴⁵

The right to freedom of expression is a cornerstone of fundamental human rights found in Article 19 of both the Universal Declaration and the ICCPR.⁴⁶

As social media platforms become the central place where public discussion happens, there is a strong debate about the role of platforms in content moderation.⁴⁷ With hate speech, false news, and media manipulation circulating on platforms like Facebook and Twitter, legislators and the public are calling for companies to address the problem.

These calls to action are met with concerns about how private companies can meaningfully determine the boundaries of speech. For example, David Kaye, the UN’s special rapporteur on the right to freedom of opinion and expression, has expressed

concern that content moderation systems could, even unintentionally, censor minority opinions and other unpopular yet critical forms of free expression.⁴⁸ The NYU's Center for Business and Human Rights argues for more governance of technology platforms but states that government intervention can also do harm.⁴⁹

The resource-intensive and relentless task of content moderation requires making difficult decisions on both standards and their subsequent application. Kaye goes on to indicate that in this murky environment, putting human rights at the center of this debate offers states, companies, and key stakeholders practical standards to guide content regulation, which would also apply when deploying automated systems.⁵⁰ While this report cannot delve far into this contentious issue,⁵¹ it is important to note that a human rights perspective informs us that few human rights are absolute. The decisions about tradeoffs involve questions around proportionality, that is, balancing the legal and social impact relative to multiple rights. A rights-based frame offers language to analyze the balance between the right to the freedom of expression with rights and freedoms such as political participation, information, assembly, association, privacy, and security.

THE DISABILITY RIGHTS APPROACH AND ACCESSIBLE DESIGN

The Convention on the Rights of Persons with Disabilities, adopted at the UN in 2006, reaffirms that anyone with a disability should be treated with human dignity and be included in the enjoyment of fundamental human rights. Disability rights have become emblematic of how technological development increases the risk to vulnerable groups. They also present a clear opportunity to enhance human rights. The convention has been signed by 161 countries and ratified by 177 countries; it reaffirms rights such as nondiscrimination and establishes principles like universal design and accessibility of communications.⁵² With the backing of international human rights, country-specific regulation, industry standards, and guidance (or pressure) from disability rights activists, AI developers could mitigate the risk of AI's disparate impacts on people with disabilities at the design stage rather than after deployment.

In a 2018 report on new technologies and AI, the Australian Human Rights Commission identified lack of accessibility as a fundamental concern.⁵³ The report observes that almost one in five Australians live with a disability and argues that these citizens' rights are protected under international human rights law and the Convention on the Rights of Persons with Disabilities, which the government of Australia ratified, along with related national laws. The commission lays out a plan that leverages international human rights law, Australian law, nonbinding guidelines, compliance frameworks, accessibility models, and necessary stakeholder consultations to address the issue.⁵⁴

The Convention on Disabilities was partly inspired by the American Disabilities Act (ADA).⁵⁵ A law backed by US enforcement mechanisms, the ADA has become influential for tech industries. Apple, for example, has become a market leader in accessibility features such as automated programs that verbalize on-screen content for the blind.⁵⁶

Yet other companies, such as Netflix, did not automatically comply with the ADA simply because it became law. It took years of advocacy and public pressure from disability rights groups and a lawsuit by the National Association for the Deaf to force Netflix to make its platform accessible by adding closed captioning to its online videos. The industry has also been motivated to change its policies on accessible design because of structured negotiations, a method of dispute resolution between advocates and companies, rather than lawsuits.⁵⁷ This underscores a broader point: Applying human rights to AI governance on an international level entails more than law and regulation. Change needs to come from additional forces, including market incentives, public awareness, local activism, and the technological innovation that would make compliance easier.

THOSE WORKING ON DISABILITY RIGHTS HAVE MADE GREAT PROGRESS – AT LEAST WITH SOME OF THE BIG TECH COMPANIES – AND THIS CAN BE A MODEL FOR HOW PRODUCT DESIGN COULD BE INTEGRATED WITH OTHER KINDS OF RIGHTS.

This precedent illustrates how AI companies and developers can use an inherent respect of human dignity and human rights to act on anticipated harms. By proactively addressing negative impacts, as in the case of accessibility for people with disabilities, developers can take steps to advance human rights. In a similar way, those working on AI can look at any number of human rights issues to anticipate both social risks and benefits—rights like health and education or issues like migration or the protection of marginalized groups in conflict zones like Myanmar. This only scratches the surface of how human rights can be applied to AI systems.

Yet fully integrating a human rights approach to building and maintaining AI systems would require change in tech industry culture and organizations. For example, we would need to see human rights integrated into product and design teams, not just in statements of corporate social responsibility. Human rights values would need to be infused into the workflow of the organization as part of the jobs of employees working on quality assurance, test suites, and product design documentation. Those working on disability rights have made great progress – at least at some of the big tech companies – and this can be a model for how to work with other kinds of rights.⁵⁸

STAKEHOLDER OVERVIEW

Recently, stakeholders have initiated a number of activities at the intersection of AI and human rights. This section provides a snapshot of the current landscape for business, government, intergovernmental organizations, civil society, and academia. The discussion will focus on AI and human rights activity in business, with short examples of activities by civil society, governments, the UN, intergovernmental organizations, and academia.

BUSINESS

WHILE AI “ETHICS” IS THE TOPIC DU JOUR FOR THE TECH INDUSTRY, HUMAN RIGHTS IS BEGINNING TO EMERGE AS AN ADDITIONAL PERSPECTIVE.

