

Derechos Digitales’ comments to the UNESCO Guidelines for regulating digital platforms: a multi-stakeholder approach to safeguarding freedom of expression and access information - Draft 2.0¹

08 March, 2023

Derechos Digitales² is an independent, not-for-profit organisation with a Latin American scope, founded in 2005, whose main objective is the development, defence and promotion of human rights in the digital environment. The mission of Derechos Digitales is to ensure the full exercise of human rights in the digital environment in Latin America, through the study, dissemination of information, and advocacy in public policies and private practices, promoting social change around respect and dignity of people.

In this document we comment and provide feedback on the “Draft 2.0 of the Guidelines for regulating digital platforms: a multi-stakeholder approach to safeguarding freedom of expression and access information”³, from the form available on the UNESCO website⁴. Earlier, we provided comments on the 1.0⁵ and 1.1⁶ drafts, participated of the UNESCO Global Conference: “Internet for Trust”⁷, in addition to publishing an opinion column about this process⁸.

¹ Comments drafted by J. Carlos Lara. Please direct comments to info@derechosdigitales.org.

² More information: <https://www.derechosdigitales.org/>.

³ Available at: https://unesdoc.unesco.org/ark:/48223/pf0000384031_locale=en

⁴ Available at: <https://forms.office.com/e/ZGHMwq416G>

⁵ Available at:

<https://www.derechosdigitales.org/wp-content/uploads/UNESCO-Regulacion-de-plataformas-digitales-Comentarios-DD.pdf>

⁶ Available at:

<https://www.derechosdigitales.org/wp-content/uploads/Derechos-Digitales-Comments-on-UNESCO-platform-guidelines-Draft-1.1.pdf>

⁷ Available at:

<https://www.derechosdigitales.org/20148/derechos-digitales-participa-de-conferencia-de-la-unesco-por-una-internet-confiable/>

⁸ Maria Paz Canales and Michel Roberto de Souza. UNESCO: An opaque process and problematic recommendations for regulating digital platforms. Derechos Digitales. Available in [English](#), [Spanish](#) and [Portuguese](#).

Questions:

12. General comments on the overall draft 2.0 Guidelines.

Please provide us with your advice and perspectives about how the draft Guidelines could be improved.

It is our view that there is substantial improvement in Draft 2.0. Although we still maintain our critical view with regards to the need for this document as encouragement for platform regulation, we recognise the effort carried out by UNESCO to collect input from different stakeholders and allow those input to influence the draft, including through the global conference in February 2023. It is our expectation that such efforts to collect an increasing array of views from different stakeholders and internet users from all around the world can continue, so those inputs will be part of future drafts.

We maintain our concerns about this process as a potential pathway for States to act as empowered or emboldened to carry out digital platform regulation in a manner that risks going against human rights. The continued framing of this effort as one to guide *platform regulation* risks becoming an incentive to introduce new regulatory efforts that may impair freedom of expression and access to information, rather than enhancing them. Separately, we maintain our scepticism with regards to the need to provide specific guidance on the independent regulatory system, as well as specific guidance around sensitive issues such as situations of heightened risks, as in the case of elections. Some of these subjects are difficult enough to deserve more focused discussion, separately from broader ideas on platform regulation. Moreover, the language in several points still requires some detailed revision.

Because of the substantive discussions that each point of the shared documents requires, it is important that feedback is not only part of future versions of the document (or its sections), but also that inputs by different stakeholders are made public. Transparency in the discussion itself is a value that highlights the relevance of this subject matter, and publicising contributions opens up opportunities to identify convergence between different stakeholders and different groups. In order to allow for

meaningful participation in this process, UNESCO should be very transparent about the opportunities for future consultation, avenues for engaging, and guidance for meaningful contributions.

13. General comments on the drafting process

Please provide us with your advice and perspectives about how the drafting process could be improved.

With regards to the process, we maintain our stance that tight deadlines, shorter times to study non-English versions, and limited participation in open debates regarding the contents of the documents themselves, do not constitute a constructive manner to address the need for broad, inclusive, transparent, participatory discussion. This extends to the need for recognition of regional consultations as a valuable way to collect input from different contexts.

We strongly believe that this process should not be subject to hard deadlines, and beyond that, it should not be subject to a specific type of output, before the full discussion has been finalised. In our understanding, UNESCO has to ensure meaningful transparency regarding the whole process, publicising the documents, making public the inputs by different stakeholders, and opening documentation for public and regional consultations.

Overall draft 2.0 Guidelines

This section provides you with an opportunity to comment on each paragraph of the Guidelines, offering specific feedback about how you believe the document might be improved.

14. Introduction

Paragraph 1

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.”² 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 shaped our mandate within the United Nations system to protect and promote freedom of expression, access to information, and safety of journalists.

While this mandate can ensure the status of UNESCO as a promoter of rights, it is important to also recognise that the range of affected human rights requires an openness to the expertise of not only different stakeholders, but the expertise of a broad range of UN institutions and experts who share mandates concerning human rights that can be fostered or affected by the use of digital platforms.

15. Paragraph 2

Building upon relevant principles, conventions, and declarations over the past decade, the UNESCO Secretariat is now developing, through multistakeholder consultations and a global dialogue, Guidelines for regulating digital platforms: a multistakeholder approach to safeguarding freedom of expression and access to information (the Guidelines).

It is important to note that the framing of this document can be very different if the purpose is to regulate digital platforms, an effort intrinsically linked to the power of the State, than if it is to safeguard freedom of expression, which can happen through other means. If the previously expressed mandate of UNESCO is linked to freedom of expression and access to information, we understand that safeguards are the better focus of this document, rather than regulation as a highly complex action by states.

16. Paragraph 3

This endeavour also builds upon UNESCO’s work in the domain of broadcast regulation over several decades and furthers the Organization’s Medium-Term Strategy for 2022–2029 (41 C/4).

Broadcasting regulation is not necessarily a good point to build upon, notwithstanding the lessons from so many decades of broadcast regulation. There are inextricable differences among different means of communication, including those related to

scarcity, infrastructure control, state or private monopolies, obligations to carry content, and much more. If narratively this effort is linked to that body of work, it risks being interpreted in a similar fashion, which is risky considering the control that states claim with regards to broadcast media.

17. Paragraph 4

In 2015, UNESCO's General Conference endorsed the ROAM principles,⁴ which highlight the importance of human rights, openness, accessibility, and multi-stakeholder participation to the development, growth, and evolution of the internet. These principles recognize the fundamental need to ensure that the online space continues to develop and be used in ways that are conducive to achieving the Sustainable Development Goals.

