Centre for International
Governance Innovation

Policy Brief No. 196 — February 2025

Applying International Human Rights Principles for AI Governance

Sabhanaz Rashid Diya

Key Points

- Despite gaining prominence, the fairness, accountability, transparency and ethics (FATE) framework in artificial intelligence (AI) governance poses significant limitations. It is inadequately defined to meet the complexities of a pluralistic world, lacks consensus on normative values underpinning it, is prone to misuse and misrepresentation, and inadvertently promotes ethics washing.
- The International Bill of Human Rights, while not devoid of criticism and implementation challenges, provides a universal foundation for building consensus around value archetypes within and between societies.
- Canada can play a critical leadership role in international AI governance through the Global Digital Compact, as well as its membership in the Group of 20 (G20) and its presidency in the Group of Seven (G7), by establishing human rights frameworks as a governance norm for AI systems.

Introduction

AI has increasingly captured policy attention, leading to its usage doubling in global legislative proceedings, from 1,247 in 2022 to 2,175 in 2023 (Stanford Institute for Human-Centered Artificial Intelligence 2024). But despite the frenetic pace of this new technology, questions persist about the normative values underpinning AI governance. Some researchers have questioned how AI systems can better align with human values, while others argue that these systems "lock in" specific value archetypes that reflect those of their developers and designers (Gabriel and Ghazavi 2021).

"Value sensitive design," particularly in technology, is focused on what people consider important in their lives, with an emphasis on ethics and morality (Friedman and Hendry 2019). However, moral discourse is saturated with disagreements. American communitarian theorist Michael J. Sandel (2011) argues that morality is centred around three major ideas — maximizing welfare, respecting freedoms and promoting virtues — and each ethical theory "points to a different way of thinking about justice." Eastern philosopher Li Zehou contends that justice is intertwined with harmony (D'Ambrosio 2016), while Nobel Prize winner and economist Amartya Sen (2012) proposes making comparative judgments

About the Author

Sabhanaz Rashid Diya is a CIGI senior fellow and the founder of Tech Global Institute, a global tech policy non-profit focused on advancing equity in design and governance of technologies in the global majority. She has advised governments in 20 countries, including leading closed-door briefings with the White House, multilateral international organizations and bilateral donors on global internet and platform governance, responsible artificial intelligence (AI) and human rights.

A computational social scientist by training, Sabhanaz has more than 17 years of experience at the intersection of technology policy, ethics and international development. She was most recently the head of public policy for Bangladesh at Meta, where she led on various regulatory and legislative issues, including privacy, online harms and algorithmic transparency. Sabhanaz also worked at the Bill & Melinda Gates Foundation, leading policy and advocacy efforts in digital identity, data governance and AI. Her career spans the private and public sectors in the United States, Asia and Africa on encryption policy, digital trade, AI applications in the global majority and internet governance. She is a visiting policy fellow at the Oxford Internet Institute, and a member of the Advisory Network of the Freedom Online Coalition.

about justice and injustice rooted in "social realizations" — the actual lives of people and the freedoms they enjoy. Depending on which perspective is taken, there may be different implications for what is considered ethical and moral in the use and governance of technologies.

This philosophical dilemma raises several questions. First, what specific ethical principles do AI systems need to adhere to, and which actors are responsible for defining these principles? Human beings hold a variety of contrasting but reasonable beliefs about moral values (Rawls 1999), so designing governance systems for AI with a single moral doctrine would involve imposing a set of values and judgments on other people who may not agree with them (Gabriel 2020). Second, can localized and contextualized rules be embedded in transnational technology architectures? Societies operate on varied public conceptions of justice (ibid.). Domestic principles of justice are siloed and cannot be packaged into transnational technology designs with the same constraints that domestic policies impose. Historically, while nation states have intervened to influence the character and use of the internet — and now AI — in different parts of the world, the technology embodies design principles that are counterproductive to efforts around localization (Leiner el al. 2009). Third, and most critically, which set of normative values should be encoded in AI systems or in their governance? The largest AI systems in the world are predominantly designed under sociopolitical and commercial conditions in economically advanced Western democracies, inherently posing the risk of imparting doctrines of morality and ethics that may not agree with those operating in other parts of the world.