Microsoft completed the first Human Rights Impact Assessment (HRIA) on AI for a major tech company, to be released in late 2018.⁵⁹ HRIAs are an established methodology for the business sector that are used to examine the impact of a product or action from the viewpoint of the rights holders—whether they be direct consumers or external stakeholders.⁶⁰ Depending on how transparent Microsoft’s HRIA report will be, others in the industry may learn how to operationalize human rights due diligence for AI products. In addition, new AI-specific approaches, such as algorithmic impact assessments, may inform traditional HRIA practices.⁶¹

Increasingly, advocates and concerned employees have been able to exert meaningful pressure on big tech companies. After political firm Cambridge Analytica surreptitiously gained access to the private data of tens of millions of Facebook users to influence their voting behavior, a number of nonprofits and investment groups sent an open letter to Facebook’s largest institutional shareholders. The letter claimed that the company “is failing to both assess and address longstanding — yet urgent — human rights problems, including critical concerns regarding civil, political and privacy rights.”⁶²

Furthermore, tech workers have organized public campaigns to pressure their employers to stop building technologies and AI that may be used by governments for social harm.⁶³ In April 2018, around 4,000 Google employees sent a letter to their CEO demanding the company cease its contract to participate in an AI development project called Maven with the US Department of Defense.⁶⁴ The letter cited “biased and weaponized AI.” Google later said it planned to not renew its contract with project Maven. In June 2018, Google released a statement of its AI principles, which said it would still work with the defense industry but will not develop AI weapons.⁶⁵ Importantly, Google states that it will not design or deploy AI technologies “whose purpose contravenes widely accepted principles of international law and human rights.”⁶⁶ In August 2018, 1,400 Google employees signed another letter expressing concern about of the company’s work on a search engine in China that would censor content.⁶⁷ While Google’s AI principles indicate some engagement with human rights, both of the Google employee letters appealed to moral and ethical concerns, in vague terms, rather than human rights.

These letters follow a trend of tech employees pressuring their companies to live up to their aspirational missions on hot button issues such as migration and refugees.⁶⁸ In June 2018, Microsoft received a letter from 300 of its employees demanding that it cancel its contract with US Immigration and Customs Enforcement (ICE), citing human rights concerns related to ICE's forced separation of migrant and refugee parents from their children at the border.⁶⁹ In July 2018, the president of Microsoft released a statement calling for government regulation of facial recognition technology with an explicit appeal to protect fundamental human rights such as freedom of expression and privacy. The statement reveals that customer requests for facial recognition have been turned down by the company when the deployments were determined to present a greater human rights risk.⁷⁰ While it is unclear if these determinations were made with a formal HRIA, they underscore the importance of developing a litmus test for whether a given AI-driven technology should be deployed in a specific context based on human rights risks.

ONE OF THE TRAGEDIES OF THE FACEBOOK MYANMAR ISSUE IS THAT THE COMPANY APPARENTLY DID NOT RESPOND ADEQUATELY TO REPEATED ATTEMPTS BY CIVIL SOCIETY, HUMAN RIGHTS GROUPS, AND ACADEMIC RESEARCHERS TO ALERT THE COMPANY THAT HATE GROUPS WERE USING THE PLATFORM TO HARM OTHER USERS.

Some tech companies are already teaming up with civil society organizations to address social harms. One of the tragedies of the Facebook Myanmar issue is that the company apparently did not respond adequately to repeated attempts by civil society, human rights groups, and academic researchers to alert the company that hate groups were using the platform to harm other users.⁷¹ What would it take for Facebook or a similar technology organization to meaningfully “hear” human rights activists and researchers? Far more work is needed to develop a functioning model where tech industry leadership listens and acts on these warning signs. Since the Senate hearings, Facebook stated that it would implement positive changes, such as hiring more staff to monitor the problem, including some with Burmese language skills, and working with local groups to identify hate speech. Nonetheless, a report by Reuters and the Human Rights Center at the University of California–Berkeley Law School has subsequently found that the platform's efforts have not effectively curtailed hate speech in Myanmar.⁷²

CIVIL SOCIETY

WHILE SOME MAJOR INTERNATIONAL HUMAN RIGHTS ORGANIZATIONS ARE STARTING TO FOCUS ON AI, ADDITIONAL ATTENTION IS NEEDED FROM CIVIL SOCIETY ON POTENTIAL RISKS AND HARMS.

It can be difficult for civil society organizations, especially smaller ones in the Global South, to find ways to engage with AI. Organizations in developing countries may therefore see the AI field dominated by powerful countries.

- In 2017, Amnesty International launched an initiative on AI and human rights, noting that “AI is built by humans and it will be shaped by human values. If we build AI systems that are a mirror to our current societies, they will be riddled with the historical biases and inequalities of our societies.”⁷³ Human Rights Watch is also building a program investigating AI impacts.
- The 2017 Global Symposium on Artificial Intelligence and Inclusion in Brazil highlighted the need to foster diverse voices in AI research, development, policy-making, and advocacy.⁷⁴
- The Digital Asia Hub in Hong Kong has been leading AI discussions in the Asia-Pacific and provides a model of research and engagement on topics important for the region.
- The WEF is continuing its council to address the future of human rights and developing a project on preparing civil society to respond to the challenges of digital and emerging technologies such as AI.⁷⁵

GOVERNMENTS

DOZENS OF COUNTRIES HAVE INITIATED NATIONAL STRATEGIES ON AI, YET HUMAN RIGHTS ARE NOT CENTRAL TO MANY OF THESE EFFORTS.

Some governments are seeking more of a regulatory stance, while others are starting to focus on the human rights impact of AI.