The development of the ROAM Principles by itself is not only a source of lessons from their content, but also from their development. We urge UNESCO to consider consultations in the same fashion as the ROAM Principles included.

18. Paragraph 5

UNESCO's 41st General Conference endorsed the principles of the Windhoek+30 Declaration⁵ in November 2021, following a multistakeholder process that began at the global celebration of World Press Freedom Day in May of that year. The Declaration recognized information as a public good and set three goals to guarantee that shared resource for the whole of humanity: the transparency of digital platforms, citizens empowered through media and information literacy, and media viability. In speaking about 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.

Derechos Digitales wishes to point out that even though the Windhoek+30 Declaration recognises information as a public good, this does not amount to hard international human rights law. At the same time, we must highlight the problematic nature the term may have. For instance, a 2013 Law in Ecuador, the “Organic Law on Communication”, declared information a public good and a constitutional right, while

establishing restrictions to its circulation in the form of requirements on all media. We consider it a problematic concept for as long as it is not properly defined in its content.

19. Paragraph 6

The focus of the Guidelines on challenges related to freedom of expression and access to information complement the Organization’s work in the areas of education, the sciences, and culture. This includes UNESCO’s Recommendation on the Ethics of Artificial Intelligence,⁶ the 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expressions,⁷ and the MONDIACULT Declaration of 2022.

20. Paragraph 7

The current version of the Guidelines was produced through a multistakeholder consultation process that began in September 2022. Draft 2.0 will be discussed and consulted during the Internet for Trust Global Conference, to be held at UNESCO Headquarters in Paris from 21 to 23 February 2023. Subsequently, a revised draft of the Guidelines will be circulated for further consultations with a view towards finalization in the months following the Conference.

Derechos Digitales wishes to point out that the referenced “multistakeholder consultation process” is not fully known or accessible in official UNESCO channels. We are not aware of consultations like one which took place in late November 2022. Transparent, inclusive and meaningful participation is a value that we highlight as part of every policymaking effort, and we expect to be able to assess the same from this process. Clarity with regards of the “further consultations”, and how they will take place in each region, is expected as part of the next version of this document.

21. The objective of the Guidelines

Paragraph 8

The aim of the Guidelines is to support the development and implementation of regulatory processes that guarantee freedom of expression and access to information while dealing with content that is illegal⁹ and content that risks significant harm to democracy and the enjoyment of human rights.¹⁰ They call for States to apply regulation in a manner consistent with international human rights

standards and Article 19 of the International Covenant on Civil and Political Rights (ICCPR).

This paragraph marks the objective of the Guidelines, in a way that seems to focus on facilitating regulation rather than safeguarding freedom of expression online.

A larger issue arises with regards to the content that is targeted here. Illegal content can justify certain actions, provided international human rights law is fully complied with. But “content that risks significant harm to democracy and the enjoyment of human rights” is a wide, open, broad, vague, ambiguous category, which is not properly narrowed down by a reference to other categories of harmful speech. There is no objective standard in international human rights law to better determine the content of this category, including the element of enjoyment of rights or harm to democracy. Finally, this risks extending restrictions of forms of expression that are strongly protected under Article 19 ICCPR and, relevantly for Derechos Digitales, under Article 13 of the American Convention on Human Rights, such as the protection of the expression of personal opinions, however shocking or offensive.

It would be best for the document to promote safeguards for freedom of expression under international human rights law, relying on the large experience of regional systems and UN experts, and promote other safeguards that do not promote more restrictions on speech.

22. Paragraph 9

The Guidelines may serve as a resource for a range of stakeholders: for policymakers in identifying objectives, principles, and processes that could be considered in policymaking; for regulatory bodies dealing with the implementation of regulation; for digital platforms in their policies and practices; and for other stakeholders, such as civil society, in their advocacy and accountability efforts.

23. Paragraph 10

The Guidelines will inform regulatory processes under development or review for digital platforms, in a manner that is consistent with international human rights standards. Such regulatory processes should be led through an open, transparent, multistakeholder, and evidence-based manner.

a. The scope of these Guidelines includes digital platforms that allow users to disseminate content to the wider public, including social media networks,

messaging apps, search engines, app stores, and content-sharing platforms. Bodies in the regulatory system should define which digital platform services are in scope, and also identify the platforms by their size, reach, and the services they provide, as well as features such as whether they are for-profit or non-profit, and if they are centrally managed or if they are federated or distributed platforms.

We value the advancement of this paragraph in terms of what is expected for regulatory processes. However, the definition of the scope in the separate paragraph is confusing, since the definition of the platforms does not coincide with the examples that follow, nor with the experience of platforms that have been subject to scrutiny in different jurisdictions, like private messaging applications.

We appreciate the characteristics listed as relevant to determine platform obligations, but those are not necessarily considered in other sections of the document, thus becoming a mention without becoming guidance. We note that the shape and functionality of platforms is in constant shift, influenced both by user experience and by design decisions of platform controllers, which means that functionality becomes more relevant than initial intent. In other words, the use of an internet service is determined not only by its initial design, but also by how it is used. This has also an effect on the dynamics of virality, individual and collective habits, geographic locations, and more, which affect how much an individual piece of content can have further “reach”.

Finally, we disagree with the mention of the “bodies in the regulatory system” that can determine the entities in scope, as it risks limiting expression via administrative decisions.

24. Paragraph 11

The Guidelines will:

- a. Enrich and support a global multistakeholder shared space to debate and share good practices about digital platform regulation to protect freedom of expression and access to information, while dealing with content that is illegal under international human rights law and content that risks significant harm to democracy and the enjoyment of human rights, 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 government and digital platforms accountable.

c. Add to existing evidence-based policy approaches that respect human rights, ensuring alignment where possible.

While these are lofty goals, we must stress that the mention of objectives in the first sections of the document is not by itself a means of guidance. There is no clarity regarding what is a “global multistakeholder shared space”, or whether it is UNESCO’s intention to host one such space regardless of existing multistakeholder efforts.

We also stress that there is no clarity with regards what is “content that is illegal under international human rights law”, since it may differ from nationally regulated content, and “content that risks significant harm to democracy and the enjoyment of human rights” which is a category we do not consider should be part of this document as a distinct subject of regulation.

