

KEY PRINCIPLES FOR HUMAN RIGHTS IMPACT ASSESSMENT OF DIGITAL BUSINESS ACTIVITIES

Authors

Cathrine Bloch Veiberg, Senior Adviser, Human Rights, Business and Tech

Graphic design: Semin Alekic

Cover image: <u>Desola Lanre-Olugun</u> on <u>Unsplash.com</u>

e-ISBN: 978-87-7570-218-3

© 2023 The Danish Institute for Human Rights Wilders Plads 8K DK-1403 Copenhagen K Phone +45 3269 8888 www.humanrights.dk

Provided such reproduction is for non-commercial use, this publication, or parts of it, may be reproduced if the author and source are quoted.

KEY PRINCIPLES FOR HUMAN RIGHTS IMPACT ASSESSMENT OF DIGITAL BUSINESS ACTIVITIES

10 KEY CRITERIA FOR ASSESSING HUMAN RIGHTS IMPACTS OF DIGITAL PROJECTS, PRODUCTS AND SERVICES

The fast pace of innovation and technological development in the digital sphere has led to a new generation of human rights challenges. Technology and non-tech companies active in the digital ecosystem that seek to improve their business conduct are confronted with a lack of guidance, tools and accumulated knowledge to refer to when assessing and addressing the human rights impacts of their digital activities.

This document outlines key criteria for human rights impact assessment (HRIA) of digital projects, products, and services. The criteria are taken from the Danish Institute for Human Rights (DIHR) guidance on human rights impact assessment of digital business activities, which was originally published in 2020 to provide guidance on HRIA of digital project product and services. The guidance was developed in close cooperation and dialogue with academia, civil society, and tech companies.

In the guidance we delve further into the unique challenges of impact assessment of digital project, product and services, including:

- The lack of geographical boundaries: Al systems, software and other digital solutions can be used anywhere, and may even be designed to be used anywhere. This poses challenges for the context analysis. In the guidance we recommend considering the risk associated with the type of digital technology itself, how it will or is applied in a given context, and the risks inherent to the context of actual/intended application.
- Stakeholder identification and engagement: The stakeholders affected by AI systems, software and other digital solutions can be difficult to identify and engage with. In some cases, it may be the users affected, in others the stakeholders affected are not users, but those affected by the conduct of users facilitated by the technology. As above this also poses problems for context analysis, and again we recommend further defining the context by considering the risk associated with the type of digital technology itself, how it will or is applied in a given context, and the risks inherent to the context of actual/intended application.

OVERVIEW OF THE 10 KEY CRITERIA

BRIEF OVERVIEW - KEY CRITERIA FOR ASSESSING HUMAN RIGHTS IMPACTS		
Key criteria for the process and content of HRIA		
Process	1. Participation	Meaningful participation of actually or potentially affected rightsholders
	2. Non-discrimination	Inclusive and gender-sensitive engagement and consultation processes.
	3. Empowerment	Capacity building of individuals and groups at risk of vulnerability or marginalisation
	4. Transparency	Transparent character of the impact assessment process to adequately engage affected or potentially affected rightsholders.
	5. Accountability	The impact assessment team is supported by human rights expertise, and the roles and responsibilities for impact assessment, prevention, mitigation, and management are assigned and adequately resourced.
Content	6. Benchmark	Human rights standards constitute the benchmark for the impact assessment. Impact analysis, assessment of impact severity and design of mitigation measures are guided by international human rights standards and principles.
	7. Scope of impacts	The assessment identifies actual and potential human rights impacts that the technology caused or contributed to-, and impacts that are directly linked to the technology. The assessment analyses cumulative impacts and legacy issues.
	8. Assessing impact severity	Impacts are addressed according to the severity of their human rights consequences.
	9. Impact mitigation measures	All human rights impacts are addressed. Where it is necessary to prioritise actions to address impacts, severity of human rights impacts is the core criterion. Addressing identified impacts follows the mitigation hierarchy of 'avoid-reduce-restore-remediate'.
	10. Access to remedy	Impacted rightsholders have avenues whereby they can raise grievances regarding the digital project, products or services, as well as the impact assessment process and outcomes.

