

The Promise and Perils of International Human Rights Law for AI Governance

Anna Su

University of Toronto, Canada

Abstract

This article considers the benefits and pitfalls of international human rights law as a component of artificial intelligence (AI) governance initiatives. It argues that (1) human rights law can serve as an authoritative resource for providing definitions to highly contested terms such as fairness or equality, (2) it can be used to address the problem of international regulatory arbitrage, and (3) it provides a framework to hold public and private actors legally accountable. At the same time, the paper considers recent critiques of human rights law and its application to AI governance, such as (1) lack of effectiveness; (2) inability to effect structural change, and finally, (3) the problem of cooptation. The article argues that while there is room for international human rights in the realm of AI governance, we should look to it with tempered expectations as to its promises and limitations.

Keywords: International human rights; AI governance; cooptation; AI nationalism; regulatory arbitrage; AI ethics.

Introduction

The use and deployment of artificial intelligence (AI) presents many challenges for human rights. Consequently, the search for an AI governance framework has led to a relatively recent proliferation of government strategies, corporate ethics codes, engineering design ethics, and international regulatory frameworks that seek to regulate its application. For purposes of this paper, AI—a term that has no consensus definition in technology and policymaking circles¹—refers to a complex information system that approximates behavior commonly understood as requiring intelligence, similar to human behavior, such as pattern recognition, logical reasoning, or language processing. I will use it interchangeably with the terms “algorithmic decision-making” and “AI-based technologies,” which refer to the use of algorithms in decision-making processes such as the delivery of health care and legal services, systems to assist in criminal sentencing or bail decisions, determination of personal credit ratings, assisting in employee hiring and retention, and the provision of medical diagnostics to name a few of its applications. Legal scholar Ryan Calo described the contemporary excitement around AI as stemming from the enormous promise of machine learning,² thanks to faster computers and the accumulation of copious amounts of data. These massive amounts of data fuel and train the algorithms that engage in these abovementioned processes.

There is no doubt that the human rights implications of this technological development are profound. Concerns about bias, lack of fairness, discrimination, the replication or even exacerbation of existing inequalities, and lack of transparency and accountability when it comes to algorithms are, by now, well-documented.³ As AI capabilities become more powerful and sophisticated in the next few years and AI-based technologies move toward broader adoption, there is an unsurprising proliferation of literature on how to regulate, not only its use but also its design. Regulation of AI generally occurs on two

¹ Council of Europe Commissioner of Human Rights, *Unboxing Artificial Intelligence*, 5.

² Calo, “Artificial Intelligence Policy.”

³ Pasquale, *The Black Box Society*; Eubanks, *Automating Inequality*; O’Neil, *Weapons of Math Destruction*; OHCHR, *The Right to Privacy in the Digital Age*.



Except where otherwise noted, content in this journal is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/). As an open access journal, articles are free to use with proper attribution. ISSN: 2652-4074 (Online)

conceptual levels.⁴ On the technical level, AI safety focuses on questions surrounding how AI is built. The Institute of Electrical and Electronic Engineers (IEEE) for example, issued *Ethically Aligned Design* (2019), a crowdsourced statement on general principles that should govern the ethical design and application of AI-based technologies.⁵ Among other things, *Ethically Aligned Design* is intended to ensure that any AI-powered systems will not infringe on internationally-recognized human rights.⁶ Similar proposals that fall under varying rubrics for beneficial AI,⁷ ethical AI,⁸ explainable AI,⁹ and responsible AI¹⁰ are largely technical attempts to promote the values of fairness, privacy, and non-discrimination for physical systems such as self-driving cars or software-based systems used in algorithmic decision-making systems.

The governance of artificial intelligence in contrast, focuses on the institutions and contexts in which AI is built and used. A variety of entities ranging from supranational organizations to national governments to private sector think tanks have issued several statements that set standards and comprise a web of “soft law” on the use and development of AI. Tech companies themselves have come up with a slew of ethical guidelines that they have adopted with regard to their own products. Governments have issued national strategies to guide the development and deployment of AI by companies situated within their jurisdictions. Consider the *OECD Principles on AI*, which intends to “promote artificial intelligence that is innovative and trustworthy and that respects human rights and democratic values.”¹¹ Endorsed by 42 governments, the guidelines address not only the technical makeup but also the broader policy and regulatory environment of AI, albeit confined to a limited geographic region. For instance, one of the recommendations is the creation of an OECD AI Policy Observatory, which would function as a clearinghouse of strategies, policies, and initiatives among the relevant stakeholders.¹²

To be sure, these levels are not entirely distinct. In fact, these levels are more often than not simultaneously addressed within a single policy statement or declaration. The European Commission’s *Guidelines for Trustworthy AI*, for instance, addresses both the technical robustness and safety of AI development as well as its sociopolitical impacts on human rights. In a 2017 speech, Salil Shetty, the secretary-general of Amnesty International, concluded that “there are huge possibilities and benefits to be gained from artificial intelligence if human rights is part of the core design and use of this technology.”¹³ A two-tiered approach to AI regulation along these lines offers distinct focal points for expert and stakeholder discussions without expecting everyone to absorb and understand otherwise inscrutable information while maintaining an open door for necessary interdisciplinary conversations. As Urs Gasser wrote, “the scale, heterogeneity, complexity and degree of technological autonomy of AI systems require new thinking about policy, law and regulation.”¹⁴ Thus, regulating AI can only be effective if these AI-based systems can be analyzed using a layered model that includes the relevant sets of actors at the international, national, and industry levels.

This paper focuses on the latter dimension of AI regulation and examines the place of international human rights law (IHRL) in the existing strategies authored by international bodies, national governments, corporations, and non-profit bodies. Not all of these strategies refer to or include explicit references to human rights law or principles. Most of them are self-adopted ethical guidelines or norms based on a variety of sources to mitigate the risks and challenges of AI-based systems, as well as to identify and take advantage of the opportunities brought about by AI-based systems. Since 2018, academic and policy literature from a number of disciplines has emphasized the importance of a human rights-based approach to AI governance.¹⁵ That means identifying and mapping the risks to recognized human rights, obliging governments to incorporate their human rights obligations in their national strategies, and even applying IHRL itself. To illustrate, in May 2018, a group of academics and civil liberties groups called on states and companies to meet their existing responsibilities to safeguard human rights in the Toronto Declaration.¹⁶ However, save for a few exceptions, what that approach concretely looks like remains a question, as does how to properly assess whether and to what extent it is beneficial to follow that approach in the first place.

⁴ See generally Gasser, “A Layered Model for AI Governance.” Gasser and Almeida offer three (3) layers: namely, the technical, ethical, and social/legal layers of AI governance. This paper only considers two, by combining the ethical/social/legal into a single layer.

⁵ IEEE, *Ethically Aligned Design*. The 2020 update included a chapter entitled “Extended Reality in Autonomous and Intelligent Systems.”

⁶ IEEE.

⁷ Future of Life Institute, “AI Principles.”

⁸ Winfield, “Machine Ethics.”

⁹ Hagras, “Toward Human-Understandable, Explainable AI.”

¹⁰ Peters, “Responsible AI.”

¹¹ OECD, “Recommendation of the Council.”

¹² High-level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*.

¹³ Amnesty International, “Artificial Intelligence for Good.”

¹⁴ Gasser, “A Layered Model,” 5.

¹⁵ For example, see Risse, “Human Rights and Artificial Intelligence”; Latonero, “Governing Artificial Intelligence”; Donahoe, “Artificial Intelligence and Human Rights”; McGregor, “International Human Rights Law”; van Veen, “Artificial Intelligence”; Desierto, “Human Rights.”

¹⁶ Amnesty International, “The Toronto Declaration.”

Legal scholars recently argued that international human rights law offers an appropriate organizing framework in the design, development, and application of algorithms that takes into account human rights in order to promote algorithmic accountability in what turns out to be one of the early and comprehensive attempts to apply IHRL to algorithmic decision-making processes.¹⁷ Others similarly suggest that an international human rights framework provides the most promising set of standards for ensuring that AI systems are ethical.¹⁸ Nathalie Smuha argued that we already need to go beyond calls for a human rights approach and go straight to putting in place the essential ingredients of such an approach.¹⁹ In a speech made after visiting US tech companies, the United Nations (UN) High Commissioner for Human Rights, Michelle Bachelet, stated: “We cannot expect Big Tech to self-regulate effectively, nor do I believe we would want them to. The onus is on both technology businesses and governments—and also civil society—to work together to identify effective and equitable policies.”²⁰

Building on these recent scholarly efforts to flesh out the application of IHRL and examine its possibilities as a governing framework for AI-based technologies, in this paper it is argued that, first, notwithstanding the variety of corporate AI ethics statements, national AI strategies and guidelines proposed by numerous public–private partnerships, these statements leave many gaps that can be uniquely addressed by human rights law. For example, IHRL can serve as an authoritative resource for providing definitions to highly contested terms such as fairness or equality. Second, IHRL can also be used to address the asymmetries or inequalities engendered by so-called AI nationalism and the corresponding problem of regulatory arbitrage. Finally, it provides a workable framework to hold public and private actors legally accountable in ways that are not possible with self-governing corporate ethics codes.