25. Paragraph 12

The Guidelines will contribute to ongoing UN-wide processes, such as the implementation of the proposals in “Our Common Agenda,” including the development of the Global Digital Compact, 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 integrity in public information. 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) in 2025.

We value and appreciate greatly that evolving versions of these Guidelines recognise existing and ongoing UN processes that touch upon platform regulation and connect different subjects related to cyberspace. We must also stress that since the referenced processes do not refer to this effort by UNESCO, it is unclear how these different efforts connect, or how this contribution will be recognised in those other processes, or how those other processes will be reflected in the Guidelines’ content.

Derechos Digitales wishes to also ask whether the mention of the 2025 IGF is an intentional attempt at setting a finalisation date for this process, given that other UN processes might be over before that IGF meeting.

26. Structure of the Guidelines

Paragraph 13

The Guidelines start by setting out the overall approach to regulation. They continue by outlining the responsibilities of different stakeholders in fostering an environment for freedom of expression, access to information, and other human rights. 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.*

As with the paragraphs setting the high-level goals of this effort, the description of the document itself belongs to other informative materials and not the Guidelines themselves.

27. Paragraph 14

Then the Guidelines propose some preconditions that should be considered in the establishment of an independent regulatory system, regarding its constitution, powers, and external review.

This should not be part of these Guidelines, considering highly contingent and context-specific factors that can hardly be anticipated or generalised.

28. Paragraph 15

Finally, it describes the areas where digital platforms should have structures and processes in place to fulfil the objective of the regulation.

The language seems to betray the stated goals of the document if the purpose is a “regulation” directly, which may differ between jurisdictions. We recommend that the language better reflect the objectives of the document.

29. Paragraph 16

It is important to underscore that this document should be considered in its

entirety. The adoption or implementation of specific provisions on their own will not be sufficient to achieve the regulatory goals.

As with the previous paragraph, there is a mention of “regulatory goals” that are not part of UNESCO’s mandate or jurisdiction, and which could hardly set boundaries to efforts by States. We recommend that the language better reflect the objectives of the document.

30. Approach to regulation

Paragraph 17

The goal of any regulation of digital platforms that intends to deal with illegal content and content that risks significant harm to democracy and the enjoyment of human rights should include guaranteeing freedom of expression, the right to access information, and other human rights. This goal should be established in law and be drawn up after an open, transparent, multistakeholder, and evidence-based process.

Unlike previous versions, reference is made to “any regulation” without explicitly mentioning co-regulation or self-regulation until this point, which may lead to assuming a reference to State regulation. We are concerned that UNESCO is here trying to commandeer countries’ legitimately defined regulatory objectives or purposes, instead of attempting to establish safeguards for the exercise of human rights.

31. Paragraph 18

Regulation should focus mainly on the systems and processes used by platforms, rather than expecting the regulatory system to judge the appropriateness or legality of single pieces of content. Any specific decisions about the legality of specific pieces of content should follow due process and be open to review by a judicial body, following the three-part test on legitimate restrictions to freedom of expression as laid out in the ICCPR, and where relevant, the six-point threshold for defining criminal hatred that incites to discrimination, hostility, or violence outlined in the Rabat Plan of Action.

We value the focus on “systems and processes” rather than on content. The paragraph itself joins several different subjects, including judicial review and the

factors that are part of the Rabat Plan of Action. These are at least three or four distinct points that should be addressed separately.

32. Paragraph 19

Within regulation, digital platforms are expected to be transparent about the systems and processes used to moderate and curate content on their platforms and how those systems and processes fulfil the goal of regulation. If the established goal is not being fulfilled, the regulatory system should have the power to require the digital platform to take further action, as described in paragraph 46(f). The regulator will expect digital platforms to adhere to international human rights standards in the way they operate and to be able to demonstrate how they are implementing these standards and other policies contained in their terms of service.

It is troubling that this paragraph keeps a mention of a “regulator”, probably as a remnant of previous versions. As highlighted during the February conference and in previous contributions, mentions to “the regulator” become problematic if there is assertion or expectations of actions disregarding the proper conditions for their existence.

33. Paragraph 20

Alongside the regulation of digital platforms, it is essential that key media and information literacy skills for users are promoted, including by the platforms themselves. This enables users to engage critically with content and technologies, navigate a rapidly evolving media and information landscape marked by the digital transformation, and build resilience in the face of related challenges.

34. Paragraph 21

The current approach taken by these Guidelines is one of co-regulation, implying that the State, on the one hand, provides a legal framework that enables the creation, operationalization, and enforcement of rules, and self-governing bodies, on the other hand, create rules and administer them, sometimes through joint structures or mechanisms. This should be done in accordance with international human rights law and under the public scrutiny of civil society organizations, journalists, researchers, and other relevant institutions in a system of checks and balances.

If the Guidelines focus on process, UNESCO should not preempt debates or legitimate decisions on the approach each jurisdiction may adopt, let alone for the creation of “self-governing bodies”. Self-regulation, even if desired legitimately, is not necessarily an ideal solution in an abstract manner considering all the potential evolution of digital platforms. The reference to international human rights law in relation to such vaguely defined notions is a confusing addition.

35. Enabling environment

Paragraph 22

To accomplish the goal of regulation, all stakeholders involved have a role in sustaining an enabling environment for freedom of expression and the right to information, while dealing with content that risks significant harm to democracy and the enjoyment of human rights.

36. Paragraph 23

Creating a safe and secure internet environment for users while protecting freedom of expression and access to information is not simply an engineering question. It is also a responsibility for societies as a whole and therefore requires whole-of-society solutions.

This is a justification paragraph and not part of operational Guidelines. It could be part of a preamble or additional informative documents and not part of the document.

37. States' duties to respect, protect, and fulfil human rights

Paragraph 24

States have a particular duty to promote and guarantee freedom of expression and the right to access information, and to refrain from censoring legitimate content.

Derechos Digitales wishes to stress that there is no category for “legitimate content” in international human rights law, and that the prohibition on prior censorship (in the case of the Inter-American Human Rights system, for instance) is not limited to such a category.

38. Paragraph 25

A key element of an enabling environment is the positive obligation to promote universal and meaningful access to the internet. In 2011, 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.”

Although the central idea of this paragraph is one that in broad terms Derechos Digitales agrees with, it should not be part of this document.

39. Paragraph 26

Moreover, it is a responsibility of the State to be transparent and accountable about the requirements they place upon digital platforms.

