THE DANISH INSTITUTE FOR HUMAN RIGHTS

F,

PHASE 4: IMPACT PREVENTION, MITIGATION AND REMEDIATION

GUIDANCE ON HRIA OF DIGITAL ACTIVITIES

0

PHASE 4: IMPACT PREVENTION, MITIGATION AND REMEDIATION

Acknowledgments: This 2020 version of the Guidance was written by Emil Lindblad Kernell and Cathrine Bloch Veiberg, with research assistance from Claire Jacquot. We would also like to thank Eva Grambye, Elin Wrzoncki, Tulika Bansal and Ioana Tuta for their input and support. The Guidance has been developed on the basis of the HRIA Guidance and Toolbox. We also received invaluable input from several individuals and organisations who contributed their expertise, reflections and time on a voluntary basis, for which we are deeply thankful. We wish to extend our sincere thanks to: Rikke Frank Jørgensen from the Danish Institute for Human Rights; Richard Wingfield from Global Partners Digital; Jan Rydzak from Ranking Digital Rights; Jason Pielemeier and David Sullivan from Global Networking Initiative (GNI); Molly Land from University of Connecticut School of Law; Alex Warofka from Facebook; Nicole Karlebach from Verizon; Alexandria Walden from Google; Michael Karimian from Microsoft; Patrik Hiselius from Telia; Dunstan Allison-Hope from BSR; Margaret Wachenfeld from Institute for Human Rights and Business (IHRB); Mark Hodge and Lene Wendland from OHCHR and the B-Tech project; Isabel Ebert, University of St. Gallen, on behalf of the German Institute for Human Rights; Ephraim Percy Kenyanito from ARTICLE 19; Louise Kjær from Bluetown; and Elias Aboujaoude, Professor of Clinical Psychiatry at the Stanford University School of Medicine. In addition, the DIHR would like to recognise the collaboration with Lorna McGregor and Sabrina Rau from the Human Rights and Big Data Project (HRBDT) at Essex University, as DIHR's academic partner in the project, with a particular focus on Data Privacy Impact Assessments.

The contribution of expert reviewers does not represent their endorsement of the content.

© 2020 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 author and source are quoted.

At DIHR we aim to make our publications as accessible as possible. We use large font size, short (hyphen-free) lines, left-aligned text and strong contrast for maximum legibility. For further information about accessibility please click www.humanrights.dk/accessibility

CONTENT

IMPACT	PRE	VENTION, MITIGATION AND REMEDIATION	5
1.1	PLANNING AND RESOURCING FOR HUMAN RIGHTS IMPACT		
	M	ANAGEMENT	6
1.2	DE	EVELOPING ACTIONS TO ADDRESS IMPACTS AND EX	ERCISING
	LE	VERAGE	10
1.2	2.1	LEVERAGE IN RELATION TO THIRD PARTIES	13
1.2	2.2	EXAMPLE HRIA FINDINGS AND MITIGATION MEAS	SURES 17
1.3	Μ	ONITORING	24
1.3	3.1	PARTICIPATORY MONITORING	28
1.4	AC	CCESS TO REMEDY AND OPERATIONAL LEVEL GRIEVA	ANCE
	Μ	ECHANISMS	30
END NO	TES		34

This document contains the Phase 4: Impact Prevention, Mitigation and Remediation section of the Guidance on Human Rights Impact Assessment of Digital Activities (the Guidance).

You can access the full version of the Guidance at:

https://www.humanrights.dk/publications/human-rights-impact-assessmentdigital-activities

A NOTE ON THIS VERSION

This first version of the Guidance on Human Rights Impact Assessment (HRIA) of Digital Activities (the Guidance) is based on DIHR materials and experiences, input from expert reviewers and practitioners, the UN Guiding Principles on Business and Human Rights and international human rights instruments, as well as public domain sources on impact assessment.

The preparation of this Guidance included a workshop in Denmark in November 2019, during which 20 expert reviewers participated in a discussion on human rights impact assessment of digital activities i.e. digital projects, products and services.

It is anticipated that in 2020-2021, a Phase II of the project will focus on applying the Guidance in practice, the gathering and sharing of learning, and subsequently updating the Guidance based on experiences from practice.

As HRIA of digital activities is an emerging practice, this Guidance seeks to provide support to those working with HRIA of digital projects, products and services, but also to contribute to a platform for dialogue about HRIA practice and standards in the 'digital' business and human rights field. In this context, we welcome comments from stakeholders on the Guidance and on experiences with using it.

Please send comments, questions and suggestions to:

Emil Lindblad Kernell <u>emke@humanrights.dk</u> and Cathrine Bloch Veiberg <u>cph@humanrights.dk</u>

Funding

Creation and publication of this guidance has been made possible by general operating funds received from the Danish Ministry of Foreign Affairs.

IMPACT PREVENTION, MITIGATION AND REMEDIATION

WHAT HAPPENS IN PHASE 4?

In the impact prevention, mitigation and remediation phase, the business, HRIA team, and stakeholders collaborate to create a plan for preventing, mitigating and remediating negative human rights impacts. All human rights impacts need to be addressed, with the most severe impacts taking priority. Rightsholders and/or their proxies should be meaningfully involved in planning, enacting and monitoring impact prevention, mitigation and remediation efforts.

Planning for effective impact prevention, mitigation and remediation (impact management, for short) is an integral part of the HRIA process. Allocating time and resources for developing a detailed impact management plan at the outset of the HRIA is very helpful for facilitating this.

In determining what actions should be taken to address identified impacts, mitigation plans should focus primarily on avoiding and reducing negative human rights impacts. Businesses should exercise leverage to address impacts that involve third parties such as government actors, peers and other actors in the digital ecosystem (developers, commissioners, investors, users etc.). As human rights impacts relate to a variety of business functions and related activities, it is also important to consider how different units within the business might be involved in human rights impact management (e.g. legal team, compliance team, policy team, technical staff, procurement, legal, data ethicists, data protection officer, sustainability team).

