GLOBAL NETWORK INITIATIVE

Policy Brief on Government Interventions in AI

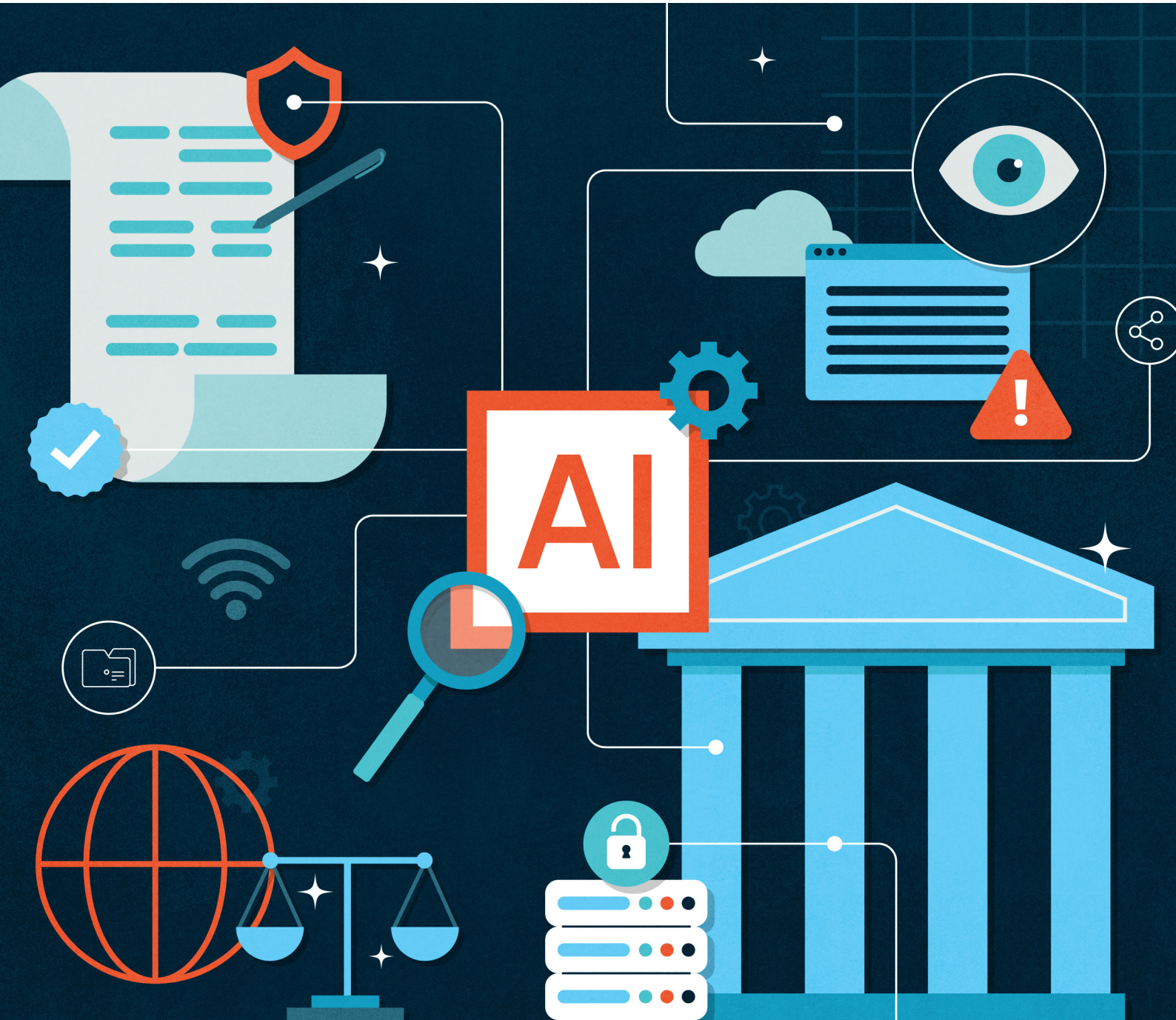

Table of Contents

Executive Summary

This Policy Brief ("Brief"), developed by the Global Network Initiative (GNI) and its multistakeholder AI Working Group (AIWG), establishes a framework for understanding the human rights implications of government interventions in artificial intelligence (AI), with a particular focus on the rights to freedom of expression, privacy, and non-discrimination.

This analysis is informed by the diverse perspectives and expertise of the AIWG and GNI's broader membership, as well as GNI's 17+ years of experience in working collaboratively with tech companies, civil society, investors, and academics to understand and guide responsible business conduct.[1] The Annex to this Brief unpacks the GNI framework and examines its applicability to AI-related corporate conduct in more detail.

The Brief articulates a taxonomy of five broad categories of government interventions ("hard" and "soft" governance, investment, procurement, and informal influence) across three segments of the AI value chain (AI infrastructure, development, and deployment). It provides illustrative examples of each category of government intervention in the context of each aspect of the value chain. The Brief unpacks the international human rights to freedom of expression, privacy, and non-discrimination at a high-level before concluding with recommendations for governments, companies and civil society.

This analysis reveals both opportunities to advance human rights and risks of rights violations from government interventions in AI. Illustrative examples from diverse regions demonstrate this duality: interventions such as mandatory human rights assessments, risk-tiered taxonomies of use-cases with corresponding levels of required risk management, privacy laws, investments to enhance public access, and rights-protecting procurement guidelines can promote and protect human rights. By contrast, overbroad censorship mandates, discriminatory surveillance, restrictive export controls, and the absence of rights-protecting laws and regulations and related state capacity lead to negative rights impacts and entrenched inequalities.

Using international human rights law to ground government interventions in AI, as well as responses to them, facilitates the protection of rights and helps in building public trust, thereby guiding future AI developments toward inclusive and sustainable progress.

[1] For more information on GNI's framework for responsible tech company conduct and unique assessment process, see here and here.

SECTION 1

Introduction

The Global Network Initiative ("GNI") is the leading multistakeholder forum for accountability, shared learning, and collective advocacy on government and company policies and practices at the intersection of technology and human rights. GNI sets a global standard for responsible company decision-making to promote and advance freedom of expression and privacy rights across the technology ecosystem.