In addition, apart from the possibilities, the paper also considers recent critiques of human rights and how reliance on IHRL could end up reproducing some of the problems it is meant to address. Some of the reasons IHRL as law should not apply to the overall algorithmic cycle draw from the same wellspring of critiques that are addressed to the human rights project in general, particularly its lack of effectiveness and enforcement and the inability to effect structural change. Ultimately, it is not clear whether IHRL could sufficiently address the root problem associated with AI and algorithmic decision-making if we conceive of the latter as an expression of so-called neoliberal managerialization (i.e., a system of public or private governance that prioritizes freedom and efficiency above all others) and thus help entrench neoliberal economic and social policies.²¹ At best, as Samuel Moyn argues, human rights is a powerless companion to the widening gap between rich and poor.²² One reason is that the modern human rights regime—at least as it is practiced today—is arguably not adept at untangling complex social phenomena, including the one in which AI is currently embedded. Related to this is the concern echoed by some legal scholars that the corollary use of AI-powered applications to protect human rights (e.g., monitoring or evaluating compliance) would privilege certain rights that are more susceptible to a quantitative measurement. But more importantly, using IHRL to regulate AI exacerbates what is already prevalent in human rights monitoring circles in which fealty to mandated metrics and indicators can be mistaken for compliance and subsequently elide debate and discussions over underlying substantive issues. This, in turn, transforms the actors—in this case, the companies creating these technologies—into self-regulating subjects, which weakens whatever regulatory bite that IHRL supposedly provides. In any case, the goal is not to claim that there is no room for IHRL in the realm of AI governance—there is—but that we should do so with a clear eye and tempered expectations as to its promises and limitations.

I. The Promise of International Human Rights Law

In some respects, many of the problems and risks associated with the rise of AI and algorithmic decision-making are not new. Various technological innovations in the past have profoundly changed the labor market, for instance. But the use of AI and algorithms offers revolutionary potential in the way our world is presently run and ordered. At its simplest, algorithms are pieces of code that identify likely patterns and generate particular outcomes, often without any human intervention. The biggest benefit of AI—specifically its subset of machine learning—as well as the source of the biggest concern around it is the ability to autonomously learn and produce results. AI and algorithms in the machine-learning context work by identifying patterns using a trove of data. Based on those patterns, they spew results to certain questions. As law and tech scholar Jonathan Zittrain explained, “Provide a neural network with labelled pictures of cats and other, non-feline objects, and it will learn to distinguish cats from everything else.”²³ The problem, of course, is that algorithms, especially the deep learning sort, “learn” inside a black

¹⁷ McGregor, “International Human Rights Law.”

¹⁸ Yeung, “AI Governance by Human Rights.”

¹⁹ Smuha, “Beyond a Human Rights-Based Approach.”

²⁰ OHCHR, “‘Smart Mix’ of Measures Needed.”

²¹ Waldman, “Power, Process”; Alston, Report of the Special.

²² Moyn, “A Powerless Companion.”

²³ Zittrain, “The Hidden Costs of Automated Thinking.”

box, to the extent it would be almost impossible even for its own designers to figure out the process by which it arrived at a result.²⁴ The opacity of these algorithmic processes thus raises concerns that are relevant for the protection of individual rights and the public interest.

It is important not to exaggerate the state of technology when it comes to AI at the moment. The much-touted “strong AI” or “artificial general intelligence” is still very much in the far future.²⁵ Today, when we speak of AI as an umbrella term, we are actually referring to artificial narrow intelligence in which algorithms perform discrete functions however exceptional. Nevertheless, it does not mean that these algorithms and AI-based technologies are not performing important tasks. One of the more prominent examples is the use of algorithmically produced risk assessments to aid judges in criminal sentencing decisions.²⁶ Indeed, in December 2018, the European Commission for the Efficiency of Justice issued guidelines on the application of AI in judicial systems,²⁷ noting that respect for human rights and non-discrimination remain the most fundamental principles governing the use of this technology. It also guarantees that the use of AI in the judiciary should not be in contravention with the European Convention of Human Rights. Another example is the use of automated decision-making in immigration and refugee determination proceedings in Europe and Canada.²⁸ In the United States, algorithms are now being used to support not only sentencing but also bail decisions, predictive policing, employee hiring and retention, credit score evaluations, and medical diagnostics.²⁹ While there is considerable variation in terms of the accuracy of these AI-based applications at the moment—for instance, AI-assisted medical diagnostics or facial recognition are somewhat accurate, while predictive systems pertaining to criminal recidivism or job success fall somewhat on the dubious end of the spectrum³⁰—the ubiquity in the use of AI has prompted a broad interdisciplinary conversation on its human rights impacts in recent years. The scale, opacity, and adaptability of AI-powered systems give rise to understandable concerns that may not have been present with respect to other technological innovations that came before them.

In recognition of these risks and opportunities, the field of AI ethics and governance has emerged and grown exponentially in the last five years. As of April 2020, there have been 167 statements in total.³¹ What is notable is that this field is not only being driven by technology companies, technical professional associations, and civil society organizations but also by national governments and intergovernmental organizations. One glaring reality, however, is the lack of AI-specific regulation that will fill the gaps created by existing rules. Instead, we have a hodgepodge of policy statements, guidelines, national strategies, corporate ethics statements, and principles. This is not an accident. For one, many technology companies (mostly based in the United States) believe self-regulation or market-driven governance is the best approach to AI and all the resulting technologies around it.³² Because of the prevalent notion from both industry and government that law is slow to adapt and, indeed, even tends to stifle innovation,³³ many believe that the industry itself is in the best position to develop the standards and rules that will guide innovation while at the same time taking into account and minimizing risks to the public welfare. This state of affairs has led some to believe that tech companies engage in a kind of ethics-washing, that is, a self-interested adoption of appearances of ethical behavior that includes the formation of ethics councils, hiring of in-house philosophers or ethicists, and the funding of work on “fair” machine-learning systems, which elides a broader questioning and criticism of these systems’ impact on society.³⁴ Consider, for example, the adoption of Google’s *AI at Google* principles. They were only released in June 2018,³⁵ leading many to accuse the company of issuing them only as a reaction to the public backlash to its contract with the US Defense Department involving Project Maven, an AI project that studies imagery with the aim of improving drone strikes in the battlefield.³⁶

At the moment, the European Union remains the clear governmental frontrunner when it comes to advancing the global discussion on the ethical and social implications of AI. The European Commission (EC) in particular emphasizes the importance

²⁴ Zittrain, *The Black Box Society*.

²⁵ See, for example, Bostrom, *Superintelligence*; Brooks, “An Inconvenient Truth About AI.”

²⁶ Chohlas-Wood, “Understanding Risk Assessment.”

²⁷ CEPEJ, “European Ethical Charter.”

²⁸ For Canada, see, for example, The Citizen Lab, *Bots at the Gate*; for Europe, see, for example, Molnar, *Technological Testing Grounds*; Molnar, “Technology on the Margins.”

²⁹ See, for example, Price, “Artificial Intelligence”; Thomas, “Medicine’s Machine Learning Problem.”

³⁰ See, for example, Narayanan, “How to Recognize AI Snake Oil.”

³¹ AlgorithmWatch, “AI Ethics Guidelines Global Inventory,” <https://inventory.algorithmwatch.org/database> (last accessed June 30, 2022).

³² See, for example, Sherman, “Oh Sure, Big Tech Wants Regulation—on Its Own Terms”; Ochigame, “The Invention of ‘Ethical AI,’” Marchant, “Soft Law Governance of Artificial Intelligence.”

³³ See Calo, “Artificial Intelligence Policy”; Wadhwa, “Laws and Ethics Can’t Keep Pace with Technology.”

³⁴ Powles, “The Seductive Diversion of ‘Solving’ Bias.”

³⁵ Pichai, “AI at Google.”

³⁶ Vincent, “Google Promises Ethical Principles”; Shane, “How a Pentagon Contract.”

of ensuring an appropriate and legal framework to strengthen European values. Consequently, in its *Ethics Guidelines for Trustworthy AI*, a high-level EC expert group wrote:

We believe in an approach to AI ethics based on the fundamental rights enshrined in the EU Treaties, the EU Charter and IHRL. Respect for fundamental rights, within a framework of democracy and the rule of law, provides the most promising foundations for identifying abstract ethical principles and values, which can be operationalized in the context of AI.³⁷

Indeed, the EC recently proposed a legal framework in the form of the *Artificial Intelligence Act (AIA)* that would ensure that AI-based systems within the EU, among others, (1) are safe and respect existing law on fundamental rights and (2) enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems.³⁸ Outside the Union, it remains to be seen what similar regulation would look like. It may well be the case that the EU example could serve as a template for other jurisdictions to follow in coming up with their own concrete regulatory efforts. Indeed, some commentators observe that the proposed AIA is Europe's attempt to set a global AI standard as it is intended to extend beyond EU borders. Until that is operationalized, the common reference to the "protection of human rights norms" across these AI ethics codes, national strategies, and statements allows IHRL to serve as a meta-framework for AI governance and regulation. As several researchers have noted, IHRL, in particular, is well-suited for this purpose for the reasons articulated below.