Once the adverse human rights impacts have been identified (see Phase 3) and an impact management plan has been created, it is important to follow-up on whether and how the actions to avoid and address the identified impacts are implemented and that they effectively address the impacts. Access to remedy is a key component of impact management. The role of operational-level grievance mechanisms in impact management, both as a resource to identify impacts as well as a means to address any grievances associated with the HRIA process itself, should be considered.

KEY QUESTIONS ADDRESSED IN THIS SECTION:
What can contribute to effective planning and resourcing for human rights impact management?

- What types of actions is a business expected to take in response to the different identified impacts?
- What is the role of leverage in impact management?
- How should a company engage with third-parties to address identified impacts?
- What is participatory monitoring and how can it be applied in impact management?
- What is the role of an operational-level grievance mechanism in human rights impact management?

11 1 PLANNING AND RESOURCING FOR HUMAN RIGHTS IMPACT MANAGEMENT

Impact management involves designing and implementing measures to avoid and address impacts through prevention, mitigation and remediation (hereinafter, simply, impact management). To ensure that the HRIA contributes to effectively addressing the human rights impacts that are identified, it is essential that adequate resources are allocated to impact mitigation, as well as for monitoring effectiveness, addressing unanticipated impacts, and resolving grievances. These considerations should be clearly outlined in an impact management plan. For examples of HRIA findings and mitigation measures, see section 1.2.3 of this chapter.

It is important to involve rightsholders and duty-bearers in the development of the impact management plan and its implementation, as relevant and appropriate. In order for rightsholders to meaningfully participate in impact management, whether through participatory monitoring (see Chapter 1.3.1) or other various means, it is important that they have the capacities needed to meaningfully participate. In some contexts this may be the case, however, in

many scenarios **the company will need to engage in capacity building initiatives to facilitate meaningful rightsholder participation** in impact management.

See cross-cutting Stakeholder Engagement section for more information around rightsholder engagement, capacity needs and common challenges to engagement (e.g. potential risks to rightsholders for engaging, stakeholder fatigue, and legal obstacles).

In the HRIA process, resources and approaches for impact management should be considered and accounted for from the outset, including through steps such as:

Impact management step	Example action
Ensuring that the development of an impact management plan is an integral part of the HRIA process	Providing for the development of an impact management plan in the TOR for the HRIA (see Phase 1 for more on TOR development).
Developing a detailed impact management plan that assigns specific persons to the implementation of the mitigation measures (e.g. legal data engineers, sustainability team, sales and procurement), and ensuring that the people assigned have the relevant skills, time, management, support and other resources necessary to effectively implement the mitigation measures.	For example, ensuring that the sustainability team is tasked with human rights capacity-building activities for data engineers and software developers, who in turn are responsible for assessing the digital products they develop for certain specific human rights risks.
Developing the impact management plan collaboratively, involving: individual end-users and other rightsholders, and/or their proxies; government actors; and other relevant organisations and experts.	Workshops with digital rights groups and bilateral engagement with other businesses in the digital ecosystem can help gain stakeholder and rightsholder buy-in for the proposed mitigation measures. Impact management plans could also be shared with stakeholders engaged during earlier stages of the HRIA process, including rightsholders, for their input.

	If impacts on women's rights have been identified, a local NGO or CSO working on women's rights might be involved in impact mitigation planning and implementation, and if children may be impacted UNICEF could be engaged.
Ensuring that the impact management measures are based on and build on the human rights indicators that have been established in the scoping phases.	See Data Collection and Context Analysis, Chapter 1.4.
Integrating different mitigation measures into the relevant management plans and systems of the business.	This may include: developing Key Performance Indicators (KPIs) that incentivise early, extensive and meaningful stakeholder engagement; integrating human rights in existing development review processes; and introducing Know Your Customer (KYC) processes in sales of potentially harmful digital products and services in order to better control and limit future use cases.
Ensuring that the business commits to dedicating adequate and appropriate resources for the implementation of impact mitigation measures and ongoing impact management, including through assigning adequate budget, time and human resources to impact management.	This may include: in response to a finding in the HRIA, the company could allocate resources to recruit and train additional content moderators that speak a variety of local languages to detect and remove hate speech posts on a social media platform, which have the potential to lead to offline violence
Taking a multi-disciplinary and cross-functional approach to impact management . Often, departments within the business which oversee data ethics and privacy, community relations, social	This may include creating a (or tasking an already existing) cross-functional working group to monitor the implementation of the HRIA impact management plan and its effectiveness.

This may include, establishing monitoring structures with individual end-users, rightsholder proxies, digital rights groups and others.
This may be done through involving national agencies responsible for data- related issues (i.e. data protection authorities or data ethics councils) when addressing impacts associated with data collection and sharing, or aligning impact mitigation strategies with national data strategies where possible and as appropriate.
This may include identifying that internet shutdowns is a major issue that no single telecommunications company can address, whereas collective action from all telecommunications companies in a country to influence the government may be more successful.
See Chapter 1.4.

human rights impacts throughout and beyond the HRIA process.

12 2 DEVELOPING ACTIONS TO ADDRESS IMPACTS AND EXERCISING LEVERAGE

In developing actions to address the human rights impacts that have been identified, several points should be considered:¹

	impacts should be adressed as a matter of priority, as explained in Phase Analysing Impacts.
	+
F	The identification of actions to avoid and/or address the identified potential and actual impacts should involve the rightsholders who are (potentially) impacted, or their proxies, as well as relevant duty-bearers and other relevant parties.
	+
	The mitigation hierarchy applied should be compatible with international human rights standards and principles.
	+
t	The kind of involvement in the adverse impacts (cause, contribute, linked to) does not dicatate which impacts ought to be prioritised, but will imply different types of preventive or mitigation measures , and
	+
6	For impacts that the business contributes to or is directly linked to, the extent of leverage that the business can exercise in addressing the

Box 1, below, outlines some points to consider for developing a human rightscompatible mitigation hierarchy.