The rapid development of artificial intelligence ("AI") is increasingly influencing the information environment with associated impacts on human rights, including freedom of expression and privacy.[2] While there is substantial literature on the conduct of AI risk assessments, much of it focuses on broadly defined concepts of AI safety, AI ethics, and responsible AI, which are not grounded in international human rights norms and do not provide clear, consistent guidance on identifying and mitigating human rights risks.[3] Other analyses provide taxonomies of government interventions in AI, but again without focusing on human rights.[4] Meanwhile, resources related to human rights and AI often do not focus on the significant roles played by governments in the development of relevant national and international AI landscapes.[5]

And yet governments around the world are increasingly intervening in the infrastructure, development, and deployment of AI systems, with significant implications for human rights. Governments influence AI through policies, regulations, deployments, and investments that can either safeguard or undermine a broad range of human rights, including but not limited to the rights to freedom of expression, privacy, and non-discrimination. The framework provided by international human rights law provides many benefits, including a shared, global set of norms to evaluate interventions by state actors and to hold them accountable.[6] A rights-based approach helps ensure that government action supports transparency, due process, accountability,

[2] For the purposes of this Brief, GNI will use the OECD definition of AI as "...a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment." For the avoidance of doubt, this Brief is not limited to Generative AI.

[3] For example, NIST's AI Risk Management Framework or the UK Government's AI Management Essentials Tool (AIME), among others.

[4] For example, CFR's Taxonomy for Navigating the Global Landscape of AI Regulation

[5] For example, ECNL on the rights impacts of LLMs in content moderation, The Future of Free Speech and CDT's Artificial Intelligence & The First Amendment, CDT's AI Governance Lab and BSR's Human Rights Across the Generative AI Value Chain. Some literature from GNI members do provide a focused analysis of some government interventions in AI, such as Article 19's Red Lines for AI on government surveillance.

[6] Pielemeier, J., "AI & Global Governance: The Advantages of Applying the International Human Rights Framework to Artificial Intelligence," UN University Center for Policy Research (Feb. 26, 2019).

protecting individuals from harm,promoting global trust in AI governance, and providing the necessary legal recourse and remedies should individuals be subjected to any harm.

GNI, through its multi-stakeholder AI Working Group ("AIWG"), has produced this Policy Brief ("Brief") to help clarify and articulate the human rights implications of government interventions in AI, and put forward rights-respecting recommendations for government, company, and civil society stakeholders. This Brief draws upon responsible business conduct frameworks, insights from GNI's multi-stakeholder membership, and GNI's seventeen years of work seeking to uphold freedom of expression and privacy across the tech sector.

This Brief presents a high-level framework for understanding government interventions. It is intended to help government officials across various branches, levels, and roles, as well as other stakeholders, understand the implications of government action on the rights to freedom of expression, privacy, and non-discrimination, so that decisions can be designed and scoped in accordance with human rights principles. All government actions should be evaluated across the full range of cultural, civil, economic, labor, political, and social rights. While this Brief focuses on freedom of expression, privacy and non-discrimination, which GNI considers to be particularly salient to AI, the interpretative principles outlined in Section 5 can and should help inform broader human rights analyses.

The Brief is informed and substantiated by scenarios that were publicly available or foreseeable as of late 2025. While it does not provide in-depth or conclusive analysis on any such examples or scenarios, the Brief is intended to help inform such analysis by GNI, its members, and other key actors going forward. In this Brief, we use hyperlinks when referring directly to specific examples or concepts, while using footnotes where additional explanation or description is needed.

SECTION 2

The AI Value Chain

Understanding where and how government interventions occur within the AI ecosystem requires a clear understanding of the AI value chain. While a number of models for explaining the AI value chain exist, GNI has chosen to use the following, simplified three-part AI value chain for the purpose of this Brief:[7]

Value Chain Stage	Description	Example of Company Actors
Infrastructure 	Encompasses the broad range of physical and digital inputs necessary for AI products and services to be developed and deployed, including hardware (e.g., chips/GPUs), infrastructure (e.g., cloud and data centers, hosting), and inputs (e.g., data, energy, and scientific knowledge).	**Semiconductors:** NVidia, Broadcom, Google, Meta, etc. **Data Centers:** Amazon, Google, Microsoft, Cloudflare, etc. **Host:** Hugging Face, GitHub, etc. **Energy:** Utility companies
Development 	Upstream actions taken to design, test, prepare, and produce and update AI products/services, including: collecting, refining/curating, and labeling training data; designing, testing, training, and evaluating models; and work to identify and tailor AI tools to specific use cases/ business models.	**Data:** Scale AI, Appen, iMerit, etc. **Model Developers:** OpenAI, Google, Anthropic, Meta, Baidu, Alibaba, etc.

[7] For example, BSR: Human Rights Across the Generative AI Value Chain, OECD: Artificial Intelligence & Responsible Business Conduct, French Treasury: The Artificial Intelligence Value Chain

Deployment[8] 	Downstream actions and scenarios that can take place after AI services/products are put into use at a commercial or public scale. This would cover both intended, prescribed uses, as well as unintended uses, whether or not they are explicitly proscribed by a developer, deployer, or government. Also covers initial deployment and updates thereof.	**Model Deployers:** Most model developers also deploy, as do numerous GNI members, third parties and smaller companies, as well as individual end users.

[8] The definition of deployers here is distinct from the EU AI Act's definitions, which excludes personal (i.e. non-professional) AI use.

SECTION 3

Types of Government Intervention in AI

Various types of government interventions in AI could impact the extent and manner in which AI services are developed, deployed, and used, leading to commensurate impacts on rightsholders. This section of the Brief outlines five types of known and foreseeable government interventions in AI, including actions and inactions:[9]

Type of Intervention	Description
"Hard" governance 	Binding legal, regulatory, and judicial measures that establish conditions for, align liability around, or otherwise govern the use and impact of AI nationally or in specific sectors. These actions are mostly domestic but can also be extraterritorial in their intended or unintended impact (e.g., export controls, binding treaty commitments, court orders). They can be AI-specific (e.g. EU AI Act), or general but applicable to AI (e.g. copyright laws, data protection laws).
"Soft" governance 	Actions that governments may or may not take to establish, support, or incentivize (positively or negatively) the development of non-binding standards, guidelines, or principles directed specifically toward AI. These actions can be taken domestically or multilaterally.
Investment 	The use of public resources and/or the incentivization/ coordination of private investments (e.g., public-private partnerships) that facilitate or direct AI-related research and development (e.g., funding scientific research), establish infrastructure and technical conditions for AI-development

[9] CFR's Taxonomy for Navigating the Global Landscape of AI Regulation divides hard and soft governance interventions in a similar manner

	(e.g., developing compute capabilities, aka “sovereign AI”), or support particular sub-sectors or actors within the larger AI-sector (e.g., providing seed capital to start-ups or investing in national champions, or providing subsidies or tax incentives for deployments).
Procurement & usage 	Ways that governments can use their purchasing power to incentivize or otherwise impact the development of AI products and services, as well as the various ways that governments themselves may set conditions for the deployment of AI by the public sector, including via private sector actors.
Diffuse & informal interventions 	Actions that governments take that (i) are not specifically focused on but are likely to impact the development and deployment of AI (e.g., energy policy, educational investments), as well as those that (ii) are specifically focused on AI but are more informal and often less-transparent (e.g., commercial diplomacy on behalf of domestic companies, “revolving doors,” jaw-boning).

