SECTION 5

The Human Rights Lens

While different types of companies will have different types of human rights risks, all companies have a responsibility to respect human rights. In order to determine which rights their activities may impact and how, the UNGPs call on governments and companies to consider a full suite of rights recognized under widely ratified human rights conventions and treaties (the so-called International Bill of Rights) as the starting point for their analysis. This is especially important in the context of AI, given its broad application across a wide range of contexts, including healthcare, education, financial services, law enforcement, retail, transportation infrastructure, and many more.

This brief focuses on government interventions that may affect privacy, freedom of expression, and non-discrimination. This aligns with GNI's focus and is the segment of the AI and human rights field where GNI is best placed to comment. Human rights are interdependent and interrelated, so adverse impacts on privacy, freedom of expression, and non-discrimination can have implications for a broad range of other rights. While the Universal Declaration of Human Rights and many of its progeny were developed before the advent of digital technologies, their respective provisions on freedom of expression all share language emphasizing that this right must apply "through any media" and "regardless of frontiers." The UN Human Rights Committee in its General Comment No. 34 (GC34) has subsequently clarified that, under the International Covenant on Civil and Political Rights (ICCPR), "[a]ny restrictions on the operation of websites, blogs or any other internet-based, electronic or other such information dissemination system, including systems to support such communication, such as internet service providers or search engines, are only permissible to the extent that they are compatible with [Article 19] paragraph 3." The UN Guiding Principles on Business and Human Rights ("UNGPs") stipulate that, "[i]n meeting their duty to protect [human rights], states should . . . [e]nsure that . . . laws and policies governing the creation and ongoing operation of business enterprises . . . do not constrain but enable business respect for human rights."

5.1

High Level Analysis

5.1.1 Freedom of Expression

Article 19 of the Universal Declaration of Human Rights, as well as Article 19 of the International Covenant on Civil and Political Rights ("ICCPR"), together with accompanying interpretation by the UN Human Rights Committee (primarily through GC34) and other human rights sources, provide an authoritative basis for interpreting the impact of government interventions in AI. Interpretation of ICCPR Article 19 centers around the so-called "three-part test," using the principles of legality, legitimacy, and necessity/proportionality.[24]

The principle of "legality" focuses on the processes by which states act to restrict freedom of expression, as well as the manner in which such restrictions are articulated. As such, it reflects concepts of notice and transparency that are fundamental to the rule of law. According to the Human Rights Committee, any intervention impacting freedom of expression must be prescribed by law, be publicly accessible, and formulated with sufficient precision to enable individuals to regulate their conduct accordingly (see GC34 para. 25).

The separate principle of "legitimacy" insists that laws restricting expression can only be justified in order to achieve specific, enumerated purposes. Article 19(3) of the ICCPR describes these as "respect for the rights or reputations of others" and "the protection of national security or of public order, or of public health or morals." Meanwhile, Article 20 states that "propaganda for war" and "advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence" shall be prohibited by law. While international law gives states significant room to determine what sorts of activities can be understood to sufficiently impact these purposes so as to justify restrictions, that discretion is not unlimited (see GC34 para. 26).

The final principle of necessity requires states seeking to restrict expression to "demonstrate in specific and individualized fashion the precise nature of the threat, and the necessity and proportionality of the specific action taken, in particular by establishing a direct and immediate connection between the expression and the threat." (see GC34 para 35) The term "proportionality," which is best understood as an element of "necessity" but at times is referenced as a stand alone limiting principle, limits restrictive laws to those that are "appropriate to achieve

[24] A more detailed analysis of these principles can be found in GNI's "Content Regulation & Human Rights Policy Brief," (2020).

their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function." (see GC34 para. 34)

5.1.2 Privacy

Protections against arbitrary or unlawful interference with privacy are established in the UDHR, the ICCPR, and most regional human rights treaties. According to various UN sources, the same legality, necessity, and proportionality considerations discussed above also apply with respect to government interventions that impact the right to privacy. In addition, international practice emphasizes that any interference with privacy must be accompanied by effective safeguards, such as independent oversight, access to remedies, and protection against arbitrary or discriminatory application, particularly in the context of surveillance and data retention regimes.

5.1.3 Non-Discrimination

ICCPR Article 2 and UDHR Article 2 prohibit discrimination based on race, sex, political opinion, or other protected characteristics.[25] This also requires governments not to cause discrimination. According to GC31, this duty also applies extraterritorially where the state has effective control.

[25] This includes preventing bias across all types of interventions that impact both public and private sector actors

5.2

Analyzing the Human Rights Impacts of Government Interventions in AI

5.2.1 Infrastructure

Legality: Due to their capital and time-intensive nature, infrastructure-related interventions often require cooperation between executive and legislative branches. Transparency and procedural provisions associated with budget, procurement, and export-control decisions can also help such interventions meet the notice and due process elements of the legality test. However, as is the case across all levels of the AI ecosystem, these elements are often harder to demonstrate and satisfy in the context of soft governance and diffuse or informal interventions.

