GNI Submission to OHCHR's Consultation on Protecting Human Rights Defenders in the Digital Age

Note on structure of this submission:

This submission follows the structure of the consultation issued by the Office of the United Nations High Commissioner for Human Rights (OHCHR). Relevant consultation questions and contextual information have been reproduced for ease of reference, and those to which the Global Network Initiative (GNI) provides responses are highlighted. Responses appear directly beneath those questions and are presented in bullet-point format where appropriate.

Background

Digital technologies have transformed both the work of human rights defenders (HRDs) and the nature of the threats and attacks they face. HRDs operate at the forefront of public engagement, increasingly relying on digital tools for communication, monitoring, documentation and advocacy. As these tools evolve, so does the nature of the threats and attacks HRDs face – impacting their safety online and offline. In April 2025, the Human Rights Council adopted resolution 58/23 mandating OHCHR to conduct consultations to assess the risks created by new and emerging technologies to HRDs and to identify effective practices to address these risks across different geographical contexts. It also requested OHCHR to prepare a report on the outcomes of these consultations, which may include recommendations on due diligence and improved responses to digital technology-related risks faced by HRDs.

Key questions and types of input/comments sought

Inputs are sought to gather information on the ways in which the work and safety of HRDs are affected by digital technologies. We will look at overarching trends which we have identified as particularly relevant to protection of HRDs: how digital technologies affect the communications, privacy and safety of HRDs; how legislative and regulatory measures impact digital spaces; and how companies have responded to identified risks affecting HRDs on their platforms and services.

We would appreciate receiving inputs in response to some or all of the questions listed below.

1. Legislative and regulatory measures

- What impacts have recent trends in legislative and regulatory efforts at local, regional and global levels – including, for example, on information integrity, online safety and

cybercrime – had on the work and safety of HRDs offline and online?

- Online safety regulations in the UK and Australia are setting precedents through "Safety by Design" and "Duty of Care" models that increasingly mandate age verification technologies, identity verification requirements, proactive content monitoring, and, in some proposals, client-side scanning of encrypted services. On the whole, these approaches, while well intentioned, create risks for human rights defenders (HRDs), and have not undergone sufficient independent technical and human rights assessments regarding their necessity, proportionality, effectiveness, and risks — particularly regarding data storage, security, and potential misuse.

  Blanket age or identity verification frameworks can be especially invasive for at-risk communities, including whistleblowers, activists, LGBTQ+ persons, and HRDs who rely on anonymity and secure communications for their safety. Proposals such as client-side scanning of end-to-end encrypted services risk weakening encryption protections, thereby exposing HRDs to surveillance, retaliation, and data breaches. Beyond their domestic impact, such regulatory models risk legitimising intrusive compliance practices globally, particularly in jurisdictions where safeguards are weak or absent.

  The Global Network Initiative (GNI) has observed similar regulatory patterns emerging in countries such as Vietnam (under Decree No. 147/2024/ND-CP (formerly Decree 72)) and Mexico (under the Telecommunications Act, the General Population Act, the Intelligence Act, and the Disappeared Persons Act, and the National Guard and Public Administrations Act), where proactive monitoring obligations and real-identity verification requirements are mandated by law. When such measures are adopted in democratic contexts without strong safeguards, they risk normalising practices that may be replicated in more repressive environments.

  Recent regulatory developments concerning youth access to social media also raise potential risks for HRDs. Beginning with Australia, proposals to ban social media access for individuals under the age of 16 have prompted similar legislative initiatives in at least 41 other countries, including Indonesia, Spain, Greece, Ireland, South Korea, Türkiye, the United Arab Emirates, and Vietnam. While framed as child protection measures, such blanket prohibitions are likely to drive increased reliance on age verification technologies by online platforms. These systems often require the collection and processing of sensitive personal data and may undermine anonymity online. In addition, broad access bans risk restricting young human rights defenders' ability to access vital online

communities, support networks, information resources, and emergency services, which in some contexts play a critical role in their safety and participation in civic space.

- Similarly, cybercrime legislation is rapidly expanding worldwide, particularly in the wake of the adoption of the United Nations Convention Against Cybercrime. While international cooperation to address genuine cybercrime is important, overbroad and ambiguous cybercrime regulations pose serious risks to HRDs, which are particularly acute in the context of digital transnational repression. These risks are especially pronounced for women HRDs and journalists, who may also face technology-facilitated gender-based violence (TFGBV) as part of digital targeting. For instance, in Southeast Asia, state actors regularly target journalists, HRDs, and exiled dissidents with the help of opaque and covert regulatory frameworks, sometimes in ways that appear coordinated or mutually reinforcing. Emerging international legal frameworks risk formalising and legitimising these practices. To illustrate the mechanics of this, consider, for example, the Communications and Multimedia Act (CMA) in Malaysia which grants sweeping powers for the government to compel content removal and access communications under vague prohibitions, as well as the recently endorsed ASEAN Guidelines on the Governance of Digital Platforms, which risks facilitating coordinated regional enforcement based on expansive definitions of illegal content. Together, these measures risk lowering the threshold for cross-border identification and targeting of critics.