SECTION 4

Taxonomy of Government Involvement Across the AI Value Chain

Based on the framing described in Sections 2 and 3, the taxonomy in this section will be structured according to the table below, in value-chain order. The sub-sections of each part of the taxonomy are shown within the table:

Stages in AI value chain	"Hard" governance	"Soft" governance	Investment	Procurement & usage	Diffuse & informal interventions
Infrastructure	4.1.1	4.1.2	4.1.3	4.1.4	4.1.5
Development	4.2.1	4.2.2	4.2.3	4.2.4	4.2.5
Deployment	4.3.1	4.3.2	4.3.3	4.3.4	4.3.5

The descriptive content in each sub-section is supplemented with citations to existing proposed or documented actions by governments across the world. These citations are not exhaustive, but were chosen to represent relevant, high-profile examples from both the Global North and Global South based on open-source investigations conducted in mid-2025. The analysis of rights impacts (Section 5) and recommendations (Section 6) also draws upon the examples discussed in this section.

4.1

AI Infrastructure

4.1.1 “Hard” Governance in AI Infrastructure

Governments are increasingly securing AI semiconductor supply chains and restricting technology transfers for national security purposes, as well as promoting self-sufficiency, and economic development. Measures deployed include mandates for public investments in AI infrastructure (which will be discussed under “Investments in AI infrastructure”), and conditional local content requirements.[10]

Additionally, some countries have imposed export controls on advanced semiconductor equipment and high-performance AI chips, primarily on national security grounds.[11] In some cases, these approaches have been expanded in order to address concerns around circumvention, where potentially intrusive chip tracking mechanisms have been explored.[12] In addition, some countries have cited the illegal use of restricted technologies to justify broad bans on certain commercial semiconductors.[13]

4.1.2 “Soft” Governance in AI Infrastructure

While most AI infrastructure interventions tend to involve legislation, regulation, investments (see subsection below), or some combination of all three, some governments have adopted a soft policy approach to AI infrastructure. For example, the UK has developed AI Growth Zones to stimulate the rapid deployment of AI data centres, while its National Semiconductor Strategy calls for the UK to retain its leading position in semiconductor R&D and design, both without the use of dedicated investment or legislative instruments.

[10] For example, in the US and China

[11] Examples include controls on advanced lithography equipment from the Netherlands and Japan, the U.S. October 7, 2022 Export Administration Regulations update, the U.S. AI diffusion rule, and Taiwan’s ban of exports to specific Chinese companies.

[12] The emergence of Deepseek raised concerns in the United States about circumvention of controlled semiconductors via third countries, prompting the Commerce Department’s recommendations on diversion risks and legislative proposals like the Chip Security Bill, which could introduce chip tracking mechanisms.

[13] For instance, the U.S. has restricted use of Huawei Ascend AI semiconductors due to alleged illegal use of U.S.–origin technologies.

Beyond the legal mandates for local sourcing of technology mentioned earlier, officials in China have also informally pressured local companies to source locally produced AI semiconductors. In relation to data infrastructure, governments hosting data enrichment workers, such as Kenya, have participated in ILO-led national social dialogue to promote decent work conditions for such workers.

4.1.3 Investments in AI Infrastructure

Governments worldwide have committed public resources for the domestic development of AI hardware and software, semiconductor manufacturing capabilities, and local AI datacenters. Selected examples include the development of an indigenous GPU as part of the INDIAai initiative, financing for semiconductor manufacturing in Europe in the European Chips Act (and similarly in the U.S. Chips Act), South Korea's $1 billion fund for local semiconductor manufacturers, Brazil's $4 billion fund to (among others) invest in AI infrastructure, and Canada's AI Compute Challenge fund to spur private-sector data centres in Canada that specialize in AI compute.

Governments also provide in-kind investments (including subsidies) to incentivise local infrastructure development, for example, the US facilitates leasing federal sites owned by the Defense and Energy departments with rapid access to large amounts of clean energy for AI data centers and new clean power facilities, in return for sourcing an "appropriate share" of American-made semiconductors.

In addition, governments have invested in various forms of "public compute" targeting local researchers and SMEs, such as Canada's AI Compute Access Fund that funds private cloud AI compute, publicly funded AI supercomputer facilities in Japan, or public compute facilities and subsidised open GPU marketplaces as part of the INDIAai initiative.

4.1.4 Procurement and Use of AI Infrastructure

In order to promote national AI semiconductor supply chains, governments in China and Russia impose local content requirements in state-procured AI infrastructure, especially semiconductors. Furthermore, public tenders for compute under India's INDIAai Mission need to comply with local content requirements under its "Make in India" initiative.

4.1.5 Diffuse and Informal Interventions in AI Infrastructure

Beyond the formal interventions mentioned above, some governments use informal methods to influence the direction of public and private sector decision-making. For example, when allegations emerged that Singapore was being used to trans-ship export-controlled NVIDIA GPUs to China, a government minister clarified that Singapore is not legally bound to enforce the unilateral export controls of other countries, but that companies operating in Singapore are expected to consider such regulations when relevant, effectively placing informal expectations on Singapore-based trading firms.

4.2

AI Development

4.2.1 "Hard" Governance in AI Development

Governments globally are increasingly regulating how AI models are developed, with risk management requirements being the most common thread among such regulations. The EU AI Act, the world's first, comprehensive, dedicated AI regulation, mandates data governance (Art. 10), and transparency (Art. 13) for high-risk systems, while classifying some generative AI models as posing with systemic risk (Art. 51) triggering additional obligations (Art. 55).[14] Similar laws are being drafted in Brazil (and other Latin American nations) and have passed in South Korea. Other countries, like the UK, have taken lighter or sector-specific approaches, or, like the U.S., have remained largely hands-off or have taken a sub-national approach.