40. Paragraph 27.a

Specifically, States should:

a. Respect the requirements of Article 19(3) of the ICCPR: any restrictions applied to content should have a basis in law, have a legitimate aim, and be necessary and proportional, ensuring that users’ rights to freedom of expression, access to information, equality and non-discrimination, autonomy, dignity, reputation, privacy, association, and public participation are protected.

41. Paragraph 27.b

Specifically, States should:

b. Provide an effective remedy for breaches of these rights.

42. Paragraph 27.c

Specifically, States should:

c. Ensure that any restrictions imposed upon platforms consistently follow the high threshold set for defining legitimate restrictions on freedom of expression, on the basis of the application of Articles 19 and 20 of the ICCPR.

43. Paragraph 27.d

Specifically, States should:

d. Be open, clear, and specific 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 Article 19 of the ICCPR.

44. Paragraph 27.e

Specifically, States should:

e. Refrain from disproportionate measures, particularly prior censorship and internet shutdowns, under the guise of combatting disinformation or any other reason inconsistent with the ICCPR.

45. Paragraph 27.f

Specifically, States should:

f. Refrain from imposing a general monitoring obligation or a general obligation for digital platforms to take proactive measures to relation to illegal content. 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 illegal content.

46. Paragraph 27.g

Specifically, States should:

g. Refrain from 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, as this may have a chilling effect on freedom of expression.

47. Paragraph 27.h

Specifically, States should:

h. Promote media and information literacy, including in digital spaces, as a complementary approach to regulation with the aim of empowering users. This should draw upon the expertise of media and information literacy experts, academics, civil society organizations, and access to information institutions.

48. Paragraph 27.i

Specifically, States should:

i. Ensure that the regulatory system with responsibilities in this area is structured as independent and has external review systems in place (see paragraphs 47–49) such as legislative scrutiny, requirements to be transparent and consult

with multiple stakeholders, and the production of annual reports and regular audits.

49. The responsibilities of digital platforms to respect human rights

Paragraph 28. A

Digital platforms should comply with five key principles:

a. Platforms respect human rights in content moderation and curation. They have content moderation and curation policies and practices consistent with human rights standards, implemented algorithmically and through human means, with adequate protection and support for human moderators.

50. Paragraph 28.b

b. Platforms are transparent, being open about how they operate, with understandable and auditable policies. This includes transparency about the tools, systems, and processes used to moderate and curate content on their platforms, including in regard to automated processes.

51. Paragraph 28.c

c. Platforms empower users to understand and make informed decisions about the digital services they use, including helping them to assess the information on the platform.

52. Paragraph 28.d

d. Platforms are accountable to relevant stakeholders, to users, the public, and the regulatory system in implementing their terms of service and content policies, including giving users rights of redress against content-related decisions.

53. Paragraph 28.e

e. Platforms conduct human rights due diligence, evaluating the risks and impact on human rights of their policies and practices.

This could be better detailed and subject to further discussion, as due diligence and impact assessments are difficult subjects.

54. Paragraph 29

To follow these principles, there are specific areas on which digital platforms have a responsibility to report to or act before the regulatory system. These areas are described in paragraphs 50–105.

55. The role of intergovernmental organizations

Paragraph 30

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, including by providing technical assistance, monitoring and reporting human rights violations, developing relevant standards, and facilitating multistakeholder dialogue.

We fail to understand why this is not an explicit reference to UNESCO's role in following up the recommendation set by this document.

56. The role of civil society and other stakeholders

Paragraph 31

Every stakeholder engaged with the services of a digital platform as a user, policymaker, watchdog, or by any other means, has an important 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 every regulation should take a multistakeholder approach; a broad set of stakeholders should also be engaged in oversight.

We fail to understand why the first half of this paragraph is not part of a preamble or ancillary documents. There is no clarity on what a multi-stakeholder “approach” means.

57. Paragraph 32

Civil society plays a critical role in understanding the nature of and countering abusive behaviour online, as well as challenging regulation that unduly restricts freedom of expression, access to information, and other human rights.

We fail to understand why this is not referencing civil society groups or organisations.

58. Paragraph 33

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 oversight of how the regulatory system is working. Independent institutions and researchers can support risk assessments, audits, investigations, and other types of reports on platforms' practices and activities.

We must stress that identifying patterns is not the purview of “researchers” as much as it is the result of research work, regardless of who conducts it.

59. Paragraph 34

Media and fact-checking organizations have a role in promoting information as a public good and dealing with content that risks significant harm to democracy and the enjoyment of human rights on their own platforms.

60. Paragraph 35

Engineers, data scientists, and all the technical community involved also have a role in understanding the human rights and ethical impacts of the products and services they are developing.

61. Paragraph 36

All of these stakeholders should have an active role in consultations on the operation of the regulatory system.

We stress that stakeholder participation is part of the whole policymaking cycle, including prior assessments as well as monitoring and evaluation, not just consultations.

62. The regulatory system

Paragraph 37

There are vastly different types of bodies involved in online regulation throughout

the world. They range from existing broadcast and media regulators who may be asked to take on the role of regulating content online, to newly established dedicated internet content regulators or communications regulators given an extended remit. There may also be overlap in some states with advertising or election bodies, or with information commissioners or national human rights institutions. Some regulators exist independently of the government while others are constituted as government agencies.¹⁵ Recognising the complexity of this environment, these Guidelines are meant to be generally applicable to any system of regulation, irrespective of its specific modalities, and accept that local contexts will impact how regulation is enacted and implemented.

We recognise the effort to acknowledge different regulatory contexts and systems, although the examples are not very fortunate and the explanation in general should not necessarily be part of the Guidelines. It would be better to maintain a broad idea of the characteristics of any regulatory system that should implement the recommendations as allowed by law. Because in general this is a sensitive topic in different contexts, this whole section should be subject to focused discussion.

63. Paragraph 38

Whatever form it takes, any process that establishes a regulatory system for digital platforms should be open and transparent and include multistakeholder consultation. Additionally, achieving the goal of regulation requires the existence of an independent regulatory system that allows regular multistakeholder consultation on its operation.

64. Paragraph 39

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 (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.”

This should be part of explanatory documents and not part of the Guidelines themselves.

65. Paragraph 40

The proposal below is divided into three sections: the constitution of an independent regulatory system, its powers, and suggested provisions for review.

