

# Guidelines for the Governance of Digital Platforms

Safeguarding freedom of expression and access to information  
through a multistakeholder approach



Published in 2023 by the United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO 2023

ISBN 978-92-3-100620-3



This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (<http://en.unesco.org/open-access/terms-use-ccbysa-en>).

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The ideas and opinions expressed in this publication are those of the authors; they are not necessarily those of UNESCO and do not commit the Organization.

**Cover, Typeset and Graphic Design:** Luiza Maximo

**Illustrations:** Plastic Horse/Grand Matter

Printed by UNESCO

*Printed in Paris*

# SHORT SUMMARY

## Guidelines for an Internet for Trust

Safeguarding freedom of expression and the right to information while dealing with dis- and misinformation, hate speech, and conspiracy theories requires a multistakeholder approach.

This is the reason why UNESCO, the leading UN agency for the promotion and protection of freedom of expression and to information, is launching Guidelines for the Governance of Digital Platforms.

The Guidelines outline a set of duties, responsibilities and roles for States, digital platforms, intergovernmental organizations, civil society, media, academia, the technical community and other stakeholders to enable the environment where freedom of expression and information are in the core of digital platforms governance processes.

The Guidelines were produced through a multi-stakeholder consultation that gathered more than 10,000 comments from 134 countries. These global-scale consultations fostered inclusive participation, ensuring a diversity of voices to be heard, including those from groups in situation of marginalization and vulnerability.

Cultivating an Internet of Trust is a shared responsibility among all stakeholders. It calls upon us all to sustain an enabling environment for freedom of expression and the right to information.



10,000  
COMMENTS FROM

134  
COUNTRIES

# **Guidelines for the Governance of Digital Platforms**

Safeguarding freedom of expression and access to information  
through a multi-stakeholder approach

## 6 Foreword

## 8 The objective of the Guidelines

## 11 Introduction

## 14 Structure of the Guidelines

16 Enabling environment

24 The governance system

35 Principle 1. Platforms conduct human rights due diligence

37 Principle 2. Platforms adhere to international human rights standards, including in platform design, content moderation, and content curation

42 Principle 3. Platforms are transparent

46 Principle 4. Platforms make information and tools available for users

48 Principle 5. Platforms are accountable to relevant stakeholders

50 Context-specific provisions

## 54 Conclusion

## 56 Appendix

# **Foreword by the Director-General of UNESCO, Audrey Azoulay**

## **Preserving Freedom of Expression and Access to Information: Principles for a Multistakeholder Approach to the Governance of Digital Platforms**

September 2023

In 2023, 60% of the global population, or 4.75 billion people, are using social media platforms to express, inform, and affirm themselves.

The digital realm, a space of freedom and a new forum for expression and debate, interweaves our social relationships, identities, and lives.

These platforms have become amplifiers for champions of equality and freedom – giving voice to the voiceless, offering a haven for diverse forms of expression.

However, these same social networks, whose name holds so much promise, all too often become bubbles of isolation, cocoons of misinformation, which sometimes foster conspiracy theories and extreme violence.

As virtual spaces for social interaction, they are beholden to algorithms designed to monopolize our attention, inadvertently favouring misinformation and hate speech by prioritizing clicks over certainty, probability over proof.

Yet if we can no longer distinguish fiction from reality, falsehood from truth, the foundations of our societies crumble. Democracy, dialogue, and debate – all essential to address major contemporary challenges – become impossible.

Faced with the global nature of these issues, we need to develop consistent responses around the world, and avoid the fragmentation of regulations or approaches that compromise human rights.

It is precisely this global challenge to which UNESCO must rise, for it is at the core of our mandate.

Since our Organization was created, it has worked to advance “the mutual knowledge and understanding of peoples,” notably through “the free flow of ideas by word and image,” as underlined in our Constitution.

This commitment led UNESCO to publish guidelines for broadcasting regulation in 2005. More recently, our Recommendation on the Ethics of Artificial Intelligence,

adopted in 2021 by our 193 Member States, established a humanist framework for the evolution of this technology.

Staying true to its values and history, UNESCO has worked to develop the groundbreaking guidelines presented in this publication. They seek to combat misinformation and hate speech, while promoting transparency and freedom of expression on platforms.

These efforts have been steered by the *Windhoek+30 Declaration of 2021*, whose principles were adopted by all UNESCO Member States. The Declaration identified three pillars of action: championing platform transparency, ensuring media viability, and fostering critical thinking among users.

This endeavour, which culminates with these guidelines, is the result of extensive consultations, enriched by over 10,000 comments, making it one of the most comprehensive consultations conducted by the United Nations. The Internet for Trust conference, organized by UNESCO in February 2023, alone brought together over 4,000 stakeholders from 134 countries.

These guidelines propose fair, clear, and shared measures: online moderators in all languages, including indigenous ones; greater transparency of platforms and their financing, with better risk assessment; the establishment of independent regulators; the promotion of critical thinking; support for gender equality; and, above all, the safeguarding and strengthening of freedom of expression, cultural diversity, and other human rights.

Looking beyond the current realities of digital platforms, this text also addresses future challenges, especially those posed by generative artificial intelligence.

UNESCO is committed to assisting Member States, civil society, and major digital players in embracing this text, so that platform operations fully align with our values and international human rights standards.

Let us remain focused on our goal: combating hate speech and misinformation while preserving freedom of expression. This is not a contradiction. By bolstering access to free and reliable information, we also enhance freedom of thought and expression.

In the words of Hannah Arendt, “Freedom of opinion is a farce unless factual information is guaranteed and the facts themselves are not in dispute.”

# Guidelines for the Governance of Digital Platforms

Safeguarding freedom of expression and access to information through a multistakeholder approach

October 2023

## The objective of the Guidelines

1. Building upon relevant principles, conventions, and declarations, UNESCO developed—through multistakeholder consultations and a global dialogue—the present document: *Guidelines for the Governance of Digital Platforms: Safeguarding freedom of expression and access to information through a multistakeholder approach* (the Guidelines).<sup>1</sup>
2. The aim of the Guidelines is to safeguard the right to freedom of expression, including access to information, and other human rights in digital platform governance, while dealing with content that can be permissibly restricted under international human rights law and standards. By extension, digital platform governance that is grounded in human rights would further promote cultural diversity, cultural expression, and cultural diverse content.<sup>2</sup> The Guidelines outline a human rights-respecting governance system and promote risk-mitigation and system-based processes for content moderation and curation. These Guidelines highlight overarching principles that should be followed in all governance systems that impact freedom of expression and access to information on digital platforms—independently of the specific regulatory arrangement and the thematic focus, as long as those arrangements are aligned with the provisions established in these Guidelines.

1. The original version of this document is in English.

2. UNESCO 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions, articles 1 and 4. Under the Convention, “cultural content” refers to the symbolic meaning, artistic dimension, and cultural values that originate from or express cultural identities. Moreover, “cultural expressions” are those expressions that result from the creativity of individuals, groups, and societies, and that have cultural content.



3. The Guidelines recognize that the application of rules and regulations in every governance system must adhere to international human rights standards, including with Article 19 (3) of the *International Covenant on Civil and Political Rights* (ICCPR),<sup>3</sup> which provides that any restriction to freedom of expression must be provided by law, pursue a legitimate aim as set out in the provision, and must be necessary and proportionate; as well as Article 20 of the ICCPR and other international standards, particularly the authoritative interpretations of these treaties' provisions by the UN Human Rights Committee, international and regional human rights courts, and the Rabat Plan of Action on the prohibition of advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence.<sup>4</sup>
4. The Guidelines focus on protecting and promoting human rights standards and enabling the existence of a plurality of platforms, including de-centralized ones, and an ecosystem that has a diversity of content standards and moderation systems.
5. The Guidelines may serve as a resource for a range of stakeholders: for policymakers in identifying legitimate objectives, human rights principles, and inclusive and participatory processes that could be considered in policymaking; for regulatory and other governance bodies dealing with the implementation and evaluation of policies, codes of conducts, or regulation; for digital platforms in their policies and practices; and for other stakeholders, such as civil society, in their advocacy and accountability efforts. News media can also benefit from these Guidelines in their ongoing efforts to hold powerful actors accountable.
6. The Guidelines are designed to inform both governance processes specific to the management of content on digital platforms, and governance processes that are already being implemented in other areas that may have an impact on the exercise of freedom of expression and access to information and diverse cultural content and should be considered in light of changes in the digital environment (such as elections, data protection, and antitrust regulations). Depending on the issue and the jurisdictional context, such governance processes may take the form of a combination of complementary pillars—self-regulation, co-regulation, and statutory regulation—structured in a manner consistent with international human rights standards (see “The governance system” section below). Such

---

3. *International Covenant on Civil and Political Rights* (ICCPR), 1966. <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>.

4. The Guidelines should be read in harmony with all core international human rights instruments. These are outlined at: <https://www.ohchr.org/en/core-international-human-rights-instruments-and-their-monitoring-bodies>.

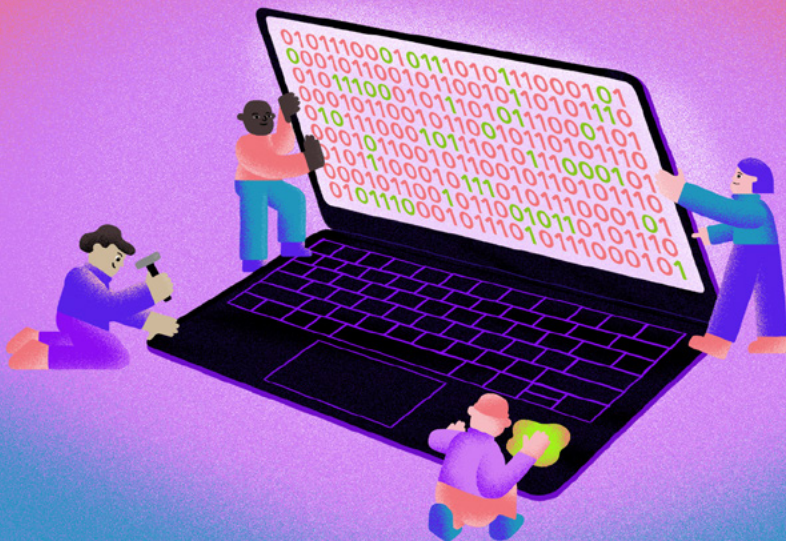
governance processes should be led in an open, transparent, multistakeholder, proportional, and evidence-based manner. To this end, these Guidelines should be a living document, subject to periodic reviews and updates, including to consider the lessons learned from their implementation, as well as subsequent technological changes and impacts.