Bearing these considerations in mind, this policy brief argues that ethics as a sole framework for AI governance in a pluralistic world may be insufficient and ill-defined and might inadvertently result in co-option of one ethics value archetype over others. Different interpretations of "ethical" and "fair" AI systems rely on different contextual assumptions (Friedler, Scheidegger and Venkatasubramanian 2021), resulting in inconsistencies in their effects on specific population subgroups (Whittlestone et al. 2019). Moreover, such interpretations risk being top-down and focused on "the machine" by reflecting the values of those who design and deploy AI systems, thereby undermining the diversity of people's

experiences on the receiving end of these systems. This policy brief proposes that *universal* human rights principles — despite their limitations — must provide a pragmatic way forward by making human conditions central to the impact of AI technologies, while still addressing pluralism, political fragmentation and cross-culturalism.

Limitations of Governing Globally through Algorithmic Fairness and Ethics

Limited Scholarship in AI Ethics Exacerbates Historical Harms

While the field of ethics in computing is not new, it has recently gained increased attention due to the growing discourse on the societal implications of machine learning and AI technologies. Despite this field's long-standing history, a mere seven percent of the 186 machine-learning courses surveyed across the top 20 US universities offer ethics-specific training (Saltz et al. 2019). Although FATE research in AI has advanced substantially over the past several years, the broader FATE research community is lacking in terms of geographic and cultural diversity (Freire, Porcaro and Gómez 2021). The sparseness of ethics content within technical machine-learning/AI courses is further compounded by a dearth of philosophical and ethical perspectives outside of the United States and Western Europe (referred to as "the West" in this brief). This results in recommendations that are primarily situated within a limited view of dominant English-speaking and Western European cultural values (Schwartz 2006; Prabhakaran et al. 2022). Consequently, this content responds to injustices and social discrimination particular to Anglo-Saxon political environments, while failing to adequately consider the systemic injustice and discrimination faced by communities in the Global Majority, as well as economically, racially or ethnically marginalized groups in the West. While many scholars in AI ethics are immigrants or children of immigrants themselves, and their work sometimes strives to address these systemic disparities, there remains a noticeable gap in research representing the full spectrum of sociopolitical realities worldwide. This limited scope risks perpetuating a narrow understanding of ethics in AI governance.

Researchers have previously highlighted how neoliberal AI systems perpetuate data extractionism, stifling Indigenous innovation and making Global Majority communities dependent on North American or Western European software and infrastructure, which are rooted in colonial business models (Birhane 2020). These business models risk disenfranchising communities already prone to exploitation and underdevelopment due to centuries of imperialism (Kwet 2019). Algorithmic fairness methodologies assume a complete representation of people and phenomena in the data set; however, despite more than 50 percent of India's population being online, models tend to overfit for "digitally rich groups," namely, urban and male populations (Sambasivan et al. 2021). Similarly, statistical fairness assumes that user data corresponds one-to-one with people. But this does not account for the multilayered power dynamics experienced in many South Asian households, where men — who have more access to "devices, documentation and mobility" — often respond on behalf of women during national data collection initiatives, with data instruments attributing these responses to women (ibid.). Without substantive engagement with the politics, values and conditions of non-Western populations, the field of algorithmic fairness can be naive and myopic, while exacerbating historical harms (ibid.).

Values pluralism exists both *between* regions and countries and *within* countries and communities. For example, in Anglocentric society, neoliberal conceptions of the ethical use of creating, distributing and preserving data and knowledge give autonomy to the individual, while some Native American communities assume that data is part of a group's overall identity, with "certain members hav[ing] a duty to keep the [knowledge] on behalf of the group" and safeguard its privacy (Tsosie 2007). Within tribal cultures, individuals are born into "an integrated network of family, kinship, social and political relations," and individual rights exist within these normative frameworks and relationships (Clinton 1990).

Lack of Consensus on AI Fairness and Ethical Values Undermines Their Efficacy and Implementation

While there is a global push to adopt the FATE framework for AI governance, varying interpretations of these principles complicate their adoption and implementation. An analysis of more than 200 guidelines and recommendations worldwide for AI systems finds that "transparency," "explainability," "reliability" and "fairness" are among the most cited principles (Corrêa et al. 2023). Some practitioners argue that the overlap of these principles signals a positive convergence of values, suggesting a shared commitment to ethical AI. However, in practice, there is significant divergence in how these principles are interpreted, the value systems they represent, and the specific issues, domains or actors they are meant to address (Jobin, Ienca and Vayena 2021).