BOX 1: THE MITIGATION HIERARCHY

The majority of mitigation hierarchies in different kinds of impact assessment methodologies take the following approach:

- Avoid: making changes to the project, product or service to avoid the impact.
- Reduce: implementing actions to minimise the impacts.
- Restore: taking actions to restore or rehabilitate to the conditions that existed prior to the impact.
- Compensate: compensating in kind or by other means, where other mitigation approaches are neither possible nor effective.

In broad terms, a similar approach can be adopted in HRIA of digital activities, i.e. an approach that always prioritises avoiding, and if this is not possible, considers ways to reduce and mitigate impacts. However, from a human rights perspective, there are <u>three things which warrant attention</u> when adapting the above approach for HRIA:

- 1. Any measures taken must themselves be compatible with international human rights standards, as well as a human rights-based approach.
- 2. **Remediation should be explicitly included.** This includes understanding and explaining that compensation and remediation are not synonymous, and that compensation should not be the default remedy.
- 3. Human rights impacts cannot be subject to 'offsetting', as compared to, for example, environmental impacts which can be offset (for more information, see chapter 1.3 in the Phase 3 section of the Guidance).

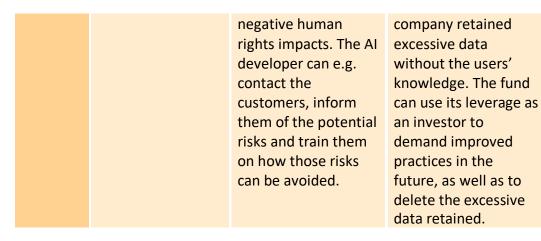
In **determining what type of action** to take to address a particular impact, there will be differences depending on whether the business has *caused, contributed to* or is *directly linked to* the impact. The type of action will also differ if the impacts are cumulative due to systemic issues that a collective of actors are best placed to address.

Note, however, that all impacts regardless of the type of involvement should be addressed, and that the prioritisation of impact mitigation measures is based on severity and not on 'proximity'—a very severe impact that a company is linked to may warrant immediate action, whereas a less severe impact that it contributes to can be addressed at a later stage. See Chapter 1.2 in the Phase 3 section of the Guidance for more on severity.

Table A, below, provides an overview for determining appropriate business responses for each of the different types of impacts.

TABLE A: DETERMINING APPROPRIATE ACTIONS TO ADDRESS THE IMPACTS IDENTIFIED			
Type of impact	Impacts <i>caused</i> by the business	Impacts to which the business contributes	Impacts <i>directly linked</i> to a business's operations, products or services through its business relationships, contractual and non- contractual
Required actions	Take necessary steps to cease and prevent the impact. Provide for, or collaborate in, remediation for actual impacts caused.	Take necessary steps to cease or prevent contribution to the impact, including through exercising leverage and taking steps to increase leverage if this is needed. Provide for, or cooperate in, the remediation of adverse impacts.	Exercise existing leverage to prevent or mitigate the impact. Increase and exercise leverage if existing leverage is inadequate. The business is not required to provide for remediation, although it may take a role in doing so.
Example	A bank that is using algorithmic credit risk scoring that is indirectly discriminating against ethnic minorities should immediately stop using the system until any issues have been fixed.	An Al developer develops an automated decision- making algorithm for 'efficient hiring' and markets it to business customers. Some of its customers misuse the algorithm in a way that means that the Al developer contributes to	A private equity fund invests in a biotech company operating in a country without data protection laws. Following its due diligence, the fund recommends changes in the company's data protection practices. However, upon an external audit it is found that the

TABLE AV DETERMINING APPROPRIATE ACTIONS TO ADDRESS THE IMPACTS



In determining appropriate actions to address identified impacts that a company is **directly linked to** through its business relationships, the UN Guiding Principles suggest that the following factors should be considered:

- The severity of the situation
- The business's leverage over the entity/entities concerned (see Chapter 1.2.1 below)
- How crucial the relationship is to the digital project, product or service, and
- Whether terminating the relationship with the entity itself would have adverse human rights consequences.

121 1 LEVERAGE IN RELATION TO THIRD PARTIES

'Leverage' is considered to exist where a business has **the ability to effect change in the practices of another entity that causes harm**—i.e. "leverage is a company's ability to influence the behaviour of others."² **If the business has leverage, it is expected to exercise it. If the business lacks leverage, it is expected to seek ways to increase it**, for example, by offering capacity building or other incentives to the third party to address the impact.

It is *important to remember the following* in relation to leverage:

- 1. Severity is relevant for determining the <u>order of priority</u> in which the identified impacts should be addressed.
- Leverage becomes relevant for determining <u>how</u> to address impacts that the business contributes to or is directly linked to through its business relationships.
- 3. The **absence of leverage does not absolve a business from responsibility** to address the impacts that have been identified.

Table B, below, gives an overview of some examples of different types of leverage and how leverage might be exercised.

ADDRESS HUMAN RIGHTS IMPACTS				
Examples of types of leverage	Examples of exercising leverage			
Traditional commercial leverage — leverage that sits within the activities the company routinely undertakes in commercial relationships, such as through contracting.	 Include human rights standards in contracts— purchasing and sales contracts as well as licensing and other forms of agreements. Retain a right to assess compliance with human rights standards included in sales contracts. For developers: Include information in contracts about ways in which the digital product or service should not be used. For developers: Put in place an application process for the sale of the product or service and develop an 'allow list' of acceptable customers. This can serve to 'gate' the use of the product or service so that only certain planned use-cases are accepted. For companies and state actors procuring digital products or services: Include human rights standards in pre-qualification criteria in bidding processes, in order to ensure that bids without human rights considerations are not competitive. For a state actor this could include the procurement of e.g. digital health service platform or the development of a contact tracing app. 			
Broader business leverage—leverage that a company can exercise on its own through activities that are not routine or typical in commercial relationships, such as capacity building.	• Build the capacity of business partners to meet the responsibility to respect human rights—e.g. developers of algorithms that assist in capacity building efforts of those buying the algorithms to address potential bias issues and to ensure other human rights are not impacted in the application, or the other way around.			