7. These Guidelines are designed to contribute in a practical way to broader efforts to realize a human-centred model for digital governance. They are also part of the broader toolkit of actions necessary to achieve sustainable development. They will:
  - a. Encourage and contribute to the development of global multistakeholder networks and common spaces to debate and share good practices about digital platform governance, gathering different visions and a broad spectrum of perspectives.
  - b. Serve as a tool for all relevant stakeholders to advocate for human rights-respecting regulation and to hold governments and digital platforms accountable.
  - c. Advance evidence-based and human rights-based policy approaches.
  - d. Encourage as much worldwide convergence as possible in platform governance policies to avoid internet fragmentation.
8. The Guidelines seek to contribute to and be informed by ongoing UN-wide processes, such as the implementation of the proposals in “Our Common Agenda.” This includes the development of the Global Digital Compact,<sup>5</sup> the preparation of the UN Summit of the Future to be held in September 2024, and the creation of a Code of Conduct that promotes information integrity on digital platforms.<sup>6</sup> The Guidelines will also feed into discussions about the upcoming 20-year review of the World Summit on the Information Society (WSIS) and the Internet Governance Forum (IGF). This text has further benefited from and aims to contribute to initiatives led by other international governmental organizations, including those of a regional scope.

---

5. See Our Common Agenda Policy Brief 5, issued by the UN Secretary-General, with cross references to the process leading to these Guidelines: [https://digitallibrary.un.org/record/4011891/files/%5EEOSG\\_2023\\_5%5E--EOS-G\\_2023\\_5-EN.pdf](https://digitallibrary.un.org/record/4011891/files/%5EEOSG_2023_5%5E--EOS-G_2023_5-EN.pdf).

6. See Our Common Agenda Policy Brief 8, issued by the UN Secretary-General, with cross references to the process leading to these Guidelines: <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf>.



## Introduction

9. In November 1945, UNESCO was created with the mission of “contributing to peace and security by promoting collaboration among nations through education, science and culture in order to further universal respect for justice, for the rule of law and for the human rights and fundamental freedoms which are affirmed for the peoples of the world.”<sup>7</sup>
10. UNESCO’s global mandate, which includes the promotion of “the free flow of ideas by word and image,” has guided the Organization’s work for nearly 80 years—as a laboratory of ideas, a clearing house, a standard-setter, a catalyst and motor for international cooperation, and a capacity-builder. This history has also shaped the Organization’s mandate within the United Nations system to protect and promote freedom of expression, access to information, and safety of journalists, both off-line and online.
11. UNESCO’s ongoing work and commitment is to ensure that digital platform governance protects and promotes freedom of expression, access to information and diverse cultural content, and other human rights for all, including groups in situations of vulnerability and marginalization.<sup>8</sup>

7. *Constitution of the United Nations Educational, Scientific and Cultural Organization*, Article 1. <https://www.unesco.org/en/legal-affairs/constitution#article-i---purposes-and-functions>.

8. “Groups in situations of vulnerability and marginalization” refers to children and adolescents; persons with disabilities; migrants, refugees, and asylum-seekers; LGBTI persons; and older persons.

12. This endeavour draws lessons from UNESCO's decades of work in the domain of broadcast regulation, as any governmental intervention that deals with content issues—regardless of the source of the content—must always include safeguarding diversity and freedom of expression and access to information as an ultimate goal. The Guidelines also contribute to the implementation of the Organization's Medium-Term Strategy for 2022–2029 (41 C/4).<sup>9</sup>
13. In 2015, UNESCO's General Conference endorsed the Internet Universality ROAM principles, which highlight the importance of human **Rights, Openness, Accessibility, and Multistakeholder** participation in the development, growth, and evolution of the internet.<sup>10</sup> **These principles recognize the fundamental need to ensure that online spaces continue to develop and be used in ways that are conducive to achieving the Sustainable Development Goals.**
14. A multistakeholder approach to the development and application of shared principles, norms, rules, decision-making procedures, and programmes that shape the evolution and use of the internet has underpinned the overall strategy adopted by the UN system, including UNESCO, since the World Summit on the Information Society (2003 and 2005), and was reaffirmed by the UN General Assembly during the ten year review process in 2015:

We reaffirm, moreover, the value and principles of multi-stakeholder cooperation and engagement that have characterized the World Summit on the Information Society process since its inception, recognizing that effective participation, partnership and cooperation of Governments, the private sector, civil society, international organizations, the technical and academic communities and all other relevant stakeholders, within their respective roles and responsibilities, especially with balanced representation from developing countries, has been and continues to be vital in developing the information society.<sup>11</sup>

15. **UNESCO's 41<sup>st</sup> General Conference endorsed the principles of the Windhoek+30 Declaration in November 2021, following a multistakeholder process that began**

9. Strategic Objective 3 of the Medium-Term Strategy is to build inclusive, just, and peaceful societies, including by promoting freedom of expression. Strategic Objective 4 is to foster a technological environment in the service of humankind through the development and dissemination of knowledge and skills and ethical standards. <https://unesdoc.unesco.org/ark:/48223/pf0000378083>.

10. UNESCO. "Internet Universality Indicators." <https://www.unesco.org/en/internet-universality-indicators>.

11. UN General Assembly. 2015. "Outcome document of the high-level meeting of the General Assembly on the overall review of the implementation of the outcomes of the World Summit on the Information Society." 70/125. [https://unctad.org/system/files/official-document/ares70d125\\_en.pdf](https://unctad.org/system/files/official-document/ares70d125_en.pdf).

at the global celebration of World Press Freedom Day in May of that year.<sup>12</sup> The Declaration asserted that information is a public good and set, among the goals, three steps to guarantee information as a shared resource for the whole of humanity: the transparency of digital platforms, citizens empowered through media and information literacy, and media viability. In promoting the vision of information as a public good, UNESCO recognizes that this universal entitlement is both a means and an end for the fulfilment of collective human aspirations, including the 2030 Agenda for Sustainable Development. Information empowers citizens to exercise their fundamental rights, supports gender equality, and allows for participation and trust in democratic governance and sustainable development, leaving no one behind.

16. The focus of the Guidelines on challenges related to freedom of expression and access to information and diverse cultural content in the digital environment complements the Organization's work in the areas of education, the sciences, and culture. This includes but is not limited to UNESCO's Recommendation on the Ethics of Artificial Intelligence,<sup>13</sup> which calls for international and national policies and regulatory frameworks to ensure that emerging technologies benefit humanity as a whole, and the 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions<sup>14</sup> and its "Guidelines on the Implementation of the Convention in the Digital Environment." Those Guidelines promote "respect for fundamental freedoms of expression, information and communication, and for privacy and other human rights as pre-requisites for the creation, distribution and access to diverse cultural expressions [including] artistic freedom as a corollary to freedom of expression, the social and economic rights of authors and artists working in the digital environment and the connectivity of all partners with partners of their choice."<sup>15</sup> The focus of these Guidelines also complements the MONDIACULT Declaration of 2022, which calls for "substantial regulation of the digital sector, notably of the major platforms," for the benefit of online cultural diversity and fair access to content for all.<sup>16</sup>

12. UNESCO. 2021. *Windhoek+30 Declaration: Information as a public good*. <https://unesdoc.unesco.org/ark:/48223/pf0000378158>.

13. UNESCO. 2021. "Recommendation on the Ethics of Artificial Intelligence." <https://unesdoc.unesco.org/ark:/48223/pf0000380455>.

14. UNESCO. 2005. *2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions*. <https://en.unesco.org/creativity/convention>.

15. UNESCO. 2017. "Guidelines on the Implementation of the Convention in the Digital Environment." <https://unesdoc.unesco.org/ark:/48223/pf0000370521.page=92>.

16. UNESCO. 2022. "UNESCO World Conference on Cultural Policies and Sustainable Development – MONDIACULT 2022 – Final Declaration." [https://www.unesco.org/sites/default/files/medias/fichiers/2022/10/6.MONDIACULT\\_EN\\_DRAFT%20FINAL%20DECLARATION\\_FINAL\\_1.pdf](https://www.unesco.org/sites/default/files/medias/fichiers/2022/10/6.MONDIACULT_EN_DRAFT%20FINAL%20DECLARATION_FINAL_1.pdf).



# Guidelines

## Structure of the Guidelines

- 17.** The Guidelines start by describing the enabling environment needed to safeguard freedom of expression, access to information, and other human rights, while ensuring an open, safe, and secure environment for digital platform users and non-users. The Guidelines outline the responsibilities of different stakeholders in this regard. This includes:
  - a.** States' duties to respect, protect, and fulfil human rights.
  - b.** The responsibilities of digital platforms to respect human rights.
  - c.** The role of intergovernmental organizations.
  - d.** The role of civil society, media, academia, the technical community, and other stakeholders in the promotion of human rights.
- 18.** The Guidelines then set out the basic principles for the governance system of digital platforms with a multistakeholder and human rights-based approach. This section sets out complementary self-regulatory, co-regulatory, and statutory regulatory arrangements, as well as criteria that can be used for defining the scope of companies covered by statutory regulation.
- 19.** Then, the Guidelines identify media and information literacy, as well as respect for cultural diversity, as a common responsibility of all stakeholders involved in the governance of digital platforms.



20. Finally, they describe the areas where digital platforms should have systems and processes in place to assess risk; to curate and moderate content based on international human rights standards and respect for cultural diversity as defined by UNESCO 2005 Convention; to empower users through media and information literacy; and to be accountable through reporting mechanisms and redress in order to safeguard freedom of expression, access to information, and other human rights.
21. It is important to underscore that the different areas covered by these Guidelines (as identified in paragraphs 17–21 above) must be considered as a whole. Safeguarding freedom of expression and access to information and diverse cultural content requires consideration of all of the elements previously described.