Further, the lack of consensus on the content of normative values underpinning AI ethics has created a disconnect between AI developers and the communities they aim to serve. AI ethics is often treated as a technical discipline marked by insufficient engagement with the lived experiences of the end user. A survey of 146 research papers on analyzing bias in natural language processing models found limited engagement on the ways in which bias is harmful, why it is harmful and who is harmed by it (Blodgett et al. 2020). Although a particular automated decision-making system can improve the overall accuracy of decisions in recruitment or loan applications, for example, it can still discriminate against a specific subgroup of the population (Whittlestone et al. 2019). A beneficial AI system can save lives while also using personal data in ways that violate a community's notion of privacy (ibid.). The absence of a shared vocabulary on AI ethics widens the gap between research, policy practice, innovation and communities.

Existing governing frameworks and ethical codes for AI often contain "abstract and vague concepts," such as the idea of "fair AI," without offering specific, actionable steps to implement these principles (Mittelstadt 2019). They also fail to address fundamental normative and political tensions embedded in these concepts (ibid.). For example, the Montréal Declaration[1] cites "well-being" and "sustainable development" as two principles to develop responsible AI, but does not address the potential for contradiction between these concepts. On one hand, an AI system can improve human productivity through the automation of cumbersome administrative tasks, which directly affects human well-being, while on the other hand simultaneously contributing to carbon emissions and environmental degradation (Hogan 2015). In other words, AI can promote the well-being of one party or subject at the expense of another.

Moreover, existing translational tools and methods for converting AI ethics to practice are either too flexible, rendering them ineffective at addressing harms, or too strict, in that they are unresponsive to context (Morley et al. 2021). A survey among AI practitioners finds 91 percent of respondents agree that "ethically designing" AI products is important; however, 41 percent of respondents associate ethics with a narrow view of compliance with data protection regimes in which they are operating (Morley et al. 2023). It is unclear to practitioners *who* ultimately is responsible for ensuring alignment between AI product design and ethical principles (ibid.) and at which stages of the product life cycle. There is a disconnect between the demand and availability of pro-ethical design resources that are practical and enforceable, and others that are available lack engagement with specific assumptions and definitions underpinning ethical principles, making them non-actionable.

The Adoption of AI Ethics Masks Institutional Shortcomings and Evasion of Regulatory Scrutiny

The lack of definitions, specificity, actionability and consensus about AI ethics has resulted in this concept lacking measurable results and ineffectively holding AI developers accountable, despite being "buzzworthy" in international political discourse. Conversely, the prevalence of vaguely worded principles drawn from the FATE framework that advocate for self-regulation masks private interests in the policy-making process, also known as "ethics washing" (Bietti 2020). This approach avoids strict state-led interventions in the public interest in an effort to avoid public backlash. Governing AI technologies requires expertise and capabilities that do not exist in government agencies, often leading governments to allow the private sector to

1 See https://declarationmontreal-iaresponsable.com/la-declaration/.

actively participate in and shape the regulations intended to govern itself. In Russia, for example, overreliance on the expertise of domestic private organizations resulted in the Russian government intentionally avoiding regulatory obligations that may be resource intensive or misaligned with the business priorities of AI developers (Papyshev and Yarime 2024). The resulting AI ethics code was voluntary with unenforceable principles, and undermined consumer protection under the guise of technological innovation.

In 2019, Google formed an AI ethics board with no veto power over controversial products, though it was quickly dissolved after facing public backlash (Knight 2019). A survey of more than 45 generative AI systems finds that although companies frequently use the term "open source" and claim that transparency and fairness are central to their AI product life cycle, many models are at best "open weight," meaning that their internal parameters or values are trained and adjusted to learn relationships between data points. Companies often evade legal, regulatory and scientific scrutiny by withholding information about training data (Liesenfeld and Dingemanse 2024). In fact, most AI governance and ethics proposals have emerged from within the private sector and large corporations (Jobin, Ienca and Vayena 2021), raising serious concerns about ethics washing or using vague principles to distract from the actual harms caused by AI products and business models (Hu 2021). Despite espousing ethics guidelines, few companies can show tangible changes in how AI products are designed. In practice, individual employees who advocate for ethics within companies frequently lack institutional support and risk being sidelined in the face of fast-paced product launch goals (Ali et al. 2023).

China's national body on technical standards also espouses human welfare, transparency and ethics in their AI guidelines (Roberts et al. 2021). Although these principles are ostensibly similar to those supported by Western democracies and ethicists, it is critical to understand them in the context of China's politics, ideology, culture and public consensus (Webster et al. 2017). Chinese culture places greater emphasis on social responsibility and community relations than on individual rights, so the data and privacy policies of its companies will likely come with problematic exemptions for the collection and use of personal data to "advance public interest" for security and health reasons.