a) Defining the Content of Rights

In a Berkman Klein Center report on AI governance,³⁹ existing AI governance frameworks—both ethical and rights-based approaches—highlight the principles of privacy, accountability, safety and security, transparency, fairness and non-discrimination, human control of technology, professional responsibility, and the promotion of human values. As evident from the report, each of these principles invites a myriad number of interpretations. Consider the example of fairness. Non-discrimination and prevention of bias in the training of algorithms are among the most high-profile concerns surrounding the use of AI-based technologies, particularly the possibility that AI would only replicate existing patterns of bias under the veneer of technological objectivity. But as computer scientist Arvind Narayanan explains, fairness has (at least) 21 definitions—each of which finds itself laden with a specific value and political context.⁴⁰

Thus, the biggest advantage of IHRL compared to existing ethical codes is that the use of terms such as fairness, non-discrimination, and the right to health and equality has been discussed, debated, and developed by a variety of stakeholders in the human rights community, ranging from local grassroots organizations that advocate on behalf of particular rights to regional human rights tribunals such as the European Court of Human Rights where judges provide authoritative interpretations. At the very least, the universe of potential definitions is circumscribed to an intelligible degree, even though the level of detail may still not satisfy its most ardent critics. At the same time, these principles are malleable and flexible enough that they have been adapted in a variety of local settings through regional and domestic court interpretations. That is, they have gained a kind of acceptance through local adoption and interpretation. This addresses understandable concerns about the lack of universality of these rights; that is, they are some form of continued imposition of Western hegemony. For instance, legal scholar Evelyn Douek notes that, in the context of online content moderation, one of the main criticisms that global platforms have encountered has been their lack of attention to different demands of varying contexts.⁴¹ Moreover, and perhaps more importantly, this circumscribed universe of definitions carries an air of legitimacy as well as normative authority that none of these corporate ethical codes could hope to possess. As former UN special rapporteur on freedom of expression David Kaye noted, "it would offer a globally recognized framework for designing [the] tools [to accommodate varied interests] and a common vocabulary for explaining their nature, purpose and application to users and State."⁴² Whereas ethical codes are more often than not formulated in-house within corporate organizations with minimal consultation from the communities these applications will most likely affect, the content of these human rights has been the product of tension and engagement between and among its stakeholders for a number of years. The most important consequence of recognizing these rights, as opposed to more ambiguous ethical norms, is that it would present some clarity about the necessary threshold that would trigger obligations and thus help designers of these AI-powered applications *ex ante* or even its users *ex post* as they hold companies accountable.⁴³

³⁷ AI HLEG, "Ethics Guidelines for Trustworthy AI"; see also Muller, *The Consequences of Artificial Intelligence*.

³⁸ European Commission, "Proposal for a Regulation."

³⁹ Fjeld, "Principled Artificial Intelligence."

⁴⁰ Narayanan, *21 Fairness Definitions and Their Politics*.

⁴¹ Douek, 27

⁴² Kaye, *Report of the Special Rapporteur*.

⁴³ This is, for example, the main argument in Yeung, "AI Governance by Human Rights."

b) *The Problem of AI Nationalism*

The term ‘AI nationalism’ was first coined by English investor Ian Hogarth at a 2017 conference on how to make AI serve democratic ends. In an extended essay on the subject, Hogarth argued that “rapid progress in machine learning will drive the emergence of a new kind of geopolitics.”⁴⁴ That states recognize this is evident from the plethora of national AI policy statements and strategies to claim and stake their positions. Not that different from a nuclear arms race, world leaders recognize how progress in AI has acquired strategic significance except that geopolitical leadership in the AI race is not merely about spearheading technological advances but also laying out the normative values that will end up encoded in these applications and thus shape the way how they work. Therefore, governments understandably seek to advance the interests of companies headquartered within their territory, such as Baidu in China or Google in the United States.⁴⁵ Russian leader Vladimir Putin once remarked that “whoever became the leader in the field would rule the world.”⁴⁶ Similarly, in separate interviews with *WIRED* magazine, then US president Barack Obama and French President Emmanuel Macron acknowledged the importance of AI as a tool of soft power.⁴⁷ Macron, in particular, pointed to the two contrasting models of AI development in the United States and China and emphasized that if France wants to defend its way of dealing with privacy, its preference for individual freedom and to have a say in the development of the AI revolution, it has to be among the leaders.⁴⁸ Indeed, the French national strategy, for instance, emphasizes that AI must help us promote our fundamental rights, improve social cohesion and strengthen solidarity.⁴⁹ No wonder China, already the world leader in AI research in terms of the number of research papers produced on deep learning, views AI as a key component of its “Made in China 2025” industrial transformation program.⁵⁰ As Macron alluded to in his *WIRED* interview, Chinese technology companies are far more coupled to state policy than those in the UK or the United States. But it is still telling that China’s first ever national statement on AI ethics also referred to respecting “human rights and the fundamental interests of humankind” as one of the fundamental ethical norms. It also emphasized human control and oversight of AI-based systems in order to promote fairness, privacy and safety as well “improve human well-being.”⁵¹

In this scenario, if AI nationalism is a kind of AI arms race in terms of state investment in and promotion of technology development in pursuit of political independence, it also means that the states that engage in this race in varying capacities would not necessarily have the same incentives to come up with a shared set of rules that will govern the creation and use of such technology. Of course, there is nothing about this scenario that ordains it to be viewed as an arms race.⁵² Indeed, it does no one any favors if it continues to be viewed as such. Furthermore, it is far from clear what exactly constitutes “winning” in AI, but a divide between AI-capable and AI-non-capable states is not such a farfetched scenario and certainly one that can have important and long-lasting human rights consequences.

International human rights law can be useful in preventing this help prevent and fill the gap as a shared, coordinating meta-framework to address the existing asymmetries in the governance of AI-based technologies and thus address, even if only partially, the problem of regulatory arbitrage.⁵³ It should be considered part and parcel of what is now being termed digital constitutionalism—that is, a “constellation of initiatives that have sought to articulate a set of political rights, governance norms and limitations on the exercise of power on the Internet.”⁵⁴ If the global AI revolution is not just about developing the technology but also establishing and shaping the normative values behind it, then it is clear that IHRL should be given an important role in ensuring that the rights of all affected, regardless of one’s location in the world. For one, while self-imposed ethics codes certainly have a role to play in regulating the creation and use of AI,⁵⁵ they do not serve as an adequate oversight mechanism as debates about fairness, accountability, and transparency do not necessarily encompass all the fundamental issues about AI. Many of them also only mention values such as privacy and human dignity or extol the values of responsible or ethical AI in the most general way possible so as to render them almost meaningless.

⁴⁴ Hogarth, “AI Nationalism.”

⁴⁵ See, for example, Moore, “Trump’s Techno-Nationalism.”

⁴⁶ Vincent, “Putin Says”.

⁴⁷ Dadich, “Barack Obama Talks AI”; Thompson, “Emmanuel Macron Talks to WIRED.”

⁴⁸ Thompson, “Emmanuel Macron Talks to WIRED.”

⁴⁹ Thompson, “Emmanuel Macron Talks to WIRED.”

⁵⁰ Li, “Is China Emerging.”

⁵¹ Yi, “The Ethical Norms.” The official Chinese text is available from http://www.most.gov.cn/kjbgz/202109/t20210926_177063.html. See also UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, adopted November 23, 2021

⁵² Pascal, “Artificial Intelligence.”

⁵³ See generally High-level Panel on Digital Cooperation, *The Age of Digital Interdependence*; Beduschi, “International Migration Management.”

⁵⁴ Redeker, “Towards Digital Constitutionalism?” For a more EU-centric elaboration, see De Gregorio, “The Rise of Digital Constitutionalism.”

⁵⁵ Ayling & Chapman, “Putting AI Ethics to Work”; Jobin, *The Global Landscape*.”

Moreover, many of the ethics statements and guidelines were issued in Europe and North America. In a recent study on the global landscape of ethics guidelines, it was reported that African and Latin American countries are not represented.⁵⁶ While that might seem unsurprising as most technology companies are based in the West, it bolsters the view that this scenario is but an instance of what a sociologist has termed “digital colonialism.”⁵⁷ If we are to conceive of the online ecosystem as a global public square,⁵⁸ it is unconscionable to exclude the voices of those who will be disproportionately impacted in the event of any adverse consequences, especially given existing inequalities in terms of access to any remedies and resources in the Global South.⁵⁹ IHRL provides an avenue by which these voices can be or are given a seat at the table and amplified through its institutions and mechanisms, as well as shape the definitions and parameters of these rights.