Legitimacy: Most infrastructure-level interventions are justified broadly on national security and/or economic development grounds. These justifications often meet the legitimacy principle. However, it is important to ensure that the "race" to compete geopolitically, militarily, and economically isn't used by government actors as a pretext or blank check to justify interventions that are not rights respecting or are susceptible to politicized implementation.

Necessity / Proportionality: Infrastructure-level interventions tend to have indirect and diffuse impacts on freedom of expression and privacy, which can make it harder to establish a "direct and immediate connection" between the action and any related restriction. The breadth of the potential downstream impacts of such actions on both freedom of expression and privacy nevertheless tend to justify particularly careful proportionality analysis, in order to understand whether such actions and their likely intended and unintended consequences can be considered the "least restrictive" means for achieving relevant policy objectives, in other words, that no less rights-intrusive measure could achieve the same policy objectives.

Example: Export Controls

Impact on Freedom of Expression: While export controls on national security grounds often have a local legal basis, export controls can have significant unintended consequences, including but not limited to restricting access to computing power and scientific capacity by people in countries unassociated with the national security concern in question. In some cases, export controls have also led to retaliatory policies from targeted nations, which may impact scientific development and freedom of expression of the source nations.[26] Although export restrictions on national security grounds are often targeted at specific nations, collateral impacts on the citizens of the target nations and in some cases third countries (including, in some scenarios, those in the country imposing the restriction) may be relevant when determining the proportionality of a measure. These concerns are generally ameliorated in situations where the policy justification for export controls is tied directly to human rights objectives, such as enhancing privacy or limiting surveillance.[27]

Impact on Freedom of Privacy: The aforementioned efforts to trace chip origins to prevent diversion may compromise security and privacy if user devices become trackable or vulnerable to security backdoors.[28]

Impact on Non-Discrimination: Export controls, foreign model usage restrictions, and local sourcing requirements target certain countries or companies, and can not only result in restricting access by individuals in target states to controlled technologies but can also institutionalize geopolitical bias while stigmatizing decisions related to specific technologies, nations, companies, and workers within the source state – all of which may impact the right to non-discrimination.[29] Such selective regulation may also undermine trust and cooperation in international AI governance, further increasing the divide in the development and use of AI technologies, especially in the nations subject to such controls.

[26] Thereby potentially violating International Covenant on Economic, Social and Cultural Rights (ICESCR) Article 15(1)(b)

[27] Jennifer Brody, "How Stronger Export Controls Can Better Protect Human Rights," Freedom House (8 Feb. 2024).

[28] Luke O'Grady, "Congress' Proposed Chip Security Act Threatens to Create New Cyber Vulnerabilities in U.S. Semiconductors," Center for Cybersecurity Policy and Law (15 July 2025).

[29] For example, influencing decisions on research collaborations with Chinese institutes and companies in the UK and the US

5.2.2 Development

Legality: By contrast with infrastructure-focused interventions, government interventions at the development stage can be more directly targeted at achieving certain expressive or surveillance outcomes. As such, it is important that such efforts are authorized and conducted pursuant to valid, duly enacted, and clear laws and regulations. It is also vital that the methods for carrying out such actions are transparent and rule-of-law compliant.

Legitimacy: The same types of legitimate objectives (economic development, national security, sovereignty) are often deployed to justify all kinds of government interventions across the AI value chain. However, where those actions have foreseeable (even if unintended), direct, negative impacts on human rights, the burden becomes stronger on governments to more explicitly justify these actions and explain how it is trying to avoid or mitigate those impacts. In this sense, the legitimacy analysis is reinforced by the necessity principle's insistence that governments engage in the exercise of analyzing likely impacts in order to ensure that the proposed action is narrowly tailored and appropriate to the intended purpose.

Necessity / Proportionality: Government actions targeting the AI development stage are more likely to produce direct impacts than those directed toward infrastructure. At the same time, by virtue of their relatively upstream nature, these actions can have broad impacts, especially as they pertain to innovation, strategic business decisions, product dissemination, and competition. Government approaches at this stage that are designed to allow for experimentation, flexibility, and adaptation may be more consistent with the goal of protecting human rights; while those that mandate specific ideologies or political perspectives (e.g. by making requirements related to model inputs and outputs) are more likely to result in human rights harms. In general, government actions that deepen uncertainty and ambiguity regarding expectations and consequences related to AI development, while leveraging heavy penalties or threats, are more likely to result in human rights harms.

Example: AI (human rights) Risk Assessments Mandates

Requirements for AI model developers to conduct risk assessments typically serve legitimate purposes, especially when they are grounded in international human rights. Some examples of potentially disproportionate rights impacts from the presence or absence of risk assessment mandates are illustrated below:

Impact on Freedom of Expression: Overbroad risk assessment regulations not fully grounded in international human rights norms can negatively impact freedom of expression. For example, in China, developers may be required to censor content that should be protected under IHRL, as a result of mandatory "risk assessments" undertaken to ensure compliance with "core socialist values". Conversely, the absence of rights-protecting risk assessment regulations can also negatively impact freedom of expression, for example by allowing models to be developed that fail to anticipate and address downstream impacts such as over- or under-moderation of content. The likelihood of preventing, mitigating, and remedying such harms, is also exacerbated where models lack transparency or explainability, which in turn can have a chilling effect on freedom of expression.