Regulations relevant to AI are supplemented by cross-sectoral (i.e. not AI specific) legislation covering topics such as data localisation, copyright, data protection in multiple jurisdictions, for example, GDPR and the DSA. Amidst concerns that AI regulations could hamper innovation, some jurisdictions have introduced AI regulatory sandboxes, sectoral in the UK and Singapore, cross-sectoral in South Korea and Norway.[15] [16] In some jurisdictions, AI is being regulated under existing regulations, such as in India, while the need for dedicated regulations is explored.

Liability regimes for AI have been discussed globally but are still evolving. While countries like China place strict liability for AI-generated content on developers under its Interim Measures (which has subsequently been enforced in court rulings), discussions in the EU have revolved around the now-withdrawn AI Liability Act, which would have introduced a presumption of causality requiring AI providers to prove that their system did not cause alleged harms. In some jurisdictions, some are seeking to use existing negligence and product liability laws to impose liability, as is the case in the EU under its updated Product Liability Directive.

[14] A high risk system is defined in Art. 6 of the EU AI Act

[15] A controlled environment that allows developers to test and validate innovative AI systems under regulatory supervision for a limited time, aiming to foster innovation while identifying and mitigating potential risks.

[16] There been been discussions within academia and civil society on ensuring such sandboxes are inclusive and promote responsible AI development

Some governments seek to regulate training data. The EU AI Act (Art. 53(1)(c)) applies copyright obligations to model training datasets, and the UK is drafting similar rules. Beyond copyright, France, and Italy have taken GDPR-based legal actions against unauthorized use of personal data for model training, while China requires training datasets to be embedded with "core socialist values."

Governments also regulate model outputs in some jurisdictions. For example, watermarking of AI-generated content is required in the EU AI Act (Art. 50) and imposes additional obligations (Art. 55) on general-purpose AI models deemed to have systemic risk (Art. 51), as well as in China and India. Additionally, output filtering features in some jurisdictions: the EU AI Act (Art. 52) targets illegal content, while China and Russia impose stricter controls based on their own definitions of illegal content (which have requirements for political and ideological alignment that are not present in rights-protecting jurisdictions). Some EU jurisdictions (e.g. Italy) have taken a more prescriptive approach in their implementation of the EU Act by introducing criminal offenses and mechanisms for content traceability and authenticity. In addition, China specifically requires models to be licensed ahead of deployment as an additional way of ensuring compliance with censorship.

Finally, as with AI infrastructure, national security concerns have also prompted export controls: for example, China restricts certain AI model exports, while the U.S. has considered export restrictions on model weights.

Many in civil society and even some governments argue that current laws, as well as related enforcement and remedy mechanisms, are insufficient to address emerging AI risks.[17] These gaps can create challenges for accountability and transparency, meaning that the lack of strong AI regulation can ultimately harm human rights.

4.2.2 "Soft" Governance in AI Development

In addition to or in lieu of relying on binding regulations, **some jurisdictions have enacted principle-based approaches to build trust and mitigate risk in model development.** For instance, Australia's AI Ethics Principles offer high-level, non-binding guidance on fairness, privacy, and accountability, emphasizing co-regulation between government, industry, and civil society. Meanwhile, Singapore's AI Verify provides voluntary testing and benchmarking tools for responsible AI, and its Model AI Governance Framework proposes an AI governance framework emphasizing interoperability and accountability. Similar initiatives exist in Japan (AI Guidelines for

[17] See examples from US and pan-African civil society

Business), India (AI Safety Institute), and the EU, which released a voluntary AI code of practice that complements the EU AI Act, and the Inter-American Development Bank's (IDB) fAIr LAC+ platform in Latin America and the Caribbean.[18]

In relation to **training data**, the applicability of existing copyright laws in the context of AI training data is currently being contested in multiple jurisdictions. The UK is exploring a pioneering collective licensing initiative, supported, though not led, by the government, to ensure that authors are fairly compensated when their works are used to train AI models. This approach aims to balance the quality and diversity of training data with the protection of intellectual property rights and the livelihoods of content creators.

4.2.3 Investments in AI development

Some governments are investing in the development of AI models in local languages to support inclusive AI development. India's Bhashini mission under the Ministry of Electronics & IT (MEITy) funds natural language translation tools across local languages, while the AI4Bharat initiative has received government funding to develop datasets to train and build AI systems in local languages. In Africa, governments partner with NGOs to fund speech-data collection and transcription services in local languages, while some Latin American countries are collaborating to launch Latam-GPT in September, the first large-scale AI language model designed to capture the region's cultural diversity and linguistic nuances. Meanwhile, South Korea's government-run AI Hub provides resources to spur private sector development of Korean language AI models.

Beyond investments in local language AI datasets and tools, governments are also **investing in or planning to invest in domestic foundation AI models**, for example, in France (Mistral), Germany (OpenGPT-X), Switzerland (unnamed), India (planned), and China (WuDao).

4.2.4 Procurement and Usage in AI Development

While governments typically procure and use models developed by the private sector (which is covered in the "Procurement & Usage in AI deployment" section below), there is some evidence of **governments influencing the development of models through their use.** Canada's Directive

[18] While not a government, the IDB influences Latin American and Caribbean governments by conditioning loans and cooperation on policy guidelines and by providing technical expertise to address capacity gaps.

on Automated Decision-Making requires federal institutions to conduct Algorithmic Impact Assessment, thereby influencing the development of models to minimise the risks assessed in relevant tools. A more direct example is DARPA's XAI Program, where the need for explainability in AI models used in operational military contexts (for legal and ethical compliance) resulted in the Explainable AI Toolkit, which contains a variety of tools and resources to help users, developers, and researchers understand outputs from machine learning models beyond military applications. In Chile, public-sector AI projects must follow public tendering AI standards set by ChileCompra, the government's procurement agency under the Ministry of Finance, which may impact the long term development of such systems.

4.2.5 Diffuse and Informal Interventions in AI Development

Governments regularly host AI forums to highlight domestic innovation and AI achievements in a bid to influence global AI model development.[19] Such forums can also achieve political objectives. For example, the World Artificial Intelligence Conference (WAIC) in Shanghai demonstrated local developers' resilience and advancement despite U.S. export controls, while Saudi Arabia's Global AI Summit focused on "harnessing AI for the good of humanity".