To be sure, it is not enough to say that IHRL applies to the use and design of each AI-based or AI-powered technology. In fact, its actual application is far from straightforward. These general principles have to be further translated into particular rules, processes, and procedures tailored to a particular technology or application, for instance, online content moderation,⁶⁰ facial recognition,⁶¹ or medical diagnostics.⁶² But there has to be a necessary convergence at some level between platform-specific rules and the principles and values as well as the specific case law articulated throughout the international human rights regime. Otherwise, the existing asymmetries will create a distressing path dependency for future attempts at regulation.

c) *Providing a Framework for Public–Private Accountability*

Unlike other kinds of revolutionary technologies, such as nuclear technology, the race to develop AI and AI-powered applications is not a state monopoly. But it is not solely a private endeavor either. As discussed in the previous section, state involvement in AI as a regulator, investor, or promoter of so-called national champions occur on a spectrum. In places such as the United States, the UK, and Canada, AI development is still largely spearheaded by the private sector. Google, for instance, still dominates the global research output related to AI. Most of the funding for AI startups in the United States is raised through private investors, with a whopping US\$7.4 billion dollars raised in the second quarter of 2019 alone.⁶³ By contrast, there is much more cooperation between the state and private companies in China. In 2017, Chinese government entities have announced billions of dollars in AI investment, while Chinese AI startups have received 48 percent of global AI venture funding, outpacing the United States for the first time.⁶⁴

Far from just a case of a government supporting or championing its own companies, the inverse is also true. The Chinese government’s use of research developed jointly by US and Chinese researchers, which has evolved into commercial products, has attracted scrutiny. SenseNets, for example, is part of China’s national surveillance program intended to fight crime and prevent disasters. Its main products, which focus on facial recognition, crowd analysis and human verification, have been key in the construction of what amounts to a mass surveillance laboratory in the province of Xinjiang, where millions of minority Uighur Muslims have been sent for “re-education.”⁶⁵ In February 2018, MIT had forged a research partnership agreement with SenseTime, which held a 49-percent stake in SenseNets. The corporate structure surrounding SenseNets also involves robust exchanges of expertise and resources between the United States and China, thus giving rise to moral and dual-use national security questions. In other words, there might be circumstances in which private companies are pursuing projects that are in tension with or even outright opposition to the goals of the state in which it is headquartered.

International human rights law is an existing framework that addresses the responsibilities of both states and private actors, such as transnational corporations. Even though IHRL is primarily addressed to states, there have been several efforts to oblige private actors to ensure that their products and services do not violate accepted human rights norms such as those enumerated in the international Bill of Rights, the International Covenant on Civil and Political Rights, as well as the International Covenant on Economic, Social and Cultural Rights. In 2011, UN Special Representative John Ruggie issued the *Guiding Principles on Business and Human Rights*, which articulated the distinct roles of states and business enterprises in the protection of human rights. The *Guiding Principles*, in particular, operate on a three-pillar framework referred to as “Protect, Respect and Remedy,” whose purpose is to prevent, mitigate, and redress business-related human rights abuses.⁶⁶ In particular, it expects corporations

⁵⁶ Jobin, “The Global Landscape”

⁵⁷ Kwet, “Digital Colonialism.”

⁵⁸ Hamilton, “Governing the Global Public Square.”

⁵⁹ Dias Oliva, “Content Moderation Technologies.”

⁶⁰ See, for example, Sander, “Freedom of Expression.”

⁶¹ Murray, “Using Human Rights Law.”

⁶² See, for example, World Health Organization, “Ethics and Governance.”

⁶³ Arnold, “Tracking AI Investment: Initial Findings from Private Markets.”

⁶⁴ Loucks, “Future in the Balance?”

⁶⁵ Doffman, “China Is Using Facial Recognition”; see also Human Rights Watch, “China’s Algorithms of Repression.”

⁶⁶ Ruggie, Report on the Issue of Human Rights.

to “avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur,” as well as “seek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts.”⁶⁷

Even though the *Guiding Principles* are not legally enforceable, corporations have largely operationalized them insofar as they have adopted human rights due diligence processes and impact assessment mechanisms in internal corporate procedures. Admittedly, many of these actions have been voluntary in nature. A recent attempt to build on the *Guiding Principles* and create a legally binding instrument to impose direct liability on and hold multinational enterprises accountable for human rights violations is ongoing.⁶⁸ Further, the new *Hague Rules on Business and Human Rights* provide for arbitration as a private dispute resolution mechanism that focuses on human rights issues.⁶⁹ Whatever the result of these efforts, it is clear that these are unfolding amidst heightened attention toward the human rights impact of transnational corporate activities, especially on vulnerable communities around the world. Technology companies are no different from other kinds of multinational businesses in this regard. As far as access is concerned, existing domestic, regional, and international human rights institutions provide already familiar fora for vulnerable groups and their representatives to navigate and make themselves heard.

Moreover, even with the aforementioned enforcement problems, IHRL, or to be more specific, a domestic legal regime that is informed by IHRL, is nonetheless superior to a corporate ethics-driven system, as discussed in a previous section in part because corporate ethics are either too general and vague or encapsulate a narrow view of ethics that do not take into account the complexity and diversity of the world that AI-based technology would serve and affect. In other words, corporate ethics, while important, do not go far enough. Ethics statements represent a corporation’s voluntary attempt to establish ethical boundaries for its products and services. Another more cynical view of these statements is that they represent corporate attempts at ethics theater. That is, a corporation might adopt a practice that it pays mere lip service to or even fabricate its interest in observing responsible practices when it comes to developing AI technologies. In any event, while the incorporation of these ethical principles, such as fairness or explainability or concern for user privacy,⁷⁰ is important and perhaps even ultimately beneficial to the consumer, it operates largely on the technical level of AI governance. On its own, more often than not, it does not concern itself with broader, structural issues. Perhaps one exception is Cisco Systems, one of the world’s biggest manufacturers of networking hardware, and its *Human Rights Position Statements*, which unusually catalog the human rights impacts of AI in accordance with specific provisions of the Universal Declaration of Human Rights. According to the statement, well-designed AI, for instance, can help companies address challenges such as hate speech and child exploitation imagery, although it can also put at risk an individual’s right to freedom of expression.⁷¹ More importantly, it lays out concrete steps that the company is taking in order to mitigate or address those risks, including facilitating increased dialogue between the technical professions and the human rights community. At the least, this is a good illustration of how corporate ethics statements and IHR can be mutually reinforcing.

Finally, IHRL not only can help define terms but, more importantly, engage in the organized coordination of an increasingly complex web of corporate ethics statements, national AI strategies, as well as more comprehensive AI governance principles issued by private think tanks and public–private partnerships. It provides an overarching framework for redress and accountability, as well as the necessary institutions that have the power to investigate, monitor, and address human rights issues that arise from the use of AI-powered applications. Making room for IHRL thus obviates, even if it cannot entirely address, many of the problems associated with the privatization of AI governance in particular and digital governance in general.

II. Why Not International Human Rights Law?

Given the ubiquity of the calls for the application of IHRL or a human rights approach and the apparent necessity for these calls, it might seem odd to consider the question of the limitations of IHRL. But it is important to acknowledge first and foremost that IHRL is not a panacea for any ills that relate to the use of AI-based technologies. Even as we recognize that IHRL is an integral piece of digital constitutionalism, it will be helpful to consider the ways in which IHRL will not get us to the desired outcome in matters of general AI governance and thus temper our expectations as to its promises.

⁶⁷ OHCHR, *Guiding Principles*, annex para. 13.

⁶⁸ OIEGWG, *Legally Binding Instrument*.

⁶⁹ Center for International Legal Cooperation, *The Hague Rules*. For one type of critique, see Columbia Center on Sustainable Investment, “Re: Elements for Consideration.”

⁷⁰ See, for example, IBM Design for AI, “Everyday Ethics for AI.”

⁷¹ Cisco, “Cisco Human Rights Position Statements.”

a) Problem of Enforcement

As one might have surmised from the voluntary nature of the human rights obligations of these transnational corporations, enforcement of these obligations can be a challenge. Despite lip service paid to the importance of incorporating human rights in their business practices, it is not unlikely to expect that corporations will prioritize their commercial interests over human rights obligations in cases of conflict. So, what happens in the case of a breach? Domestic litigation is the most familiar remedy. In the United States, there is, at present, a small body of case law where public interest groups or lawyers have sued the government for the deployment of AI-based systems.⁷² For example, in *Arkansas Dept of Human Services v. Ledgerwood*, the State of Arkansas introduced an algorithm that drastically reduced nursing care in a disabled home in order to save money. As a result, severely disabled Medicaid recipients were left alone without access to food, toilets, and medicine for hours on end. A federal court enjoined the use of the algorithm. In the UK, a British court recently ruled that the use of facial recognition technology (using datasets trained by AI) by police has violated human rights and data protection laws, although it did not suspend the use of all such technology.⁷³

But this remedy is unavailable outside of jurisdictions with robust legal systems or when the effects of these applications are felt in places where legal remedies are illusory. The truth is that there is no global governance framework for the use of these kinds of technologies, save for voluntarily-adopted corporate ethics codes. IHRL is also quite limited in this regard. Beyond an ability to stigmatize, international enforcement is nonexistent, although regional protection schemes such as the European human rights regime might fare better. For example, if it is the case that some technology companies have been proven to be engaged in systematic human rights violations, say, SenseNets in the case of facial recognition to police Uighur Muslims in the Xinjiang Province of China, then there is no clear legal remedy, save for mobilizing public opinion against the practice or the technology. Another example is the use of automated technologies in the border control and migration management context, which subjects mostly stateless persons and refugees to human rights violations, particularly in providing advanced tools to strengthen existing unlawful border control practices.⁷⁴ In November 2020, the UN Special Rapporteur on Contemporary Racism and Xenophobia published a report on how these technologies are being deployed to advance xenophobic and racially discriminatory ideologies.⁷⁵ The report noted a “trend in immigration surveillance where predictive models use AI to forecast whether people with no ties to criminal activity will nonetheless commit crimes in the future.”⁷⁶ Finally, another report by the UN Special Rapporteur on Extreme Poverty released in October 2019 highlighted the use of AI-based applications in facilitating a move toward “a detached bureaucratic process and away from one premised on the right to social security”⁷⁷ in the administration of social welfare programs. It noted that the World Bank and regional development organizations have promoted the use of digital identity documents as part of the embrace of the digital welfare state without taking into account privacy and cyber-security concerns. In these cases, especially in countries located in the Global South, the problem of international enforcement, as in the previous examples, will appear once again.