  Similar patterns are evident elsewhere. In Zambia, domestic frameworks have enabled disproportionate surveillance and enforcement powers; in Pakistan, amendments to the Prevention of Electronic Crimes Act have been criticised for granting expansive authorities capable of silencing dissent and compelling intermediary compliance with overbroad demands; and in Canada, the proposed Online Harms Act has raised concerns regarding remote government access to data and stringent takedown obligations that may incentivise over-removal of content and weaken privacy protections. Across contexts, broadly framed cybercrime and online safety laws, coupled with enhanced cross-border cooperation mechanisms, risk being used to restrict HRDs' rights to freedom of expression, association, and privacy.

  Across many jurisdictions, these legislative and regulatory frameworks also lack specific safeguards and exceptions applicable to protected categories such as journalists and human rights defenders. While the objectives of certain measures — including cybercrime legislation — may in some cases be justified, it is essential that such frameworks incorporate appropriate safeguards,

proportionality standards, and narrowly tailored carve-outs to avoid unnecessary or unjustified interference with human rights. Equally important is the existence of independent and competent institutional mechanisms responsible for the enforcement of these laws. In practice, enforcement bodies may lack the technical expertise required to assess complex digital issues, may be subject to political influence, or may operate without sufficient transparency and accountability. In several contexts, judicial oversight and review mechanisms are also being weakened or bypassed, increasing the risk of arbitrary or disproportionate decisions that may negatively affect HRDs.

It is also important to recognise AI's dual nature. While AI-enabled surveillance and generative systems can be weaponised against HRDs, the same technologies have the potential to enhance human rights documentation, digital forensics, OSINT, and verification of citizen journalism, particularly in conflict or high-risk settings. Multimodal AI, such as Gemini 3's frame-by-frame video analysis, transforms video from a passive medium into structured, queryable datasets, enabling more precise and rapid documentation of abuses. At the same time, generative AI introduces new risks: deepfakes and manipulated media can target HRDs, discredit their work, and exacerbate the "liar's dividend," undermining the evidentiary credibility essential to their advocacy. Women journalists, public figures, and HRDs are disproportionately affected by AI-facilitated technology-based gendered attacks, including non-consensual sexual imagery and false attribution of fabricated content.

- What legal or regulatory instruments and institutional procedures are commonly used to restrict the rights to freedom of expression, association and privacy of HRDs online?
- How have legislative and regulatory efforts in one country or region impacted similar legal and regulatory measures in other countries or regions?
    - See response to the previous question.
- For each of the questions listed above, please provide information on the national, regional or international laws or regulations referred to, case examples and other relevant illustrative data.

2. Digital communications

- Which risks do internet shutdowns, network interferences, geo-blocking or other forms of restrictions of connectivity and communications pose to HRDs' work and safety?

- GNI has observed a significant increase in the use of network disruptions[1] by governments during periods of political instability, protests, elections, or conflict. These measures take various forms, including nationwide internet shutdowns, information blackouts, whitelisting regimes, social media and website blocking, and technical interference such as TLS/SSL protocol failures. As seen during the communications blackouts in Bangladesh in 2024, Nepal in 2025, and the ongoing communications restrictions in Iran and Gaza, such measures create significant risks to HRDs and severely impair their ability to document violations, communicate securely, and share information domestically and internationally. They also isolate defenders from emergency assistance, legal counsel, medical support, and broader protection networks, while limiting the ability of companies, civil society, and international actors to assess and respond to emerging human rights risks.

  Governments frequently justify these disruptions on grounds of public safety or national security, and in some contexts frame them as necessary to compel compliance with domestic regulatory frameworks. However, these blanket measures are rarely necessary or proportionate and often exacerbate insecurity and mistrust. Government-ordered shutdowns and blocking measures place companies in acute human rights dilemmas, particularly where domestic legal obligations conflict with international human rights standards. Such orders often lack transparency, independent oversight, or meaningful avenues for appeal, constraining companies' ability to mitigate harm.