Furthermore, there is evidence that **governments have influenced and attempted to influence the actions of AI companies without the use of formal legislation.** Examples include the Biden-Harris Administration securing voluntary commitments from AI companies, and the Sunak-led UK government securing an agreement for the UK's AI Safety Institute to test models pre-launch, among others.[20] [21] While these initiatives have been framed largely in positive, safety/rights-enhancing terms, governments can also use the threat of regulatory action or the withholding of benefits to "jawbone" providers and inappropriately coerce them into changing their models to meet governmental objectives or expectations. Such informal interventions can also include outright threats linked to protectionism, which can impact company behaviour, such as the Trump administration's trade-related threats to the EU due to the implementation of the Digital Services Act, which may impact the compliance of U.S.-based companies operating in these jurisdictions.

[19] For example, the AI Impact Summit in India and the AI Action Summit in France

[20] Specifically, ensuring the safety and security of AI products before public release, prioritizing cybersecurity, and building public trust through mechanisms like watermarking and transparent reporting on capabilities and risks.

[21] Specifically, providing the UK's AI Safety Institute with prerelease access to OpenAI, Google DeepMind, and Anthropic's most advanced AI models for safety testing.

4.3

AI Deployment

4.3.1 "Hard" Governance in AI Deployment

Governments globally are adopting diverse approaches to regulation intended to influence AI deployment, ranging from outright prohibitions in certain sensitive areas to outcome-related mandates that deploy varying forms of enforcement and liability apportionment. In some cases, governments are attempting to restrict when and where AI can be used. For example, the EU AI Act (Art. 5) prohibits use cases with disproportionate rights impacts, as well as in certain defined circumstances (e.g., elections and CSAM). It also defines specific use cases as high risk (Art. 6), which require additional assessment and mitigation measures, including risk-based fundamental rights impact assessments (Art. 27) in certain limited instances. Furthermore, companies could be required to conduct human rights due diligence for their EU AI deployments under the CSDDD (Art. 1).

However, to date, it has been more common for governments to affirmatively mandate the use of AI. For instance, India's Intermediary Guidelines require the use of AI tools to proactively detect and remove misinformation, deepfakes, and illegal content; the UK's Online Safety Act mandates the use of "proactive technologies" for content moderation; and Vietnam requires the use of AI for rapid "toxic content" detection and takedown.

Some countries restrict access to tools like Deepseek on government devices due to national security concerns. Others, such as Russia, Turkey, and China, block certain AI services entirely due to national security concerns, efforts to control information ecosystems, or to ensure alignment with domestic laws on content and data sovereignty.

Pre-deployment and ongoing model evaluations are mandated in some jurisdictions. California's vetoed SB 1047 represented the first state-level attempt in the US to establish deployment restrictions based on third-party safety audits for large AI models.[22] California has since passed SB 53, which places transparency, safety and accountability obligations on frontier AI developers, which has a downstream impact on deployers. Similarly, the EU AI Act (Art. 43) mandates third-party evaluations for high-risk systems, while New York State requires third-party assessments of

[22] While these requirements technically begin prior to deployment, given that such mandates often include ongoing evaluation, we see them as fitting in the "deployment," rather than the "development," phase of the AI lifecycle.

hiring algorithms. In China, model security self-assessments have to be filed with the Cyberspace Administration of China (CAC) before deployment.

Finally, **documented law enforcement demands for AI chatbot logs** have also started emerging in some jurisdictions, such as the US and UK, and it is likely that requests for user data, as well as law enforcement use of AI for surveillance purposes more broadly, will emerge in other jurisdictions as well.

4.3.2 "Soft" Governance in AI Deployment

Some governments attempt to guide rights-respecting AI use through voluntary frameworks. Singapore's cross-sectoral AI Verify and its financial sector-specific Fairness, Ethics, Accountability, Transparency (FEAT) Principles promote responsible use through fairness audits and transparency measures. Similar cross-sectoral and sector-specific guidance is provided by the U.S. NIST AI Risk Management Framework and the Pan-Canadian AI for Health (AI4H) Guiding Principles for organizations deploying AI systems.

4.3.3 Investments in AI Deployment

Governments are increasingly investing in AI deployments to stimulate innovation and responsible adoption. In Singapore, the Infocomm Media Development Authority launched Gen AI Sandbox 2.0, which provides grants to businesses piloting the deployment of Gen AI solutions. Similar initiatives are seen in Canada's AI Compute Access Fund, France's bpifrance AI fund, South Korea's AI Voucher Program, and various similar programmes in China, all of which offer financial support to help small and medium-sized businesses deploy innovative AI applications, and in some cases, AI literacy programmes.[23]

4.3.4 Procurement and Usage in AI Deployment

Governments around the world are actively deploying AI to attempt to enhance the delivery of public services. Examples include the Indian government's use of AI in agriculture, Rwanda's use of AI-enabled triage tools in healthcare, Singapore's use of its Virtual Intelligent Chat Assistant

[23] For China, examples include the national AI fund and local initiatives such as the Shanghai AI subsidy scheme

(VICA) for public service delivery, the UK National Health Service's predictive analytics to identify need for early social care interventions, and the Office of the Federal Chief Information Officer's extensive documentation of AI use cases by federal agencies in the US .

In some jurisdictions, **public service delivery extends into the use of AI (especially facial recognition) for surveillance** that underpins law enforcement. Beyond well-documented use cases in China, AI-based surveillance is extensively used in Israel and Singapore for surveillance purposes, impacting the right to privacy and other rights in all three cases.

To ensure **rights-protecting public sector use,** the UK has developed procurement guidelines to ensure ethical AI adoption in government operations, while a U.S. federal memorandum provides guidance, which, among other things, places requirements to identify risks of high-impact AI use cases through an AI Impact Assessment. Canada mandates the use of Algorithmic Impact Assessments for public sector deployments of AI in automated decision making.

Such guidance can also have broader consequences. For example, the Trump Administration's July 2025 AI Action and Executive Order on Preventing Woke AI in the Federal Government mandates that procurement guidelines be updated to ensure that only AI systems deemed objective and free from "top-down ideological bias" are eligible for government contracts. The lack of clarity as to how to objectively define bias and demonstrate its absence has raised concerns about the practicality of this requirement, as well as its consistency with freedom of expression.