## Enabling environment

22. All stakeholders share responsibility for sustaining an enabling environment for freedom of expression, access to information, and other human rights while ensuring there is an open, safe, and secure environment for digital platform users and non-users.<sup>17</sup>
23. Creating such an enabling environment is not simply an engineering question. It is also an endeavour that calls for the engagement of societies as a whole and therefore requires whole-of-society solutions. All relevant stakeholders in every governance system should take action to enable the exercise of the right to freedom of expression of groups in situations of vulnerability and marginalization, women and girls, and indigenous communities, as well as of journalists, artists, human rights defenders, and environmental defenders, for example. **All members of society have a role to play to make the internet safe, to challenge violent or threatening behaviours,** to respect the rights of others in exchanges online, to respect the diversity of cultural content, and to be aware of inherent biases in societies.
24. Children have a special status given their unique stage of development, limited or lack of political voice, and the fact that negative experiences in childhood

17. The words “safe” and “safety” in these Guidelines are conceived as the conditions in which individuals are able to trust that their human rights—including the right to free expression and access to information—are protected.



can result in lifelong or transgenerational consequences.<sup>18</sup> Thus, while the protection of freedom of expression and access to information applies for all individuals, governments and digital platforms should also recognize their specific responsibilities toward children<sup>19</sup> within the governance systems. Every stakeholder should uphold high ethical and professional standards when it comes to children's engagement in the digital environment, including the protection and promotion of children's freedom of expression and access to information.

25. All stakeholders involved in digital platform governance should foster and, when applicable, fund collaborative responses—involving civil society organizations, journalists' networks, and researchers—to gain more granular knowledge about content that could be permissibly restricted under international human rights law and standards and the responses to protect and support women and girls, groups in situations of vulnerability and marginalization, journalists, artists, human rights defenders, indigenous communities, and environmental defenders.

## States' duties to respect, protect, and fulfil human rights

26. States should respect and promote human rights, including the right to freedom of expression and the right to access information. Restrictions on freedom of expression are permissible only under the conditions established by Articles 19 (3) and 20 of the ICCPR. States have positive obligations to protect human rights against unjustified interferences by private actors, including digital platforms, as they have the responsibility to create a regulatory environment that facilitates platforms' respect for human rights, and to provide guidance to the digital platforms on their responsibilities.
27. Moreover, States have an obligation to be fully transparent and accountable about the requirements they place upon digital platforms, ensuring legal certainty and legal predictability, which are essential preconditions for the rule of law.
28. Specifically, States should:

---

18. See UN Committee on the Rights of the Child (2013), "General comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights," par. 4. See also "General comment No. 25 (2021) on children's rights in relation to the digital environment." <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>.

19. For most purposes, children are generally understood to be people below the age of 18.

- a. Promote universal and meaningful access to the internet and guarantee net neutrality.<sup>20</sup>**
- b. Ensure that all children have equal and effective access to the digital environment in ways that are meaningful for them, and take all measures necessary to overcome digital exclusion.<sup>21</sup>**
- c. Direct resources and accelerate efforts to close the digital divide, fill data gaps, remove other barriers faced by groups in situations of vulnerability and marginalization, and fulfil all women's and girls' right to access to information.**
- d. Strengthen civic space and promote free, independent, and pluralistic media, and support independent research around online speech, content moderation and curation, and platform accountability.**
- e. Guarantee strong protections for journalists (including women journalists), human rights defenders, and whistleblowers, and consider supporting transparent self-regulatory mechanisms by media that promote and protect the highest standards of professionalism.**
- f. Guarantee strong protections for artists, recognizing the importance of their works for the renewal of cultural production and the promotion of cultural diversity, and consider that they are at the very heart of the cultural fabric of society.**
- g. Guarantee digital platform users' rights to freedom of expression, access to information, equality, and non-discrimination, as well as protecting users' rights to privacy, data protection, association, and public participation.**
- h. Adopt laws, grounded in international human rights standards, and ensure their effective implementation to prohibit, investigate, and prosecute online gender-based violence.<sup>22</sup>**
- i. Ensure that any restrictions imposed upon platforms consistently follow the high**

20. In the "Joint Declaration on Freedom of Expression and the Internet," the special international mandates on freedom of expression indicated: "Giving effect to the right to freedom of expression imposes an obligation on States to promote universal access to the Internet." Adopted 1 June 2011, par. 6(a). <http://www.law-democracy.org/wp-content/uploads/2010/07/11.06.Joint-Declaration.Internet.pdf>.

21. See "General Comment No. 25 (2021) on children's rights in relation to the digital environment." <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>.

22. See "A/76/258: Gender justice and freedom of expression - Report of Special Rapporteur on the promotion and protection of freedom of opinion and expression." "All legal measures to restrict gender-based hate speech or gendered disinformation should comply with the three-part test of legality, necessity and proportionality, and legitimate objectives, as set out in Article 19 (3) of the Covenant. Criminalization should be avoided except in the most egregious cases of advocacy that constitutes incitement." <https://www.ohchr.org/en/documents/thematic-reports/a76258-gender-justice-and-freedom-expression-report-special-rapporteur>.

threshold set for restrictions on freedom of expression, based on the application of Articles 19 (3) and 20 of the ICCPR, respecting the conditions of legality, legitimate aim, necessity, and proportionality.

- j.** Strongly discourage—including through measures such as professional codes of conduct—public officials from disseminating disinformation, including gendered disinformation;<sup>23</sup> misinformation; and intimidating or threatening the media. Further, prohibit expressions amounting to advocacy of national, gender-based, racial, or religious hatred that constitute incitement to discrimination, hostility, or violence, as prohibited under international human rights law, in line with the UN Strategy and Plan of Action on Hate Speech.
- k.** Be transparent and disclose all information about the type, number, and legal basis of requests they make to digital platforms to take down, remove, and block content. States should be able to demonstrate how this is consistent with Articles 19 (3) and 20 of the ICCPR.
- l.** Promote media and information literacy to enhance positive engagement with the platforms and develop online safety skills, including in digital spaces, with the aim of empowering users, in particular groups in situations of vulnerability and marginalization. This should include promoting knowledge about the rights to freedom of expression, privacy, equality, access to justice and knowledge about means of complaint and redress, as well as drawing upon the expertise of media and information literacy experts, libraries, academics, civil society organizations, and access to information institutions.
- m.** Ensure that any regulatory authority that deals with digital platforms content management, regardless of the thematic, is structured as independent, shielded from political and economic interests, and has external review systems in place (see paragraphs 68–73 of these Guidelines). Such review systems may include legislative and judicial scrutiny, as well as requirements to be transparent and consult with multiple stakeholders, and to produce annual reports and be subject to periodic external audits. This would also involve establishing clear rules on the competence and authority of the judicial branch.
- n.** Ensure that the regulatory authorities have sufficient resources and the capacity to make assessments in line with the objectives of these Guidelines.

---

23. Idem, par. 21: “Gendered disinformation is also on the rise. While it is a subset of gender-based violence, it has some distinct characteristics, using ‘false or misleading gender and sex-based narratives against women, often with some degree of coordination, aimed at deterring women from participating in the public sphere. It combines three defining characteristics of online disinformation: falsity, malign intent, and coordination.’”

- o.** Recognize that any governance systems should draw from the expertise of human rights experts, academics, and civil society organizations, as well as recognized good practices from other governance systems.
- p.** Encourage international cooperation, including triangular and South-South cooperation, among regulatory authorities and judicial actors, promoting the exchange of good practices and knowledge.

**29.** States should refrain from:

- a.** Imposing measures that prevent or disrupt general access to the dissemination of information, online and off-line, including internet shutdowns.
- b.** Imposing a general monitoring obligation or a general obligation for digital platforms to take proactive measures in relation to content considered illegal in a specific jurisdiction or to content that could be permissibly restricted under international human rights law and standards. Digital platforms should not be held liable when they act in good faith and with due diligence, carry out voluntary investigations, or take other measures aimed at detecting, identifying, and removing or disabling access to content that is prohibited under Article 20 of the ICCPR or that has been restricted in terms of Article 19 (3) of the ICCPR.
- c.** Subjecting staff of digital platforms to criminal penalties for an alleged or potential breach of regulations in relation to their work on content moderation and curation.

## The responsibilities of digital platforms to respect human rights

**30.** Digital platforms should comply with five key principles:

- a. Platforms conduct human rights due diligence**, assessing their human rights impact, including the gender and cultural dimensions, evaluating the risks, and defining the mitigation measures.
- b. Platforms adhere to international human rights standards, including in platform design, content moderation, and content curation.** Platforms should follow relevant international human rights standards, including the UN Guiding Principles on Business and Human Rights. Design should ensure non-discrimination and equal treatment and that harm is prevented; content moderation and curation policies and practices should be consistent with human rights standards, whether these practices are implemented through automated or human means, with

knowledge of local languages and linguistic context as well as respect for cultural diversity, and adequate protection and support for human moderators.

- c. Platforms are transparent** and open about how they operate, with understandable and auditable policies as well as multistakeholder-designed metrics for evaluating performance. This includes transparency about the tools, systems, and processes used to moderate and curate content on their platforms, including in regard to algorithmic decisions and the results they produce.
- d. Platforms make information accessible** for users to understand the different products, services, and tools provided, and to make informed decisions about the content they share and consume. Platforms provide information and enable users' actions in their own languages and consider users' age and disabilities.
- e. Platforms are accountable to relevant stakeholders**—including users, the public, and actors within the governance system—in implementing their terms of service and content policies. They give users the ability to seek appropriate and timely redress against content-related decisions, including both users whose content was taken down or moderated and users who have made complaints about content.

- 31.** Platforms should apply these principles in every jurisdiction where they operate, ensuring necessary resources and capacities to timely and effectively serve users.
- 32.** To follow these principles, there are specific areas on which digital platforms have a responsibility to report or act before actors within the governance system, in accordance with international human rights standards. These areas are described in paragraphs 85–129 of these Guidelines.

## The role of intergovernmental organizations

- 33.** Intergovernmental organizations, in line with their respective mandates, should support relevant stakeholders in guaranteeing that the implementation of these Guidelines is in full compliance with international human rights law and standards. This support should include providing technical assistance, monitoring and reporting human rights violations, developing relevant standards, facilitating multistakeholder dialogue, and nurturing networks.
- 34.** Intergovernmental organizations and national regulatory agencies may create modalities for engagement to further develop and share good practices. Such engagement may include sharing emerging insights and regulatory trends, and supporting or making suggestions to national regulators to refine

institutional standards and methods to safeguard freedom of expression and access to information. Such modalities should work to reduce risks of internet fragmentation, as well as provide tools that allow for prior assessment of the impacts of regulation on the functioning of the internet as a whole.