Sources: These criteria are based on a literature review including sources on human rights impact assessment, stakeholder engagement, social impact assessment and the human rights-based approach. See Danish Institute for Human Rights, "Human rights impact assessment guidance and toolbox": https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-toolbox

GUIDING QUESTIONS

The criteria can be used in several ways:

- By human rights practitioners and consultants conducting impact assessments of digital projects, products or services.
- **Companies**, in particular staff who are responsible for commissioning and overseeing impact assessments, whether those businesses are developing the digital projects, products or services themselves or are buying them.
- Public entities that are procuring and using digital products and services—in relation to e.g. e-governance initiatives, digital health services, automated decision-making in court systems, and other forms of public service delivery.
- Financial institutions, including development finance institutions, institutional investors, private equity funds, providing financial support towards the development of digital projects, products and services.

- National human rights institutions exercising their mandate to promote and protect human rights can use these principles when advising the government and other stakeholders on impact assessment laws (e.g. in relation to mandated privacy impacts assessments, data protection impact assessments, algorithmic impact assessments etc.), policies and practice, to ensure that the adoption of a human rights-based approach and international human rights standards are reflected.
- Government departments and state institutions that are responsible for providing guidance to businesses on respecting human rights, or setting standards for due diligence and impact assessment in relation to digital projects, products or services, could draw on the principles for information on how human rights might be better reflected in such guidance and standards.

10 KEY CRITERIA FOR ASSESSING HUMAN RIGHTS IMPACTS - GUIDING QUESTIONS

Key criteria for the process and content of HRIA

1. Participation

Meaningful participation of actually or potentially affected rightsholders

- Have a broad range of rightsholders been engaged in the impact assessment, including vulnerable groups (in person or virtually; directly, or through representatives or proxies)?
- Have the rights and involvement of rightsholders throughout the digital ecosystem been considered (e.g. individual end-users, those potentially impacted by the design and other individuals that are not users but that nonetheless may be negatively impacted (non-users), such as those who may be subject to offline violence after their personal information is shared online)?
- Have rightsholders, or their proxies, been involved throughout the impact assessment process, including during early phases of the impact assessment such as: design of the impact assessment process; development of terms of reference for the assessment; impact scoping; and prioritisation of critical issues to be considered by the assessment?
- Have rightsholders, duty-bearers and other relevant parties been involved in designing measures to address impacts (e.g. through prevention, mitigation and remediation) and follow-up to evaluate the effectiveness of these measures?
- Have rightsholder representatives or representative organisations, or rightsholder proxies, been included in consultation and engagement, including consideration of the legitimacy of their claim to represent the relevant individuals and/or groups?
- Is engagement and participation in the impact assessment guided by the local context, including through using the impacted individuals' preferred mechanisms (e.g. modes of communication) where possible?
- Is the assessment process being undertaken at particular times that allow diverse participation of key stakeholder groups (e.g. when women are not at work and young people are not at school)?
- Does the impact assessment provide for ongoing dialogue between rightsholders, duty-bearers and other relevant parties (e.g. through collaborative problem analysis and design of mitigation measures)?
- To the extent digital and virtual means of engagement are utilised (e.g. online consultations and surveys) have accessibility issues been assessed, particularly with regard to the most vulnerable rightsholders? E.g. if the most vulnerable do not have physical access to internet or if internet data is prohibitively expensive, they will not be able to participate.

2. Non-discrimination

Inclusive and gender-sensitive engagement and consultation processes.

Guiding Questions:

- Has impact assessment consultation and engagement involved persons of all genders, including through gender-sensitive engagement methods as necessary (e.g. through holding women-only meetings with female HRIA team members)?
- Have steps been taken to ensure that the modes of engagement and participation address any barriers that may be faced by vulnerable and marginalised individuals (e.g. by offering transport or holding meetings in culturally appropriate locations, and considering 'technology barriers' for older persons or persons with disabilities)?
- Have the vulnerable or marginalised individuals and groups in the given context been identified and considered, (e.g. by considering discrimination, resilience, poverty factors etc.)?
- Have the needs of vulnerable and marginalised individuals been identified in stakeholder mapping and engagement planning?