66. Constitution

Paragraph 41

Any regulatory system—whether comprised of a single body or multiple overlapping bodies—which assesses applications or performs inspectorial, investigative, or other compliance functions over how digital platforms conduct content moderation and curation, needs to be independent and free from economic or political pressures.

67. Paragraph 42

The regulatory system must have sufficient funding to carry out its responsibilities effectively. The sources of funding must also be clear, transparent, and accessible to all and not subject to the decisions of the regulator(s).

68. Paragraph 43

Officials or members of the regulatory system should:

- a. Be appointed through a participatory and independent merit-based process.*
- b. Be accountable to an independent body (which could be the legislature, an external council, or an independent board/boards).*
- c. Have relevant expertise in international human rights law.*
- d. Deliver a regular 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 conflict of interest and declare any gifts or incentives.*
- f. After completing the mandate, not be hired or provide paid services to those who have been subject to their regulation, and this for a reasonable period, in order to avoid the risk known as “revolving doors.”*

There is confusion here regarding which requirements refer to entities and which to their members, and the expertise expected or obligations created. This should be subject to further consultation, to better differentiate which of these requirements are targeting expertise and independence as a precondition, and which is rather aimed at ensuring transparency and accountability as obligations.

69.Powers

Paragraph 44

The regulatory system should primarily focus on the systems and processes used by digital platforms to moderate and curate content, rather than making judgements about individual pieces of content. The system should also look at how digital platforms promote freedom of expression and access to information and the measures it has established to deal with illegal content and content that risks significant harm to democracy and the enjoyment of human rights.

70. Paragraph 45

The regulatory system should have the power to assess applications or perform inspectorial, investigative, or other compliance functions over digital platforms to fulfil the overarching goals to protect freedom of expression and access to information, while moderating illegal content and content that risks significant harm to democracy and the enjoyment of human rights, in a way consistent with the provisions of Article 19 of the ICCPR.

71. Paragraph 46.a

To fulfil the goal of regulation, the regulatory system should have the following powers:

a. Establish standardized reporting mechanisms and formats. Ideally, reports should be made annually in a machine-readable format.

As in previous versions, the Guidelines describe “powers” of a system which could only become the powers of regulating entities. It must be stressed that advocating for the creation of powers that can affect the exercise of freedom of expression can risk becoming a way of restricting expression by means that are not in accordance with the principle of legality. The language should be subject to further consultation.

72. Paragraph 46.b

Commission off-cycle reports if there are exigent emergencies, such as a sudden information crisis (such as that brought about by the COVID- 19 pandemic) or a specific event which creates vulnerabilities (for example, elections or protests).

73. Paragraph 46.c

Summon any digital platform deemed non-compliant with its own policies or failing to protect users. Any decision by the regulator should be evidence-based, the platform should have an opportunity to make representations and/or appeal against a decision of non-compliance, and the regulatory system should be required to publish and consult on enforcement guidelines and follow due process before directing a platform to implement specific measures.

This power is ambiguously defined, and conflates different ideas on decisions, appeals and self-regulation.

74. Paragraph 46.d

Commission a special investigation or review by an independent third party if there are serious concerns about the operation or approach of any platform or an emerging technology when dealing with illegal content or content that risks significant harm to democracy and the enjoyment of human rights.

75. Paragraph 46.e

Establish a complaints process that offers users redress should a platform not deal with their complaint fairly, based on the needs of the public they serve, the enforcement powers they have in law, their resources, and their local legal context.

The proposal could advance to promote the legitimacy to question decisions of the platforms before the regulator when it comes to cases involving systemic failures or collective harm to users. Here, the proposal could go forward to state that organised civil society will have legitimacy to collectively challenge such decisions, as a measure to facilitate access to justice.

76. Paragraph 46.f

Oversee the fulfilment by the digital platforms of the five principles detailed in these guidelines, taking necessary and proportional enforcement measures, in

line with international human rights law, when platforms consistently fail to implement these principles.

77. Review of the regulatory system

Paragraph 47

There should be a provision for a periodic independent review of the regulatory system, conducted by a respected third party, reporting directly to the legislature.

78. Paragraph 48

Any part of the regulatory system should act only within the law in respect of these powers, respecting fundamental human rights—including the rights to privacy and to freedom of expression. It should be subject to review in the courts if it is believed to have exceeded its powers or acted in a biased or disproportionate manner.

79. Paragraph 49

Decisions on eventual limitations of specific types of content must be allowed to be reviewed by an independent judicial system, following a due process of law.

80. Responsibilities of digital platforms

Paragraph 50

Digital platforms should respect human rights and adhere to international human rights standards in accordance with the UN Guiding Principles on Business and Human Rights.

81. Paragraph 51

According to the five principles set above, digital platforms are expected to have structures and processes in place and should be accountable to the regulatory systems, in line with the powers described above, in the following areas:

82. Principle 1. Platforms respect human rights in content moderation and curation

Content moderation and curation policies and practices

Paragraph 52

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

83. Paragraph 53

The content moderation and curation policies of digital platforms should be consistent with the obligations 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.

84. Paragraph 54

Content moderation and curation structures and processes should be applied consistently and fairly across all regions and languages.

85. Paragraph 55

No distinction should be made between content that is similar or between users. However, content moderation decisions should, in a transparent manner, take into account the context, the wide variation of language nuances, and the meaning and linguistic and cultural particularities of the content.

86. Paragraph 56

Digital platforms should—in policy and practice—ensure whenever they become aware of the availability of illegal content that they act with due diligence and in accordance with international human rights standards. At a minimum, they should ensure that there is quick and decisive action to remove known child sexual abuse materials or other explicit and severe illegal content which is not contextually dependent.

87. Paragraph 57

It would be expected that illegal content be made unavailable solely in the geographical jurisdiction where it is illegal.²⁰ Identification of illegal content should be interpreted consistently with international human rights law to avoid unjustified restrictions on freedom of expression.

It is unclear why the paragraph starts with “it would be expected”, instead of directly recommending reducing the reach of restriction decisions.

88. Paragraph 58

Platforms should be able to demonstrate to the regulatory system about the measures they carry out to detect, identify, or remove illegal content.

89. Paragraph 59

In the case of other content that risks significant harm to democracy and the enjoyment of human rights, digital platforms should systematically assess the potential human rights impact of such content and take action to reduce vulnerabilities and increase their capacities to deal with it. For instance, companies should be able to demonstrate to the regulatory system the measures that they have in place if such risk is identified. These could be by, for example, providing alternative reliable information, indicating concerns about the origin of the content to users, limiting or eliminating the algorithmic amplification of such content, or de-monetizing from advertising revenue.