## The role of civil society and other stakeholders

35. Every stakeholder engaged with the services of a digital platform as a user, policymaker, watchdog, or by any other means has a vital role to play in supporting freedom of expression, access to information, and other human rights. Toward this end, the process of developing, implementing, and evaluating regulation that impacts content on digital platforms should take a multistakeholder approach. A broad set of stakeholders should also be engaged in oversight, including those representing groups in situations of vulnerability and marginalization, as well as journalists, artists, human rights defenders, and environmental defenders.
36. Civil society plays a critical role in understanding the nature of and countering harmful content and behaviour online, particularly that directed toward all groups in situations of vulnerability and marginalization, women and girls, journalists, artists, human rights defenders, and environmental defenders. Civil society also plays a significant role in monitoring and reporting on government laws, policies, and regulatory actions that impact human rights. They are key in bridging the gap between the digital governance ecosystem and the people in general.
37. Independent researchers have a role in identifying patterns of abusive behaviour and where the possible root causes could be addressed; researchers should also be able to provide independent scrutiny of how the governance system is working. Independent institutions and researchers can support human rights due diligence, including gender assessments, audits, investigations, and other types of reports on platforms' practices and activities. Researchers should be able to collect and analyse disaggregated data based on gender and other relevant intersecting factors (such as race, ethnicity, age, socioeconomic status, disability, etc.). This helps identify disparities, biases, and differential impacts of digital platforms on different groups in vulnerable and marginalized situations.
38. Media outlets, fact-checking organizations, and the professionals within these institutions are important stakeholders and have a role in promoting the enjoyment of freedom of expression, access to information, and other human rights, while performing their watchdog function. Therefore, it is necessary to involve the media and its professionals in the regulatory process, recognizing

their role as active participants in positively contributing to the digital information ecosystem. A constructive relationship between digital platforms and credible news sources will enhance the role of digital platforms in providing information in the public interest.

39. Educators and caregivers have a critical role in helping young people and learners of all ages understand the wider digital environment, including how to look for and understand credible information and how to respectfully engage with others online. There is also a role in providing lifelong learning as technology changes rapidly.
40. Engineers, data scientists, and the technical community involved in building platform services and products have a role in understanding the human rights, risks, and ethical impacts of the products and services they are designing and developing.<sup>24</sup>
41. All of these stakeholders should have an active role in consultations on the development and operation of the governance system. Collaboration and dialogue among stakeholders should be fostered. Constructive discussions and deliberations should take place to exchange ideas, knowledge, and perspectives. Establishing working groups, task forces, or advisory committees provides opportunities for active participation in shaping regulatory proposals.

---

24. See OHCHR report on the relationship between technical standard setting and human rights, A/HRC/53/42.



## The governance system

42. The digital governance ecosystem consists of an array of diverse stakeholders, bodies, and regulatory arrangements throughout the world. While some existing governance systems, such as in the case of elections or data protection, should be interpreted and considered in accordance with the changes and challenges that the digital age entails, new governance systems are also being created in various contexts to directly regulate digital platforms. In any case, these regulatory mechanisms might have profound implications for freedom of expression and access to information and diverse cultural content online.
43. The present Guidelines highlight overarching principles that can be applied, as relevant, to diverse processes that touch on the governance of content on digital platforms, regardless of form or field. They indicate that a comprehensive governance system can effectively leverage various complementary regulatory arrangements to address the challenges faced by different stakeholders in the digital ecosystem.
44. The Guidelines call for a multistakeholder approach to the governance of digital platforms. This approach could incorporate aspects such as: identifying all relevant stakeholders (including the platforms that are within its scope), encouraging inclusive participation, ensuring balanced representation, ensuring transparency and accountability, fostering collaborative decision-making and dialogue, facilitating an iterative process, coordinating implementation efforts among stakeholders, and conducting periodic evaluation and review.



45. Depending on the context, the accountability and compliance mechanisms for the governance of digital platforms may include complementarity and convergence within different regulatory arrangements, such as:
- a. Self-regulatory structures and mechanisms, where rules may be overseen and enforced by non-state actors, like industry-wide bodies or social media councils.
  - b. Co-regulatory structures and mechanisms where, in some cases, codes of conduct may be granted legal force then serve as regulation.
  - c. Statutory regulatory frameworks in which one or more independent regulators make final decisions on setting rules for platforms.
46. Recognizing the complexity of this environment, these present Guidelines are designed to apply to a wide range of forms of governance. It is important to note that statutory regulatory frameworks may be needed in some domains to address areas unsuitable for self- and co-regulatory mechanisms. Such frameworks must always ensure the independence of the official regulatory authorities and, in line with the aim of the Guidelines, should always safeguard human rights.

## Principles of governance systems

47. First, **transparency** should be a common overarching principle. In all governance systems, digital platforms are expected to be transparent about the terms, systems, and processes they use to moderate and curate content on their platforms, as well as on any human rights due diligence in line with the provisions of these Guidelines and the UN Guiding Principles on Business and Human Rights. They should be able to explain how their systems and processes fulfil their terms of service and effective implementation thereof, and if these are consistent with human rights international standards.
48. Governance systems and procedures external to the platforms should also be transparent. Any external regulatory action should be proposed, openly and widely debated, and finally executed under public oversight, with open and clear delineation of the remit and responsibilities for decisions.
49. Second, a common regulatory principle is that the **checks and balances** between different interests should be formally institutionalised. Governance systems should always have a multistakeholder approach across all forms of regulation and combinations thereof. This means providing for broad and inclusive participation among all stakeholders that can best represent divergent

interests and values, including diverse gender and intersectional perspectives. Multistakeholder participation should be meaningful in terms of representation and in creating, applying, monitoring, and reviewing the governance processes (rules, principles, and policies). Public awareness campaigns, targeted outreach, respect for cultural diversity, and the use of inclusive language and formats in governance processes can facilitate effective participation.

50. Third, governance processes should be **open and accessible** to all stakeholders, particularly the groups impacted by a proposed structure or type of regulation. Public consultations, open hearings, and online platforms should be utilized to provide opportunities for public input and feedback. The concerns of groups in situations of vulnerability and marginalization, as well as women and girls, should be adequately represented in the decision-making process.
51. The governance system should ensure that digital platforms actively engage with children, protect their freedom of expression and other rights, apply appropriate safeguards, and give their views due consideration when developing products and services.
52. Governance systems should also promote dialogue with media, including for the investment in independent news media, and support the media ecosystem by making data available and supporting actions to bolster media sustainability, diversity, and plurality.
53. Fourth, inclusion of **diverse expertise** should be a common feature of all regulatory arrangements. The governance system requires that stakeholders have the necessary capacity through training and regulatory instruments to understand human rights frameworks and to consider technological developments. They should have the capacities and technical knowledge to make informed decisions and to apply these Guidelines. Every governance system should be encouraged to report to the public and assess the risks and opportunities associated with new and emerging technologies.
54. Stakeholders within governance systems should share regulatory expertise and knowledge across jurisdictions. National, regional, and global governance systems should be able to cooperate and share practices in order to achieve the goal to safeguard freedom of expression, access to information, and other human rights, while also addressing content that could be permissibly restricted under international human rights law and standards.
55. Fifth, the governance system should ensure that digital platforms are engaged in **protecting and promoting cultural diversity and the diversity of cultural**

**expressions** in the creation, production, distribution, dissemination, access, and enjoyment of cultural goods and services online, including by ensuring their fair discoverability and representation.

## Accountability and compliance

- 56. Regulatory arrangements should be effective and sustainable, taking into account the available local resources and the main priorities needing attention (for example, whether to address primarily issues around elections, public health, advertising, or data protection, etc.). Independent oversight is needed for all forms of regulation. The process for developing regulation should be open, transparent, and evidence-based.
- 57. Digital platforms deemed non-compliant with their own policies or failing to fulfill their duties to safeguard freedom of expression and access to information while dealing with content that could be permissibly restricted under international human rights law and standards, in accordance with the five principles described in paragraphs 85–129, should be held accountable to relevant bodies within the governance system and subject to proportionate enforcement measures with necessary procedural safeguards.
- 58. **Self-regulation systems** can be complementary and converge with other forms of regulation. They should include **independent periodic mandatory audits** that assess digital platforms' compliance with self-regulatory codes, policies, or norms. Such audits should not be directly funded by the industry or individual digital platforms, although levies on these entities can help cover the costs of such exercises. Nor should audits be conducted by any person or entity that would have or appear to have a conflict of interest. The terms and the results of the audit should be available for public comment.
- 59. **Co-regulatory structures** should provide a legal framework that enables the environment for freedom of expression, access to information, and other human rights. In co-regulation, the regulatory role should be shared between industry and other stakeholders, and the government or the official independent regulatory authorities or bodies. The role of the relevant public authorities includes recognition of the co-regulatory scheme, auditing of processes, and funding the scheme (possibly through levies on platforms). Co-regulation should allow for the possibility of state-enforced penalties like fines, in the event that agreed objectives are not being met.

- 60. Statutory regulation of digital platforms addressing issues that might impact freedom of expression should be considered only when there is independence in decision-making of the regulatory authorities involved in its implementation.** Such regulation should focus on systems and processes for content moderation and curation, rather than determining the legality of individual pieces of content, and must have a basis in law (i.e., be sufficiently defined), pursue a legitimate aim under Article 19 (3) of the ICCPR, and be necessary and proportionate.
- 61.** The multistakeholder approach in statutory regulation should be reflected in an arrangement by which:
- a. Relevant State authorities, including official independent regulatory authorities, set the legitimate aim of the regulation through participatory and inclusive legislative processes.**
  - b. Digital platforms report publicly to official regulatory authorities.**
  - c. Civil society organizations, artists, independent researchers, and other relevant institutions provide inputs into rulemaking, contribute to oversight, and achieve the necessary checks and balances through institutionalised involvement and scrutiny.**
- 62.** Any specific decisions about the legality of specific pieces of content should follow due process and be open to review by an impartial and independent judicial body.
- 63.** In all cases, assessments regarding content should follow the three-part test on legitimate restrictions to freedom of expression as laid out in Article 19 (3) of the ICCPR,<sup>25</sup> and the prohibition of advocacy to hatred that constitutes incitement against discrimination, hostility, or violence as laid out in Article 20 (2) of the ICCPR; including, as applicable, the six-point threshold test for defining such content outlined in the Rabat Plan of Action.<sup>26</sup>
- 64.** Every statutory regulatory intervention should be evidence-based, proportionate, and include procedural safeguards, including by ensuring the platform access to all facts and considerations upon which the decision is made. This process should involve multiple stakeholder groups, considering a broader view of the sustainability, effectiveness, and impact of the intervention. The call for an

25. UNESCO. 2021. "The Legitimate Limits to Freedom of Expression: the Three-Part Test." <https://www.youtube.com/watch?v=Wq8fVtHPDag>.