Impact on Privacy: The absence of laws and regulations can allow AI models to integrate unchecked capability to collect, process, and share personal data without adequate safeguards, increasing the risk of products being used for downstream surveillance, as well as increasing the threat surface for cybersecurity and data breaches. Meanwhile, strict liability or inconsistent and/or politicized enforcement of such laws can lead to self-censorship by model developers and result in unfair competition. AI risk assessments can help protect user privacy with respect to both model inputs and outputs, while offering developers an important degree of flexibility in product design. Furthermore, without risk assessments, developers may overlook how models can be attacked to reveal personal information from their training data.

Impact on Non-Discrimination: Like the impact on privacy above, AI risk assessments generally help to protect the right to non-discrimination, while the absence of such assessments can lead to unaddressed systemic biases that, when deployed into automated decision-making systems, can lead to discriminatory outcomes in areas such as law enforcement, hiring, access to healthcare, and content moderation.

5.2.3 Deployment

Legality: Government interventions at the AI deployment stage are simultaneously easier to justify and more susceptible to abuse for ideological, political, or other inappropriate purposes (see example below). Given their proximity to and likelihood of impacting end uses of AI, it is especially important that these actions are clearly authorized, narrowly scoped, and carefully deployed.[30] The government's responsibility for any resulting negative human rights harm is most directly established where the government itself is the one that causes that impact through its own use of AI.

For individuals to be able to understand and navigate these boundaries, restrictions must clearly and precisely define both what is prohibited and who can be held responsible for failing to enforce the prohibition. Any vagueness or ambiguity can cause individuals to refrain from exercising their rights and lead intermediaries to be overly aggressive in censoring expression for fear of being held in violation of the law.

Legitimacy: Given the focus of many of the examples cited in Section 4.3 on regulating content and conduct produced through, with, or by AI, it is worth emphasizing the risk of such actions creating chilling effects. Whenever expression is prohibited, the mere possibility of being accused of violating the law or being subject to costly court proceedings can cause individuals not to express themselves and companies to refrain from facilitating expression.

Necessity / Proportionality: Government restrictions on expressive uses of AI (e.g. through direct censorship, strict liability, or the prosecution of AI users/uses) must be clearly articulated and narrowly tailored. This is especially important in the context of laws that outsource the enforcement of speech regulation to private actors of varying sizes, business models, and capacities. As the Human Rights Committee explained in GC34, laws regulating speech "may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution."

This concern does not prohibit governments from apportioning liability to AI developers, deployers, or users for narrowly and clearly defined harms. Indeed, it is incumbent on governments to identify when and how such liability attaches, in order to provide all actors with the notice and predictability that they need to be able to conduct themselves appropriately in accordance with the law. It is also critical to ensure that any party that is harmed has access to appropriate remedies, as well as that anyone accused of being responsible for harm is guaranteed appropriate due process. As the UNGPs make clear, the responsibility for guaranteeing appropriate and meaningful remedy applies to both states and companies.

[30] In other words, that such interventions are legal, legitimate, and necessary/proportionate

Example: AI in Surveilliance

Due to its sensitive nature, the specific uses of AI in government surveillance may not be fully transparent, but the use of surveillance technologies must nevertheless be authorized and governed by local laws.[31]

Impact on Freedom of Expression: The use of AI in surveillance—such as facial recognition—can generate a chilling effect on freedom of expression and other rights, as individuals may self-censor or alter their behavior out of fear of being monitored, (mis)identified, or (mis)targeted.

Impact on Privacy: In rights-protecting jurisdictions, the existence of rights-protecting laws and legal frameworks, including robust data protection and privacy laws may help safeguard citizens from privacy infringements, including from overbroad surveillance (such as the ban on facial recognition in law enforcement by many US jurisdictions). Conversely, the lack of such laws may enable unchecked collection, processing, and sharing of personal data by governments and private actors, increasing the intrusiveness of surveillance, raising the impact of data breaches, and other violations of individuals' privacy rights.[32]

Impact on Non-Discrimination: Surveillance can lead to profiling based on protected characteristics, resulting in discriminatory treatment from law enforcement, exclusion from services, targeted law enforcement actions, or social stigmatization.

[31] Various legal bases for mass surveillance in multiple jurisdictions are detailed in this Human Rights Watch article. Meanwhile, efforts are under way to increase transparency, e.g. the EU AI Act Annex III (law enforcement use cases defined as a high risk system) and Article 13 (greater transparency for high risk systems).

[32] See this CSIS source for a discussion of how data privacy should be protected in responsible AI