



AI, Human Rights, and the Evolving Regulatory Environment

Summary note

Workshop hosted on 24 October 2023 in Brussels
by the [UN Human Rights B-Tech Project](#) and the [Global Network Initiative](#)

Regulatory requirements around AI are emerging across sectors and jurisdictions, along with broader questions around [responsible AI governance](#). While there has been substantial discussion around responsible use and ethics of AI, to date these conversations have failed to sufficiently embed human rights and leverage the UN Guiding Principles on Business and Human Rights (UNGPs), which should be used to inform and drive policy coherence across the evolving regulatory environment on AI.

As the United Nations High Commissioner for Human Rights recently noted in an [open letter](#) to European Union institutions:

“Assessing adverse impacts on human rights is a core component of human rights due diligence, both for private and public sector use of AI, and follows from international human rights law and the UN Guiding Principles on Business and Human Rights. OHCHR’s work with leading AI companies in the context of the UN Human Rights B-Tech project demonstrates that assessing adverse impacts of AI on people is the “art of the feasible”. The fact that some companies at the forefront of AI development are endorsing and implementing a rights-based approach to risk management points to fundamental rights as a promising foundation for rights-respecting AI practices also in the EU AI Act context.”

This workshop hosted by the UN Human Rights B-Tech Project and the Global Network Initiative explored the status of AI regulation including how the UNGPs can be used by regulatory bodies to: 1) develop rights-respecting AI regulation; 2) drive policy coherence across AI legislation and the regulatory ecosystem more broadly; and 3) inform AI risk assessment and mitigation methodologies.

1. The UNGPs can help policymakers develop rights-respecting AI regulation

Participants explored how the UNGPs can be used to develop rights-respecting AI regulation. This included discussion around key UNGPs elements that should inform AI regulation, such as stakeholder engagement, the identification and management of risks across the value chain through human rights due diligence, and the design and review of relevant corporate systems and processes related to risk identification and management.

The recent mapping by the Danish Institute for Human Rights across EU regulatory initiatives, [How do the pieces fit in the puzzle? Making sense of EU regulatory initiatives related to business and human rights](#), demonstrates that the UNGPs are not specifically mentioned in key regulatory initiatives relevant to AI, but the concept of due diligence is. The EU Digital Services Act (DSA) is an exception in the sense that it mentions the UNGPs in recital 47. The process of human rights due diligence (HRDD) is incorporated in Art. 9 of the draft EU AI Act and in its proposed fundamental rights impact assessment (FRIA) provision, which is currently being debated in trilogue at the time of writing. Stakeholder engagement remains a challenge – while it is central to HRDD, it is only mentioned in recitals as “where relevant” and “as appropriate”. The draft AI Act has a focus on addressing the impact of the technology on users, but as we know impacts of technology go beyond users and can extend more broadly to individuals, communities, and societies. Another challenge relates to the way remedy is structured in the draft AI Act. European National Human Rights Institutions have raised concerns that, as structured in the Act, national competent authorities will not be able to provide sufficient remedy and therefore have [called for more focus in the draft EU AI act](#) and need to set up clear and accessible grievance mechanisms.

Turning the focus to Latin America, a draft report on AI regulation emerging across Latin American countries was presented. The report highlights a shift towards countries putting in place mandatory regulations, as indicated by the mapping of 22 regulatory instruments across nine nations, with 70% of them currently under parliamentary discussion. Most of the regulations have a strong human rights focus with some of them focused more on specific rights such as privacy and non-discrimination. Notably, some countries like Colombia take a more comprehensive approach by establishing human rights standards that span civil, political, social, and environmental rights. While transparency standards are integral to most regulations, certain countries, like Chile and Brazil, are forging ahead by creating dedicated oversight bodies such as national councils and commissions. These structures aim to provide clarity for companies operating in the AI landscape. Yet, while most regulations that are emerging incorporate and reference human rights, there is little to no commitment to requiring companies to undertake human rights due diligence. Equally, the regulations lack dedicated mechanisms for remedy and those that mention remedy lack clarity on how individuals can access the same.

Participants agreed that it would be beneficial for AI regulation globally to reference and align with the UNGPs, both in regulatory and supporting policy measures. It is essential to include a human rights-based approach – in both the process of developing and regarding the substantive components of AI regulation.

2. Applying the UNGPs to drive policy coherence across AI legislation and the regulatory ecosystem more broadly

Participants compared and contrasted the different requirements of AI-relevant regulations, including focus, scope, and approach. Participants focused in particular on if and how human rights and the international standards of business conduct (UNGPs and OECD Guidelines) have been incorporated. Examples included the EU Digital Services Act (which mandates risk management obligations for online platforms), the draft EU AI Act, the US National Institute of Standards and Technology (NIST) Trustworthy & Responsible AI framework, and the Brazilian Draft AI Bill.

There was wide agreement that grounding regulatory approaches to AI, including generative AI, in the international human rights framework and existing corporate conduct expectations will help ensure that the human rights of individuals are protected and will [build policy coherence at the global level](#). In practice, regulatory approaches (including laws, policies, and standards) that reinforce the business responsibility to respect human rights would include:

- Consideration of the full range of human rights when assessing human rights risk and conducting human rights due diligence, due to the wide range of applications of generative AI;
- Consistent application of due diligence practices to assess and mitigate the most salient risks that particular products, services, or technologies pose to human rights;
- Consideration of the full value chain of AI technologies (i.e., both upstream supply chains and downstream end-uses) while focusing attention on the distinct ways that companies are involved in actual or potential adverse impacts;
- A coherent and comprehensive suite of incentive-based policy instruments and enforcement provisions to accompany the implementation of regulation;
- Transparency and process-oriented legislation clarifying the expectations that businesses should meet (legislation should ensure appropriate notice, due process, and non-discrimination in relevant corporate systems and processes);
- Meaningful stakeholder engagement, both by government and business actors;
- Measures enabling easy and direct access to effective remedy and redress.