26. UNESCO. 2021. "The Rabat Plan of Action on the Prohibition of Incitement to Hatred." <https://www.unesco.org/archives/multimedia/document-5554-eng-3>.

evidence-based process cannot be an excuse to delay necessary regulatory actions to protect human rights.

65. All relevant stakeholders, including the platforms, should have an opportunity to make representations and/or appeal against a decision of non-compliance. The regulatory system should be required to publish and consult on enforcement measures and follow due process before directing a platform to implement specific measures.

## Defining the digital platforms within the scope of regulation

66. When defining the digital platforms that should be in the scope of statutory regulation, the regulatory authorities should identify those platforms that have relevant presence, size, and market share in a specific jurisdiction. These should be determined through an independent assessment of the risk they pose to human rights, including of groups in situations of vulnerability and marginalization, as well as to democratic institutions.<sup>27</sup> The definition of the scope should protect the right to privacy and not result in the weakening of protections for encryption or other privacy-protecting technologies.
67. Reflecting regional and jurisdictional realities, the following criteria can be taken into account to identify the companies in scope:
  - a. **Size and reach**, with a focus on platforms that are most likely to have an impact on a significant portion of the population and/or on groups in situations of vulnerability and marginalization.
  - b. **Market share**, recognizing the considerable influence of dominant platforms on the entire information ecosystem. The application of the Guidelines should avoid penalising start-ups and new entrants while ensuring that the digital platforms with the most potential impact are covered in a proportional manner. While all platforms are expected to follow the general principles, the specific reporting obligations under paragraphs 85–129 of these Guidelines may apply primarily to the larger platforms that have greater capacity to comply with them.
  - c. **Functionality and features**, cognizant of the relevant differences that distinct services have regarding visibility, influence over, and directionality of content. Risk can be determined by the platform's user base, forms of ownership, business

27. A supplemental guide for identifying the systemic risk of platforms may be developed as a companion for operationalizing these Guidelines.

model, functionality, and features, such as real-time posting, potential for virality, volume, velocity of distribution, verisimilitude, and the extent to which content can be posted without a content moderation process.<sup>28</sup>



## Characteristics of independent regulatory authorities

68. In statutory regulation, official regulatory authorities, though constituting part of the executive state apparatus, should be wholly independent of the government and be primarily accountable to legislatures for fulfilment of their mandates.<sup>29</sup> This applies to existing regulatory bodies that have a legitimate interest in content on platforms (such as electoral management bodies, advertising authorities, child protection entities, data and privacy commissions, competition bodies, etc.), as well as any new dedicated or coordinating regulatory instances that may be established.
69. With regard to all statutory bodies engaging in platform regulation, either solely or jointly, periodic review<sup>30</sup> should be performed by an independent body reporting directly to the legislature. Statutory interventions should also be subject to review in the courts if authorities are believed to have exceeded their powers, acted unreasonably, or acted in a biased or disproportionate manner.
70. Official regulatory authorities need to be independent and free from economic, political, or other pressures. Their power and mandate should be set out in law. They should also comply with international human rights and promote gender equality standards.

28. "Risk-based regulations are based on the assessment by the rule/standard-setter of the risks relevant to their mandate, and the appropriate level of intervention required in accordance with the level of risk. If an actor performs low-risk activity, the regulation would be accordingly streamlined, providing for lower compliance requirements." [https://www3.weforum.org/docs/WEF\\_Pathways\\_to\\_the\\_Regulation\\_of\\_Crypto\\_Assets\\_2023.pdf](https://www3.weforum.org/docs/WEF_Pathways_to_the_Regulation_of_Crypto_Assets_2023.pdf).

29. The World Bank stated that the key characteristic of the independent regulator model is decision-making independence. A guiding document on broadcast regulation commissioned by UNESCO (Salomon, Eve. *Guidelines for broadcasting regulation*. 2006) also highlighted that "an independent authority (that is, one which has its powers and responsibilities set out in an instrument of public law and is empowered to manage its own resources, and whose members are appointed in an independent manner and protected by law against unwarranted dismissal) is better placed to act impartially in the public interest and to avoid undue influence from political or industry interests." For the complete references, see the appendix to these Guidelines.

30. The review should place particular emphasis on how decisions of the regulatory system may affect the enjoyment of human rights.

- 71.** Official regulatory institutions must have sufficient funding and expertise to carry out their responsibilities effectively. The sources of funding must also be clear, transparent, and accessible to all, and not subject to the governmental discretion.
- 72.** Governing officials or members of the official regulatory institutions working on the issue of content on platforms should:
- a.** Be appointed through a participatory, transparent, non-discriminatory, and independent merit-based process.
  - b.** Be accountable to an independent body (which could be the legislature, judiciary, an external council, or an independent board/boards).
  - c.** Include relevant expertise in international human rights law and the digital ecosystem.
  - d.** Deliver an annual public report to an independent body—ideally the legislature—and be held accountable to it, including by informing the body about their reasoned opinion.
  - e.** Make public any possible conflicts of interest and declare any gifts or incentives.
  - f.** After completing the mandate, for a reasonable period, not be hired or provide paid services to those who have been subject to their regulation, in order to avoid the risk known as “revolving doors.”
- 73.** The official regulatory authorities should be able to request that digital platforms provide periodic reports on the application of their terms of services, and take enforcement action against digital platforms deemed non-compliant with their own policies or failing to fulfil their responsibilities to safeguard freedom of expression and access to information and diverse cultural content. They should be able to establish a complaints process and issue public recommendations that may be binding or non-binding and be empowered to issue transparent and appropriate directives to the platforms for the promotion and respect of human rights, based on international human rights standards.



74. Media and information literacy covers a broad range of skills that allow users to think critically about the information they interact with online. Media and information literacy should be addressed specifically through the governance system to ensure that all stakeholders, including digital platforms, are effectively playing their part.
75. Media and information literacy will be most effectively achieved when stakeholders within the governance system share a common vision and work collaboratively to achieve it through sharing knowledge and resources. Media and information literacy programmes should be responsive to the availability of existing and emerging media and information technologies so that citizens can fully benefit from their use to actively participate in their societies.
76. Media and information literacy programmes should put an emphasis on the empowerment of users and ensure that they have the skills and knowledge that will enable them to interact with content critically and effectively in all forms of diverse media and with all information providers—including schools, universities, research institutions, libraries, archives, museums, media companies, publishers, statistical entities, and more. When media and information literacy programmes only emphasize protection or digital safety skills, they may lead to excessive restrictions placed on the use of digital platforms. However, they should prioritise specific steps that users can take, based upon best practices published by UNESCO and other international bodies, to identify content that could be permissibly restricted under international human rights law and standards.
77. Media and information literacy programmes should promote cultural diversity, social inclusion, and global citizenship, and aim to reduce the “participation gap” between people who are engaged in the creation and critical use of media and information content and those who are not. Media and information literacy programmes should also promote gender equality and women’s empowerment and provide opportunities for participation by groups in situations of vulnerability and marginalization.
78. Governments should always consider the promotion of media and information literacy, including online safety skills, for users, especially all groups in situations of vulnerability and marginalization, as well as women and girls. This enables

31. See UNESCO’s “Media and information literacy: Policy and strategy guidelines.” <https://unesdoc.unesco.org/ark:/48223/pf0000225606>.



users to engage critically with content and technologies, navigate a rapidly evolving media and information landscape marked by digital transformation, promote human rights, and build resilience in the face of related challenges.

79. Governments should disseminate information and conduct awareness-raising campaigns on the rights of the child in the digital environment, including their right to freedom of expression, focusing in particular on those whose actions have a direct or indirect impact on children. They should facilitate educational programmes for children, parents and caregivers, the general public, and policymakers to enhance their knowledge of children's rights in relation to the opportunities and risks associated with digital products and services. Such programmes should include information on how children can benefit from digital products and services and develop their media and information literacy, including digital skills.
80. Platforms should establish a clear and public strategy to empower users and promote a favorable online environment that safeguards freedom of expression and access to information through media and information literacy, including online safety education. There should be a specific focus within the digital platform on how to improve the digital literacy of all users, especially groups in situations of vulnerability and marginalization, with thought given to this in product development teams.
81. Digital platforms should allocate adequate resources to improve media and information literacy of all users, including digital literacy about the platform's own products and services, as well as relevant processes. This should especially focus on improving users' understanding of the ways that a given platform presents, curates, recommends, and/or flags content (also connected to the steps outlined under Principles 3 and 4, below) and specific steps users can take to themselves identify content that could be permissibly restricted under international human rights law and standards.
82. Platforms should train their product development teams on media and information literacy, including online safety, from a user empowerment perspective and based on international standards, and put in place both internal and independent monitoring and evaluation mechanisms.
83. Both governments and digital platforms should implement media and information literacy programmes in close collaboration with organizations and diverse experts independent of the platforms, including but not limited to: public authorities responsible for media and information literacy, academia,

civil society organizations working with groups in situations of vulnerability and marginalization, researchers, librarians, teachers, specialized educators, journalists, artists, and cultural professionals. Specific measures should be taken for users and non-users and audiences from groups in situations of vulnerability and marginalization, as outlined in the many UNESCO tools available on media and information literacy.

84. Governments and digital platforms should collaborate and ensure that users understand their rights online and off-line, including the role of media and information literacy in the enjoyment and protection of the rights to freedom of expression and access to information.



## Principle 1. Platforms conduct human rights due diligence

### *Human rights safeguards and risk assessments*

- 85.** In any kind of regulatory arrangement, digital platforms should be able to demonstrate the systems or processes they have established to ensure ongoing human rights due diligence, including human rights and gender impact assessments,<sup>32</sup> as well as risk mitigation measures.<sup>33</sup> These systems should be reviewed periodically and the review should be made public.
- 86.** In line with international human rights standards, including the UN Guiding Principles on Business and Human Rights, platforms should conduct periodic risk assessments to identify and address any actual or potential human rights impacts of their operations. When implementing human rights risk assessment

32. See the 18 October 2021 “Statement by Irene Khan, Special Rapporteur on the promotion and protection of freedom of opinion and expression.” In line with the UN Guiding Principles on Business and Human Rights, “social media companies should carry out regular human rights and gender impact assessments to identify and mitigate systemic risks affecting women and gender nonconforming people. They should make platforms safe and gender-inclusive, and in line with international human rights standards, adopt effective safety policies and tools, ensure meaningful transparency, including of algorithms, and provide adequate remedies.” <https://www.ohchr.org/en/statements/2022/02/statement-irene-khan-special-rapporteur-promotion-and-protection-freedom-opinion>.