TABLE B: EXAMPLES OF EXERCISING AND INCREASING LEVERAGE TOADDRESS HUMAN RIGHTS IMPACTS

TABLE B: EXAMPLES OF EXERCISING AND INCREASING LEVERAGE TOADDRESS HUMAN RIGHTS IMPACTS

Examples of types of leverage	Examples of exercising leverage	
	• Ensure that purchasing teams have the capacity to discuss potential human rights risks and issues in their conversations with suppliers tasked with developing the products or services.	
Leverage together with business partners ³ — leverage created through collective action with other companies in or beyond the same industry	 Work with business partners to establish common pre-qualification criteria for specific bidding processes. Work with peers to develop capacity-building materials together on the appropriate use of the digital products or services developed. Engage with peer companies in the same sector who may be facing similar issues to share lessons learned and identify and co-create possible solutions for the sector. This may include: social media platforms developing standard, public methodologies, in collaboration with one another around disinformation around election results. Important note: ensure that collaboration stays clear of anti-competitive behaviour. 	
Leverage through bilateral engagement— leverage generated through engaging bilaterally and separately with one or more other actors, such as government actors, business peers, international	• Engage government actors on identified issues that are likely to need policy responses rather than individual company actions due to their systemic nature. For example, a development finance institution (DFI) investing in digital transition projects in a country should engage with the government in case it has found that it may be linked to negative impacts on the right to privacy due to inexistent data protection regulations in the country.	

TABLE B: EXAMPLES OF EXERCISING AND INCREASING LEVERAGE TOADDRESS HUMAN RIGHTS IMPACTS

Examples of types of leverage	Examples of exercising leverage
organisations and/or CSOs.	
Leverage through multi- stakeholder collaboration—leverage generated through collective, collaborative action with business peers, governments, international organisations and/or NGOs or CSOs.	 Develop shared standards for developers and/or those procuring digital products or services through multi-stakeholder initiatives, thereby enhancing the credibility of the standards. Use the business' brand and reputation to convene relevant stakeholders (from industry peers and government actors to civil society and academics) to address identified systemic issues that are beyond one single entity to solve. For example, to conduct collective public policy advocacy around the need for clear laws and regulations, based on international human rights principles, to address harmful content online.⁴ Collectively encourage governments and international institutions to adopt policies, practices, and actions that are consistent with identified relevant human rights.⁵ For example, telecommunications companies and other stakeholders issuing joint public statements on network and service
	shutdowns. ⁶

Source: Adapted from Shift (2013), *"Using Leverage in Business Relationships to Reduce Human Rights Risks"*, New York: Shift, p.14-24.

122 2 EXAMPLE HRIA FINDINGS AND MITIGATION MEASURES

Table C, below, provides examples of HRIA findings and potential mitigation measures by issue area.

TABLE C: EXAMPLES OF FINDINGS AND MITIGATION MEASURES			
Issue area	HRIA findings	Recommendations/mitigation measures	
Misinformation	Synthetic speech technologies using deep learning, a powerful machine learning technique, can create very credible synthetic speech with limited audio recordings from real individuals. Manipulated speech can be 'weaponised' to spread misinformation and can thereby harm both individuals and society at large by e.g. further increasing social divisions and distrust in governments.	Participate in efforts to establish industry standards on how and when synthetic speech can and/or should be used in order to avoid potential severe human rights impacts. Restrict ('gate') the sales of the technology to ensure that the use of synthetic speech is used in circumstances where the risks related to misinformation are minimal. Implement a customer application process to define which clients will (and will not) be allowed to buy the digital product based on their ability to protect the right to privacy, based on a specific risk assessment process.	
Privacy	The large amount of information collected by a company through its digital platform (e.g. a search engine) may create incentives for law enforcement agencies to request	Redesign the product so that it collects only as much information as is justified and necessary, with a strong focus on data minimisation rather than data anonymisation.	

TABLE C: EXAMPLES OF FINDINGS AND MITIGATION MEASURES			
Issue area	HRIA findings	Recommendations/mitigation measures	
	personal data or access to large pools of data following a court order.	Develop vetted processes for reviewing and responding to government requests, including that requests should only be honoured if there's a sufficient legal basis and that they should be interpreted as narrowly as possible. ⁷ Produce and publish periodic transparency reports detailing government requests and responses. Publish periodic human rights updates to the public and provide best practice guidance, advice, and training to customers on how they can use the technology in a manner that reduces the risks related to the collected data, if requests are made by law enforcement agencies.	
Discrimination	A company's social media platform is found to host content discriminating against rightsholders based on characteristics such as gender, LGBT status, national origin, ethnicity, language, pregnancy, or disability—e.g.	Recruit and train moderators belonging to the rightsholder groups that are discriminated against. Take proactive efforts to reduce discrimination risk by providing training to the platform's moderators and operators working on the development of the technology.	

TABLE C: EXAMPLES OF FINDINGS AND MITIGATION MEASURES			
Issue area	HRIA findings	Recommendations/mitigation measures	
	through reduced career opportunities when the platform and the data available there is used by third parties in their hiring processes. The same platform is also found to have an unequal benefit of use between men and women in certain countries.	Initiate (or participate in) country-focused multi- stakeholder dialogues on the future of anti-discrimination protections in an era of the Internet and social media.	
	Brands and other advertisers on a digital platform target specific groups of internet users in ways that are discriminatory.	 Limit advertisers' ability to narrowly segment audiences in ways that may result in a discrimination of specific rightsholder groups. Prohibit advertisers from targeting ads for jobs, housing, and education according to users' assumed age, sex, race, or other protected category. Monitor algorithmically generated advertising categories to ensure that they do not facilitate negative impacts on human rights. 	
	A developer of an automated decision- making algorithm has been informed	Provide information and guidance to customers on how they should deploy and make use of the algorithm in order to avoid biased outcomes.	