This is not to suggest that IHRL is completely ineffectual but rather just more limited than widely assumed in its ability to bring about particular desired outcomes. As renowned human rights lawyer and scholar Philip Alston wrote:

we need to reflect on how better ensure effective synergies between international and local human rights movements. ... There will be times when only international groups can function effectively; but there will also be situations in which exclusively international advocacy will be ineffective and perhaps counterproductive.⁷⁸

IHRL provides both with the necessary vocabulary and institutions to carry out their respective advocacies. Whether this suffices remains to be seen.

b) Inability to Effect Structural Change

The inability of human rights to effect structural change has been a subject of perennial academic scholarship. A constant feature of this critique is the emphasis of human rights and IHRL by extension on the redress of individual rights, thus eliding and, in some cases, even exacerbating systemic problems such as material inequality.⁷⁹ Particularly, discussions of inequality in IHRL are often confined within socioeconomic rights litigation and the resulting outcomes. Likewise, the problem of racism

⁷² See Richardson, *Litigating Algorithms 2019 US Report*.

⁷³ Kirka, “UK Court Says Face Recognition.”

⁷⁴ See, for example, Molnar, *Technological Testing Grounds*.

⁷⁵ Achiume, *Contemporary Forms of Racism*.

⁷⁶ Achiume, paragraph 53.

⁷⁷ Achiume, paragraph 50.

⁷⁸ Alston, “The Populist Challenge to Human Rights.”

⁷⁹ Moyn, *Not Enough*; Marks, “Human Rights and Root Causes.”

has been narrowed down to the harm of racial discrimination, which reflects the impoverished state of contemporary discourse.⁸⁰

In the context of the human rights and AI discussion, it is easy to see how one can fall into a similar trap by looking at ways the current literature focuses on the individual human rights impact of AI-based applications. To be sure, one can start from the principle that an individual right is fundamental, say, the right to health, and use that as a basis to construct a human rights-informed framework, say, building institutions and accountability mechanisms to promote and protect that right. More often than not, however, the language and structure of human rights itself—and this goes as well for the abovementioned critique on the lack of effectiveness—hampers efforts that are aimed at systematic changes. The way that human rights litigation at both the domestic and regional court levels has unfolded in different parts of the world, for example, with respect to the right to medicine,⁸¹ has, indeed, for the most part, ignored the larger structural forces that eventually undermine the needs of the global poor. Studies have shown, for example, that social rights enforcement in the Global South has tended to benefit middle- or upper-class groups rather than the poor.⁸² Part of the problem is that human rights—as law, discourse, or practice—has evolved to be understood as premised on the free market assumption that IHRL is intended to protect the inherent dignity and autonomy of individuals. Thus, as Barrie Sander observes in the context of social media governance,⁸³ such conceptions tend to adhere to a form of abstract individualism that neglects power asymmetries between individuals and other more powerful actors in the relevant ecosystem. Litigation involving AI-based facial recognition systems such as that used in the British case of *Bridges* often features a right to individual privacy or a right to private life (under Section 8 of the European Court of Human Rights) that can be used as a shield against unjustified state intrusion, but it fails to capture the bigger picture shaped by a surveillance economy, which enables the ubiquity of this technology.⁸⁴ Another example involves the use of AI in digital welfare systems that are already in use in a number of countries, where pursuing the right to social security, for example, or the right to fairness in government decisions for access to social benefits can detract from criticizing the broader neoliberal edifice within which these digitized welfare systems are built.⁸⁵

A counterpoint is that it is not impossible to make IHRL compatible with a more structural understanding of it, that is, characterized by an openness to positive state intervention to safeguard public and collective values and one that tries to take into account imbalances of power in a given ecosystem.⁸⁶ Thus, awareness of the broader ecosystem in which an affected right operates is key for courts and other authoritative interpreters of laws and regulations, especially as it involves new or emerging technologies. Last December 2021, the Canadian federal privacy commissioner, as well as three provincial privacy authorities, ruled that Clearview AI's facial recognition technology resulted in the mass surveillance of Canadians and, thus, violated federal and provincial laws governing personal information.⁸⁷ Clearview AI, a US company, rejected the view that Canadian laws applied to it because it has no connection to Canada and that consent was required because the information they gathered was publicly available. But Canadian regulators rejected this argument because they found that the company had actively collected images of Canadians and marketed its services to law enforcement agencies in Canada.⁸⁸ This ruling showed the possibility that IHRL can concern itself with more than just individual rights. Indeed, the notion that human rights can and do produce structural effects insofar as requiring institutional changes on the part of the relevant state-party has already been recognized in a number of cases decided by the European Court of Human Rights, most notably with respect to the right to life and the prohibition of torture or inhuman/degrading treatment.⁸⁹

c) *The Problem of Cooptation*

If AI-based systems are indeed what Ari Waldman argues to be “social, political and economic expressions of neoliberal managerialization,”⁹⁰ that is, borrowing from Julie Cohen,⁹¹ an organizational system of public or private governance that prioritizes freedom and efficiency above all other values, then IHRL is not going to be the emancipatory tool that its advocates might expect that it is. Rather, it is easily coopted into an ethos that prizes efficiency and market solutions onto social problems, which can then be tracked and monitored by certain indicators. Indeed, some observers argue that IHRL already sits comfortably

⁸⁰ Spain Bradley, “Human Rights Racism.”

⁸¹ Kapczynski, “The Right to Medicines.”

⁸² Landau, “The Reality of Social Rights Enforcement.”

⁸³ Sander, “Democratic Disruption.”

⁸⁴ See generally Zuboff, *The Age of Surveillance Capitalism*.

⁸⁵ Alston, Report of the Special Rapporteur, paragraph 8.

⁸⁶ See also Douek, “The Limits of International Law” (suggesting the need for a more systemic view of rights).

⁸⁷ OPC, “PIPEDA Findings #2021-001.”

⁸⁸ OPC, “PIPEDA Findings #2021-001.”

⁸⁹ Leloup, “The Concept of Structural Human Rights”; Smuha, “Beyond the Individual”; Cohen, “Affording Fundamental Rights.”

⁹⁰ Waldman, “Power, Process.”

⁹¹ Waldman, citing Cohen, *Between Truth and Power*.

alongside the inequalities that such an ethos generates.⁹² One such mechanism in the toolkit to monitor and assess human rights compliance is the use of indicators. A likely risk is that superficial compliance with these metrics and the achievement of “having been measured” could become an end in itself while the underlying social reality is ignored. In addition, as Yarden Katz writes, AI creates and promotes a governance-by-numbers regime that offers several opportunities for manipulation and control.⁹³

In the field of corporate accountability and human rights, for example, the Global Reporting Initiative (a framework that incorporates legal standards such as UN international human rights conventions for corporations to report on their environmental, social, and economic performance) has been criticized for promoting box-ticking compliance, transferring decision-making authority to third-party technical experts that has since become an assurance industry, and the distortion of public values into numbers.⁹⁴ For instance, consider the Bisha mining facility in Ethiopia operated by Nevsun Resources, a Canadian mining company. Nevsun has a corporate human rights policy that incorporates the UN *Guiding Principles on Business and Human Rights*. In 2015, an external auditor carried out a human rights impact assessment (HRIA) at the mine and concluded that “the risk of child labor at the Bisha mine is remote.”⁹⁵ But these impact assessments can easily be a matter of lip service. In 2020, the Canadian Supreme Court rejected Nevsun’s argument that Eritrean employees alleging torture should pursue their claim in Eritrea, thereby opening the doors for the parent company to be sued for human rights abuses in Canadian courts.⁹⁶

Surveying the current AI regulation literature, there has been much discussion on the conduct and use of human rights impact assessments or, to be more specific, algorithmic risk assessments as part of a due diligence process to evaluate the risks *ex ante* and *ex post* of an AI-based system or even as part of technical efforts to bake human rights considerations into the design of these technologies.⁹⁷ In 2019, the UN human rights commissioner stated that “the risk of discrimination linked to AI-driven decisions – decisions that can change, define or damage human lives – is all too real. This is why there needs to be systematic assessment and monitoring of the effects of AI systems to identify and mitigate human rights risks.”⁹⁸ Indeed, algorithmic auditing conducted by third-party companies is a popular policy suggestion nowadays to counteract possible charges of bias, even though this route has well-documented cautionary tales.⁹⁹ AI registers employ a similar concept and contain information on how and where cities are using AI/ML, which data and algorithms they use, and what ethical principles are employed to mitigate biases and risks.¹⁰⁰ If used effectively, these assessments can indeed offer a number of benefits, such as measuring accountability to existing standards and norms, assess compliance with policies and specific targets, and evaluate performance with respect to stated objectives. Such an impact assessment can help decision-makers before deploying AI-based surveillance systems (e.g., those used for the prevention and detection of crime)¹⁰¹ by taking into account the different considerations courts themselves use in evaluating compliance with prevailing human rights instruments. They can also facilitate an efficient processing of information by reducing the costs and resources devoted to decision-making.