Similarly, Bangladesh, India, Mexico, Pakistan and Uruguay include ethics in their national AI strategies, but fail to also provide clear definitions or actionable measures. In extreme cases, policy makers interpret ethics through the lens of protectionism and upholding national sovereignty; in other words, these policies are misused to omit or revise aspects of science or history to align with political values that undermine the very premise of ethics and accountability.

The International Bill of Human Rights as an Engine for Consensus Building

In 1948, the international community adopted the Universal Declaration of Human Rights (UDHR) to protect human dignity following the Second World War. In 1966, the United Nations added two more treaties — the International Covenant on Economic Social and Cultural Rights and the International Covenant on Civil and Political Rights. Together, these three documents form the International Bill of Human Rights,[2] ratified by 170 UN member states. The bill provides a universal set of normative principles with legal precedent, offering a framework to guide and operationalize AI systems in ways that can safeguard fundamental rights. Broad support for this human rights framework among governments, organizations, industry and civil society offers a foundation for addressing AI's impact on human lives (Latonero 2018).

A human rights framework can serve as a starting point for achieving what many ethical frameworks aim to accomplish, but often fail to deliver. Shifting AI governance discourse from a focus on what machines can or cannot do to their impact on human lives, conditions and rights will enable policy makers to build broader consensus. This approach can have the following benefits:

→ **Empowering communities as rights holders:** The vagueness and elasticity of AI ethics

2 See www.ohchr.org/en/what-are-human-rights/international-bill-human-rights.

principles often means their content can be shaped by anyone based on individual preferences and political interests. In contrast, the UDHR provides an analytical framework to resolve tensions between rights and societal interests through structured reasoning and evaluation (Yeung, Howes and Pogrebna 2020). These rights, enshrined to individuals as bearers of equal status and dignity, empower communities as owners and holders of specific protections and freedoms, rather than as mere "recipients" and "stakeholders" of vague ethics standards. This framework demands a focus on whose rights are at stake, what those rights are and the potential risks that are entailed (Prabhakaran et al. 2022). The UDHR includes the rights to life, freedom and safety (article 3); justice (article 8); and privacy (article 12), as well as freedom of movement (article 13), thought (article 18), expression (article 19), and the right to work and equal pay (article 23), all of which are salient in informing the design and governance of AI systems. For example, human rights principles of non-discrimination and privacy can be applied to algorithmic discrimination and surveillance, as seen in facial recognition systems that produce culturally ingrained biases against people of colour (Buolamwini and Gebru 2018) and disproportionately target Black people and others from communities of colour (Turner Lee and Chin-Rothmann 2022).

→ **Assigning explicit responsibilities to states, private industry and individuals:** International human rights law places states as primary duty-bearers, responsible for safeguarding the fundamental freedoms of their populations (Rawls 1993). In cases where domestic law is lacking, these laws carry moral and normative significance. The 2011 UN report, *Guiding Principles on Business and Human Rights*, extends specific responsibilities to businesses, including technology companies, to identify, prevent and mitigate human rights risks posed by their products and services (United Nations Human Rights Office of the High Commissioner 2011). Developers of products and services must make an informed effort to understand the implications on rights holders and respond through an iterative and consultative process of evaluation, review and mitigation based on a clearly defined human rights framework.

→ **Establishing norms for assessing AI impacts:** The 2011 UN report outlines human rights assessments as an integral part of human rights due diligence, and takes into account geographies, resources, products and people in ensuring if businesses are fulfilling the recommended guidelines. The international human rights framework provides comprehensive norms for AI systems throughout their life cycles, ensuring conformity in design, assessment, testing, promotion, sales, licensing and distribution (Yeung, Howes and Pogrebna 2020; Business for Social Responsibility 2021). Meanwhile, human rights law imposes obligations on states to prevent violations, actively promote these rights and establish appropriate mechanisms for remedy and justice. A rights-respecting assessment would also impose a duty on both states and businesses to safeguard the rights and dignity of the oft-hidden workforce, predominantly from Global Majority regions, that is often responsible for labelling, moderating and maintaining AI technologies.

→ **Facilitating a shared vocabulary across product life cycles:** Different sets of normative values can be embedded throughout the stages of an AI product's life cycle, resulting in contradictions and opacity. These stages can include the values intrinsic to data collection and selection processes; sociopolitical values encoded during the design and development of the product; normative values dictating how data is trained; and policy or political values that determine the product's deployment and distribution. The UDHR provides a consistent framework with well-established definitions that can be operationalized throughout the AI development and deployment process.