TABLE C: EXAMPLES OF FINDINGS AND MITIGATION MEASURES		
Issue area	HRIA findings	Recommendations/mitigation measures
	that its algorithm is used in ways that lead to highly biased outcomes.	
Children's rights	Publishers and digital marketing agencies active on a digital platform works with their clients to target children by inviting them to access inappropriate content or communicating with them in ways that may result in real world harm, such as psychological harm, trafficking, or sexual exploitation.	Develop service specific terms that limit underage access to the particular product or service, and that address children rights. Monitor the adherence to the terms via user-testing or other relevant mechanisms. Engage with child rights organizations such as UNICEF and local organizations to discuss and better understand these concerns. Integrate child rights considerations into all appropriate corporate policies and management processes. Develop safer and age appropriate online environments on the platform Actively support the primary role of parents in evaluating and minimizing risks of harm to their children. Encourage adults to be involved in children's consumption and use of online content, so that they can assist and guide children.

TABLE C: EXAMPLES OF FINDINGS AND MITIGATION MEASURES		
Issue area	HRIA findings	Recommendations/mitigation measures
		Engage a child rights organization to engage with parents and communities and jointly develop trainings and awareness materials (for the parents, caretakes, teachers and for company staff) on child violence online or violence that is otherwise related to online activities.
Freedom of expression	A company is found to be involved in the overly extensive removal of content (e.g. content intended to expose rather than encourage human rights violations), content restrictions demanded by government, and the potential blocking of the platform by governments. A digital communication platform has become the major source of information and expression in a country, leading to the responsibility of the company to preserve freedom of speech.	Provide training on human rights to moderators, in order to help moderators balance the need to respect freedom of speech online. Get involved in or initiate a debate regarding the government policies on information regulation and create new opportunities to increase the realization of human rights.

TABLE C: EXAMPLES OF FINDINGS AND MITIGATION MEASURES		
Issue area	HRIA findings	Recommendations/mitigation measures
Hate speech	A company's platform is found to be used by certain groups to spread rumors and hate speech, to incite violence against other individuals (especially marginalized communities).	Fund research on how hate speech is spread in relevant countries and act upon relevant findings. The creation of "counter hate-speech" content may contribute to reduce risks. Advocate and collaborate with national human rights institutions on the elaboration of anti-hate speech regulations in domestic law.

In Box 2, below, you can read about how a process can be structured to identify future impacts and integrating mitigating measures in the design of a digital product or service.

BOX 2: IDENTIFYING MITIGATING MEASURES FOR THE DESIGN STAGE OF A PRODUCT OR SERVICE – HUMAN RIGHTS BY DESIGN

Contribution by Dunstan Allison-Hope, BSR.

A significant challenge when assessing human rights impacts and identifying mitigating measures in the digital sphere is the **interplay between the design of a product or service by a technology company and how it is applied or used in real life**, whether by individuals, enterprise customers, or governments and states.

Companies today deploy increasingly sophisticated 'privacy by design' processes that integrate privacy considerations during key milestones in product design and development so that products are more privacy protective when deployed in the real world. A similar approach can be taken for the full range of potential human rights impacts, and **preventive measures can be taken early in the lifecycle of digital products or services**.

Privacy by design principles include **being proactive rather than reactive by anticipating and preventing** privacy invasive events before they happen, and embedding privacy into the design and architecture of products, services, and systems, not bolting them on after the fact. These principles align strongly with the spirit and intent of human rights due diligence as well as with HRIAs conducted during the development stage of digital projects, products or services.

According to BSR a 'human rights by design' process includes three core elements:

- **Cross functional:** Bring insights from a range of professional communities business and human rights teams, product managers, research and design teams, and sales and marketing teams—to fully integrate human rights considerations into the design, development, and sale of new products, services, and technologies.
- Inclusive: In the technology industry, companies identify "personas" to represent the different user types that might use a product or service, and design with their needs in mind. A "human rights by design" approach would identify "personas" from a diverse range of backgrounds—especially personas from groups or populations that may be at heightened risk of vulnerability or marginalization and understand how their rights may be impacted during product or service application or use. When a human rights by design process is informed by a HRIA this will include meaningful engagement with real users and potentially impacted non-users. (See chapter 1.3 in cross-cutting Stakeholder Engagement section)

• **Expansive and forward looking:** A "human rights by design" approach seeks to identify a wide range of potential impacts that might arise in the future and is not constrained by impacts that are well known today. The futures methodology example in Phase 3 of the Guidance explains this approach in more detail.

A 'human rights by design' approach can enhance the product design process by ensuring that respect for human rights is deliberately integrated throughout and that more rights-respecting design choices can be made, as part of efforts to mitigate identified potential impacts. Conducting a robust HRIA at the outset of a process of human rights by design can inform such future adaptations.

See also: Global Network Initiative, *"The GNI Principles at Work: Public Report on the Third Cycle of Independent Assessments of GNI Company Members 2018/2019"*, p. 35: <u>https://globalnetworkinitiative.org/wp-content/uploads/2020/04/2018-2019-PAR.pdf</u>

13 3 MONITORING

Once adverse human rights impacts have been identified and an impact management plan has been finalised, it will be **important to follow up on whether the actions to address the identified impacts are implemented and whether they effectively address the identified impacts**. Planning for the monitoring of impact mitigation measures should therefore be an integral component of a HRIA and be included in the impact management plan.

It is important that any planning for monitoring <u>considers the following</u> <u>questions</u>:

- What, precisely, is to be monitored?
- When—e.g. how long after finalised HRIA report—should monitoring activities occur?
- How often—at what intervals?
- Who, internal and/or external, should conduct the monitoring activities?