33. Human rights impact assessments should include all human rights that companies’ policies may impact. This includes civil and political rights such as freedom of expression, access to information, and privacy, as well as economic, social, and cultural rights, the right to be free from violence, and the right to participate in public life, among others.

processes, digital platforms should consider how any product or service impacts user behaviour beyond the aim of user acquisition or engagement.

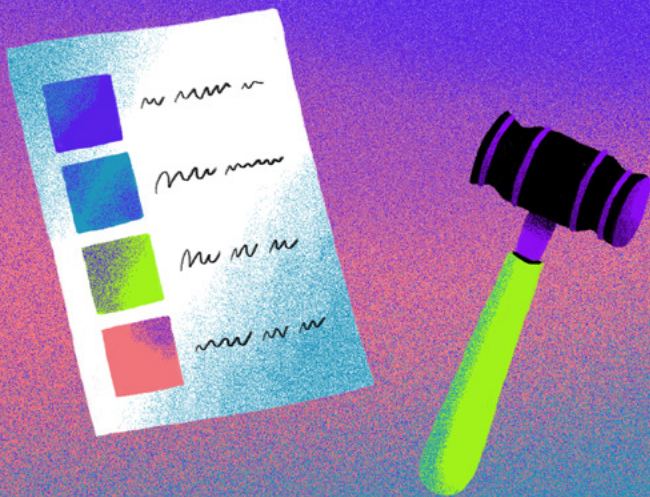
- 87.** Risk assessments should be an anchor for decision-making within digital platforms, informing the approach of the design and operation of their services, and the mitigations they deploy to address residual risk and to safeguard human rights, non-discrimination, and equal treatment. Moreover, responsibilities for risk management should be clearly specified and owned at the most senior levels and risk management activities should regularly be reported to senior decision-makers.
- 88.** At a minimum, human rights and risk assessments should take place:
- a.** Prior to any significant design changes, major policy decisions (including those related to the advertising system, if applicable), changes in operations, or new activity or relation/partnerships.
  - b.** Regularly, to protect the rights of all groups in situations of vulnerability and marginalization, as well as women and girls, journalists, artists, human rights defenders, and environmental defenders.<sup>34</sup>
  - c.** Ahead of electoral processes to protect their integrity.<sup>35</sup>
  - d.** In response to emergencies, crises, or conflict or significant changes in the operating environment.<sup>36</sup>
- 89.** During the human rights due diligence process, platforms should ensure meaningful engagement with a variety of stakeholders to identify specific risks for groups in situations of vulnerability and marginalization. It is critical that digital platforms are open to expert and independent input on how these assessments are structured.
- 90.** From the beginning, platforms should create spaces to listen, engage, and involve users, including those who have experienced harassment or abuse, their representatives, and users from groups in situations of vulnerability and marginalization, as well as women and girls, journalists, artists, to inform platform policies and processes. This could include ways to identify and counter content that could be permissibly restricted under international human rights law and standards, as well as opportunities and solutions to address the assessed risks.

---

34. See Context-specific provisions, par. 126.

35. See Context-specific provisions, par. 127–37.

36. See Context-specific provisions, par. 138–40.



## Principle 2. Platforms adhere to international human rights standards, including in platform design, content moderation, and content curation

91. Digital platforms should ensure that human rights and due process considerations are integrated into all stages of the design process, as well as in content moderation and curation policies and practices.

### *Design processes*

92. The design of new products, as well as the content moderation and curation policies of digital platforms, should be consistent with the responsibility of corporations to respect human rights, as set out in the UN Guiding Principles on Business and Human Rights and other established international human rights standards.
93. Digital platforms should ensure non-discrimination and equal treatment in their design processes, as well as in their content moderation and curation policies, practices, and systems. This encompasses addressing biases, stereotypes, and discriminatory algorithms or content moderation practices that affect women and girls, as well as groups in situations of vulnerability and marginalization,

including indigenous communities. There should be an expectation that digital platforms ensure that all users, regardless of their background or abilities, can participate fully and engage with their services.

## ***Content moderation and curation policies and practices***

94. Content moderation and curation systems, including both automated and non-automated components, should be reliable and effective and at a scale appropriate to the volume of content being moderated, in all jurisdictions where the platform operates. This includes pursuing accuracy and non-discrimination in detection methods. Content moderation and curation should be applied consistently with international human rights law and standards, notably not to infringe on freedom of expression and cultural diversity.
95. Content moderation decisions across all regions and languages should, in a transparent manner, take into account the context, the wide variation of language nuances impacting meaning, and linguistic and cultural particularities of the content.
96. Platforms operating in multilingual environments should ensure that human and automated content moderation is available in all major languages spoken in that environment (at a minimum), at a scale appropriate to the volume of content.
97. Digital platforms should ensure that there is quick and decisive action to remove known child sexual abuse materials or live-streaming of acts of terror, in respect for the rights of all individuals, including groups in situations of vulnerability and marginalization. Platforms should nonetheless ensure that such content, which may be vital in the investigation and prosecution of crime, is not deleted, but rather preserved and securely safeguarded for use by law enforcement agencies and researchers as appropriate.
98. As outlined above, it is the State's responsibility guarantee the right to freedom of expression and ensure that any restrictions of content are consistent with international human rights law and standards, particularly Articles 19 (3) and 20 of the ICCPR. However, digital platforms should be able to demonstrate that any action taken when moderating and curating content has been conducted in accordance with their terms of services and community standards and should report accurately to the governance system or to the independent judicial system when applicable on performance vis-à-vis their responsibilities and/or plans.

- 99.** When considering measures to restrict content, platforms should take into account the conditions on legitimate restrictions to freedom of expression as laid out in Article 19 (3) of the ICCPR, and the prohibition of advocacy to hatred that constitutes incitement against discrimination, hostility, or violence as laid out in Article 20 (2) of the ICCPR, including the six-point threshold test for defining such content outlined in the Rabat Plan of Action.
- 100.** Once digital platforms identify content that could be permissibly restricted under international human rights law and standards, they should take measures such as: providing alternative reliable information; indicating concerns about the origin of the content to users; limiting or eliminating the algorithmic amplification of such content, with due attention to content reflecting gender biases or gender-based violence; de-monetizing content from advertising revenue; or removing/taking down the content.<sup>37</sup>

## ***Human content moderation***

- 101.** Human content moderators, whether employed by platforms directly or hired as outside contractors through outsourced roles, should be adequately trained, fluent in the language(s) used on the platforms and familiar with local linguistic and cultural contexts, evaluated, vetted, and psychologically supported. Platforms should further put in place well-funded and well-staffed support programmes for content moderators to minimize harm caused to them through their reoccurring exposure to violent or disturbing content while at work. The number of human moderators employed should be adequate to the complexity and volume of content they are expected to deal with.
- 102.** Platforms should also be explicit about whether they partner with third-party content moderation service providers, outside organizations, or experts to help them make decisions, particularly in countries or regions where the platform itself has little local knowledge. In doing so, platforms should always follow due diligence and refrain from revealing partners in situations in which there is a risk for their safety.

---

37. Applied measures should always be proportional to the legitimate objective they seek to protect. Removal and take downs of content and account suspension or blocking should be the last possible resort and should be used as uttermost means in uttermost cases.



## *Use of automated systems for content moderation and curation*

- 103.** Where appropriate, digital platforms should commission regular external audits, with binding follow-up steps, of the automated and human tools used for content moderation, curation, and recommender mechanisms for their precision, accuracy, and for possible bias or discrimination across different content types, languages, cultures, and contexts; they should also review their linguistic capacity and the consistent use across jurisdictions. As outlined in paragraph 87, they should also commission regular independent assessments of the impacts of their advertising systems on human rights, cultural diversity, and gender equality. The results of these reviews should be made public.<sup>38</sup>
- 104.** Digital platforms should have in place systems and processes to identify and take necessary action, in line with the provisions of these Guidelines, when any aspect of the design of the platform's services could result in the amplification of content that could be permissibly restricted under international human rights law and standards.
- 105.** Platforms should also ensure that curation and recommender systems, including both human and automated tools, do not amplify content that could be permissibly restricted under international human rights law and standards.
- 106.** Platforms should also ensure that content that could be permissibly restricted under international human rights law and standards is not amplified by automated curation or recommender mechanisms simply due to these mechanisms' linguistic limitations.
- 107.** Digital platforms should be able to explain to the governance system about the use and impact of the automated systems, including the extent to which such tools affect the data collection, targeted advertising, and the disclosure, classification, and/or removal of content, including artistic and election-related content.
- 108.** Digital platforms should provide users with options to adjust content curation and moderation systems. Users should be given the ability to control the content they see, and they should be able to easily understand how they can access diverse sources and viewpoints around trending topics. Platforms could also be required

---

38. One option is for independent audits and assessments to be done in accordance with global standards, and ideally verified by an independent body so that they could use the same reports regardless of the regulatory arrangement.



to give users options to manage the collection of personal data and the extent to which content recommenders respond to explicit or inferred preferences.

- 109.** Digital platforms should not use personal data obtained directly from children or obtained indirectly or inferred about children from other sources for profiling.

## **Notice**

- 110.** Digital platforms should notify users when their content is removed and the reason behind it. This would allow users to understand why that action on their content was taken, the method used (through automated means or after human review), and under which platform rules action was taken. Digital platforms should also have processes in place that permit users to appeal such decisions (see paragraphs 125–28). This provision may vary with the size of enterprise, and with the degree to which there are effective redress procedures for users to appeal against actions.



## Principle 3. Platforms are transparent

- 111.** Digital platforms should regularly report to the public and the governance system on how they adhere to the principles of transparency and explicability, and how they perform relative to their terms of services and community standards. This includes their responses to government demands for information or content removal.<sup>39</sup> The implementation of this provision may need to vary in practice based on company size, to limit the burden on smaller companies and start-ups.
- 112.** Transparency should be meaningful—the information provided should be as clear and concise as possible, and as detailed and complex as necessary. Transparency is not simply the provision of legal texts or a data dump, but about providing stakeholders with the information they need to make informed decisions.
- 113.** The transparency standards presented in these Guidelines can be considered as a minimum that should be met by all companies within the scope of any governance system.