- Few states have enacted legislation that would direct the human rights impact of AI, yet the European Union has demonstrated an interest in regulating technology companies with an appeal to rights-based principles. The EU’s General Data Protection Regulation (GDPR) establishes new protections for European citizens’ rights around data protection and privacy, which impacts any organization collecting European residents’ data.⁷⁶

- A Council of Europe study notes that “there is growing concern at the political and public level globally regarding the increased use of algorithms and automated processing techniques and their considerable impact on the exercise of human rights.”⁷⁷ Moreover, in 2018, the Council of Europe’s Commissioner for Human Rights argued for safeguarding human rights in the era of AI, particularly the rights of privacy and equality and freedoms of expression and assembly.
- In June 2018, Canada and France called “for the creation of an international study group that can become a global point of reference for understanding and sharing research results on artificial intelligence issues and best practices.”⁷⁸ Global Affairs Canada’s Digital Inclusion Lab has been leading discussions on AI and human rights, and Canada’s Treasury Board is exploring an Algorithmic Impact Assessment for its procurement practices, which includes how these systems impact individual liberty rights.⁷⁹
- In 2018, the Australia Human Rights Commission launched a project to directly address the human rights impact of AI and emerging technologies, which includes a robust engagement of international human rights law and may serve as a guide for other countries.⁸⁰
- In 2017, New York City passed a law that aims to help ensure that algorithms used by city agencies are transparent, fair, and valid by setting up a task force to make recommendations on algorithmic regulation, transparency, and bias. While these rules apply only to New York and do not appeal to human rights directly, this move to regulate AI may become a model for other cities.⁸¹

UNITED NATIONS

THE UN HAS YET TO SUSTAIN A FOCUS ON AI FROM A RIGHTS PERSPECTIVE⁸² WITH SOME NOTABLE EXCEPTIONS, PARTICULARLY FROM UN INDEPENDENT INVESTIGATORS, SPECIAL RAPORTEURS, AND THE SECRETARY GENERAL’S STRATEGY ON NEW TECHNOLOGY.

- As noted, UN investigators found evidence that Facebook was used to exacerbate hate and violence in Myanmar, despite the platform’s use of algorithms and AI to identify such hate speech.⁸³
- The UN special rapporteur on the right to freedom of opinion and expression David Kaye is expected to present a report in late 2018, which investigates the impact of AI and the responsibilities of tech companies to protect human rights. After his official 2017 visit to the US, UN special rapporteur on extreme poverty and human rights Philip Alston urged actors to give more attention to

the role of automation and robotization on job insecurity, as well as to examine possible solutions like universal basic income.⁸⁴

- The UN has addressed the pressing issue of autonomous weapons. In addition to international humanitarian law of conduct in war, there is a strong human rights dimension, such as the right to security of person for civilians caught in conflict, which needs to be further examined.⁸⁵
- In September 2018, The UN Secretary General released a strategy on new technologies that seeks to align the use of technologies like AI with global values found in the UN Charter, the UDHR, and international law.⁸⁶

INTERGOVERNMENTAL ORGANIZATIONS

INTERGOVERNMENTAL ORGANIZATIONS MAY PLAY AN INFLUENTIAL ROLE, INCLUDING THE ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD), WHICH IS PREPARING GUIDANCE RELATED TO AI FOR ITS 36 MEMBER COUNTRIES.⁸⁷

The OECD's 2011 Guidelines for Multinational Enterprises aligns with the UN Guiding Principles for Human Rights by calling for companies to protect human rights in the countries they operate in, conduct human rights due diligence, and provide a mechanism for accountability. Notably, the OECD guidance requires a system of National Contact Points, for which governments in member countries appoint representatives that will hear grievances about company conduct.⁸⁸ The National Contact Point system is a potential example of a mechanism for redress. Although the OECD produces “soft law” that is non-judicial and nonbinding, it could provide a forum to address human rights impacts that arise from AI deployed by companies.⁸⁹

ACADEMIA

MORE WORK CAN BE DONE TO BRIDGE ACADEMICS IN HUMAN RIGHTS LAW, SOCIAL SCIENCE, COMPUTER SCIENCE, PHILOSOPHY, AND OTHER DISCIPLINES IN ORDER TO CONNECT RESEARCH ON THE SOCIAL IMPACT OF AI, NORMS AND ETHICS, TECHNICAL DEVELOPMENT, AND POLICY.

- Scholars at Harvard University have published a report on AI and human rights for the Canadian government; have called for an urgent human rights agenda for AI; and have argued that any emerging model for AI governance “must be situated in and interact with existing institutional frameworks of applicable laws and policies, particularly human rights.”⁹⁰
- The University of Essex has urged the House of Lords of the United Kingdom Select Committee on AI that “a human rights-based approach should sit at the centre of the development and use of artificial intelligence, enabling a more holistic, consistent, universal, and enforceable approach.”⁹¹
- In June 2018, Stanford University’s Global Digital Policy Incubator held a cross-disciplinary conference on AI, design, and human rights.⁹² The conference included a conversation between the UN High Commissioner for Human Rights and a founder of the Silicon Valley organization OpenAI. The discussion highlighted the need for more translation work between human rights proponents and technologists.



CONCLUSION

Today's AI debates are searching for principles to govern emerging and future technological systems for the common good. If the “good” involves upholding human dignity, then the international human rights system is fit for purpose. If AI researchers, developers, and designers work to protect and respect fundamental human rights, they could open the path for broad social benefit. To disregard human rights would be to close off that path.

IF AI RESEARCHERS, DEVELOPERS, AND DESIGNERS WORKED TO PROTECT AND RESPECT FUNDAMENTAL HUMAN RIGHTS, IT COULD OPEN THE PATH FOR SOCIAL BENEFIT. TO DISREGARD HUMAN RIGHTS WOULD BE TO CLOSE OFF THAT PATH.

Of course, human rights have limitations. While international human rights law and principles cover a broad class of risks and harms, they are not equipped to address all of the known and unknown concerns pertaining to AI. There will be instances where AI systems have negative social impacts that are not identifiable or anticipated in terms of human rights. While the human rights systems are supported globally by deliberative bodies such as the UN and have gained legitimacy through the treaty system and national laws, there are still many critics. The very terms “human rights” and “human dignity” have long histories replete with intense controversies about their intrinsic philosophical and political value.⁹³ And reflecting on our current geopolitical moment, the outgoing UN High Commissioner for Human Rights has warned that the universal rights system is under attack from chauvinistic nationalism that promotes self-interest over the common good.⁹⁴

Even if we bring human rights to the center of AI governance discussions, gaps would remain between rights, principles, design and development, deployment, and usage. Integrating any desirable value into a sociotechnical system is a perennial challenge.⁹⁵ In essence, a human rights approach to AI would need to be fully integrated into the organizational contexts of technologists, academics, and researchers, as well as in the social contexts of users. In doing so, we may also find that AI might influence the evolution of human rights and dignity in the future.