In addition to providing information on whether the impact mitigation measures are effective, and making any necessary adjustments if they are not, ongoing monitoring provides an opportunity to identify unforeseen impacts.

Involving rightsholders, duty-bearers and other relevant parties in impact monitoring, as appropriate in the given context, can provide valuable opportunities for strengthening accountability and building trust between different parties. It can also provide a way to involve the necessary expertise or contribute to building the capacity of the stakeholders involved in impact management. See Table D, below, for more information about key questions that practitioners may reflect on when developing a monitoring plan.

TABLE D: QUESTIONS TO CONSIDER WHEN DEVELOPING A MONITORING PLAN	
Question	Description
Why are you monitoring?	This clarifies the purpose of the monitoring plan. For example, is the monitoring being conducted to address the public's uncertainties and concerns related to the digital project, product or service? Is it to gather data periodically? Did the assessment find a potential impact that needs ongoing monitoring in order to ensure that the project, product or service is not causing or contributing to an actual human rights impact?
What will be monitored?	 A monitoring plan can be set-up to measure several potential or actual impacts. For example, it may be set-up to monitor internal processes and/or staff, third party staff or impacts related to individual users or other rightsholders. Some examples to monitor relating to potential impacts may include: Monitor the number of complaints received by the company in relation to the digital project, product or service in question. Monitor use-cases of the digital product or service by third-parties, to see whether there are any unintended use-cases. Monitor the outputs and predictions made by an algorithm that is meant to assist decision-making and check for biased outcomes. Monitor news stories and civil society reporting concerning the company's digital project, product or service and/or similar products or services. Monitor the number of complaints related to impacts on specifically vulnerable and marginalised groups. Monitor the effectiveness of a company grievance mechanism, with reference to the UNGPs effectiveness criteria.

TABLE D: QUESTIONS TO CONSIDER WHEN DEVELOPING A MONITORING PLAN

Question	Description
	 Monitor user satisfaction with human rights related engagement activities.
How will it be monitored?	Consider the type of monitoring being undertaken, and the necessary data that needs to be gathered in order tomonitor successfully. Will it be primarily qualitative data? Is it more complex, which may require specialised technologies and technical capacity? Can user-testing and surveys be used? Is it best to engage with digital rights groups?
Who will participate and what will be their level of participation?	 Both external stakeholders with a high level of expertise and members of various rightsholder groups may participate in the monitoring activities. Depending on what is being monitored, the level of complexity will determine the given approach as well as the necessary capacity building that may be needed. Methods of participation may vary, and may include: Collection of data Data interpretation Communicating results
How often will it be monitored?	The frequency and duration of monitoring should be noted. Will the monitoring occur weekly, monthly, quarterly, biannually, etc.? It is also important to decide on the frequency of data collection for monitoring purposes as well as the sample size necessary to ensure an effective study.
Who is responsible for the monitoring?	 This looks to the governance structure,⁸ the actual roles, procedures and organisation for the management of the monitoring plan. While the company can take the lead in monitoring, it may want to look to other options when considering issues relating to capacity, credibility, community trust and independence, such as: Involving an NGO or CSO specialising in internet freedom, digital rights, on the rights of specific rightsholder groups, privacy or data protection specialists etc.

TABLE D: QUESTIONS TO CONSIDER WHEN DEVELOPING A MONITORING PLAN

Question	Description
	 Involving academics, e.g. those focusing on systemic issues of digital platforms, business models, internet freedom or data protection. Involving a religious institution that has insight into e.g. hate speech and harassment against the group it represents. Involving a relevant state or government agency, e.g. a data protection authority or ombudsman office. Taking a multi-stakeholder approach, inviting various of the organisations or institutions mentioned above.
How will the data be used?	 The data collected may be used for many purposes, for example: Mitigating potential or actual human rights impacts. Ensuring compliance with regulatory requirements. Ensuring a human rights-based approach to monitoring, including by focusing on increased transparency. Creating awareness and educating. Addressing public perceptions on the digital project, product or service. Building capacity of the potentially affected rightsholders and company.
How will the company respond to monitoring findings?	 The plan should include structured responses for different findings from the monitoring. For example: If severe human rights impacts are found these should be escalated to senior management immediately. If specific mitigation measures are found to be effective this should be recorded for future learning on continual improvement. If specific mitigation measures are found to be ineffective the monitoring plan should require this to be brought to the attention of relevant staff, so that new mitigation measures can be promptly designed and implemented.

TABLE D: QUESTIONS TO CONSIDER WHEN DEVELOPING A MONITORING PLAN

Question	Description
How will results be presented to stakeholders?	In case of participatory monitoring, given its public and inclusive nature the results should also be made public. This means that data should be available and accessible to those who participated in the monitoring and to other interested stakeholders; and that the monitoring design, which describes the methods, the data collecting process, the process of interpreting the data, and actual findings as well as conclusions, should be available. In general, communication about monitoring processes and
	findings should be responsive to the specific context, i.e. in relevant local languages and through rightsholders' preferred modes of communication.
Funding ⁹	Like governance, the funding of a specific monitoring program is another important area to consider. Funding should be adequate in order to ensure effectiveness. The dilemma is that while the company is often expected to pay for the costs involved in monitoring, this funding may also bring issues of credibility, lack of independence and transparency by outside organisations. Having in place a multi-stakeholder or independent governance structure, or review panel, that is involved in administering the funding for monitoring activities may go some way to addressing this issue.

Source: Adapted from the Office of the Compliance Advisor/Ombudsman (CAO) for the International Finance Corporation (IFC) Multilateral Investment Guarantee Agency (MIGA) and Members of the World Bank Group (2008), "Participatory Water Monitoring: A Guide for Preventing and Managing Conflict": https://www.commdev.org/publications/participatory-water-monitoring-a-guide-for-preventing-and-managing-conflict-advisory-note/ [Accessed July 30, 2020]; see also Facebook "Oversight Board Charter", https://www.oversightboard.com/governance/

131 1 PARTICIPATORY MONITORING

One strategy for facilitating the participation of different stakeholders is participatory monitoring. Box 3, below, provides an overview of participatory monitoring.