3. Empowerment

Capacity building of individuals and groups at risk of vulnerability or marginalisation

- Does the assessment process include sufficient time for capacity building to allow individuals and groups to be meaningfully involved (e.g. to first present the digital products or services in a way that the audience understands, and to follow-up later with the same groups when they have had time to discuss and organise, in order to receive feedback and potential concerns)?
- Do rightsholders have access to independent and competent legal, technical and other advice as necessary? If not, does the impact assessment include provisions for making such support available?
- Does the impact assessment provide for capacity building of rightsholders to know their rights (e.g. by thoroughly explaining the right to privacy before explaining how the digital product or service will be developed to ensure respect for the same right), as well as of duty-bearers to meet their human rights duties?
- Does the impact assessment provide particular attention to vulnerable or marginalised individuals and groups in engagement and participation activities (e.g. by allowing sufficient time and resources to facilitate the inclusion of these individuals)?

4. Transparency Transparent character of the impact assessment process to adequately engage affected or potentially affected rightsholders.

Guiding Questions:

- Does the impact assessment process provide for information sharing between stakeholders at relevant and regular intervals?
- Is the information about the digital project, product or service available to participating stakeholders adequate for giving a comprehensive understanding of potential implications and human rights impacts (e.g. information on intended use-cases, potential misuse and measures to address it, application and functioning of a service)?
- Are HRIA findings and impact management plans (action plans) publicly communicated to the greatest extent possible (e.g. published, with any reservations based on risk to rightsholders or other participants clearly justified)? Is there a firm top-level management commitment in place with regard to transparency before the start of the HRIA process?
- Are the phases of the impact assessment, including timeframes, communicated to relevant stakeholders in a clear and timely manner?
- Does communication and reporting take into account and respond to the local context? For example, is information made available in relevant languages and formats, in non-technical summaries and in physical and/or web-based formats that are accessible to stakeholders?
- 5. Accountability

The impact assessment team is supported by human rights expertise, and the roles and responsibilities for impact assessment, prevention, mitigation, and management are assigned and adequately resourced.

- Is responsibility for the implementation, monitoring and follow-up of mitigation measures assigned to particular individuals or functions within the company (e.g. data engineers are tasked with changing the design to limit potential misuse)?
- Are sufficient resources dedicated to undertaking the HRIA, as well as implementing the impact management plan (i.e. adequate time, as well as financial and human resources)?
- Are relevant duty-bearers meaningfully and appropriately engaged in the impact assessment process, including in impact prevention, mitigation and remediation (e.g. data protection authorities are engaged since some systemic impacts can best be dealt with through data protection policies and regulation)?
- Does the HRIA draw on the knowledge and expertise of other relevant parties, in particular human rights actors (e.g. digital rights groups working on right to privacy, fair machine learning etc.)?
- Does the HRIA team have the relevant inter-disciplinary skills and expertise (including human rights, technical, legal, language and local knowledge) to undertake the HRIA in the given context and with regard to the specific product or service (e.g. data engineers and software developers might need to be involved)?
- Have efforts been made to include local individuals, including women, in the impact assessment team, if appropriate?

6. Benchmark

Human rights standards constitute the benchmark for the impact assessment. Impact analysis, assessment of impact severity and design of mitigation measures are guided by international human rights standards and principles.

Guiding Questions:

- Are international human rights standards and principles used as the benchmark for the assessment?
- Is the impact assessment addressing the full scope of relevant human rights? If certain human rights are excluded from the assessment, is the basis for this reasonable, as well as explicitly noted and explained in the impact assessment?
- Is the scoping, data collection, analysis of actual and potential impacts, and design of mitigation measures guided by the substantive content of human rights?
- 7. Scope of impacts

The assessment identifies actual and potential human rights impacts that the technology caused or contributed to-, and impacts that are directly linked to the technology. The assessment analyses cumulative impacts and legacy issues.