Near-term work in this area should focus on how a human rights approach could be practically implemented through policy, practice, and organizational change. Further to this goal, this report offers some initial recommendations:

- Technology companies should find effective channels of communication with local civil society groups and researchers – particularly in geographic areas where human rights concerns are high – in order to identify and respond to risks related to AI deployments.
- Technology companies and researchers should conduct HRIAs throughout the life cycle of their AI systems. Researchers should reevaluate HRIA methodology for AI, particularly in light of new developments in algorithmic impact assessments. Toolkits should be developed to assess specific industry needs.
- Governments should acknowledge their human rights obligations and incorporate a duty to protect fundamental rights in national AI policies, guidelines, and possible regulations. Governments can play a more active role in multilateral institutions, like the UN, to advocate for AI development that respects human rights.
- Since human rights principles were not written as technical specifications, human rights lawyers, policy makers, social scientists, computer scientists, and engineers should work together to operationalize human rights into business models, workflows, and product design.
- Academics should further examine the value, limitations, and interactions between human rights law and human dignity approaches; humanitarian law; and ethics in relation to emerging AI technologies. Human rights and legal scholars should work with other stakeholders on the tradeoffs between rights when faced with specific AI risks and harms. Social science researchers should empirically investigate the on-the-ground impact of AI on human rights.
- UN human rights investigators and special rapporteurs should continue researching and publicizing the human rights impacts resulting from AI systems. UN officials and participating governments should evaluate whether existing UN mechanisms for international rights monitoring, accountability, and redress are adequate to respond to AI and other rapidly emerging technologies. UN leadership should also assume a central role in international technology debates by promoting shared global values based on fundamental rights and human dignity.

NOTES

- 1 “Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems (version 2),” IEEE, December 2017, http://standards.ieee.org/develop/indconn/ec/ead_brochure_v2.pdf. The IEEE’s “Ethically Aligned Design” provides the organizing ideas for a series of proposed technical standards governing ethical AI. While this document is still in development, it represents over two years of stakeholder consultations and a milestone in bridging technologists to human rights principles. It should be noted, however, that this document is not an official code of conduct or formal code of ethics for IEEE members.
- 2 “Statement by Mr. Marzuki Darusman, Chairperson of the Independent International Fact-Finding Mission on Myanmar, at the 37th session of the Human Rights Council,” United Nations Office of the High Commissioner for Human Rights, March 12, 2018, <https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=22798&LangID=E>.
- 3 “Facebook, Social Media Privacy, and the Use and Abuse of Data,” US Senate Committee on the Judiciary, April 10, 2018, <https://www.judiciary.senate.gov/meetings/facebook-social-media-privacy-and-the-use-and-abuse-of-data>; “Transcript of Mark Zuckerberg’s Senate Hearing,” *The Washington Post*, April 10, 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?utm_term=.b0a2f734d2dc; “Facebook, Inc. (FB) First Quarter 2018 Results Conference Call,” Facebook, April 25, 2018, https://s21.q4cdn.com/399680738/files/doc_financials/2018/Q1/Q1-18-Earnings-call-transcript.pdf.
- 4 “AI for Good Global Summit 2018,” ITU, last accessed August 26, 2018, <https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx>.
- 5 While the reputational risks are clear, there is a debate on whether or not corporations can be held legally accountable under international human rights law. For an overview that suggests that corporations are accountable, see International Human Rights Law and Corporations and Sufyan Droubi, “Transnational Corporations and International Human Rights Law,” *Notre Dame Journal of International and Comparative Law* vol. 6, issue 1 (September 2016), <https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1048&context=ndjicl>.
- 6 *The Guiding Principles on Business and Human Rights*, United Nations, June 16, 2011, https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.
- 7 Other industry guidelines, also known as “soft” law, include the OECD Guidelines for Multinational Enterprises (2011), <http://mneguidelines.oecd.org/guidelines> and “ISO 26000:2010,” International Organization for Standardization, last reviewed and confirmed in 2014, <https://www.iso.org/standard/42546.html>.
- 8 Christiaan van Veen and Corinne Cath, “Artificial Intelligence: What’s Human Rights Got To Do With It?,” *Data & Society Points*, May 14, 2018, <https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5>.

9 See Madeleine Elish and danah boyd, *Situating Methods in the Magic of Big Data and Artificial Intelligence*. *Communication Monographs*, September 19, 2017, <https://www.tandfonline.com/doi/abs/10.1080/03637751.2017.1375130>.

10 See, e.g., Kristian Lum and William Isaac. *To Predict and Serve? The Royal Statistical Society*, October 7, 2016, <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x>; Solon Barocas and Andrew D. Selbst, “Big Data’s Disparate Impact,” *California Law Review* 104, 671 (September 2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899##; Arvind Narayanan, “Translation Tutorial: 21 Definitions of Fairness and Their Politics,” presented at the Conference on Fairness, Accountability, and Transparency, New York, April 18, 2018, <https://www.youtube.com/watch?v=wqamrPkF5kk>.

11 According to the Royal Society definition: “Machine learning systems are set a task, and given a large amount of data to use as examples of how this task can be achieved or from which to detect patterns. The system then learns how best to achieve the desired output. It can be thought of as narrow AI: machine learning supports intelligent systems, which are able to learn a particular function, given a specific set of data to learn from.” *Machine Learning: The Power and Promise of Computers That Learn by Example*, The Royal Society (April 2017): 19, <https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf>.

12 It is important to note that bias in a technical sense is not necessarily negative or wrong. For example, we may want to introduce bias in a model to correct for inequitable outcomes for an under-represented group. Then again, it is difficult to separate statistical bias from social bias in our current debates. For a brief discussion of the nuances of bias in AI discussions see *AI Now 2017 Report*, https://ainowinstitute.org/AI_Now_2017_Report.pdf.