4.3.5 Diffuse and Informal Interventions in AI Deployment

Legislative and executive office holders in the U.S. have attempted to influence the behaviour of AI deployments without the use of formal legislative actions. For example, an Attorney General in Missouri threatened to investigate AI companies due to alleged political bias, while representatives of Congress sent a letter of concern to xAI due to antisemitic and violent posts on Grok. Both actions highlight growing informal bipartisan pressure over AI companies' content moderation practices.

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

SECTION 6

Recommendations

To Governments

Rights-Based AI Governance

An overarching recommendation is for states to adopt a rights-based AI governance framework, ensuring that human rights principles are embedded throughout the development and use of AI systems.[33] At a high level, this includes enacting complementary laws, regulations and institutions that enable the protection, respect, and remedy of rights throughout the AI value chain. This may include mandating risk-based human rights assessments (supported by meaningful external stakeholder engagement) and related preventative and mitigation measures across the AI value chain, mandating state- and non-state based remedy mechanisms, and actively participating in multilateral and multi-stakeholder efforts to shape global AI governance and advocate for the ongoing protection of human rights.[34] The specific recommendations below explore thematic areas and risks highlighted earlier in this Brief.

Restriction of Information

Government mandates related to the inputs and outputs of AI models can both impact human rights.[35] Conditions on inputs that are designed to restrict model outputs are likely to have disproportionate and unintended consequences. While limitations on outputs can be more narrowly tailored, they should focus on content that is illegal. Given the challenges that exist with "re-training" models, governments should be especially careful to design legal and regulatory frameworks so that they avoid creating impacts on rights that will be difficult to remedy retrospectively.

[33] Several GNI members have published thought leadership and have advocated for rights-based AI governance, including Global Partners Digital (GPD) and Article 19, as well as non-members such as Chatham House and Access Now, and multilateral organisations such as B-Tech

[34] Such as the Global Partnership for AI, the Council of Europe Framework Convention on AI, the Global Digital Compact, the World Summit for Information Systems, the G20 AI Dialogues

[35] The potential negative rights impacts of overbroad use of AI in content moderation has been documented extensively, including by ECNL

In addition, governments should be cautious about shifting legal liability for AI-generated content to intermediaries, as this may incentivize over-removal and over-censorship. In line with the legality requirement articulated above, government interventions must clearly define prohibited content and conduct, and allow determinations of responsibility for illegal content to be adjudicated by independent judicial bodies in conformity with due process norms.

Surveillance

Governments using AI technologies to acquire and/or analyze personal data (including biometric data) must ensure that these activities are properly authorized under public and clear legal frameworks, and that appropriate transparency, independent oversight, and remedy/accountability mechanisms exist to guard against misuse. These same safeguards are necessary when governments acquire data from companies that manage AI tools or services, whether through legal requests or via commercial procurement of data. In addition, governments are encouraged to:

- Allow users to interact with AI products or services in ways that protect their identity, including through the use of encryption;
- Avoid requirements that compel or enable tracking, tracing, or proactive monitoring of user activity by companies;
- Minimizing data collection, processing, storage, and retention requirements;[36] and
- Implement rights-protecting data protection laws to ensure users have appropriate awareness and control of their data, as well as access to remedy where their data is misused.
- Refrain from accessing user data, whether directly or indirectly through demands to third parties, without meeting appropriate safeguards.

Export Controls

While international human rights law permits restrictions on freedom of expression and privacy on national security grounds, as noted previously, export controls on critical infrastructure, as well as models themselves, can have unintended and/or disproportionate impacts. It is therefore recommended that any export controls be as targeted as possible and that governments applying such controls:

- Incorporate human rights into export controls policy, including establishing processes to routinely engage with civil society on export controls,

[36] For example, requiring global AI providers to host application or user data locally. Please see page 28 of GNI's Content Moderation Policy Brief

- Strengthen export controls on technologies with an unequivocal dual use to nations with documented human rights violations, including AI-assisted surveillance and censorship technologies,[37]
- Review and implement processes and technologies to more precisely control use cases that meet specific security objectives, instead of blanket export controls on entire nations and their rightsholders,[38] and
- Continue scientific exchange and collaboration on AI technologies to promote cross-jurisdiction understandings and collaborations around risks and empower rights-respecting uses.

Sovereign AI[39]

Governments investing in sovereign AI should ensure that such initiatives are grounded in rights-based governance frameworks, such as those mentioned earlier in this Brief. Specifically, governments should consider evaluating actions that restrict access to information, limit expression, and violate user privacy in line with the three-part test.

Recommendations related to sovereign AI initiatives include:

- Ensure AI models use inclusive datasets that represent minority languages and inputs from marginalized communities;
- Facilitate equitable and rights-respecting access to AI, through open APIs, affordable tools, and AI literacy programs, to narrow digital divides, promote equitable scientific advancement, and empower vulnerable groups;[40] and
- Prioritize opportunities for economic inclusion in AI investments, particularly in underdeveloped or underserved regions.[41]

[37] A position advocated for by Freedom House and Human Rights Watch, among others

[38] As noted by various academics, think tanks, and industry representatives, blanket bans may expedite the development of indigenous capacity in targeted jurisdictions, thereby negating the impacts of export bans, while provoking retaliatory measures that may impact rightsholders in the source nation.

[39] While there are various definitions of "sovereign AI," we refer here to NVidia's definition: "Sovereign AI refers to a nation's capabilities to produce artificial intelligence using its own infrastructure, data, workforce and business networks."

[40] For example, multiple access initiatives such as Canada's AI Compute Access Fund and Singapore's GenAI Sandbox for SMEs.

[41] For example, China's Eastern Data, Western Compute (EDWC) initiative illustrates how infrastructure and AI capabilities can be strategically directed to reduce regional disparities

Public Sector Use Cases

Governments deploying AI in the public sector should ensure that such initiatives are grounded in rights-based governance frameworks, such as those mentioned earlier in this Brief.[42] Specific incremental recommendations include:

- Prohibition of public sector use cases with a strong likelihood of significant and/or irremediable rights impacts;[43]
- Develop mitigations for use cases with lower risks of impact on human rights;[44]
- Consider this guidance in the context of public sector service delivery that involves AI-enabled automated decision making;
- Maintaining a public inventory of AI use cases across government agencies;[45] and
- Implementing remedy mechanisms related to public sector uses of AI.[46]
- Mandatory and meaningful engagement of external stakeholders, especially civil society and affected communities

[42] See this report from the Ada Lovelace Institute on public sector AI procurement, which recommends clearer, consolidated guidance, defined terminologies, stronger governance, built-in ethical and transparency safeguards, public engagement, and support for local government capacity and accountability.