90. Human content moderation

Paragraph 60

Human content moderators should be adequately trained, sufficiently staffed, fluent in the language concerned, vetted, and psychologically supported. Platforms should further put in place well-funded and -staffed support programmes for content moderators to minimize harm caused to them through their reoccurring exposure to violent or disturbing content while at work. Where possible and when it would not negatively impact human rights or undermine adherence to international norms for freedom of expression, human moderation of content should take place in the country or region where it is published to ensure close awareness of local or national events and contexts, as well as fluency in the language concerned.

91. Paragraph 61

The platform should also be explicit about whether it partners with outside organizations or experts to help it make decisions, particularly in countries or regions where the platform itself has little local knowledge. In so doing, they should always follow the “do no harm principle” and refrain from revealing partners in situations in which revealing these partners may present risks for their safety.

92. Use of automated systems for content moderation and curation

Paragraph 62

Digital platforms should commission regular external audits of machine learning tools utilised for content moderation for their precision, accuracy, and for possible bias or discrimination across different content types, languages, and contexts. They should also commission regular independent assessments of the impacts of automated content moderation tools on human rights. The results of these reviews should be reported to the regulatory system.

As with previous versions, we are concerned that the focus on specific tools leaves out similar concerns regarding other forms of moderation, singling out specific but undefined technologies. We are also wondering why this is focused on external audits on how some tools work regarding specific characteristics, instead of broader ideas of human rights impacts.

93. Paragraph 63

Digital platforms should commission regular external audits of machine learning tools utilised for automated curation and recommender mechanisms –designed to enhance user engagement –for their precision, accuracy, and for possible bias or discrimination across different content types, languages, and contexts. They should also commission regular independent assessments of the impacts of these mechanisms on human rights. The results of these reviews should be reported to the regulatory system.

94. Paragraph 64

Digital platforms should have in place systems and processes to identify and take necessary action, in line with the provisions of these guidelines, when automated curation and recommender mechanisms –designed to enhance user engagement –result in the amplification of content that risks significant harm to democracy and human rights.

95. Paragraph 65

Users should be given the ability to control the algorithmic curation and recommender mechanisms used to suggest content to them. Content curation

and recommendation systems that provide different sources and include different viewpoints around trending topics should be made clearly available to users.

96. Paragraph 66

Finally, digital platforms should notify users when their content is removed or subject to content moderation and why. This would allow users to understand why that action on their content was taken, the method used (algorithmic 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 89-91).

97. Principle 2. Platforms are transparent

Paragraph 67

Digital platforms should report to the regulatory system on how they fulfil the principles of transparency, explicability, and reporting against what they say they do in their terms of services and community standards. The meaning of transparency depends upon the audience. For users, it can mean, for example, understanding how the platform finds and presents information and collects their data; for regulators, it can mean information needed to verify the way in which digital platforms' business operations may impact democracy and human rights, and if terms of service and community standards are consistently and fairly applied; and for researchers, it can mean understanding the impact of the services on society in general.

We fail to understand why this paragraph includes a discussion on the meaning of transparency that should be part of additional explanatory documents and not necessarily as part of the guidance from the Guidelines, while at the same time limiting reports to the regulatory system instead of the now-common practice of public, periodic transparency reports.

98. Paragraph 68

The regulatory system and digital platforms should understand transparency as meaningful transparency. Transparency is not simply the provision of legal texts or a data dump—it should be understood as providing stakeholders with the information they need to make informed decisions.

99. Meaningful transparency

Paragraph 69

The effectiveness of digital platforms' transparency mechanisms should be independently evaluated through qualitative and empirical quantitative assessments to determine whether the information provided for meaningful transparency has served its purpose. Reports should be made available to users on a regular basis.

100. Transparency in relation to terms of service

Paragraph 70 a.

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: (a) Any measures used to moderate and curate content, set out in platforms' terms of services.

101. Paragraph 70 b.

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: (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.

102. Paragraph 70 c.

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: (c) Information about the reasons behind any restrictions imposed in relation to the use of their service, publicly available in an easily accessible format in their terms of service.

103. Transparency in relation to content moderation and curation policies and practices

Paragraph 70 d.

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: (d) How content is moderated and curated, including through algorithmic (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.

104. Paragraph 70 e.

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: (e) Any change in content moderation and curation policies should be communicated to users periodically.

105. Paragraph 70 f.

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: (f) 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

106. Paragraph 70 g.

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: (g) Any safeguards applied in relation to any content moderation and curation that are put in place to protect freedom of expression and the right to information, including in response to government requests, particularly in relation to matters of public interest, including journalistic content.

107. Paragraph 70 h.

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: (h) Information about the number of human moderators employed and the nature of their expertise in local language and local context, as well as whether they are in-house staff or contractors.

108. Paragraph 70 i.

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: (i) How personal data is used and what treatment is made of users' personal data, including personal and sensitive data, to make algorithmic decisions for purposes of content moderation and curation.

109. Transparency in relation to user complaints mechanisms

Paragraph 70 j.

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: (j) Information relevant to complaints about the removal, blocking, or refusal to block content and how users can access the complaints process.

110. Transparency and commercial dimensions

Paragraph 70 k.

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: (k) Information about political advertisements, including the author and those paying for the ads; these advertisements should be retained in a publicly accessible library online.

111. Paragraph 70 l.

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: (l) Practices of advertising and data collection.

112. Paragraph 70 m.

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: (m) Information which allows individuals to understand the basis on which they are being targeted for advertising.

113. Paragraph 71

Many regulatory regimes require broader and more granular transparency standards than those outlined here. The standards presented in these Guidelines can be considered as a baseline from which regulatory regimes can elaborate further.

114. Data access for research purposes

Paragraph 72

Digital platforms should provide access to non-personal data and anonymised data for vetted researchers that is necessary for them to undertake research on content to understand the impact of digital platforms. This data should be made available through automated means, such as application programming interfaces (APIs), or other open and accessible technical solutions allowing the analysis of said data.

115. Paragraph 73

They should provide access to data to undertake research on illegal and harmful content such as hate speech, disinformation, misinformation, and content which incites or portrays gender-based violence; such data should be disaggregated for the purpose of investigating impacts on specific populations. There need to be additional safeguards to protect the privacy and personal data of users, as well as businesses' proprietary information, trade secrets, and respect of commercial confidentiality.