That said, subjecting AI to scrutiny and evaluating whether its design or implementation violates certain rights makes it susceptible to the same risks associated with the audit culture endemic in the structure of IHRL. Technology companies may satisfy themselves with producing human rights-compliant products in accordance with some predetermined metrics. To illustrate, Facebook commissioned an external auditor to conduct an HRIA of its operations in Myanmar as a result of a UN investigation that concluded that the social media company had played a “determining role” in stirring hate against the Rohingya minority.¹⁰² Predictably, the HRIA concluded that Facebook did not cause or contribute to any human rights harm in the country. But researchers Mark Latonero and Aaina Agarwal found that the HRIA did not adequately assess the most salient human rights impacts of Facebook’s presence and product in the country, although they nevertheless recognize the value of HRIAs in addressing the consequential risks and impacts of AI-based systems, recommending that these assessments should feature

⁹² Moyn, “A Powerless Companion”; Kapczynski, “The Right to Medicines”; Whyte, *The Morals of the Market*.

⁹³ Katz, “Manufacturing an Artificial Intelligence Revolution.”

⁹⁴ Sarfaty, “Measuring Corporate Accountability.”

⁹⁵ LKL International Consulting Inc., *Human Rights Impact Assessment of the Bisha Mine in Eritrea; Nevsun Resources Ltd v Araya*, 2020 SCC 5 (Can.).

⁹⁶ Hofmeister, “Penny Wise and Pound Foolish” (suggesting that Canadian mining companies must measure and address human rights impacts early in a project).

⁹⁷ Reisman, “Algorithmic Impact Assessment”; Diakopoulos, “Principles for Accountable Algorithms”; for a critical examination of AIAs see Moss, “Governing with Algorithmic Impact Assessments”; Latonero, “Human Rights Impact Assessments for AI”; OHCHR, *Artificial Intelligence Technologies, “Human Rights, Democracy.”*

⁹⁸ OHCHR, *The Right to Privacy in the Digital Age*.

⁹⁹ Ng, “Can Auditing Eliminate Bias from Algorithms?”

¹⁰⁰ Modjeska, “AI Registers.”

¹⁰¹ Murray, “Using Human Rights Law.”

¹⁰² BSR, *Human Rights Impact Assessment*.

analysis that encompasses both technical and social factors.¹⁰³ Moreover, as this example shows, this audit culture makes it possible that the very actors that are supposedly regulated by an external authority end up regulating themselves, which substantially weakens the regulatory power of such a regime.¹⁰⁴

While it is clear that the use of algorithmic human rights impact assessments can help identify and evaluate risks, establish strategies to mitigate those risks, and help articulate the rationale for the automated decision-making system,¹⁰⁵ what is not so obvious is that they can also be gamed so as to enable decision-makers to wash their hands of the result. That is, even if they do work as intended, the bigger concern is that it can also be mobilized as a justificatory digital trail to push back against claims of unfair harm from those adversely affected by the algorithm. Thus, a focus on documentation and process can mistake these metrics for actual compliance and, thus, obscure the underlying substantive values of fairness, equality, and human dignity that can be eroded by the large-scale use of these systems. This means that expanding the use of indicators can submerge questions of political struggles over what human rights mean and what constitutes compliance into technical questions of measurement. Such a governance-by-numbers approach makes any reliance on IHRL as a means to effectively regulate AI, in some cases, arguably as weak as relying on self-adopted ethics codes notwithstanding the veneer of law. In any case, the debate over these assessments does not even address the first-order question of whether some of these AI-based systems should be built in the first place.¹⁰⁶

Conclusion

Rules do not get developed or implemented in a vacuum but rather arise from particular political, social, and cultural contexts. If we want to understand AI development not simply as an American, Canadian, or European project but rather ultimately as a global one but with local approaches, then the question of how the concomitant rules are formed to govern it should be seen in a similar light. The biggest contribution of IHRL is to allow for these kinds of localized approaches to be developed while providing a general framework for some kind of global coordination. For better or worse, the human rights frame as the principal language of global justice informs and will continue to inform many of the current and future debates on AI regulation and governance around the world. In the past four years alone, there has been an explosion of calls for the design and implementation of algorithms to respect human rights from different sectors. Civil society organizations, grassroots advocacy groups, and many of those critical of the increasing power of tech companies, as well as their corresponding lack of accountability, principally wield their arguments in the language of human rights, as do technology companies who profess fealty to human rights in the design of their products. Thus, it is imperative that technologists, engineers, philosophers, and other stakeholders in the AI space are knowledgeable about its power, possibilities and, more notably, limits.

¹⁰³ Latonero, "Human Rights Impact Assessments for AI."

¹⁰⁴ McGrogan, "The Population and the Individual."

¹⁰⁵ Latonero, "Human Rights Impact Assessments for AI," 12–13.

¹⁰⁶ See, for example, AHEG, First Draft of the Recommendation. The UNESCO document assumes the use of AI as beneficial for humanity but disappointingly skips the more important question of whether it should even be deployed in certain instances.

Bibliography

- Achieme, E. Tendayi. *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*, A/75/590. New York: United Nations, 2020. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/304/54/PDF/N2030454.pdf?OpenElement>
- Ad Hoc Expert Group (AHEG) for the Preparation of a Draft Text of a Recommendation on the Ethics of Artificial Intelligence. *First Draft of the Recommendation on the Ethics of Artificial Intelligence, SHS/BIO/AHEG-AI/2020/4 REV.2*. New York: United Nations, 2020. https://unesdoc.unesco.org/in/rest/annotationSVC/Attachment/attach_upload_feb9258a-9458-4535-9920-fca53c95a424
- Alston, Philip. *Report of the Special Rapporteur on Extreme Poverty and Human Rights*, A/74/48037. New York: United Nations Human Rights Office of the High Commissioner, 2019. https://www.ohchr.org/Documents/Issues/Poverty/A_74_48037_AdvanceUneditedVersion.docx
- Alston, Philip. “The Populist Challenge to Human Rights.” *Journal of Human Rights Practice* 9, no 1 (February 2017): 1–15.
- Amnesty International. “Artificial Intelligence for Good.” Amnesty International, June 9, 2017. <https://www.amnesty.org/en/latest/news/2017/06/artificial-intelligence-for-good/>
- Amnesty International and Access Now. *The Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems*. Amnesty International and Access Now, 2018. https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf
- Arnold, Zachary, Ilya Rahkovsky, and Tina Huang. “Tracking AI Investment: Initial Findings from Private Markets.” Washington DC: Georgetown Center for Security and Emerging Technology, 2020.
- Beduschi, Ana. “International Migration Management in the Age of Artificial Intelligence.” *Migration Studies* 9, no 3 (September 2021): 576–96. <https://doi.org/10.1093/migration/mnaa003>
- Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2016.
- Brooks, Rodney. “An Inconvenient Truth About AI: AI Won’t Surpass Human Intelligence Anytime Soon.” *IEEE Spectrum*, September 29, 2021. <https://spectrum.ieee.org/rodney-brooks-ai>
- Business for Social Responsibility (BSR). *Human Rights Impact Assessment: Facebook in Myanmar*. San Francisco, CA: BSR, 2018. https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf
- Calo, Ryan. “Artificial Intelligence Policy: A Primer and Roadmap.” *U.C. Davis Law Review* 51, no 2 (2017): 399–435.
- Center for International Legal Cooperation. *The Hague Rules on Business and Human Rights Arbitration*. The Hague: Center for International Legal Cooperation, 2019. https://www.cilc.nl/cms/wp-content/uploads/2019/12/The-Hague-Rules-on-Business-and-Human-Rights-Arbitration_CILC-digital-version.pdf
- Chohlas-Wood, Alex. “Understanding Risk Assessment Instruments in Criminal Justice.” *Brookings*, June 19, 2020. <https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/>
- Cisco. “Cisco Human Rights Position Statements.” Cisco, 2018. <https://www.cisco.com/c/dam/assets/csr/pdf/Human-Rights-Position-Statements-2018.pdf>
- Cohen, Julie E. *Between Truth and Power: The Legal Constructions of Informational Capitalism*. New York City: Oxford University Press, 2019. <https://oxford.universitypressscholarship.com/10.1093/oso/9780190246693.001.0001/oso-9780190246693>.
- Columbia Center on Sustainable Investment. “Re: Elements for Consideration in Draft Arbitral Rules, Model Clauses, and Other Aspects of the Arbitral Process.” Columbia Center on Sustainable Investment, January 31, 2019. <https://ccsi.columbia.edu/sites/default/files/content/docs/publications/CCSI-Submission-on-BHR-Arbitration-Elements-Paper.pdf>
- Council of Europe Commissioner of Human Rights. *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*. Strasbourg: Council of Europe, 2019.
- Dadich, Scott. “Barack Obama Talks AI, Robo Cars, and the Future of the World.” *WIRED*, November 2016. <https://www.wired.com/2016/10/president-obama-mit-joi-ito-interview/>
- De Gregorio, Giovanni. “The Rise of Digital Constitutionalism in the European Union.” *International Journal of Constitutional Law* 19, no 1 (January 2021): 41–70. <https://doi.org/10.1093/icon/moab001>
- Desierto, Diane. “Human Rights in the Era of Automation and Artificial Intelligence.” *EJIL: Talk!*, February 26, 2020. <https://www.ejiltalk.org/human-rights-in-the-era-of-automation-and-artificial-intelligence/>
- Diakopoulos, Nicholas, Sorelle Friedler, Marcelo Arenas, Solon Barocas, Michael Hay, Bill Howe, H. V. Jagadish et al. “Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.” *FAT ML*. Accessed March 12, 2022. <https://www.fatml.org/resources/principles-for-accountable-algorithms>
- Dias Oliva, Thiago. “Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression.” *Human Rights Law Review* 20, no 4 (December 2020): 607–40. <https://doi.org/10.1093/hrlr/ngaa032>
- Doffman, Zak. “China Is Using Facial Recognition to Track Ethnic Minorities, Even In Beijing.” *Forbes*, May 3, 2019. <https://www.forbes.com/sites/zakdoffman/2019/05/03/china-new-data-breach-exposes-facial-recognition-and-ethnicity-tracking-in-beijing/>