[43] For instance, as in the EU AI Act's Article 5.

[44] For instance, the EU AI Act requires the following types of mitigations: human rights risk assessments (Article 27), third-party evaluations (Article 43), transparency (Article 13), and continuous monitoring (Article 61)

[45] For example: https://github.com/ombegov/2024–Federal-AI-Use-Case-Inventory

[46] This includes executive mechanisms such as the UK's Investigatory Powers Tribunal or the US DOJ complaint mechanism, or judiciary mechanisms

To Civil Society

Civil society has long played a crucial role in safeguarding human rights in the technology sector, and this role is even more vital in the context of AI. Civil society should continue to advocate for rights-based AI governance frameworks that embed international human rights law into both national and international AI regulations and their implementation. This includes active participation in global policy forums to ensure that human rights are central to emerging AI governance structures.

Civil society actors also engage with companies to promote rights-respecting internal AI governance frameworks. This includes: providing input into and feedback on corporate policies and practices to ensure they align with the UN Guiding Principles on Business and Human Rights; engaging with companies on their ongoing human rights due diligence efforts; conducting and publishing research on the impacts of AI-enabled products and services; and participating in accessible remedy mechanisms across the entire AI lifecycle.

In the public sector, civil society should push for public consultation, robust accountability mechanisms, and independent oversight, especially for public sector use cases deployed in high-risk contexts such as recruitment, law enforcement, benefit allocation/social services, border control, and military uses.

Civil society plays an essential role in ongoing engagement with key rightsholders—such as affected communities, journalists, and legal professionals—regarding the human rights implications of AI systems. This close involvement uniquely positions civil society to conduct research, build a credible evidence base, and document, analyze, and elevate the unintended rights impacts arising from AI deployments across both public and private sectors.

Efforts should be made to ensure that civil society and representatives who study, represent, and/or advocate for vulnerable communities or represent marginalized populations are supported (including financial resources) and listened to (including meaningfully incorporating their feedback into product development or use, and policymaking).

To Companies

All companies, including companies in the AI value chain, have a responsibility to respect their users' rights, including the rights to freedom of expression and privacy, and to avoid discriminatory impacts on marginalized groups who are disproportionately impacted by AI systems. They should comply with applicable laws while respecting internationally recognized human rights wherever they operate. In cases where national laws, regulations, or policies fall short of international standards, technology companies are expected to avoid, mitigate, or address the negative impacts of government demands and seek ways to uphold these human rights principles to the greatest extent possible. Furthermore, companies should be able to demonstrate their efforts in fulfilling these responsibilities in line with the UNGPs and the OECD Guidelines for Multinational Enterprises.

To support these efforts, the Global Network Initiative (GNI) Principles on Freedom of Expression and Privacy, along with its more detailed Implementation Guidelines, provide a comprehensive framework offering guidance to the tech sector and other stakeholders in respecting and advancing human rights worldwide. GNI creates space for companies to demonstrate and receive feedback on these efforts, supports cross-industry and multistakeholder learning, supports rights-focused advocacy, and facilitates meaningful stakeholder engagement. The Annex further unpacks how the GNI framework can apply in relation to corporate conduct and decision making related to AI.

Companies should proactively advocate for laws and regulations that align with international human rights norms, refrain from advocating for laws and regulations that are inconsistent with those norms, and engage in proactive joint public policy advocacy with civil society, multilateral organizations, industry bodies, or multistakeholder initiatives in relevant jurisdictions. Companies should conduct ongoing human rights due diligence (HRDD), including meaningful stakeholder engagement, to identify and then take action to avoid or mitigate human rights impacts related to their development and deployment of AI-related technologies, tools, and features. In addition, companies may benefit from conducting detailed human rights impact assessments (HRIA) in certain circumstances, including when developing new products or entering or exiting certain jurisdictions.[47]

As part of this HRDD, companies should understand their potential exposure to diverse forms of government demands, interventions, pressures, and restrictions. When faced with

[47] See, e.g., AI-related HRIAs conducted by Microsoft, Intel and Google

such government action, companies should assess their legality, legitimacy, necessity, and proportionality in line with international human rights law, in order to determine how best to respond. Where government interventions do not meet these criteria, companies should consider how best to push back or otherwise limit compliance, including by engaging in dialogue and advocacy through relevant multilateral or multistakeholder initiatives.

Companies are recommended to maintain transparency towards impacted users and the public in their respective local languages, including by publishing the results of HRIAs, disclosing government interventions where feasible, engaging with rightsholder representatives, and notifying impacted users in affected jurisdictions where permitted by local laws. Additionally, companies should establish grievance mechanisms in line with best practices (UNGP Articles 29 and 31) to allow users to report impacts on them or the rightsholders they represent.

Examples of AI-related, pre- and post-compliance prevention and mitigation measures could include, but are not limited to:

- Conduct impact assessments on AI functionalities (especially high risk use cases such as AI-based facial recognition) in anticipation of and in response to government use and interventions,
- Funding and otherwise supporting independent research and civil society monitoring of the human rights impacts of AI systems in affected regions, especially in contexts where government oversight is weak or absent, and
- Collaborating with governments and/or civil society to provide AI literacy programs or digital security training, especially for vulnerable populations, while supporting the development and access to rights-respecting local AI models in the same regions.

SECTION 7

Acknowledgements

GNI would like to express its gratitude to institutional and individual members of its AI Working Group for contributing their time and expertise in the production of this Policy Brief during 2025. Members of GNI's AI Working Group in alphabetical order are as follows:

Institutional Members

- BNP Paribas
- Center for Democracy and Technology (CDT)
- Centre for Communications Governance at National Law University Delhi (CCG)
- Centre for Internet and Society (CIS India)
- Cloudflare
- Digital Agenda for Tanzania Initiative (DA4TI)
- Digital Rights Foundation Pakistan
- European Center for Nonprofit Law (ECNL)
- The Future of Free Speech
- Google
- Heartland Initiative
- Internet Sans Frontieres (ISF)
- InternetBolivia
- ITS Rio
- Meta
- Women at the Table (W@TT)

Individual Members

- Courtney Radsch
- Deirdre Mulligan
- Dunstan Allison-Hope

Annex: GNI Framework in AI

GNI Implementation Guidelines Category 2: Freedom of Expression & **Category 3:** Privacy