116. Paragraph 74

Platforms should build reliable interfaces for data access. The independent regulatory system should determine what is useful, proportionate, and reasonable for research purposes.

117.Principle 3. Platforms empower users User reporting

User reporting

Paragraph 75

It is critical to empower users of digital platforms. In addition to the digital platform making information about its policies accessible in a digestible format and in all relevant languages, it should demonstrate how users can report potential abuses of the policies, whether that be the unnecessary removal of content or the presence of allegedly illegal content or content that risks significant harm to democracy and the enjoyment of human rights, or of any other content which is in breach of its policies. Digital platforms should also have the means to understand local context and local conditions when responding to user complaints and ensure that their systems are designed in a culturally sensitive way.

This paragraph, while valuable in its recommendations, conflates the delivery of information with the encouragement needed to make use of them as a form of empowerment, which is tied to the proven effectiveness of the use of platform mechanisms. The final sentence of the paragraph, with the goal of understanding local contexts, should be part of a separate consideration.

118. Paragraph 76

The user reporting system should give high priority to concerns regarding content that threatens users, ensuring a rapid response, and, if necessary, by providing a specific escalation channel or means of filing the report. This is particularly important when it comes to gender-based violence and harassment.

We consider that user reporting systems should also be subject to scrutiny, as mass reporting is regularly used as a means to restrict reach or force removal of contents or account suspensions, all of which are also forms of harassment.

119.Media and information literacy

Paragraph 77

When reporting to the regulatory system, platforms should demonstrate their overall strategy related to media and information literacy and the actions they have taken to advance on it. There should be a specific focus inside the digital

platform on how to improve the digital literacy of its users, with thought given to this in all product development teams. The digital platform should consider how any product or service impacts user behaviour beyond the aim of user acquisition or engagement.

120. Paragraph 78

Platforms should train their product development teams on media and information literacy from a user empowerment perspective, based on international standards, and put in place both internal and independent monitoring and evaluation mechanisms. They should inform the regulatory system about any relevant result of these evaluations.

121. Paragraph 79

Digital platforms should implement such measures in close collaboration with organizations and experts independent of the platforms, such as public authorities responsible for media and information literacy, academia, civil society organizations, researchers, teachers, specialized educators, youth organizations, and children's rights organizations. Specific measures should be taken for users and audiences in social or cultural vulnerability and/or with specific needs.

122. Paragraph 80

Digital platforms should be explicit about the resources they make available to improve media and information literacy, including digital literacy about the platform's own products and services, as well as relevant processes, for their users.

123. Paragraph 81

Digital platforms should also ensure that users understand their rights online and offline, including the role of media and information literacy in the enjoyment of the rights to freedom of expression and access to information. Toward this end, they could partner with independent media and information literacy experts or organizations that have relevant expertise in the thematic area, including academic and civil society organizations.

124. Language and accessibility

Paragraph 82

Major platforms should have their full terms of service available in the 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, while they have their limitations, can be deployed to provide greater language accessibility.

125. Paragraph 83

Platforms should also ensure that content that risks significant harm for democracy and human rights is not amplified by automated curation or recommender mechanisms simply due to a lack of linguistic capacity of those mechanisms.

126. Paragraph 84

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.

127.Children's rights

Paragraph 85

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 can result in lifelong or transgenerational consequences. Digital platforms should therefore also recognise their specific responsibilities toward children.

128. Paragraph 86

Where digital platforms allow use of their services by children, they should provide all children with equal and effective access to age-appropriate information, including information about their rights to freedom of expression, access to information, and other human rights. Terms of services and community standards should be made available in age-appropriate language for children and, as appropriate, be co-created with a diverse group of children; special attention should be paid to the needs of children with disabilities to ensure they enjoy equal levels of transparency as set out in the previous section.

129.Principle 4. Platforms are accountable to relevant stakeholders

Paragraph 87

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 fairly and accurately to the regulatory system on performance vis-à-vis their responsibilities and/or plans. In case of failure to comply with this provision, the regulatory system should act in accordance with the powers outlined in these Guidelines.

130.Use of automated tools

Paragraph 88

Digital platforms should be able to explain to the regulatory 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 election-related content. In case of failure to comply with this provision, the regulatory system should act in accordance with the powers outlined in these Guidelines (see paragraph 46(f)). User appeal and redress⁸⁹. There should be an effective user complaints mechanism to allow users (and non-users if impacted by specific content) meaningful opportunities to raise their concerns. This should include a clear, easily accessible, and understandable reporting channel for complaints, and users should be notified about the result of their appeal.²³ See United Nations 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,” para. 4. See also General comment No. 25 (2021) on children’s rights in relation to the digital environment.

131.User appeal and redress

Paragraph 89

There should be an effective user complaints mechanism to allow users (and non-users if impacted by specific content) meaningful opportunities to raise their concerns. This should include a clear, easily accessible, and understandable reporting channel for complaints, and users should be notified about the result of their appeal.

132. Paragraph 90

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.

133. Paragraph 91

Digital platforms should notify users and explain processes for appeal when their content is removed or expressly labelled, restricted in terms of comments or re-sharing or advertising association, 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 that action on their content was taken, the method used (algorithmic or after human review), and under which platform rules action was taken. Also, they should have processes in place that permit users to appeal such decisions.

134. Principle 5. Platforms conduct human rights due diligence

Human rights safeguards and risk assessments

Paragraph 92

Digital platforms should be able to demonstrate to the regulatory system the systems or processes they have established to ensure user safety while also respecting freedom of expression, access to information, and other human rights.

It is important that this does not hinder each platform’s own rules.

135. Paragraph 93

Platforms should conduct periodic risk assessments to identify and address any actual or potential harm or human rights impact of their operations, based on the provisions of Article 19 of the ICCPR and drawing on the principles set out in the UN Guiding Principles on Business and Human Rights.

136. Paragraph 94 a.

Apart from periodic assessments, risk assessments should also be undertaken:

a. Prior to any significant design changes, major decisions, changes in operations, or new activity or relationships;

137. Paragraph 94 b.

Apart from periodic assessments, risk assessments should also be undertaken:(b) To protect the exercise of speech by minority users and for the protection of journalists and human rights defenders

138. Paragraph 94 c.

Apart from periodic assessments, risk assessments should also be undertaken:

(c) To help protect the integrity of electoral processes

139. Paragraph 94 d.

Apart from periodic assessments, risk assessments should also be undertaken:

(c) In response to emergencies, crises, or conflict or significant change in the operating environment.