- Donahoe, Eileen, and Megan MacDuffee Metzger. "Artificial Intelligence and Human Rights." *Journal of Democracy* 30, no 2 (2019): 115–26. <https://doi.org/10.1353/jod.2019.0029>
- Douek, Evelyn. "The Limits of International Law in Content Moderation." *UCI Journal of International, Transnational and Comparative Law* 6 (October 12, 2021). <http://dx.doi.org/10.2139/ssrn.3709566>
- Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press, 2018.
- European Commission. "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts." European Commission, April 21, 2021. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
- European Commission for the Efficiency of Justice (CEPEJ). "CEPEJ European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment." European Commission for the Efficiency of Justice, December 2018. <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>
- Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI." *Berkman Klein Center Research Publication* No. 2020-1, (January 15, 2020). <https://cyber.harvard.edu/publication/2020/principled-ai>
- Future of Life Institute. "AI Principles." Asilomar AI Principles. Future of Life Institute, August 11, 2017. <https://futureoflife.org/2017/08/11/ai-principles/>
- Gasser, Urs, and Virgilio A.F. Almeida. "A Layered Model for AI Governance." *IEEE Internet Computing* 21, no 6 (November 2017): 58–62. <https://doi.org/10.1109/MIC.2017.4180835>
- Hagras, Hani. "Toward Human-Understandable, Explainable AI." *Computer* 51, no 9 (September 2018): 28–36. <https://doi.org/10.1109/MC.2018.3620965>
- Hamilton, Rebecca J. "Governing the Global Public Square." *Harvard International Law Journal* 62, no 1 (Winter 2021): 117–74.
- High-level Expert Group on Artificial Intelligence (AI HLEG). *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission, 2019. <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>
- Hofmeister, Edie. "Penny Wise and Pound Foolish." *CIM Magazine*, November 12, 2020. <https://magazine.cim.org/en/voices/penny-wise-and-pound-foolish-en/>
- Hogarth, Ian. "AI Nationalism." *Ian Hogarth* (blog), June 13, 2018. <https://www.ianhogarth.com/blog/2018/6/13/ai-nationalism>
- Human Rights Watch. "China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App." *Human Rights Watch*, May 1, 2019. <https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass>
- IBM Design for AI. "Everyday Ethics for AI." IBM, May 2019. <https://www.ibm.com/design/ai/ethics/everyday-ethics/>
- IEEE. *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*. IEEE, 2019. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf
- Jobin, Anna, Marcello Ienca, and Effy Vayena. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1, no 9 (September 2019): 389–99. <https://doi.org/10.1038/s42256-019-0088-2>
- Kapczynski, Amy. "The Right to Medicines in an Age of Neoliberalism." *Humanity: An International Journal of Human Rights, Humanitarianism, and Development* 10, no 1 (2019): 79–107. <https://doi.org/10.1353/hum.2019.0003>
- Katz, Yarden. "Manufacturing an Artificial Intelligence Revolution." Unpublished manuscript.
- Kaye, David. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, A/HRC/38/35. New York: United Nations, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement>
- Kirka, Danica. "UK Court Says Face Recognition Violates Human Rights." *Tech Xplore*, August 11, 2020. <https://techxplore.com/news/2020-08-uk-court-recognition-violates-human.html>
- Kwet, Michael. "Digital Colonialism: US Empire and the New Imperialism in the Global South." *Race & Class* 60, no 4 (April 1, 2019): 3–26. <https://doi.org/10.1177/0306396818823172>
- Landau, David. "The Reality of Social Rights Enforcement." *Harvard International Law Journal* 53, no 1 (Winter 2012): 190–247.
- Latonero, Mark. *Governing Artificial Intelligence*. New York City: Data & Society Research Institute, 2018.
- Latonero, Mark, and Aaina Agarwal. "Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar." *Carr Center Discussion Paper Series*, 2021.
- Leloup, Mathieu. "The Concept of Structural Human Rights in the European Convention on Human Rights." *Human Rights Law Review* 20, no 3 (October 2020): 480–501. <https://doi.org/10.1093/hrlr/ngaa024>
- Li, Daitian, Tony W. Tong, and Yangao Xiao. "Is China Emerging as the Global Leader in AI?" *Harvard Business Review*, February 18, 2021. <https://hbr.org/2021/02/is-china-emerging-as-the-global-leader-in-ai>

- LKL International Consulting Inc. *Human Rights Impact Assessment of the Bisha Mine in Eritrea*. LKL International Consulting Inc.: Montreal, 2015. <https://media.business-humanrights.org/media/documents/files/documents/Bisha-HRIA-Audit-2015.pdf>
- Loucks, Jeff, Susanne Hupfer, David Jarvis, and Timothy Murphy. "Future in the Balance? How Countries Are Pursuing an AI Advantage." *Deloitte Insights* (blog), May 1, 2019. <https://www2.deloitte.com/content/www/us/en/insights/focus/cognitive-technologies/ai-investment-by-country.html>
- Marchant, Gary, "Soft Law Governance of Artificial Intelligence," *AI Pulse*, 2019, <http://aipulse.org/soft-law-governance-of-artificial-intelligence/>
- Marks, Susan, "Human Rights and Root Causes," *The Modern Law Review*, 74, no 1 (January 2011): 57–78. <https://doi.org/10.1111/j.1468-2230.2010.00836.x>
- McGregor, Lorna, Daragh Murray, and Vivian Ng. "International Human Rights Law as a Framework for Algorithmic Accountability." *International and Comparative Law Quarterly* 68, no 2 (April 2019): 309–43. <https://doi.org/10.1017/S0020589319000046>
- McGrogan, David. "The Population and the Individual: The Human Rights Audit as the Governmentalization of Global Human Rights Governance." *International Journal of Constitutional Law* 16, no 4 (December 2018): 1073–1100. <https://doi.org/10.1093/icon/moy086>
- Modjeska, Natalia. "AI Registers: Finally, a Tool to Increase Transparency in AI/ML." *Towards Data Science* (blog), December 28, 2020. <https://towardsdatascience.com/ai-registers-finally-a-tool-to-increase-transparency-in-ai-ml-f5694b1e317d>
- Molnar, Petra. *Technological Testing Grounds: Migration Management Experiments and Reflections from the Ground Up*. EDRI: Brussels, 2020. <https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf>
- Molnar, Petra. "Technology on the Margins: AI and Global Migration Management from a Human Rights Perspective." *Cambridge International Law Journal* 8, no 2 (December 2019): 305–30. <https://doi.org/10.4337/cilj.2019.02.07>
- Moore, Scott. "Trump's Techno-Nationalism." *Lawfare* (blog), August 15, 2019. <https://www.lawfareblog.com/trumps-techno-nationalism>
- Moss, Emanuel, Elizabeth Anne Watkins, Jacob Metcalf, and Madeleine Clare Elish. "Governing with Algorithmic Impact Assessments: Six Observations." Paper presented at AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES), Virtual, May 14, 2021. <https://doi.org/10.1145/3461702.3462580>
- Moyn, Samuel. "A Powerless Companion: Human Rights in the Age of Neoliberalism." *Law and Contemporary Problems* 77, no 4 (2014): 147–69.
- Moyn, Samuel. *Not Enough: Human Rights in an Unequal World*. Cambridge, Massachusetts: Belknap Press, 2018.
- Muller, Cateljne. *Artificial Intelligence – The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society*. Bruxelles: European Economic and Social Committee, 2017. <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and>
- Murray, Daragh. "Using Human Rights Law to Inform States' Decisions to Deploy AI." *American Journal of International Law* 114 (2020): 158–62. <https://doi.org/10.1017/aju.2020.30>
- Narayanan, Arvind. *21 Fairness Definitions and Their Politics*. Video, 2018. <https://www.youtube.com/embed/jIXIuYdnyyk>
- Narayanan, Arvind "How to Recognize AI Snake Oil." Slides and notes. Princeton, Cambridge, MA, November 2019. <https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf>
- Ng, Alfred. "Can Auditing Eliminate Bias from Algorithms?" *The Markup*, February 23, 2021. <https://themarkup.org/ask-the-markup/2021/02/23/can-auditing-eliminate-bias-from-algorithms>
- Ochigame, Rodrigo. "The Invention of 'Ethical AI': How Big Tech Manipulates Academia to Avoid Regulation." *The Intercept*, December 20, 2019. <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>
- Organisation for Economic Co-operation and Development (OECD). "Recommendation of the Council on Artificial Intelligence." OECD, May 21, 2021. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Office of the Privacy Commissioner of Canada (OPC). "PIPEDA Findings #2021-001: Joint Investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada, the Commission d'Accès à l'Information Du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta." Office of the Privacy Commissioner of Canada, February 3, 2021. <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/>
- Office of the United Nations High Commissioner for Human Rights (OHCHR). *Artificial Intelligence Technologies and Freedom of Expression*, New York: United Nations Human Rights Office of the High Commissioner, n.d. https://www.ohchr.org/Documents/Issues/Expression/Factsheet_3.pdf
- Office of the United Nations High Commissioner for Human Rights (OHCHR). *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*. New York: United Nations Human Rights Office of the High Commissioner, 2011.