The Limitations of a Human Rights-Based Approach to AI Governance

Despite its merits and international legitimacy, the International Bill of Rights is not a panacea for AI governance and the myriad of challenges imposed by AI technologies. First, governments enforce international human rights law to varying degrees due to differences in political ideology and claims to sovereignty (Latonero 2018; International Institutions and Global Governance Program 2012). To tackle the rigidity of *law*, a human rights *framework* has emerged through interactions among UN systems, civil society, nation-states, the private sector, academia, non-governmental organizations and individuals. In reality, instead of reflecting a universal heritage, the interpretation of a human rights framework has too often been politicized; oriented around dominant Western concerns; applied selectively to countries and communities; and cheapened as contradictions between principle and practice grow in number (Kaplan 2021).

Second, the universality of this bill may be overstated: what is regarded as a human rights violation in one society may be considered lawful in another, and "Western ideas of human rights should not be imposed" (Tesón 1985). On one hand, critics argue that today's human rights discourse is controlled by advocacy organizations, lobbyists and journalists who share a similar interpretation of individualistic Western norms that undermines community or shared rights (Kaplan 2021). On the other hand, proponents of the UDHR argue that it has a strong claim to "relative universality" (Donnelly 1984): it safeguards human dignity and freedoms and can be further built upon in response to specific contexts or communities.

Third, the bill may perpetuate neo-imperialism (Wall 1998) through its politicized deployment by Western states. As political scientist Susan Waltz (2002) notes, non-state actors and Global Majority countries have led the way in "promoting the political idea of human rights" in iterations of the international human rights framework. But there are non-Western documents deserving of consideration in this framework: in Iran, for example, Talibov-i Tabrizi published *Izahat dar Khusus-i Azadi* (*Explanations Concerning Freedom*), while in China, Kang Youwei published early segments of *Datong shu* (*The Book of Great Harmony*) on liberty, freedom, equality and the rights of all humanity (ibid.). And non-Western contributions to developing the UDHR are notable; for example, India's Hansal Mehta was instrumental in changing the original language of article 1 from "all men" to "all human beings" to ensure the deliberate inclusivity of all people (ibid.).

Fourth, despite legal precedent, the international human rights framework has faced similar challenges as other ethics standards in being translated from principles to practice. Although there are ongoing efforts to conduct human rights due diligence (HRDD) on multinational companies (Business for Social Responsibility and the Global Network Initiative 2022), particularly in tech, empirical evidence on the extent of their findings and the implementation of recommendations across product life cycles remains sparse. In the past 10 years, less than 50 percent of human rights allegations raised by civil society worldwide received any response from technology companies (Business & Human Rights Resource Centre 2024). Public reporting from civil society and media sources indicates that companies have repeatedly failed at operationalizing a human rights framework in their products and business activities. A European Commission survey finds that although 37 percent of businesses surveyed were conducting HRDD, only about half of them covered the entire value chain of their business (McCorquodale and Nolan 2021). The ability to assess the current state of HRDD is heavily reliant on public disclosure by private companies, which relies on the veracity and comprehensiveness of the proffered information. Reliance on voluntary reporting of human rights practices in business has long been criticized and does not include liability or enforcement in cases of failure (ibid.).

Canada's Role in Establishing a Human Rights Framework as a Path Forward

Although this policy brief does not contend that an international human rights framework is the *only* viable path forward — and identifies several limitations of this approach — it does argue that this framework can provide a well-established starting point for a wide range of actors to build consensus around international AI governance if implemented effectively. In contrast to ad hoc ethics standards created by a siloed group of vested parties, the international human rights principles have served as a widely recognized set of governance norms for more than 70 years (Donahoe and MacDuffee Metzger 2019). To that effect, Canada has a unique opportunity to reaffirm its role and commitment to human rights in global AI policy negotiations within multilateral and multi-stakeholder processes by:

- **Promoting human rights as a foundational and cross-cutting framework for assessing the impact of AI systems and their governance:** Through mechanisms like the Global Digital Compact, Freedom Online Coalition and the forthcoming presidency at G7, Canada can reaffirm its commitment to the International Bill of Human Rights and actively promote this document as an integral instrument in mitigating AI-enabled risks while creating opportunities and advancing human dignity. Specifically, Canada can reinforce the binding nature of international human rights law and advocate for clearly articulated consequences for failure to comply with this law. It can also offer the example of the 2018 Toronto Declaration,[3] authored by Access Now and Amnesty International, that lays out multi-stakeholder and rights-respecting guidelines for protections against discrimination in machine-learning systems.
- **Encouraging public participation in international AI policy making:** Building on the co-construction process for the Montréal Declaration,[4] Canada can emphasize the importance of public consultation and co-designing to establish trust in rights-respecting AI policies. Against the backdrop of growing distrust between governments, multilateral institutions and civil society globally, Canada can intervene with a proven model of building multi-stakeholder consensus around shared AI principles.
- **Strengthening transnational and translational research to map human rights principles to ethical considerations in AI systems:** Studies on approaching AI ethics with an international human rights framework are still nascent; many scholars argue that AI ethics are too vague to be operational, and human rights law is insufficient and outdated in addressing modern technological challenges. Canada can invest in and encourage governments to support translational and transnational research to establish common vocabularies and bridge the gap between technologists, policy makers, civil society and international institutions.

Conclusion

Algorithmic fairness and ethics are critical research areas that affect how AI systems should be governed; however, they also pose risks around value pluralism and adopt vaguely worded concepts that promote ethics washing and poor implementation. The international human rights framework provides an existing universal foundation for centring AI systems on human conditions — a lens through which AI systems can be designed, developed and evaluated. Although translating human rights principles into practice remains a major area needing improvement, the adoption of such a framework compels states, private entities and non-state stakeholders to shift the discourse from machines and developers to communities and rights holders. International human rights principles carry cross-cultural legitimacy, establishing an equal footing in human dignity

3 See www.torontodeclaration.org/declaration-text/english/.

4 See https://montrealdeclaration-responsibleai.com/about/.

and agency and placing specific responsibilities on different actors across the AI value chain.

Works Cited

Ali, Sanna J., Angèle Christin, Andrew Smart and Riitta Katila. 2023. "Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs." In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 217–26. https://doi.org/10.1145/3593013.3593990.

Bietti, Elettra. 2020. "From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 210–19. https://doi.org/10.1145/3351095.3372860.

Birhane, Abeba. 2020. "Algorithmic Colonization of Africa." *SCRIPTed* 17 (2): 389–409. https://script-ed.org/article/algorithmic-colonization-of-africa/.

Blodgett, Su Lin, Solon Barocas, Hal Daumé III and Hanna Wallach. 2020. "Language (Technology) is Power: A Critical Survey of 'Bias' in NLP." *arXiv*. https://doi.org/10.48550/arXiv.2005.14050.

Buolamwini, Joy and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81: 1–15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Business & Human Rights Resource Centre. 2024. "Accusations and actions: A decade tracking tech company responses to human rights allegations." Briefing. September 16. www.business-humanrights.org/en/from-us/briefings/accusations-actions-a-decade-of-tech-company-responses-to-allegations-of-human-rights-abuse/.

Business for Social Responsibility. 2021. *Human Rights Due Diligence of Products and Services: Assessing the Downstream Value Chain*. July. www.bsr.org/reports/BSR-Human-Rights-Due-Diligence-Products-Services.pdf.

Business for Social Responsibility and the Global Network Initiative. 2022. *Human Rights Due Diligence Across the Technology Ecosystem*. September. https://eco.globalnetworkinitiative.org/wp-content/uploads/2022/11/Human-Rights-Due-Diligence-Across-the-Technology-Ecosystem_Ecosystem-Mapping_Oct2022.pdf.

Clinton, Robert N. 1990. "The Rights of Indigenous Peoples as Collective Group Rights." *Arizona Law Review* 32.

Corrêa, Nicholas Kluge, Camila Galvão, James William Santos, Carolina Del Pino, Edson Pontes Pinto, Camila Barbosa, Diogo Massmann et al. 2023. "Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance." *Patterns* 4 (10): 100857. https://doi.org/10.1016/j.patter.2023.100857.

D'Ambrosio, Paul J. 2016. "Approaches to Global Ethics: Michael Sandel's Justice and Li Zehou's Harmony." *Philosophy East and West* 66 (3): 720–38. https://dx.doi.org/10.1353/pew.2016.0068.

Donahoe, Eileen and Megan MacDuffee Metzger. 2019. "Artificial Intelligence and Human Rights." *Journal of Democracy* 30 (2): 115–26. https://dx.doi.org/10.1353/jod.2019.0029.