140. Paragraph 95

Digital platforms should be open to expert and independent input on how these assessments are structured.

141. Paragraph 96

Platforms can create spaces to listen, engage, and involve victims, their representatives, and users from minorities to identify and counter illegal content and content that risks significant harm to democracy and the enjoyment of human rights, to identify opportunities and systemic risks in order to then promote solutions and improve their policies. Consideration should be given to the creation of specific products that enable all relevant groups to actively participate in the strengthening of counter-narratives against hate speech.

The paragraph presents an assumption of the existence of identifiable victims where sometimes harm might be abstract or diffuse. There is no clarity to what “products” are in this paragraph.

142. Specific measures to fight gendered disinformation and online gender-based violence

Paragraph 97

There is considerable evidence that women in public life—including politicians, journalists, and public figures—are targeted by disinformation, fake stories, sexual harassment and threats, and incitement to violence. While some of these instances may be the result of individuals, others are the result of deliberate campaigns designed to undermine women’s participation in civil and political life, to undermine their trustworthiness, or simply drive them off the digital platform and deny their right to freedom of expression. This phenomenon is even more marked for women from racial or other minority groups. Such disinformation can all too often lead to gender-based violence. This represents a significant erosion of women’s human rights.

143. Paragraph 98 a.

To fight gendered disinformation and online gender-based violence, digital platforms should:

(a) Conduct annual human rights and gender impact assessments, including algorithmic approaches to gender-specific risk assessment, with a view to identify the systemic risks to women and girls and to adjust regulations and practises to mitigate such risks more effectively.

144. Paragraph 98 b.

To fight gendered disinformation and online gender-based violence, digital platforms should:

(b) Use privacy-enhancing technology to provide external researchers access to internal data of platforms to help identify algorithmic amplification of gendered disinformation, gender-based harassment, hate speech, and toxic speech.

145. Paragraph 98 c.

To fight gendered disinformation and online gender-based violence, digital platforms should: (c) Create dedicated engineering teams that are made up of both men and women who are specifically trained to develop algorithmic solutions to different forms of gendered disinformation, including violent and other forms of toxic speech and harmful, stereotypical content.

146. Paragraph 98 d.

To fight gendered disinformation and online gender-based violence, digital platforms should: (d) Develop and launch inclusive structured community feedback mechanisms to eliminate gender bias in generative AI and generative algorithms producing content that perpetuates or creates gendered disinformation or harmful or stereotypical content.

147. Specific measures for the integrity of elections

Paragraph 99

While electoral bodies and administrators need to ensure that the integrity of the electoral process is not affected or undermined by disinformation and other harmful practices, digital platforms should have a specific risk assessment process for any election event. Such risk assessments should also consider the users, the level of influence that advertisement messages may have on them, and the potential harm that may come out of such messages if used against specific groups, such as minorities or other vulnerable groups.

The requirement for a risk assessment for “any election event” is very broad, and does not acknowledge the large number of election events that can take place in different jurisdictions. As with other “risk assessments” in the Guidelines, these assessments, if not properly given baselines, can amount to a formal requirement that does not properly identify risks.

148. Paragraph 100

Within the assessment, digital platforms should review whether political advertising products, policies, or practices arbitrarily limit access to information for citizens, voters, or the media, or the ability of candidates or parties to deliver their messages.

149. Paragraph 101

Digital platforms should also engage with the election’s administrator/regulator (and relevant civil society groups), if one exists, 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 other relevant independent regulators maybe necessary according to the particular circumstances of each jurisdiction.

We highlight that there are different powers involved in different entities in election processes, including those that manage and administer processes, and those that adjudicate results or review tallies. Which of those has which powers may change from jurisdiction to jurisdiction. This paragraph, as well as the whole section, deserves separate, focused discussion and is likely to merit a split from the general document.

150. Paragraph 102

Digital platforms that accept political advertising should clearly distinguish such content as advertisements and should ensure in their terms of service that to accept the advertisement, the funding and the political entity are identified by those that place them.

151. Paragraph 103

The platform should retain these advertisements and all the relevant information on funding in a publicly accessible library online.

152. Specific measures in emergencies, conflict, and crisis

Paragraph 104

As a human rights safeguard, digital platforms should have risk assessment and mitigation policies in place for emergencies, crises, and conflict, and other sudden world events where content that risks significant harm to democracy and the enjoyment of human rights is likely to increase and where its impact is likely to be rapid and severe. In the case of emerging conflicts, digital platforms should be alert to this type of content, which has in many instances fuelled or even driven conflict. Measures such as fact-checking content related to the crisis should be considered.

153. Paragraph 105

Risk assessments may require digital platforms to have processes in place for cases in which a large number of simultaneous requests for action by users are made, as sometimes happens in the context of social unrest or massive violations of human rights.

154. Conclusion

Paragraph 106

Digital platforms have empowered societies with enormous opportunities for people to communicate, engage, and learn. They offer great potential for communities in social or cultural vulnerability and/or with specific needs, democratizing spaces for communication and opportunities to have diverse voices engage with one another, be heard, and be seen. But due to the fact that key risks were not taken into account earlier, this potential has been gradually eroded over recent decades.

This should be part of explanatory documents and not in the Guidelines themselves.

155. Paragraph 107

The goal of these Guidelines is to support the development and implementation of regulatory processes that guarantee freedom of expression and access to information while dealing with illegal content and content that risks significant harm to democracy and the enjoyment of human rights. They aim to enrich and support a global multistakeholder shared space to debate and share good practices about digital platform regulation; serve as a tool for all relevant stakeholders to advocate for human rights-respecting regulation and to hold government and digital platforms accountable; add to existing evidence-based policy approaches that respect human rights, ensuring alignment where possible; and contribute to ongoing UN-wide processes.

This should be part of explanatory documents and not in the Guidelines themselves.

156. Paragraph 108

The Guidelines were produced through a multistakeholder consultation process that began in September 2022. The present draft Guidelines will be the basis for the dialogue taking place during the Internet for Trust Global Conference.

This should be part of explanatory documents and not in the Guidelines themselves.

157. Paragraph 109

Consultations will continue in the following months to seek a wide diversity of voices and positions to be heard around this complex issue that requires immediate action to protect freedom of expression, access to information, and all other human rights in the digital environment.

We urge UNESCO to conduct efforts of transparency and clarity regarding the consultation process.