  Other forms of network disruption — including throttling of encrypted services or blocking of specific platforms — further undermines HRDs' ability to rely on secure communication tools. In some cases, defenders are pushed toward less secure or state-controlled alternatives, increasing exposure to surveillance and reprisals. This dynamic has been observed, for example, in Russia, where access

---

[1] *"A network disruption is the intentional, significant disruption of electronic communication within a given area and/or affecting a predetermined group of citizens. Extreme manifestations of network disruptions involve the large-scale or complete disconnection of digital communication, with the impact radius covering a local area, an administrative region, several regions, or an entire country. These extreme disruptions are often called network shutdowns, Internet shutdowns, or blackouts. Unlike technical failures, intentional disruptions are typically mandated by governments, which carry them out as either a reactive or, increasingly, a preventive measure against perceived real and potential threats. The most common objective of this kind of interference is to restrict the flow of information through digital channels, particularly social media, mobile communication, and dedicated digital communication tools (e.g. WhatsApp, Voice over Internet Protocol [VoIP] services). This is especially prevalent when rising public dissent and protests are deemed to be fueled by digital communication networks."*

Global Network Initiative, Disconnected: A Human Rights-Based Approach to Network Disruptions, by Jan Rydzak (Washington, DC: Global Network Initiative, 2018), https://globalnetworkinitiative.org/wp-content/uploads/Disconnected-Report-Network-Disruptions.pdf

to WhatsApp has been restricted. The recurrent use of shutdowns and related measures risks normalising extraordinary connectivity restrictions as routine governance tools, weakening global norms supporting an open, secure, and interoperable internet.

These challenges are not new. Longstanding OHCHR and civil society documentation has traced sophisticated targeting of HRDs, shifts in platform usage, and regulatory encroachments on freedom of expression. The persistence of these risks underscores the urgent need for robust, human rights-centred standards and protections that address both emerging technologies and structural vulnerabilities.

- What forms of technology-facilitated attacks do HRDs face on social media platforms and digital communications services? How do these online attacks intersect with offline events?
- What specific risks to HRDs emerge via online platforms and communications services in situations of armed conflict, instability and/or elections?
    - As GNI has noted, situations of armed conflict and other high-risk scenarios often change the circumstances in which HRDs operate in ways that increase the risks of documenting, publicizing, and advocating around human rights impacts. Governments often cite these circumstances as justification for exercising exceptional powers or disregarding due process and other safeguards, in order to exert additional leverage over tech companies, their staff, and their operations, which in turn can be used to enhance surveillance and targeting of HRDs. These same circumstances tend to also limit access to and availability of judicial and non-judicial forms of redress.
    - See response to previous questions.
- What specific risks do women HRDs and HRDs from groups affected by marginalisation and discrimination face on online platforms and communications services?
- How do companies' policies and practices relating to content moderation and engagement with law enforcement and government authorities affect HRDs' work and safety?
- How do advances in AI technologies exacerbate risks to HRDs' operations and presence on online platforms and communications services?
- For each of the questions listed above, where possible, please provide case examples, references to State or corporate policies, practices or initiatives, and other relevant illustrative data.

3. Digital restrictions to privacy

- What risks have emerged for HRDs with the increasing procurement, use and abuse of digital surveillance tools, including spyware and interception technologies, by State and non-State actors?
- What risks have emerged for HRDs with the expansion of biometric surveillance infrastructure and increased monitoring of public and digital spaces?
- How have technological and regulatory developments relating to encryption eased or exacerbated risks to HRDs?

    - See response to Question 1.

- How do advances in AI technologies exacerbate risks to the privacy and safety of HRDs?

    - Advances in artificial intelligence technologies are creating new risks to the privacy and safety of HRDs by enabling the large-scale aggregation, inference, and analysis of personal data across multiple digital environments. Generative AI systems in particular may amplify privacy risks by synthesising information from multiple publicly available, privately procured, or leaked datasets. In some instances, AI systems have surfaced sensitive personal information that individuals had deliberately sought to keep private by aggregating fragmented data points across sources. This illustrates how AI systems capable of inferring identities from dispersed datasets may be used to expose identifying information. For HRDs operating under pseudonyms or relying on anonymity for protection, such automated data aggregation could reveal identities, locations, or networks of association, increasing the risk of harassment, surveillance, or retaliation.

      AI-enabled tools may also introduce risks when used by HRDs themselves. Many AI applications rely on cloud-based processing and may require access to documents, communications, or device-generated data. Where sensitive information related to investigations, testimonies, or organisational networks is processed through such systems, this may create additional data trails or repositories that could become accessible to governments through legal demands, security vulnerabilities, or covert access.