GNI PRINCIPLE DESCRIPTION[48]	MATCHING IG ITEMS[49]	INTERPRETATION
2.1 Participating companies will respect and protect the freedom of expression of their users by seeking to avoid or minimize the impact of government restrictions on freedom of expression, including restrictions on the information available to users and the opportunities for users to create and communicate ideas and information, regardless of frontiers or media of communication. 3.1 Participating companies will employ protections with respect to personal information in all countries where they operate in order to protect the privacy rights of users.	2.4, 3.1(c), 3.2, 3.3	• All:[50] Conduct ongoing HRDD identifying, preventing and mitigating potential FoE and privacy impacts[51] arising from the launch and operation of AI services, including but not limited to those arising from government interventions in AI as highlighted in Section 4 of this Policy Brief ("AI risks") • All: Adopting (and where possible, publishing) policies and procedures which set out how the company will respond to government interventions on AI services • All: As highlighted in IG3.2 and IG3.3, when specific government interventions related to AI services are received, ensure appropriate documentation by the government and the company, assess legality, legitimacy and necessity, interpret narrowly, and advocate against /challenge disproportionate interventions on AI services

[48] Numbering system for Principles and sub-Principles based on Assessment Toolkit Appendix IV
[49] Based on Assessment Toolkit Appendix IV
[50] Applies to entire value chain of AI, i.e. data, hardware and software vendors
[51] While not within the scope of the GNI Principles, other fundamental rights significantly impacted by AI– e.g. discrimination due to model bias, should ideally be included within any HRDD/HRIA

GNI PRINCIPLE DESCRIPTION	MATCHING IG ITEMS	INTERPRETATION
2.2 Participating companies will respect and protect the freedom of expression rights of their users when confronted with government demands, laws and regulations to suppress freedom of expression, remove content or otherwise limit access to information and ideas in a manner inconsistent with internationally recognized laws and standards. 3.2 Participating companies will respect and protect the privacy rights of users when confronted with government demands, laws or regulations that compromise privacy in a manner inconsistent with internationally recognized laws and standards.	2.4, 3.1, 3.5	• All: In line with IG3.1: ◦ Proactive direct or indirect advocacy to governments on minimisation of AI risks arising from their interventions in line with the GNI Principles and international laws and standards on FoE and privacy ◦ Setting out company's response when governments fail to provide a written directive or adhere to domestic legal procedure within policies and procedures ◦ Advocating for rights-protecting laws and regulations relevant to AI where such laws are absent or deficient • Deployer: User disclosure / information on: ◦ Relevant laws and policies that form the basis of government interventions that AI services are subject to ◦ Company policies & procedures to respond to such government interventions ◦ Disclosure on use of AI in any systems, the limitations of such systems and methods for opting out ◦ Information on user data collection, storage, processing and retention by AI services ◦ Case-by-case user notification if governments request user information from AI services, or restrict AI services to users where legally possible

GNI Implementation Guidelines Category 4: Responsible Company Decision Making

GNI PRINCIPLE DESCRIPTION	MATCHING IG ITEMS	INTERPRETATION
4.1 Participating companies will ensure that the company board, senior officers and others responsible for key decisions that impact freedom of expression and privacy are fully informed of these Principles and how they may be best advanced.	2.1, 2.2, 2.3, 2.4, 2.12, 2.13	• All: In addition to all the requirements mentioned thusfar: ◦ Board oversight of AI risks supported by regular reporting from management ◦ Regular review and management of AI risks by Board or senior management while preserving safety and liberty of personnel ◦ Risk-based training on AI risks for Board, senior management and relevant employees ◦ Internal structures (including a senior-directed, human rights function), policies & procedures to oversee, sign-off and implement measures to manage AI risks in line with the GNI Principles ◦ Record-keeping of government interventions related to AI services ◦ Grievance mechanisms ◦ Communicating aforementioned policies & procedures company-wide and escalation procedures

GNI PRINCIPLE DESCRIPTION	MATCHING IG ITEMS	INTERPRETATION
4.2 Participating companies will identify circumstances where freedom of expression and privacy may be jeopardized or advanced and integrate these Principles into their decision making in these circumstances.	2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.12, 2.13, 3.4	• All: In addition to all the requirements mentioned thus far: ◦ HRIA (supported by algorithmic IA / ethical AI assessment, privacy IA where needed) focused on the most salient AI risks identified from HRDD in circumstances suggested in IG2.6, conducted as per IG2.7 ◦ Assessing human rights risks in the collection, storage and retention of data that is collected and used in model training and deployment
4.3 Participating companies will implement these Principles wherever they have operational control. When they do not have operational control, participating companies will use best efforts to ensure that business partners, investments, suppliers, distributors and other relevant related parties follow these Principles.	2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.11	• All: In addition to all the requirements mentioned thus far: ◦ Best efforts management of AI risks in line with the GNI Principles involving upstream partners, downstream partners and (where relevant) end users, prioritized based on salience

GNI Implementation Guidelines Category 5: Multi-Stakeholder Collaboration

GNI PRINCIPLE DESCRIPTION	MATCHING IG ITEMS	INTERPRETATION
5.1 Participants will take a collaborative approach to problem solving and explore new ways in which the collective learning from multiple stakeholders can be used to advance freedom of expression and privacy	2.7(a), 2.7(b)	• All: ◦ Ensuring diverse stakeholder consultation during HRIAs related to AI risks as defined in IG2.7(b), with follow-up on company decisions arising from feedback received
5.2 Individually and collectively, participants will engage governments and international institutions to promote the rule of law and the adoption of laws, policies and practices that protect, respect and fulfil freedom of expression and privacy	3.1	• All: Proactive advocacy to governments on minimisation of AI risks due to their interventions in line with the GNI Principles and international laws and standards on FoE and privacy

GNI Implementation Guidelines Category 6: Governance, Accountability & Transparency

GNI PRINCIPLE DESCRIPTION	MATCHING IG ITEMS	INTERPRETATION
6.1 Participants will adhere to a collectively determined governance structure that defines the roles and responsibilities of participants, ensures accountability and promotes the advancement of these Principles.	2.1	• All: Board oversight of company's AI risks
6.2 Participants will be held accountable through a system of (a) transparency with the public and (b) independent assessment and evaluation of the implementation of these Principles.	3.5	• All: In addition to all the requirements mentioned thus far, publishing HRDD/HRIA & supporting algorithmic impact assessments / ethical AI assessments where possible, and inclusion of AI services in GNI assessments where relevant

GLOBAL
NETWORK
INITIATIVE