- Office of the United Nations High Commissioner for Human Rights (OHCHR). *Legally Binding Instrument to Regulate, in International Human Rights Law, the Activities of Transnational Corporations and Other Business Enterprises*. New York: United Nations Human Rights Office of the High Commissioner, 2019.
https://www.ohchr.org/Documents/HRBodies/HRCouncil/WGTransCorp/OEIGWG_RevisedDraft_LBI.pdf
- Office of the United Nations High Commissioner for Human Rights (OHCHR). “‘Smart Mix’ of Measures Needed to Regulate New Technologies.” United Nations Human Rights Office of the High Commissioner, April 24, 2019.
<https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24509>
- Office of the United Nations High Commissioner for Human Rights (OHCHR), *The Right to Privacy in the Digital Age - Report of the United Nations High Commissioner for Human Rights Report of the United Nations High Commissioner for Human Rights*. A/HRC/48/31 New York: United Nations Human Rights Office of the High Commissioner, 2021.
https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session48/Documents/A_HRC_48_31_AdvanceEditedVersion.docx
- O’Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books, 2016.
- Pascal, Alex, and Tim Hwang. “Artificial Intelligence Isn’t an Arms Race.” *Foreign Policy*, December 11, 2019.
<https://foreignpolicy.com/2019/12/11/artificial-intelligence-ai-not-arms-race-china-united-states/>
- Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2015.
- Peters, Dorian, Karina Vold, Diana Robinson, and Rafael A. Calvo. “Responsible AI—Two Frameworks for Ethical Design Practice.” *IEEE Transactions on Technology and Society* 1, no 1 (March 2020): 34–47.
<https://doi.org/10.1109/TTS.2020.2974991>
- Pichai, Sundar. “AI at Google: Our Principles.” *Google: The Keyword* (blog), June 7, 2018.
<https://blog.google/technology/ai/ai-principles/>
- Powles, Julia. “The Seductive Diversion of ‘Solving’ Bias in Artificial Intelligence.” *OneZero* (blog), December 7, 2018.
<https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>
- Price, W. Nicholson, II. “Artificial Intelligence in the Medical System: Four Roles for Potential Transformation.” *Yale Journal of Law and Technology* 21, no 3 (February 25, 2019): 122–32.
- Redeker, Dennis, Lex Gill, and Urs Gasser. “Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights.” *International Communication Gazette* 80, no 4 (June 1, 2018): 302–19.
<https://doi.org/10.1177%2F1748048518757121>
- Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker. *Algorithmic Impact Assessment: A Practical Framework for Public Agency Accountability*. New York: AI Now Institute, April 2018.
<https://ainowinstitute.org/aiareport2018.pdf>
- Richardson, Rashida, Jason M., Schultz, and Vincent M. Southerland. *Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems*. New York: AI Now Institute, 2019.
<https://ainowinstitute.org/litigatingalgorithms-2019-us.html>
- Risse, Mathias. “Human Rights and Artificial Intelligence: An Urgently Needed Agenda.” *Human Rights Quarterly* 41, no 1 (2019): 1–16. <https://doi.org/10.1353/hrq.2019.0000>
- Ruggie, John. *Report of the Special Representative of the Secretary General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises*, John Ruggie, A/HRC/17/31, New York: United Nations Human Rights Office of the High Commissioner, 2011. https://www.ohchr.org/documents/issues/business/a-hrc-17-31_aev.pdf
- Sander, Barrie. “Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law.” *European Journal of International Law* 32, no 1 (February 2021): 159–93.
<https://doi.org/10.1093/ejil/chab022>
- Sander, Barrie. “Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation.” *Fordham International Law Journal* 43, no 4 (January 2020): 939.
- Sarfaty, Galit A. “Measuring Corporate Accountability through Global Indicators.” In *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*, edited by Benedict Kingsbury, Kevin E. Davis, and Sally Engle Merry, 103–32. Cambridge Studies in Law and Society. Cambridge: Cambridge University Press, 2015.
<https://www.cambridge.org/core/books/quiet-power-of-indicators/measuring-corporate-accountability-through-global-indicators/6653B66BAF08B83C5EF514ABEEF7B01F>
- Shane, Scott, Cade Metz, and Daisuke Wakabayashi. “How a Pentagon Contract Became an Identity Crisis for Google.” *The New York Times*, May 30, 2018, sec. Technology. <https://www.nytimes.com/2018/05/30/technology/google-project-maven-pentagon.html>
- Sherman, Justin. “Oh Sure, Big Tech Wants Regulation—On Its Own Terms.” *Wired*, January 28, 2020.
<https://www.wired.com/story/opinion-oh-sure-big-tech-wants-regulationon-its-own-terms/>
- Smuha, Nathalie. “Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea.” *Philosophy and Technology* 34 (2021): 91–104. <https://doi.org/10.1007/s13347-020-00403-w>

- Spain Bradley, Anna. "Human Rights Racism." *Harvard Human Rights Journal* 32, no 1 (January 2019). The Citizen Lab and International Human Rights Program (Faculty of Law, University of Toronto). *Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System*. Toronto: Faculty of Law, University of Toronto, 2018.
- Thomas, Rachel. "Medicine's Machine Learning Problem." *Boston Review*. Accessed March 12, 2022. <https://bostonreview.net/articles/rachel-thomas-medicines-machine-learning-problem/>
- Thompson, Nicholas. "Emmanuel Macron Talks to WIRED About France's AI Strategy." *Wired*, March 31, 2018. <https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/>
- UN Secretary General's High-level Panel on Digital Cooperation. *The Age of Digital Interdependence: Report of the UN Secretary-General's High-Level Panel on Digital Cooperation*. New York: UN, 2019. <https://digitallibrary.un.org/record/3865925?ln=en>
- van Veen, Christiaan. "Artificial Intelligence: What's Human Rights Got To Do With It?" *Data & Society: Points*, May 18, 2018. <https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5>
- Vincent, James. "Google Promises Ethical Principles to Guide Development of Military AI." *The Verge*, May 30, 2018. <https://www.theverge.com/2018/5/30/17408446/google-ai-guidelines-weaponry-military-pentagon-maven-contract>
- Vincent, James. "Putin Says the Nation That Leads in AI 'Will Be the Ruler of the World.'" *The Verge*, September 4, 2017. <https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world>
- Wadhwa, Vivek. "Laws and Ethics Can't Keep Pace with Technology." *MIT Technology Review*, April 15, 2014. <https://www.technologyreview.com/2014/04/15/172377/laws-and-ethics-cant-keep-pace-with-technology/>
- Waldman, Ari Ezra. "Power, Process, and Automated Decision-Making Symposium: Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law." *Fordham Law Review* 88, no 2 (2019): 613–32.
- Whyte, Jessica. *The Morals of the Market: Human Rights and the Rise of Neoliberalism*. Brooklyn: Verso, 2019.
- Winfield, Alan F., Katina Michael, Jeremy Pitt, and Vanessa Evers. "Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems [Scanning the Issue]." *Proceedings of the IEEE* 107, no 3 (March 2019): 509–17. <https://doi.org/10.1109/JPROC.2019.2900622>
- World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*. Geneva, Switzerland: World Health Organization, June 28, 2021. <https://www.who.int/publications-detail-redirect/9789240029200>
- Yeung, Karen, Andrew Howes, and Ganna Pogrebna. "AI Governance by Human Rights–Centered Design, Deliberation, and Oversight." In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale, and Sunit Das. Oxford: Oxford University Press, 2020. <https://doi.org/10.1093/oxfordhb/9780190067397.013.5>
- Yi, Zeng. *The Ethical Norms for the New Generation Artificial Intelligence, China*. Beijing, China: International Research Center for AI Ethics and Governance, 2021. <https://ai-ethics-and-governance.institute/2021/09/27/the-ethical-norms-for-the-new-generation-artificial-intelligence-china/>
- Zittrain, Jonathan. "The Hidden Costs of Automated Thinking." *The New Yorker*, July 23, 2019. <https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking>
- Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York City: PublicAffairs, 2019.