13 A vibrant community of academic researchers and practitioners are focused on fairness, accountability, and transparency. See, e.g., *Proceedings of Machine Learning Research*, vol. 81 (February 2018), <http://proceedings.mlr.press/v81>.

14 Universal Declaration of Human Rights, United Nations, adopted December 10, 1948, <http://www.un.org/en/universal-declaration-human-rights/>.

15 For an overview of regional human rights implementation in the Americas, Europe, and Africa, see David C. Baluarte and Christian De Vos, *From Judgment to Justice: Implementing International and Regional Human Rights Decisions*, Open Society Justice Initiative (November 2010), <https://www.opensocietyfoundations.org/sites/default/files/from-judgment-to-justice-20101122.pdf>.

16 For an overview of the challenges of implementation, see International Institutions and Global Governance Program, “The Global Human Rights Regime,” Council on Foreign Relations, May 11, 2012, Web, August 31, 2018.

17 See, e.g., Joy Buolamwini and Timnit Gebru “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” *Proceedings of Machine Learning Research* 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency; by developing a new face dataset that is “more phenotypically balanced on the basis of skin type than existing benchmarks,” the researchers

were able to demonstrate that the commercially available gender classifying algorithms they tested have the lowest accuracy on darker skinned females. See also Tom Simonite, “Machines Taught by Photos Learn a Sexist View of Women,” *Wired*, August 21, 2017, <https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>.

18 van Veen and Cath, “Artificial Intelligence: What’s Human Rights Got To Do With It?,” *supra* note 8; also see Jason Pielemeier, “The Advantages and Limitations of Applying the International Human Rights Framework to Artificial Intelligence.” *Data & Society Points*, June 6, 2018, <https://points.datasociety.net/the-advantages-and-limitations-of-applying-the-international-human-rights-framework-to-artificial-291a2dfe1d8a>.

19 “World Conference on Human Rights, 14-25 June 1993, Vienna, Austria,” Office of the United Nations High Commissioner for Human Rights, last accessed August 26, 2018, <https://www.ohchr.org/en/aboutus/pages/viennawc.aspx>.

20 *E.g.*, Joshua A. Kroll, et al., “Accountable Algorithms,” *University of Pennsylvania Law Review* vol. 165, issue 3 (2017), https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/; Borocas and Selbst, “Big Data’s Disparate Impact,” *supra* note 10; Frank Pasquale, *Black Box Society* (Cambridge, Harvard University Press, 2015); “Workshop Primer: Ethics and AI,” *AI Now*, July 7, 2016, https://ainowinstitute.org/AI_Now_2016_Primer.pdf; Alex Campolo, et al., “AI Now 2017 Report,” *AI Now*, https://ainowinstitute.org/AI_Now_2017_Report.pdf.

21 Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin, “How We Analyzed the COMPAS Recidivism Algorithm,” *ProPublica*, May 23, 2016, <https://www.propublica.org/article/how-we-analyzed-the-compass-recidivism-algorithm>.

22 Virginia Eubanks, *Automating Inequality* (St. Martin’s Press: 2018); see also Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York University Press: 2018) and Cathy O’Neil, *Weapons of Math Destruction* (Crown Random House: 2016).

23 Eubanks, *Automating Inequality*, chapter 4.

24 “Big Brother Is Watching: How China Is Compiling Computer Ratings on All Its Citizens,” *South China Morning Post International Edition*, November 24, 2015, <https://www.scmp.com/news/china/policies-politics/article/1882533/big-brother-watching-how-china-compiling-computer> and “Notice of the State Council on Printing and Distributing the Outline of the Construction of the Social Credit System (2014-2020),” The State Council, the People’s Republic of China, June 27, 2014, http://www.gov.cn/zhengce/content/2014-06/27/content_8913.htm.

25 Meg Jing Zeng, “China’s Social Credit System Puts Its People Under Pressure to Be Model Citizens,” *The Conversation*, January 23, 2018, <https://theconversation.com/chinas-social-credit-system-puts-its-people-under-pressure-to-be-model-citizens-89963>.

26 Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out — Classification and Its Consequences* (MIT Press: 1999).

27 The Human Rights Council defines discrimination as any distinction, exclusion, restriction, or

preference that is based on any ground, such as race, color, sex, language, religion, political or other opinion, national or social origin, property, birth, or other status and which has the purpose or effect of nullifying or impairing the recognition, enjoyment, or exercise by all persons, on an equal footing, of all rights and freedoms. “CCPR General Comment No. 18: Non-Discrimination,” UN Human Rights Committee (HRC), November 10, 1989, <http://www.refworld.org/docid/453883fa8.html>.

28 *How to Prevent Discriminatory Outcomes in Machine Learning*, World Economic Forum, March 12, 2018, http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf.

29 Ibid.

30 “The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning systems,” May 2018, <https://www.accessnow.org/cms/assets/uploads/2018/05/Toronto-Declaration-DOV2.pdf>.

31 Alina Polyakova and Spencer P. Boyer, “The Future of Political Warfare: Russia, the West, and the Coming Age of Global Digital Competition,” 5, Brookings, March 2018, <https://www.brookings.edu/wp-content/uploads/2018/03/the-future-of-political-warfare.pdf>.

32 Amelia Acker, “Tracking Disinformation by Reading Metadata,” *D&S Media Manipulation: Dispatches from the Field*, July 17, 2018, <https://medium.com/@MediaManipulation/tracking-disinformation-by-reading-metadata-320ece1ae79b>.

33 Jon Swaine, “Twitter Admits Far More Russian Bots Posted on Election Than It Had Disclosed,” *The Guardian*, January 19, 2018, <https://www.theguardian.com/technology/2018/jan/19/twitter-admits-far-more-russian-bots-posted-on-election-than-it-had-disclosed>.

34 “Transcript of Mark Zuckerberg’s Senate Hearing,” *The Washington Post*, supra note 3.