As technological infrastructure can replicate and perpetuate pre-existing imbalances of representation across geographies, some panelists voiced concerns about the ways in which corporations have been engaging in the debate on regulation in Brazil in comparison to Europe or the US. For example, the Brazilian draft bill text elaborated by a temporary commission of jurists rests on three pillars: principles, a risk-based approach (inspired by the draft EU AI Act), and rights. The proposal goes beyond the draft EU AI act's risk-based approach by also implementing a set of rights that applies to every individual that has been affected by an AI. This includes the right to prior information about their interactions with artificial intelligence systems; the right to an explanation of the decision, recommendation or prediction made by artificial intelligence systems; the right to challenge decisions or predictions made by artificial intelligence systems; and a right to non-discrimination and the correction of direct, indirect and direct, indirect, illegal or abusive discriminatory biases, among other rights.

Experts highlighted that regulating technology through a risk-based and principled approach provides strong advantages in coping with technology's flexibility and rapid evolution. Anchoring such risk-based regulatory approaches in human rights can foster international policy coherence. Stakeholders stressed the importance of international standards such as the UNGPs, alongside the OECD Guidelines for Multinational Enterprises on Responsible Business Conduct ([MNE Guidelines](#)). The discussion also elevated key considerations for policy coherence regarding rightsholder engagement – both with users and non-users – in risk identification and management; the importance of transparency in the process and results of risk assessments; accountability in assessment and mitigation action plans; and the consideration of impacts across the full scope of human rights and the full scope of potential applications of AI.

Beyond using the UNGPs to inform individual regulations, coherence is needed across AI legislation and the regulatory ecosystem more broadly. This is instrumental from an implementation perspective; ensuring that companies and other actors are not faced with conflicting requirements, and that a common, rights-respecting standard is established across regulatory environments.

3. Drafting AI impact assessment and mitigation methodologies in line with the UNGPs

The discussion turned towards designing impact assessment methodologies for generative AI, and how such impact assessments can be overseen through internal controls and oversight bodies. As such methodologies are likely to build on experiences with and lessons learned from existing regulatory requirements for risk/impact assessment and risk/impact mitigation measures, participants debated challenges and opportunities of existing risk assessment and mitigation provisions in EU legislation.

Among such regulatory developments, lessons-learned from existing impact assessment provisions in the DSA were discussed. The DSA introduces important new requirements for assessing systemic risks related to the functioning of Very Large Online Platforms and Very Large Online Service Providers (defined as such under the DSA), including adverse impacts on fundamental rights. However, one of the main challenges according to several participants is a lack of clear guidance or expectations on how to conduct these assessments in a meaningful way. There is also no explicit reference to a more detailed delegated act or specific standards to guide these assessments. Without clear expectations regarding the level of detail, definitions of risk/materiality, transparency, minimum requirements for consultations with external stakeholders, how to interpret and scope the assessment etc. there is a serious risk of compliance not translating into desired changes in corporate conduct and better outcomes for at-risk individuals.

While every impact assessment will be different and adapted to a specific context, service, and type of company, there is a need for a harmonized structure containing clear expectations which would ensure consistent and credible risk identification, prioritization, and facilitate future oversight and accountability in accordance with the UNGPs. The process-based nature of the UNGPs provides a framework of guidance and best practice regardless of the type of system assessed or the structure of a service or company. Stakeholders agreed that the draft EU AI Act contains some guidance on identifying and addressing adverse impacts but does not include explicit language on UNGPs alignment nor explicit references to additional guidelines.

The role of human rights and human rights expertise, and the UNGPs in particular, were perceived as key: several participants argued that impact assessments must be rooted in fundamental rights norms and standards. This is also relevant to the way that companies function internally. For example, “red teaming” exercises are increasingly used to identify impacts as comprehensively as possible. Yet these methodologies rarely mention human rights or otherwise clarify what “risks” are considered and what benchmarks are used. Red teaming can be a valuable tool in the broader HRDD toolbox but does not replace FRIAs or risk assessments under Art. 34 DSA, the draft EU AI Act, or any other future mechanisms.

Another discussed challenge was that product-level decision-making about risks, impacts and mitigations typically moves faster than the pace of extensive impact assessments proposed in

regulation. In that regard, there might be strong consequences for generative AI in particular, where the risks/impacts and good practice are evolving quickly. This requires resolving tensions between the timeline of an AI product lifecycle and the period required to meaningfully conduct a FRIA, especially since the tech is still emerging and harms are not fully known. Stakeholders from across civil society and business discussed potential avenues for product teams to adapt their practices to align with proper HRDD/FRIA practice in order to comply with the DSA, and future regulatory provisions requiring risk assessments/impact assessments.

As the potential harms of generative AI are not yet fully understood (especially beyond privacy and fairness/non-discrimination), it is especially important to engage a broad range of stakeholders to map actual and potential negative effects. Implementation of the DSA should be expanded on this point to lead by example, some stakeholders argued. One of the key conditions for risk/impact assessments to be meaningful is that they can be externally scrutinized and challenged where needed. The DSA is a step up in terms of transparency in comparison, for example, to the GDPR, where there is no requirement to publish DPIAs. Nevertheless, some participants highlighted that the publication of results only happens after the full cycle of audit is completed, which prevents meaningful contribution to risk/impact assessments as they are conducted.