BOX 3: WHAT IS PARTICIPATORY MONITORING?

Participatory monitoring can be defined as **"a collaborative process of collecting and analysing data, and communicating the results, in an attempt to identify and solve problems together".** The process includes the involvement of stakeholders across all stages of the monitoring process, and incorporates methods and indicators that are salient to the stakeholders concerned. Traditionally, companies and agencies initiate and undertake monitoring. Participatory monitoring implies a dynamic process where a range of stakeholders assume responsibility for these tasks, and learn and benefit from the results. Participatory monitoring is not only a scientific process, but also social, political, and cultural. It requires openness, a willingness to listen to different points of view, as well as a recognition of the knowledge and role of different participants, and the ability to give credit where credit is due.

It should be noted that participatory monitoring methodologies primarily come from experiences related to large-scale projects with physical footprints (e.g. extractive projects) which in comparison to digital activities relatively easy can identify the stakeholders within a defined geographical community. As such, in order for participatory monitoring in relation to digital activities to be a viable option it will likely require further discussions and dialogue.

Source: IFC (2010), "International Lessons of Experience and Best Practice in Participatory Monitoring in Extractive Industry Projects": <u>https://www.commdev.org/publications/international-lessons-of-</u> <u>experience-and-best-practice-in-participatory-monitoring-in-extractive-industry-projects-guidance-note-</u> <u>on-designing-participatory-monitoring-programs/</u> [Accessed July 30, 2020]; Joss & Bellucci (2002) "Participatory Technology Assessment: European Perspectives", University of Westminster (Centre for the Study of Democracy) & Swiss Centre for Technology Assessment.

A few notes on participatory monitoring:

- It can be a way to build understanding and trust between the different stakeholders involved in HRIA.
- It can, in particular, provide an avenue for dialogue between affected rightsholders and the business that stretches beyond a single HRIA process and feeds into ongoing HRDD. HRIA can play a role in identifying and establishing initial contact with the different rightsholders and duty-bearers who might be involved in community monitoring of the impact mitigation measures. HRIAs can also identify whether the individuals, communities and groups who are anticipated to participate in monitoring need additional capacity building.
- There is no 'one size fits all' solution, as the success of participatory monitoring schemes is very dependent upon each specific local context.¹⁰ This emphasises the importance of good context analysis and stakeholder

engagement throughout a HRIA process, which can then inform the design of any participatory monitoring to be implemented.

- It is likely to be most effective when designed and implemented at the outset of a project or early during the development of a product or service and, moreover, used throughout all stages of the project or product development cycle and not only when impacts cause user and rightsholder contention.
- If a participatory monitoring scheme is implemented in a reactive way, rightsholder groups may view it suspiciously as a tool designed to silence and co-opt dissenting voices. Therefore, the monitoring effort may suffer credibility issues and further contribute to conflict and tensions between the company and potentially impacted rightsholders.
- In some scenarios, **rightsholders and their representatives may need timeto develop the capacity and technical skills to meaningfully participate** in the monitoring. Participatory monitoring programmes should therefore include a focus on rightsholders' access and ability to participate in the process. Having access to a participatory monitoring programme without having the ability to meaningfully participate will be out of sync with a human rights-based approach. The same applies for the reverse where capacities exist, but the participatory monitoring programme is not accessible.

14 4 ACCESS TO REMEDY AND OPERATIONAL LEVEL GRIEVANCE MECHANISMS

Operational-level grievance mechanisms can have an important role to play in HRIA. Those impacted by the digital project, product or service may have grievances to raise with regard to the HRIA process itself as well as the specific impacts that have been identified. Access to remedy, of which operational-level grievance mechanisms is one component, is a core pillar of the UNGPs, which also outline eight effectiveness criteria for non-judicial grievance mechanisms (see Box 4, below). It should also be noted that industry or multi-stakeholder organizations can also be used to ensure the availability of grievance mechanisms.¹¹

BOX 4: EFFECTIVENESS CRITERIA FOR NON-JUDICIAL GRIEVANCE MECHANISMS

1. Legitimate: enabling trust from the stakeholder groups for whose use they are intended, and being accountable for the fair conduct of grievance processes

- 2. Accessible: being known to all stakeholder groups for whose use they are intended, and providing adequate assistance for those who may face particular barriers to access
- 3. **Predictable**: providing a clear and known procedure with an indicative time frame for each stage, and clarity on the types of process and outcome available and means of monitoring implementation
- 4. **Equitable**: seeking to ensure that aggrieved parties have reasonable access to sources of information, advice and expertise necessary to engage in a grievance process on fair, informed and respectful terms
- 5. **Transparent**: keeping parties to a grievance informed about its progress, and providing sufficient information about the mechanism's performance to build confidence in its effectiveness and meet any public interest at stake
- 6. **Rights-compatible**: ensuring that outcomes and remedies accord with internationally recognized human rights; and
- 7. A source of continuous learning: drawing on relevant measures to identify lessons for improving the mechanism and preventing future grievances and harms.

Operational-level mechanisms should also be:

8. **Based on engagement and dialogue:** consulting the stakeholder groups for whose use they are intended on their design and performance, and focusing on dialogue as the means to address and resolve grievances.

Much has been written about operational-level grievance mechanisms in theory and practice, including how they might be designed in collaboration with rightsholders and rightsholder groups to ensure responsiveness to the specific context.¹² Unfortunately, case studies that have analysed the effectiveness of operational-level grievance mechanisms are limited¹³ and generally concern different contexts than digital projects, products or services and are not immediately applicable to the context of such activities.

A few specific points can however be said with regard to grievance mechanisms developed in relation to digital activities, specifically in relation to the accessibility of the mechanism.