### *Meaningful transparency*

- 114.** The effectiveness of digital platforms' transparency mechanisms should be independently evaluated against international standards through qualitative

---

39. Guidance on transparency for digital platforms can be found in the 26 high-level principles set forth by UNESCO in *Letting the Sun Shine In: Transparency and Accountability in the Digital Age*. <https://unesdoc.unesco.org/ark:/48223/pf0000377231>.

and empirical quantitative assessments to determine whether the information provided for meaningful transparency has served its purpose. Reports should be made publicly available on a regular basis.

**115.** Digital platforms should publish information outlining how they ensure that human rights and due process considerations are integrated into all stages of the content moderation and curation policies and practices. This publicly available information should include:

### Transparency in relation to digital platforms' terms of service

- a. Any measures used to moderate and curate content, set out in platforms' terms of service, including, for instance, lists of banned content or users.
- b. Any information about processes used to enforce their terms of service and to sanction users, as well as government demands/requests for content removal, restriction, or promotion.
- c. Information about the reasons behind restrictions imposed in relation to the use of their terms of service should be publicly available in an easily accessible format in their terms of service.
- d. Information about the types of content that are considered prohibited or against which the digital platform will act under their terms of service, and the measures taken, including the circumstances under which the digital platform will suspend a user's account, whether permanently or temporarily.

### Transparency in relation to the implementation of content moderation and curation policies and practices

- e. How content is moderated and curated, including through automated means and human review, as well as content that is being removed or blocked under either terms of service or pursuant to government demands/requests. This should include quantitative and qualitative information about the actual outcomes, results, and impacts that these systems produce.
- f. Any change in content moderation and curation policies should be communicated in accessible formats to users periodically.

- g.** Any use made of automated means for the purpose of content moderation and curation, including a specification of the role of the automated means in the review process, and any indicators of the benefits and limitations of the automated means in fulfilling those purposes.
- h.** Any safeguards applied in relation to content moderation and curation that are put in place to protect freedom of expression and access to information and diverse cultural content—including in response to government requests—particularly in relation to matters of public interest, including journalistic, artistic, and cultural content, and intellectual property rights.
- i.** Information about the number of human moderators employed or sub-contracted and the nature of their expertise in the local language(s) and local context, as well as whether they are in-house staff or contractors.
- j.** How personal data is collected, used, disclosed, stored, and shared, and what treatment is made of users' personal data, including which personal and sensitive data is used to make algorithmic decisions for the purpose of content moderation and curation. This also includes how personal data is shared with other entities and what personal data the platform obtains indirectly, for instance, through user profiling or interoperability with other parts of the digital ecosystem.

## Transparency in relation to user complaints mechanisms

- k.** Information relevant to appeals about the removal, blocking, or refusal to block content and how users can access the complaints process. This information should include quantitative and qualitative information of appeals received, treated, accepted, and rejected, and about the results of such appeals, and information about complaints received from State officials and the actions taken.

## Transparency on digital platforms' advertising practices

- l.** For digital platforms that use advertising as part of their business model, information about political advertisements and those of public interest, including the author and those paying for the ads, should be retained in a publicly accessible library online.
- m.** Practices of advertising and data collection and results of the human rights and gender impact assessment of the advertising systems.

- n. Information which allows individuals to understand the basis on which they are shown particular advertising.
- o. Content generated exclusively by machines should be labelled as such.

## *Data access for research purposes*

**116.** Digital platforms should provide vetted researchers with access to non-personal data and pseudonymous data that is necessary to understand the impact of digital platforms. This data should be made available upon request and on an ongoing basis through automated means, such as application programming interfaces (APIs), or other open and accessible technical solutions allowing the analysis of said data.

**117.** Digital platforms are expected to provide access to non-personal data to journalist and advocacy groups when there is a public interest and the access is proportionate and necessary in a determined context. There need to be additional safeguards to protect users' privacy and personal data—such as ensuring anonymizing datasets through different measures, including de-identification and sampling before sharing—as well as businesses' proprietary information, trade secrets, and respect of commercial confidentiality.

**118.** Platforms should build reliable interfaces for data access and should provide disaggregated data based on gender and other relevant intersecting factors (such as race, ethnicity, age, socioeconomic status, disability, etc.). The governance system should determine what is useful, proportionate, and reasonable for research purposes.



## Principle 4. Platforms make information and tools available for users

### *Language and accessibility*

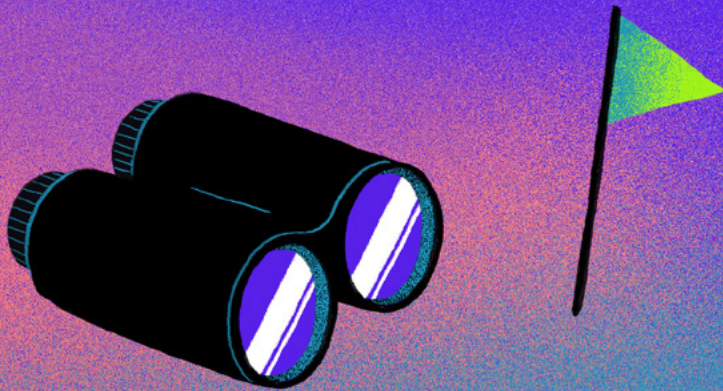
- 119.** Platforms should have their full terms of service available in the official and primary languages of every country where they operate, ensure that they are able to respond to users in their own language and process their complaints equally, and have the capacity to moderate and curate content in the user's language. Automated language translators can be deployed to provide greater language accessibility but should be monitored for accuracy due to their technical limitations.
- 120.** Platforms should ensure that reports, notices, and appeals processes are available in the language in which the user interacts with the service.
- 121.** Where digital platforms are likely to be accessed by children, they should provide all children with equal and effective access to information, and ensure the protection of their freedom of expression and privacy.<sup>40</sup> Terms of service and

40. OHCHR. 2021. "General comment No. 25 (2021) on children's rights in relation to the digital environment." <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>.

community standards should be made available in age-appropriate language for children and, as appropriate, be created with the viewpoint of a diverse group of children; special attention should be paid to the needs of children with disabilities to ensure they enjoy equal levels of access to information, as set out in the previous section.

- 122.** The rights of persons with disabilities should always be taken into account, with particular attention to the ways in which they can interact with and make complaints in relation to the platform. Platforms are expected to implement the necessary adjustments to make accessible information related to their terms of services, reports, notices, and appeals.





## Principle 5. Platforms are accountable to relevant stakeholders

### *User reporting*

- 123.** Platforms should establish reporting mechanisms for users and non-users, or third parties representing their interests, so they can report potential policy violations. Effective and accessible complaints mechanisms should be in place for members of groups in situations of vulnerability and marginalization. Digital platforms should also have the means to understand local contextual conditions when responding to user complaints, ensuring a culturally sensitive system design. Special reporting mechanisms should be established for children, designed for quick and easy use.
- 124.** The user reporting system should prioritize concerns regarding content that threatens users, ensuring rapid response, and, if necessary, provide a specific escalation channel or means of filing the report. This is particularly important when it comes to human rights violations, including gender-based violence and harassment.
- 125.** Companies should strive to prevent misuse of the reporting system through coordinated inauthentic behaviour.



## *User appeal and redress*

- 126.** Effective on-platform and external user redress mechanisms should be in place to allow users (and non-users, if impacted by specific content) to express their concerns and secure appropriate redress. This should include a clear, easily accessible, preferred, trusted,<sup>41</sup> and understandable reporting channel for complaints in their local language, with users notified about the result of their appeal.
- 127.** The appeals mechanism should follow the seven principles outlined in the UN Guiding Principles on Business and Human Rights for effective complaints mechanisms: legitimacy, accessibility, predictability, equitability, transparency, rights, compatibility, and continuous learning.
- 128.** Digital platforms should notify users and explain the appeal processes when their content is removed, expressly labelled, restricted in terms of comments or re-sharing or advertising association, or given special limits in terms of amplification or recommendation (as distinct from “organic/algorithmic” amplification and recommendation), and why. This would allow users to understand the reasons why that action on their content was taken, the method used (automated means or human review), and under which platform rules the action was taken. Platforms should also allow users to appeal such decisions and seek appropriate redress.
- 129.** Companies should work to ensure that systems for appeal and redress are not abused by coordinated inauthentic behaviour.

---

41. Ensuring user safety and compliance with international human rights law and standards.



## Context-specific provisions

*Protection of the rights of all individuals in situations of vulnerability and marginalization, women and girls, and those professionals who might be at risk because of their exercise of freedom of expression and access to information, such as journalists, artists, human rights defenders, and environmental defenders*

**130.** Digital platforms should put in place sufficient special protections for women and girls, users from groups in situations of vulnerability and marginalization, and journalists, artists, human rights defenders, and environmental defenders. To achieve this, digital platforms should:

- a.** Conduct regular human rights and gender impact assessments, including on their policies, moderation systems, and algorithmic approaches, with a view to identifying systemic risks to groups in situations of vulnerability and marginalization, women and girls, and journalists, artists, human rights defenders, and environmental defenders, and to adjust policies and practices to mitigate such risks.
- b.** Use privacy-protecting technology to provide external researchers with access to the platforms' internal data to help identify algorithmic amplification of online gender-based violence or other trends of violence arising from emerging technologies.

- c.** Create dedicated and inclusive engineering teams who are specifically trained to develop algorithmic solutions for content moderation and curation.
- d.** Develop and launch inclusive structured community feedback mechanisms to address gender, cultural, and other biases in new technologies.
- e.** Assess the human rights impact of their systems and processes for the treatment of independent news publishers and journalistic content hosted on their service.
- f.** Ensure equal treatment of independent news organizations on digital platforms.
- g.** Establish procedures to guard against the potential misuse of reporting rules and moderation mechanisms, especially misuse in bad faith designed to censor groups in situations of vulnerability and marginalization, women and girls, and journalists, artists, human rights defenders, and environmental defenders.

### ***Specific measures for electoral integrity***<sup>42</sup>

- 131.** Digital platforms should recognize their role in supporting democratic institutions by preserving electoral integrity. They should establish a specific risk assessment process for the integrity of the electoral cycle in the lead-up to and during major national election events, significant regional elections, or constitutional referendums (for instance, for the legislature or head of state in a presidential system).
- 132.** These assessments must be transparent, in line with human rights due diligence, and carried out with input from all relevant electoral stakeholders. The assessments should be conducted ahead of the electoral events in order to implement concrete measures to mitigate the identified risks. Assessments should include a gender approach, given the rise of online violence against women voters, candidates, activists, elected representatives, and electoral management officials.
- 133.** Digital platforms should make a reasonable effort to ensure that users have access to information and ideas of all kinds according to international human rights law. In particular, they should ensure that automated tools do not hinder access to election-related content and diverse viewpoints.