35 Alice Marwick and Rebecca Lewis, “*Media Manipulation and Disinformation Online*,” Data & Society, May 15, 2017, <https://datasociety.net/output/media-manipulation-and-disinfo-online/>; see also Jonathan Ong, “Architects of Networked Disinformation,” Newton Tech4Dev, February 5, 2018, <http://newtontechfordev.com/newton-tech4dev-research-identifies-ad-pr-executives-chief-architects-fake-news-production-social-media-trolling>.

36 See, e.g., Ann Cavoukian, *Privacy by Design: The 7 Foundational Principles*, Information and Privacy Commissioner of Ontario, 2011, <https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf>; “*Privacy and Data Protection by Design – From Policy to Engineering*,” European Union Agency for Network and Information Security, December 2014, <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design>.

37 Yilun Wang and Michal Kosinski, “Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation from Facial Images,” *Journal of Personality and Social Psychology* (preprint), <https://osf.io/zn79k/>.

38 For an analysis and critique, see Jake Metcalf, “‘The Study Has Been Approved by the IRB’: Gayface AI, Research Hype and the Pervasive Data Ethics Gap,” *Pervade Team*, November 30, 2017, <https://medium.com/pervade-team/the-study-has-been-approved-by-the-irb-gayface-ai-research-hype-and-the-pervasive-data-ethics-ed76171b882c>; see also Melanie Penagos, “AI Systems and Research Revealing Sexual Orientation Case Study,” AI and Human Rights Workshop, Data & Society Research Institute, April 26–27, 2018, https://datasociety.net/wp-content/uploads/2018/05/AI-Systems-and-Research-Revealing-Sexual-Orientation_Case-Study_Final_CC.pdf.

39 Jacob Snow, “Amazon’s Face Recognition Falsely Matched 28 Members of Congress with Mugshots,” ACLU, July 26, 2018, <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28>.

40 “7002 – Data Privacy Process,” IEEE Standards Association, last accessed August 26, 2018, <https://standards.ieee.org/develop/project/7002.html>.

41 Article 17 of the ICCPR states: “No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.” “International Covenant on Civil and Political Rights,” Office of the United Nations High Commissioner for Human Rights, adopted December 16, 1966, <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>.

42 “Privacy and Freedom of Expression in the Age of Artificial Intelligence,” 17, *Privacy International*, and Article 19, April 2018, <https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf>.

43 *Ibid.*

44 Adam D. I. Kramer, et al., “Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks,” *PNAS* vol. 111 no. 24 (2014), <http://www.pnas.org/content/pnas/111/24/8788.full.pdf>.

45 Taina Bucher, *If... Then* (Oxford University Press: 2018).

46 Article 19 states: “Everyone shall have the right to hold opinions without interference... Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.” “International Covenant on Civil and Political Rights,” Office of the United Nations High Commissioner for Human Rights, adopted December 16, 1966, <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>.

47 Philip M. Napoli and Robyn Caplan, “Why Media Companies Insist They’re Not Media Companies, Why They’re Wrong, and Why It Matters,” *First Monday* vol. 22, no. 5, May 1, 2017, <http://firstmonday.org/ojs/index.php/fm/article/view/7051/6124>.

- 48 *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, United Nations General Assembly, April 6, 2018, <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx>.
- 49 NYU Stern Center for Business and Human Rights, *Harmful Content: The Role of Internet Platform Companies in Fighting Terrorist Incitement and Politically Motivated Disinformation*, November 3, 2017.
- 50 *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, United Nations General Assembly, April 6, 2018, <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx>.
- 51 See, e.g., Robyn Caplan, Lauren Hanson, and Joan Donovan, *Dead Reckoning Navigating Content Moderation After Fake News* (Data & Society, 2018), <https://datasociety.net/output/dead-reckoning/>.
- 52 Convention on the Rights of Persons with Disabilities and Optional Protocol, United Nations, <http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>.
- 53 *Human Rights and Technology Issues Paper*, Australian Human Rights Commission, July 2018, <https://www.humanrights.gov.au/sites/default/files/document/publication/AHRC-Human-Rights-Tech-IP.pdf>.
- 54 Ibid.
- 55 While the ADA may have influenced the Convention on the Rights of Persons with Disabilities, President Barack Obama signed the Treaty in 2009, but the U.S. Senate has yet to ratify it. “Convention on the Rights of Persons with Disabilities,” United Nations Treaty Collection, status as of August 26, 2018, https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-15&chapter=4&clang=en.
- 56 “Accessibility,” Apple, last accessed August 26, 2018, <https://www.apple.com/accessibility>; Microsoft is also a market leader, e.g., see “AI for Accessibility,” Microsoft, last accessed August 26, 2018, <https://www.microsoft.com/en-us/ai-for-accessibility>.
- 57 Lainey Feingold, *Structured Negotiation: A Winning Alternative to Lawsuits* (American Bar Association: 2016); see also Lainey Feingold, “Shifting from Fear to Motivation when Talking about Digital Accessibility Law,” *24 Accessibility*, December 11, 2017, <https://www.24a11y.com/2017/shifting-fear-motivation-talking-digital-accessibility-law/>.
- 58 Keith Hiatt, personal interview, August 25, 2018.
- 59 “Microsoft Global Human Rights Statement,” Microsoft, last accessed Aug. 12, 2018, <https://www.microsoft.com/en-us/about/corporate-responsibility/human-rights-statement>.
- 60 See “Conducting an Effective Human Rights Impact Assessment Guidelines, Steps, and

Examples,” *Business for Social Responsibility*, March 2013, https://www.bsr.org/reports/BSR_Human_Rights_Impact_Assessments.pdf; *Human Rights Impact Assessment: Guidance and Toolbox*, The Danish Institute for Human Rights, 2016, https://www.humanrights.dk/sites/humanrights.dk/files/media/dokumenter/business/hria_toolbox/hria_guidance_and_toolbox_final_may22016.pdf_223795_1_1.pdf.