• A grievance mechanism made available in relation to digital activities **might need to meet the need of very large amounts of rightsholders**, who could be

from countries across the globe and who will therefore speak many languages. This contrasts with grievance mechanisms developed for other industries that are physically present in clearly defined geographical areas.¹⁴ While the digital activities may be global, the grievance mechanisms should be adapted to local contexts in order to be effective.

- Appropriate channels must be made available for rightsholders that are not users of the digital product or service in question (e.g. individuals that have been victims of content shared on social media platforms that led to offline violence).
- Companies should **identify all the existing communication channels and ways of interacting with users/customers that they have** (which for some digital activities may be many), in order to assess how they can be adapted to deal with human rights-related complaints and concerns.

There are numbers of ways in which grievance mechanisms specifically interact with HRIA, including:

For currently ongoing digital projects and already launched digital products and services where a grievance mechanism is already in place:

- Information from the grievance mechanism can inform the HRIA about any patterns or trends identified from the grievances that have been submitted—this could also be found through user testing and survey processes, which may not amount to operational-level grievance mechanisms but that may provide relevant information nonetheless. They are likely to provide useful information about the concerns of impacted rightsholders and interested civil society organisations.
- The HRIA can provide insights about if and how the existing grievance mechanism might need to be revised to ensure its effectiveness—by e.g. consulting vulnerable rightsholder groups on how they would like to be able to lodge complaints in the future, should they have any.

For the development of digital products and services, or ongoing running of digital projects, where a grievance mechanism is not already in place:

- The information gained through the HRIA can provide insights about how an operational-level grievance mechanism could be designed and implemented and updated to ensure responsiveness to the relevant (local) context, including e.g. by identifying any existing methods, approaches or rightsholder preferences for grievance resolution, and
- A preliminary channel for grievance resolution throughout the HRIA process should be established as part of embarking on a HRIA.

Overall, operational-level grievance mechanisms can be important for the early identification of impacts, as well as for the ongoing monitoring of the effectiveness of impact mitigation. The development, review and/or implementation of operational-level grievance mechanisms should therefore be an integral component of a HRIA process.

END NOTES

¹ UN Guiding Principle 19.

² Shift (2013), "Using Leverage in Business Relationships to Reduce Human Rights Risks", New York: Shift, p.3.

³ Remember at all times the need to collaborate with industry peers in ways that do not amount anti-competitive behaviour.

⁴ See e.g. Global Network Initiative (Oct 13, 2020), "Addressing Digital Harms AND Protecting Human Rights — GNI Shares Recommendations for Policymakers": https://globalnetworkinitiative.org/content-regulation-policy-brief/

⁵ See e.g. Global Network Initiative's "Implementation Guidelines", particularly in relation to Multi-Stakeholder Collaboration: <u>https://globalnetworkinitiative.org/implementation-guidelines/</u>

⁶ Global Network Initiative & Telecommunications Industry Dialogue (July 12, 2016), *"Joint Statement on Network and Service Shutdowns"*: <u>https://globalnetworkinitiative.org/gni-id-statement-network-shutdowns/</u> [Accessed Oct 28, 2020]

⁷ See e.g. Global Network Initiative's "Implementation Guidelines", particularly in relation to '3. Freedom of Expression and Privacy': <u>https://globalnetworkinitiative.org/implementation-guidelines/</u>

⁸ Detailed governance approaches are outlined in the Office of the Compliance Advisor/Ombudsman (CAO) *for the* International Finance Corporation (IFC) Multilateral Investment Guarantee Agency (MIGA) and Members of the World Bank Group (2008), *"Participatory Water Monitoring: A Guide for Preventing and Managing Conflict"*: <u>https://www.commdev.org/publications/participatory-water-monitoring-a-guide-for-preventingand-managing-conflict-advisory-note/</u> [Accessed July 30, 2020].

⁹ Ibid.

¹⁰ IFC (2010), *"International Lessons of Experience and Best Practice in Participatory Monitoring in Extractive Industry Projects"*: <u>https://www.ifc.org/wps/wcm/connect/db4efbd9-647c-4882-bccd-3acdc62177e6/IFC_LOE_PLNG.pdf?MOD=AJPERES&CVID=jUYC7IK</u> [Accessed July 30, 2020].

¹¹ UN Guiding Principle 30.

¹² OHCHR (2020), "Accountability and Remedy Project III: Enhancing effectiveness of non-Statebased grievance mechanisms in cases of business-related human rights abuse": <u>https://www.ohchr.org/EN/Issues/Business/Pages/ARP_III.aspx</u> [Accessed July 30, 2020]; International Commission of Jurists (2019), "Effective Operational-level Grievance Mechanisms": <u>https://www.icj.org/wp-content/uploads/2019/11/Universal-Grievance-Mechanisms-</u> <u>Publications-Reports-Thematic-reports-2019-ENG.pdf</u>; International Commission of Jurists (2019), "Companies around the world must do more to ensure effective operational grievance mechanisms practices and provide clear and transparent information": <u>https://www.icj.org/companies-around-the-world-must-do-more-to-ensure-effective-operational-grievance-mechanism-practices-and-provide-clear-and-transparent-information/</u> [Accessed July 30, 2020].

¹³ See, however, a case study on Telefónica's grievance mechanism: Global Network Initiative, "The GNI Principles at Work: Public Report on the Third Cycle of Independent Assessments of GNI Company Members 2018/2019", p. 33: <u>https://globalnetworkinitiative.org/wp-</u> <u>content/uploads/2020/04/2018-2019-PAR.pdf</u>; and BSR (2019), *"Human Rights Review: Facebook Oversight Board"*: <u>https://www.bsr.org/reports/BSR_Facebook_Oversight_Board.pdf</u>

¹⁴ See e.g. BSR (2019), *"Human Rights Review: Facebook Oversight Board":* <u>https://www.bsr.org/reports/BSR_Facebook_Oversight_Board.pdf</u>

THE DANISH INSTITUTE FOR HUMAN RIGHTS

Huyto

N

0

F.