---

42. More information can be found in UNESCO's "Elections in digital times: A guide for electoral practitioners" (2022) [<https://unesdoc.unesco.org/ark:/48223/pf0000382102>], and in the "Joint Declaration on freedom of expression in the digital age of the UN Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, and the Organization of American States (OAS) Special Rapporteur on Freedom of Expression" (2020) [https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclarationDigitalAge\\_30April2020\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclarationDigitalAge_30April2020_EN.pdf).

- 134.** As part of the assessment, digital platforms should review whether their products, policies, or practices on political advertising arbitrarily limit the ability of candidates or parties to disseminate their messages.
- 135.** Digital platforms should make a reasonable effort to address content that could be permissibly restricted under international human rights law and standards, during the electoral cycle. Promoting independent fact-checking, advertisement archives, public alerts, and other measures should be taken into consideration. Engagement with relevant official independent regulatory institutions may be necessary according to the particular circumstances of each jurisdiction.
- 136.** Digital platforms should, as relevant, be transparent about the use and practical impact of any automated tools they use, albeit not necessarily the specific coding by which those tools operate, including inasmuch as those tools affect data harvesting, targeted advertising, and the sharing, ranking, and/or removal of content, especially election-related content.
- 137.** Digital platforms should also engage with all relevant stakeholders and their governance system prior to and during an election, to establish a means of communication if concerns are raised by the administrator or by users/voters. Engagement with relevant official regulatory institutions may be necessary according to the particular circumstances of each jurisdiction.
- 138.** Digital platforms that accept advertising designed to impact the electoral cycle should clearly identify such content as political advertisements. Digital platforms' terms of service should be clear about the digital platform's responsibility to be transparent about the amount of funding, the entity providing the funds, and the advertised entity, and consistently apply equal content moderation and curation rules on such advertisements.
- 139.** Digital platforms should track the monetization of posts by political parties and individuals representing parties.
- 140.** Platforms should disclose to the public information about the specific demographics targeted by such advertising/promotions.
- 141.** Platforms should retain these advertisements and all the relevant information on funding in a publicly accessible and regularly updated online library.

## *Specific measures in emergencies, armed conflict, and crises*

**142.** As a human rights safeguard, digital platforms should conduct human rights due diligence to address crises, situations of armed conflict, and other emergencies, including public health emergencies. This due diligence should analyse the human rights impact of the companies' operations, products, services, and advertising systems, on crisis and conflict dynamics.

**143.** During armed conflicts and crises, platforms should:

- a.** Ensure that content moderation in conflict settings includes robust human review, incorporating expertise in relevant languages and local and regional contexts.
- b.** Promote fact-checking.
- c.** Establish channels for meaningful and direct engagement with relevant stakeholders, including those operating in conflict-affected and high-risk areas.
- d.** Develop cooperation with trusted partners, independent media organizations, and other reliable flaggers.
- e.** Establish early warning systems and clear escalation systems for emergency situations to help detect imminent harm to individuals' physical safety.
- f.** Implement policies to limit and track the monetization of harmful content linked to armed conflict.
- g.** Preserve all potential evidence of human rights violations or war crimes, granting access to this archived material to appropriate national or international accountability mechanisms.

**144.** Risk assessments may require digital platforms to have processes in place for cases in which many simultaneous requests for action by users are made, as sometimes happens in the context of social unrest or massive violations of human rights. The governance system should recognise existing guidance from UN agencies and experts for conducting "heightened" human rights due diligence in such scenarios.



## Conclusion

**145.** Digital platforms have empowered individuals and societies with enormous opportunities to communicate, engage, and learn. They offer great potential for groups in situations of vulnerability and marginalization, democratizing spaces for communication and opportunities to have diverse voices engage with one another, be heard, and be seen. However, the potential of these platforms has been gradually eroded over recent years due to the lack of foresight in addressing key risks.

**146.** The aim of the Guidelines is to safeguard the right to freedom of expression, including access to information and other human rights in digital platform governance, **while dealing with content that can be permissibly restricted under international human rights law and standards.** By extension, digital platform governance that is grounded in human rights would further promote cultural diversity, cultural expression, and cultural diverse content.<sup>43</sup> The Guidelines outline a human rights-respecting governance system and promote risk mitigation and system-based processes for content moderation and curation. These Guidelines highlight overarching principles that should be followed in all governance systems that impact freedom of expression and access to information on

43. UNESCO 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions, articles 1 and 4. Under the Convention, “cultural content” refers to the symbolic meaning, artistic dimension, and cultural values that originate from or express cultural identities. Moreover, “cultural expressions” are those expressions that result from the creativity of individuals, groups, and societies, and that have cultural content.

digital platforms—independently of the specific regulatory arrangement and the thematic focus, as long as those arrangements are aligned with the provisions established in these Guidelines.

- 147.** The Guidelines were produced through a multistakeholder consultation process that began in September 2022. These current Guidelines are the result of an extensive consultation process in which UNESCO received, through three open consultations,<sup>44</sup> more than 1,540 submissions with more than 10,000 comments. Different stakeholder groups from 134 countries engaged in this process. This represents one of the largest and most open consultation exercises ever conducted by UNESCO.
- 148.** Further discussions on how to operationalise the Guidelines and use them as an advocacy tool will be developed next, with the goal of protecting freedom of expression, access to information and diverse cultural content, and all other human rights in the digital environment.

---

44. The three open consultations took place between December 2022 and January 2023; between February and March 2023; and between April and June 2023.



# Appendix

## Resources

### United Nations

*Guiding Principles on Business and Human Rights.* (2011)

[https://www.ohchr.org/sites/default/files/documents/publicationsguidingprinciplesbusinesshr\\_en.pdf](https://www.ohchr.org/sites/default/files/documents/publicationsguidingprinciplesbusinesshr_en.pdf)

*The Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.* (2012)

<https://www.ohchr.org/en/documents/outcome-documents/rabat-plan-action>

UN Special Rapporteur on freedom opinion and expression: “A human rights approach to online content moderation.” (2018)

[https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/Factsheet\\_2.pdf](https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/Factsheet_2.pdf)

UN Committee on the Rights of the Child: “General comment No. 25 (2021) on children’s rights in relation to the digital environment.” (2021)

<https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>

UN Secretary-General report: “Countering disinformation for the promotion and protection of human rights and fundamental freedoms.” (2022)

<https://www.ohchr.org/sites/default/files/2022-03/NV-disinformation.pdf>

UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression: “A/77/288: Disinformation and freedom of opinion and expression during armed conflicts.” (2022)

<https://www.ohchr.org/en/documents/thematic-reports/a77288-disinformation-and-freedom-opinion-and-expression-during-armed>

UN Secretary-General’s “Our Common Agenda - Policy Brief 8: Information Integrity on Digital Platforms.” (2023)

<https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf>



UN Secretary-General's "Our Common Agenda - Policy Brief 5: A Global Digital Compact" (2023) <https://indonesia.un.org/sites/default/files/2023-07/our-common-agenda-policy-brief-gobal-digi-compact-en.pdf>

## UNESCO

Puddephatt, A. 2021. "Letting the sunshine in: Transparency and accountability in the digital age." Paris: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000377231>

UNESCO. 2021. "The Legitimate Limits to Freedom of Expression: The Three-Part Test." [video] <https://www.youtube.com/watch?v=Wg8fVtHPDag>

## References on terminology

### Content which incites or portrays gender-based violence

- UN Human Rights Council. 2018. "Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective."  
<https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/HRC/38/47&Lang=E>
- OHCHR. 2021. "Statement by Irene Khan, Special Rapporteur on the promotion and protection of freedom of opinion and expression."  
<https://www.ohchr.org/en/statements/2022/02/statement-irene-khan-special-rapporteur-promotion-and-protection-freedom-opinion>

### Cultural diversity

"Cultural diversity" refers to the manifold ways in which the cultures of groups and societies find expression. These expressions are passed on within and among groups and societies. Cultural diversity is made manifest not only through the varied ways in which the cultural heritage of humanity is expressed, augmented, and transmitted through the variety of cultural expressions, but also through diverse modes of artistic creation, production, dissemination, distribution, and enjoyment, whatever the means and technologies used.

- UNESCO. 2005. *The 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions*. Paris: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000246264?posInSet=1&queryId=85dde93e-9ef6-402b-bc9f-452e6a399382>

## Disinformation and misinformation

- “Disinformation and freedom of opinion and expression.” Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan. <https://digitallibrary.un.org/record/3925306?ln=en> \

## Hate speech

- UN. 2019. “UN Strategy and Plan of Action on Hate Speech.” <https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml>

## Independent regulator

An independent regulator has its powers and responsibilities set out in an instrument of public law and is empowered to manage its own resources; its members are appointed in an independent manner and protected by law against unwarranted dismissal. In this case, the regulator’s decisions are made without the prior approval of any other government entity, and no entity other than a court or a pre-established appellate panel can overrule the regulator’s decisions. The institutional building blocks for decision-making independence are organizational independence (organizationally separate from existing ministries and departments), financial independence (an earmarked, secure, and adequate source of funding), and management independence (autonomy over internal administration and protection from dismissal without due cause).

- Salomon, E. 2016. *Guidelines for Broadcasting Regulation*. Paris: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000144292>
- Brown, A., J. Stern, B. Tenenbaum, and D. Gencer. 2006. *Handbook for Evaluating Infrastructure Regulatory Systems*. Washington, DC: World Bank Group. <http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-6579-3>

## Regulator

A body that supervises, monitors, and holds to account a private actor.

# **Guidelines for the Governance of Digital Platforms**

Safeguarding freedom of expression and access to information through a multistakeholder approach

Drafted through extensive consultations, these Guidelines aim to safeguard the right to freedom of expression and access to information in the context of the governance of digital platforms. They present a multistakeholder approach outlining key responsibilities for States and digital platforms to fulfill their human rights obligations and define roles for intergovernmental organizations, civil society, media, academia, the technical community, and other stakeholders.

[www.unesco.org/en/internet-trust/guidelines](http://www.unesco.org/en/internet-trust/guidelines)

[internetconference@unesco.org](mailto:internetconference@unesco.org)



**unesco**

United Nations  
Educational, Scientific  
and Cultural Organization



9 789231 006203