Guiding Questions:

- Does the assessment include actual and potential impacts related to the digital project, products or services? Are these impacts categorized by: caused, contributed to, and directly linked?
- Does the assessment assess human rights impacts that the digital project, product or service is directly linked to (e.g. negative impacts caused by non-users)?
- Does the assessment consider any cumulative impacts, i.e. impacts that arise due to the aggregative or cumulative effect of multiple activities (e.g. several actors sharing limited amounts of data to one source leading to a negative impact on the right to privacy at a later stage when more data points about the same individual has been collected; or several actors taking down user-generated content, which leads to a significant negative impact on freedom of speech)?
- Does the assessment identify and address any legacy impacts associated with the digital project, product or service (e.g. previous companies have been reckless with the handling of data causing wide ranging impacts on the right to privacy and many other rights; or where a previous digital health service discriminated a minority group, which is now reluctant to access any new similar services)?
- 8. Assessing impact severity

Impacts are addressed according to the severity of their human rights consequences.

- Is the assessment of impact severity guided by relevant considerations, including the scope, scale, irremediability and interrelatedness of impacts?
- Is the assessment of severity determined with respect to the consequences for the individuals affected (as opposed to risk to the business)?
- Are the relevant rightsholders and/or their legitimate representatives or proxies involved in the assessment of impact severity? Does the assessment of severity reflect the views of the relevant rightsholders? If it does not, has that been appropriately explained?
- Has the analysis of impacts taken into account the interrelatedness of human rights, as well as the interrelatedness of social and human rights factors? (e.g. if a digital product discriminates an individual applying for and being declined a loan, this may have a corresponding impact on the rights of that individual's children to care; or if a business has insufficient safeguards in place in relation to data privacy, this may have an impact on the right to privacy but also on e.g. employees' right to freedom of association since they may not want their superiors to know their political affiliation.)

9. Impact mitigation measures

All human rights impacts are addressed. Where it is necessary to prioritise actions to address impacts, severity of human rights impacts is the core criterion. Addressing identified impacts follows the mitigation hierarchy of 'avoid-reduce-restore-remediate'.

Guiding Questions:

- Are all human rights impacts that are identified addressed?
- If it is necessary to prioritise actions to address impacts, is such prioritisation guided by the severity of human rights consequences?
- In determining mitigation measures, are all efforts made to first avoid the impact altogether, and if this is not possible, to reduce, mitigate and remediate the impact?
- Is care taken to ensure that compensation is not considered synonymous with impact mitigation and remediation?
- Does the impact assessment identify ways of exercising leverage to address any impacts which the technology contributes to or is directly linked to (e.g. through business relationships)? Where leverage does not exist, does impact mitigation include building leverage to address such impacts?

10. Access to remedy

Impacted rightsholders have avenues whereby they can raise grievances regarding the digital project,

Guiding Questions:

- Does the impact assessment identify actual impacts for which a remedy is needed? Are such impacts referred to the appropriate channels for remediation, including legal and non-legal, as appropriate?
- Have any severe human rights impacts that may constitute a legal breach been referred to the relevant legal channels (pending the consent of the rightsholders involved)? Does the project, product or service owner co-operate in legal proceedings?
- Is there an operational-level grievance mechanism in place that contributes to ongoing impact management, as well as the identification of unanticipated use-cases and impacts? If not, does the impact management plan include the establishment of such a mechanism? Does the operational-level grievance mechanism meet the eight effectiveness criteria for non-judicial grievance mechanisms that are outlined in UN Guiding Principle 31?
- Is it ensured that the operational-level grievance mechanism does not deny rightsholders access to all relevant judicial processes?
- Are the access to remedy channels responsive to the context and preferences of the rightsholders in question?

Sources: These criteria are based on a literature review including sources on human rights impact assessment, stakeholder engagement, social impact assessment and the human rights-based approach. See Danish Institute for Human Rights, "Human rights impact assessment guidance and toolbox": https://www.humanrights.dk/business/tools/human-rights-impact-assessment-guidance-toolbox