      At the same time, governments are increasingly deploying AI-enabled technologies to expand surveillance capabilities, including emotional surveillance, through facial recognition, biometric identification systems, AI-powered emotion recognition and lie-detectors, and advanced analytics

applied to telecommunications data. For example, in Bangladesh the former government led by the Awami League reportedly required telecommunications operators to install geolocation monitoring software developed by the French company Intersec, which uses advanced analytics and AI techniques to analyse telecommunications metadata and generate location intelligence. When deployed without robust legal safeguards, transparency, and independent oversight, such systems risk enabling large-scale surveillance and the identification or targeting of HRDs.

GNI has observed that the growing integration of AI into surveillance infrastructure and digital governance frameworks raises important questions about how these technologies may expand state monitoring capacities or facilitate data access demands on companies. Without clear legal limits and accountability mechanisms, these developments risk further eroding the privacy and security protections that HRDs rely upon to carry out their work safely.

- For each of the questions listed above, please describe, where possible, case examples, references to State or corporate policies, practices or initiatives, and other relevant illustrative data.

4. Corporate responses

- How are companies meeting their responsibilities to identify, assess, mitigate and respond to risks posed to HRDs on their platforms and services?
- Are existing corporate models and approaches to risk assessment, due diligence, remedial mechanisms and engagement with HRDs on protection concerns and reports of violations sufficient and/or effective?
- What challenges do civil society and companies face in ensuring corporate policies, processes and initiatives – including in relation to internal mechanisms and external engagement – adequately and effectively address the range and extent of risks faced by HRDs in the digital age?

    - Civil society organisations and companies face a range of structural and operational challenges in ensuring that corporate policies, processes, and engagement mechanisms adequately address the evolving risks faced by HRDs in digital environments. One key challenge arises from the growing complexity and fragmentation of regulatory frameworks across jurisdictions. Companies increasingly face overlapping or conflicting legal obligations relating to content moderation, data access, encryption, and platform governance. Compliance with such frameworks can create difficult operational and human rights dilemmas, particularly where domestic legal requirements conflict with international human

rights standards.

Companies face challenges from the simultaneous emergence of multiple legal and regulatory regimes worldwide, which differ in scope, implementation, and respect for human rights. Efforts to harmonise these landscapes—such as the UNESCO Guidelines on the Governance of Digital Platforms, its regional interpretations like the ASEAN Guidelines on the Governance of Digital Platforms and the Praia Model Policy Framework on Information Integrity in West Africa and the Sahel, and networks such as the Global Online Safety Regulators Network—provide reference points, but national-level adoption generally reflects local priorities and legal contexts. Policy diffusion further shapes these frameworks, for example when Sri Lanka's Online Safety Act drew on the UK's Online Safety Act as a model. GNI has raised concerns that the Sri Lankan law establishes broad and vaguely defined categories of prohibited content, imposes disproportionate penalties for online speech, and creates an Online Safety Commission with limited independence or regulatory oversight. This illustrates that policy diffusion does not automatically transfer institutional safeguards or rule-of-law protections: even when global benchmark laws influence company standards, their local implementation can diverge significantly.

Furthermore, large regimes such as the Digital Services Act in the European Union require extensive risk assessment, reporting, and compliance measures. While these obligations aim to address important harms, they can result in uneven attention across jurisdictions, leaving HRDs in regions where threats are more acute and protections weaker at greater risk. Many international frameworks can be vague and open to interpretation. It is important for international organisations to adopt standards that go beyond general principles and provide practical guidance to support legal and regulatory efforts at the national level. Stronger and more specific international standards could help ensure that companies implement safeguards equitably, supporting consistent protection for HRDs across all regions.

Globally, the political climate surrounding digital governance is increasingly affecting collaboration between companies, civil society, and independent researchers working on online harms and platform accountability. Researchers and civil society organisations examining platform practices have faced pressure including legal threats, doxxing, visa sanctions, and hostile media/communications campaigns; in the United States, debates around platform governance have at times included scrutiny of or attacks on such actors, as well as attempts to expose communications between companies and external stakeholders. Such dynamics can create a chilling effect on independent

research, responsible information-sharing, and multistakeholder engagement, potentially limiting the ability of companies and civil society to identify and respond to risks affecting HRDs online.

- What steps should companies take to improve identification, assessment and prevention of risks posed to HRDs' work and safety on their platforms and services?
- For each of the questions listed above, please describe, where possible, case examples, references to corporate policies, practices or initiatives, and other relevant illustrative data.