Donnelly, Jack. 1984. "Cultural Relativism and Universal Human Rights." *Human Rights Quarterly* 6 (4): 400–19. https://doi.org/10.2307/762182.

Freire, Ana, Lorenzo Porcaro and Emilia Gómez. 2021. "Measuring Diversity of Artificial Intelligence Conferences." *Proceedings of Machine Learning Research* 149: 39–50. https://proceedings.mlr.press/v142/freire21a.html.

Friedler, Sorelle A., Carlos Scheidegger and Suresh Venkatasubramanian. 2021. "The (Im)possibility of fairness: different value systems require different mechanisms for fair decision making." *Communications of the ACM* 64 (4): 136–43. https://doi.org/10.1145/3433949.

Friedman, Batya and David G. Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Gabriel, Iason. 2020. "Artificial Intelligence, Values, and Alignment." *Minds and Machines* 30 (3): 411–37. https://doi.org/10.1007/s11023-020-09539-2.

Gabriel, Iason and Vafa Ghazavi. 2021. "The Challenge of Value Alignment: From Fairer Algorithms to AI Safety." In *The Oxford Handbook of Digital Ethics*, edited by Carissa Véliz, 336–55. Oxford, UK: Oxford University Press.

Hogan, Mél. 2015. "Data flows and water woes: The Utah Data Center." *Big Data & Society* 2 (2). https://doi.org/10.1177/2053951715592429.

Hu, Lily. 2021. "Tech Ethics: Speaking Ethics to Power, or Power Speaking Ethics?" *Journal of Social Computing* 2 (3): 238–48. https://doi.org/10.23919/JSC.2021.0033.

International Institutions and Global Governance Program. 2012. *The Global Human Rights Regime*. Council on Foreign Relations. May. www.cfr.org/report/global-human-rights-regime.

Jobin, Anna, Marcello Ienca and Effy Vayena. 2019. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 1 (9): 389–99. https://doi.org/10.1038/s42256-019-0088-2.

Kaplan, Seth D. 2021. "The New Geopolitics of Human Rights." *PRISM* 9 (3): 76–89. https://ndupress.ndu.edu/Media/News/News-Article-View/Article/2846407/the-new-geopolitics-of-human-rights/.

Knight, Will. 2019. "Google appoints an 'AI council' to head off controversy, but it proves controversial." *MIT Technology Review*, March 26. www.technologyreview.com/2019/03/26/136376/google-appoints-an-ai-council-to-head-off-controversy-but-it-proves-controversial/.

Kwet, Michael. 2019. "Digital colonialism: US empire and the new imperialism in the Global South." *Race & Class* 60 (4): 3–26. https://doi.org/10.1177/0306396818823172.

Latonero, Mark. 2018. "Governing Artificial Intelligence: Upholding Human Rights & Dignity." *Data & Society* 38. https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

Leiner, Barry M., Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel et al. 2009. "A brief history of the internet." *ACM SIGCOMM Computer Communication Review* 39 (5): 22–31. https://doi.org/10.1145/1629607.1629613.

Liesenfeld, Andreas and Mark Dingemanse. 2024. "Rethinking open source generative AI: open-washing and the EU AI Act." In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1774–87. https://doi.org/10.1145/3630106.3659005.

McCorquodale, Robert and Justine Nolan. 2021. "The Effectiveness of Human Rights Due Diligence for Preventing Business Human Rights Abuses." *Netherlands International Law Review* 68: 455–78. https://doi.org/10.1007/s40802-021-00201-x.

Mittelstadt, Brent. 2019. "Principles alone cannot guarantee ethical AI." *Nature Machine Intelligence* 1 (11): 501–7. https://doi.org/10.1038/s42256-019-0114-4.

Morley, Jessica, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander and Luciano Floridi. 2021. "Ethics as a Service: A Pragmatic Operationalisation of AI Ethics." *Minds and Machines* 31 (2): 239–56. https://doi.org/10.1007/s11023-021-09563-w.

Morley, Jessica, Libby Kinsey, Anat Elhalal, Francesca Garcia, Marta Ziosi and Luciano Floridi. 2023. "Operationalising AI ethics: barriers, enablers and next steps." *AI & Society* 38: 411–23. https://doi.org/10.1007/s00146-021-01308-8.

Papyshev, Gleb and Masaru Yarime. 2024. "The limitation of ethics-based approaches to regulating artificial intelligence: regulatory gifting in the context of Russia." *AI & Society* 39 (3): 1381–96. https://doi.org/10.1007/s00146-022-01611-y.