61 See Andrew D. Selbst, “Disparate Impact in Big Data Policing,” 52 *Georgia L. Rev.* 109 (2017), <https://ssrn.com/abstract=2819182>; Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker, *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, AI Now Institute (April 2018), <https://ainowinstitute.org/aiareport2018.pdf>.

62 “An Open Letter to the CEOs of Facebook’s Largest Institutional Shareholders” May 2, 2018, https://static1.squarespace.com/static/57693891579fb3ab7149f04b/t/5ae9e204aa4a99634fa0a213/1525277190443/Open+Letter+to+Facebook+Investors_FinalFormat.pdf.

63 Ron Amadeo, “Google Employees Revolt, Say Company Should Shut Down Military Drone Project,” *Ars Technica*, April 4, 2018, <https://arstechnica.com/gadgets/2018/04/google-should-not-be-in-the-business-of-war-googlers-decry-pentagon-project> and “Letter of Microsoft Workers to Satya Nadella,” June 19, 2018, <https://int.nyt.com/data/documenthelper/46-microsoft-employee-letter-ice/323507fcbddb9d0c59ff/optimized/full.pdf#page=1>.

64 Daisuke Wakabayashi and Scott Shane, “Google Will Not Renew Pentagon Contract That Upset Employees,” *The New York Times*, June 1, 2018, <https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html>.

65 Sundar Pichai, “AI at Google: Our Principles,” *The Keyword*, Google, June 7, 2018, <https://blog.google/technology/ai/ai-principles/>.

66 For a further critique of Google’s principles from a human rights perspective, see Lorna McGregor and Vivian Ng, “Google’s New Principles on AI Need to Be Better at Protecting Human Rights,” *The Conversation*, June 15, 2018, <https://theconversation.com/googles-new-principles-on-ai-need-to-be-better-at-protecting-human-rights-98035>.

67 Kate Conger and Daisuke Wakabayashi, “Google Employees Protest Secret Work on Censored Search Engine for China,” *The New York Times*, August 16, 2018, <https://www.nytimes.com/2018/08/16/technology/google-employees-protest-search-censored-china.html?hp&action=click&pgtype=Homepage&clickSource=story-heading&module=first-column-region®ion=top-news&WT.nav=top-news>.

68 Mark Latonero, “Tech Companies Should Speak Up for Refugees, Not Only High-Skilled Immigrants,” *Harvard Business Review*, May 16, 2017, <https://hbr.org/2017/05/tech-companies-should-speak-up-for-refugees-not-only-high-skilled-immigrants>.

69 Letter of Microsoft Workers to Satya Nadella, June 19, 2018, <https://int.nyt.com/data/documenthelper/46-microsoft-employee-letter-ice/323507fcbddb9d0c59ff/optimized/full.pdf#page=1>.

- 70 Brad Smith, "Facial Recognition Technology: The Need for Public Regulation and Corporate Responsibility," Microsoft, July 13, 2018, <https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/>.
- 71 Libby Hogan, "Myanmar Groups Criticise Zuckerberg's Response to Hate Speech on Facebook," *The Guardian*, April 5, 2018, <https://www.theguardian.com/technology/2018/apr/06/myanmar-facebook-criticise-mark-zuckerberg-response-hate-speech-spread>.
- 72 Steve Stecklow, "Hatebook: Why Facebook Is Losing the War on Hate Speech in Myanmar," *Reuters*, August 15, 2018, <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>.
- 73 Salil Shetty, "Artificial Intelligence for Good," Amnesty International, June 9, 2017, <https://www.amnesty.org/en/latest/news/2017/06/artificial-intelligence-for-good/>.
- 74 David Talbot, Levin Kim, Elena Goldstein, and Jenna Sherman, "Charting a Roadmap to Ensure Artificial Intelligence (AI) Benefits All," Berkman Klein Center, November 30, 2017, <https://medium.com/berkman-klein-center/charting-a-roadmap-to-ensure-artificial-intelligence-ai-benefits-all-e322f23f8b59>.
- 75 "Preparing Civil Society for the Fourth Industrial Revolution," World Economic Forum, <https://www.weforum.org/projects/preparing-civil-society-for-the-fourth-industrial-revolution>.
- 76 "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)," EUR-Lex, last accessed August 26, 2018, <https://eur-lex.europa.eu/legal-content/EN/TX-?qid=1532348683434&uri=CELEX:02016R0679-20160504>.
- 77 Council of Europe, *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, March 2018, <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>.
- 78 "Canada-France Statement on Artificial Intelligence," Government of Canada, last modified July 6, 2018, http://international.gc.ca/world-monde/international_relations-relations_internationales/europe/2018-06-07-france_ai-ia_france.aspx?lang=eng.
- 79 Digital Inclusion Lab, Global Affairs Canada, *Artificial Intelligence and Human Rights: Towards a Canadian Foreign Policy* (unpublished draft), 2018; Michael Karlin, "A Canadian Algorithmic Impact Assessment," March 18, 2018, <https://medium.com/@supergovernance/a-canadian-algorithmic-impact-assessment-128a2b2e7f85> (note this is a personal blog post rather than a government statement).
- 80 "Human Rights and Technology," Australian Human Rights Commission, May 22, 2018, <https://www.humanrights.gov.au/our-work/rights-and-freedoms/projects/human-rights-and-technology>.