Prabhakaran, Vinodkumar, Margaret Mitchell, Timnit Gebru and Iason Gabriel. 2022. "A Human Rights-Based Approach to Responsible AI." Preprint, *arXiv*. https://doi.org/10.48550/arXiv.2210.02667.

Rawls, John. 1993. "The Law of Peoples." *Critical Inquiry* 20 (1): 36–68. www.jstor.org/stable/1343947.

———. 1999. *Collected Papers*. Edited by Samuel Freeman. Cambridge, MA: Harvard University Press.

Roberts, Huw, Josh Cowls, Jessica Morley, Mariarosaria Taddeo, Vincent Wang and Luciano Floridi. 2021. "The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation." *AI & Society* 36: 59–77. https://doi.org/10.1007/s00146-020-00992-2.

Saltz, Jeffrey, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar et al. 2019. "Integrating Ethics within Machine Learning Courses." *ACM Transactions on Computing Education* 19 (4): 1–26. https://doi.org/10.1145/3341164.

Sambasivan, Nithya, Erin Arnesen, Ben Hutchinson, Tulsee Doshi and Vinodkumar Prabhakaran. 2021. "Re-imagining Algorithmic Fairness in India and Beyond." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 315–28. https://doi.org/10.1145/3442188.3445896.

Sandel, Michael J. 2011. *Justice: What's the Right Thing to Do?* New York, NY: Farrar, Straus and Giroux.

Schwartz, Shalom H. 2006. "A Theory of Cultural Value Orientations: Explication and Applications." *Comparative Sociology* 5 (2–3): 137–82. www.researchgate.net/publication/304824378_A_theory_of_cultural_value_orientations_Explication_and_applications.

Sen, Amartya. 2012. "Values and justice." *Journal of Economic Methodology* 19 (2): 101–8. https://doi.org/10.1080/1350178X.2012.683601.

Stanford Institute for Human-Centered Artificial Intelligence. 2024. "Chapter 7: Policy and Governance." In *Artificial Intelligence Index Report 2024*, 1–50. https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report-2024_Chapter_7.pdf.

Tesón, Fernando R. 1985. "International Human Rights and Cultural Relativism." *Virginia Journal of International Law* 25: 869–98. https://ir.law.fsu.edu/articles/30/.

Tsosie, Rebecca. 2007. "Cultural Challenges to Biotechnology: Native American Genetic Resources and the Concept of Cultural Harm." *Journal of Law, Medicine & Ethics* 35 (3): 396–411. https:// doi.org/10.1111/j.1748-720X.2007.00163.x.

Turner Lee, Nicol and Caitlin Chin-Rothmann. 2022. "Police surveillance and facial recognition: Why data privacy is imperative for communities of color." Brookings. April 12. www.brookings.edu/articles/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color/.

United Nations Human Rights Office of the High Commissioner. 2011. *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*. New York, NY: United Nations. www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.

Wall, Christopher. 1998. "Human Rights and Economic Sanctions: the New Imperialism." *Fordham International Law Journal* 22 (2): 577–611. https://ir.lawnet.fordham.edu/ilj/vol22/iss2/7.

Waltz, Susan. 2002. "Reclaiming and rebuilding the history of the Universal Declaration of Human Rights." *Third World Quarterly* 23 (3): 437–48. https://library.fes.de/libalt/journals/swetsfulltext/13640294.pdf.

Webster, Graham, Rogier Creemers, Elsa Kania and Paul Triolo. 2017. "China's Plan to 'Lead' in AI: Purpose, Prospects, and Problems." DigiChina, Stanford Cyber Policy Center, Stanford University. August 1. https://digichina.stanford.edu/work/chinas-plan-to-lead-in-ai-purpose-prospects-and-problems/.

Whittlestone, Jess, Rune Nyrup, Anna Alexandrova, Kanta Dihal and Stephen Cave. 2019. *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. London, UK: Nuffield Foundation.

Yeung, Karen, Andrew Howes and Ganna Pogrebna. 2020. "AI Governance by Human Rights – Centered Design, Deliberation, and Oversight: An End to Ethics Washing." In *The Oxford Handbook of Ethics of AI*, edited by Markus D. Dubber, Frank Pasquale and Sunit Das, 77–106. New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.5.

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Research Director, Transformative Technologies Tracey Forrest
Director, Program Management Dianna English
Program Manager Jenny Thiel
Publications Editor Christine Robertson
Publications Editor Susan Bubak
Graphic Designer Sepideh Shomali

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

CIGI