- 81 Julia Powles, “New York City’s Bold, Flawed Attempt to Make Algorithms Accountable,” *The New Yorker*, December 20, 2017, <https://www.newyorker.com/tech/elements/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable> and “A Local Law in Relation to Automated Decision Systems Used by Agencies,” The New York City Council, <http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>.
- 82 van Veen and Cath, “Artificial Intelligence: What’s Human Rights Got To Do With It?,” supra note 8.
- 83 Statement by Mr. Marzuki Darusman, Chairperson of the Independent International Fact-Finding Mission on Myanmar, at the 37th session of the Human Rights Council, United Nations Human Rights Council, March 12, 2018, <https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=22798&LangID=E>.
- 84 *Report of the Special Rapporteur on Extreme Poverty and Human Rights, United Nations Human Rights Council*, March 22, 2017, http://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/35/26.
- 85 “Pathways to Banning Fully Autonomous Weapons,” United Nations Office for Disarmament Affairs, October 23, 2017, <https://www.un.org/disarmament/update/pathways-to-banning-fully-autonomous-weapons/> and Ted Piccone, “How Can International Law Regulate Autonomous Weapons?,” Brookings Institution, April 10, 2018, <https://www.brookings.edu/blog/order-from-chaos/2018/04/10/how-can-international-law-regulate-autonomous-weapons/>.
- 86 “UN Secretary-General’s Strategy on New Technologies,” United Nations, <http://www.un.org/en/newtechnologies/images/pdf/SGs-Strategy-on-New-Technologies.pdf>.
- 87 The OECD has previously released guidance on big data. See “OECD Guidelines for Multinational Enterprises,” OECD, <http://mneguidelines.oecd.org/guidelines> and “Big Data: Bringing Competition Policy to the Digital Era,” OECD, November 2016, <http://www.oecd.org/competition/big-data-bringing-competition-policy-to-the-digital-era.htm>.
- 88 “Implementing the OECD Guidelines for Multinational Enterprises: The National Contact Points from 2000 to 2015,” OECD, 2016, <http://mneguidelines.oecd.org/OECD-report-15-years-National-Contact-Points.pdf>.
- 89 *OECD Guidelines for Multinational Enterprises: IV. Human Rights*, OECD, 2011, <http://mneguidelines.oecd.org/2011humanrights.pdf>.
- 90 See F. Raso, V. Krishnamurthy, et al., “Artificial Intelligence and Human Rights: Opportunities & Risks,” Berkman Klein Center for Internet & Society Research Publication (forthcoming); Urs Gasser and Virgilio A.F. Almeida, “A Layered Model for AI Governance,” Harvard University, November 2017, <https://dash.harvard.edu/bitstream/handle/1/34390353/w6gov-18-LATEX.pdf?sequence=1>; see separately Matthias Risse, *Human Rights and Artificial Intelligence: An Urgently Needed Agenda*, Harvard Kennedy School, May 2018, <https://research.hks.harvard.edu/publications/getFile.aspx?Id=1664>.

91 Submission to the House of Lords Select Committee on Artificial Intelligence by the Human Rights, Big Data and Technology Project – Written Evidence (AIC0196), The Human Rights, Big Data and Technology Project, September 6, 2017, <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69717.html>.

92 “Human-Centered AI: Building Trust, Democracy, and Human Rights by Design,” Stanford Center on Democracy, Development, and the Rule of Law, <https://cddrl.fsi.stanford.edu/global-digital-policy-incubator/content/human-centered-ai-program-and-schedule>.

93 Christopher McCrudden, *Understanding Human Dignity* (Oxford University Press: 2014).

94 Opening Statement and Global Update of Human Rights Concerns by UN High Commissioner for Human Rights Zeid Ra’ad Al Hussein at 38th Session of the Human Rights Council, United Nations Human Rights Council, June 18, 2018, <https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=23206&LangID=E>.

95 For more on the socio-technical gap, see Mark Ackerman, *The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility*, *Human Computer Interaction*, 2000, <http://web.eecs.umich.edu/~ackerm/pub/00a10/hci.final.pdf>.

ACKNOWLEDGMENTS

The author would like to express his appreciation to the individuals who generously reviewed this report: Aaina Agarwal, Miranda Bogen, Corinne Cath, Tim Engelhardt, Madeleine Elish, Eimear Farrell, Iason Gabriel, Janet Haven, Michael Karimian, Hibah Kamal-Grayson, Maroussia Levesque, Jake Metcalf, Carly Nyst, Maria Sapignoli, Brittany Smith, and Cristiaan van Veen. Keith Hiatt lent his keen insight on technology and disability rights. Zachary Gold made invaluable contributions and Melanie Penagos provided key research support. Patrick Davison provided astute editorial guidance and critical feedback, and the entire team at Data & Society supported this project in innumerable ways. Sue Glueck sparked key ideas that continue to inspire this work.

This report was animated by the ideas generated at the workshop, Artificial Intelligence & Human Rights, held at Data & Society in April 2018. Workshop attendees came from across sectors including, Global Affairs Canada, USAID, New York City government, The Lisbon Council, OECD, United Nations Office of the High Commissioner for Human Rights, Accenture, Microsoft, DeepMind, Google, Facebook, Gensler Research Institute, Carnegie Mellon University, Cornell University, Oxford Internet Institute, Princeton University, New York University, University of California Berkeley, Max Planck Institute, Data & Society, Digital Asia Hub, Global Network Initiative, Business for Social Responsibility, World Economic Forum, Human Rights Watch, Privacy International, Article 19, AccessNow, Amnesty International, IEEE, Open Society Foundations, Ford Foundation, and The Rockefeller Foundation.

Following the workshop, a body of published work emerged from Aubra Anthony, Corinne Cath & Christiaan van Veen, Elizabeth Eagan, Sherif Elsayed-Ali, Jason Pielemeier, Enrique Piracés, and Ben Zevenbergen which was highly instructive. Also foundational were conversations with Dan Bross, Steve Crown, Eileen Donahoe, Hannah Hilligoss, Mark Hodge, Dunston Allison Hope, Alexa Koenig, Vivek Krishnamurthy, Dinah PoKempner, and Filippo Raso. The inclusion of the organizations and individuals above does not convey an endorsement of this report and any errors remain the author's alone.

For further correspondence about this report please contact Mark Latonero at mark@datasociety.net.

DATA & SOCIETY

Data & Society is an independent nonprofit research institute that advances new frames for understanding the implications of data-centric and automated technology. We conduct research and build the field of actors to ensure that knowledge guides debate, decision-making, and technical choices.

www.datasociety.net

@datasociety