



## **Access Now Submission to the UNESCO Consultation on “Internet for Trust - Towards Guidelines for Regulating Digital Platforms for Information as a Public Good”**

### **Guidance for regulating digital platforms: a multistakeholder approach Draft 1.1 - Access Now Comments**

20 January 2023

#### **Introduction**

Access Now welcomes the opportunity to participate in the UNESCO consultation on “Guidance for regulating digital platforms: a multistakeholder approach” (Guidance). Access Now, a United Nations (UN) Economic and Social Council (ECOSOC) accredited organization, routinely engages with the UN in support of our mission to extend and defend digital rights of individuals and communities at risk around the world.<sup>1</sup> Since our founding in 2009, we monitor the abuse and misuse of new and emerging technologies in ways that threaten the realization of fundamental human rights, including freedom of expression and opinion.

This submission serves as our procedural and substantive comments on the Guidance.<sup>2</sup> At the outset, we commend UNESCO for its ongoing efforts to advance freedom of expression and access to information, online and offline, and for recognizing the impact that digital platforms and their governance wield on our fundamental rights. Norms, laws, and regulations governing digital platforms require review for their effectiveness and relevance in light of the rapid spread of new and emerging technologies, and multilateral institutions such as UNESCO, in partnership with civil society and other stakeholders, can play an important role in these reforms.

Our inputs consist of two sections. First, we provide comments on UNESCO’s procedure for engaging with civil society. Second, we provide direct edits to the text’s second draft along with commentary. It is important to note that while this submission draws upon examples from various regions worldwide, these examples are non-exhaustive, and do not represent the lived experiences of all persons and communities at risk. The submission therefore calls for the need to further examine the impact of these emerging issues in greater detail and in consultation with communities at risk.

#### **1. Comments on the procedural aspect of the UNESCO consultation**

1. We recognize the efforts that UNESCO has made with the intention to create a global multi-stakeholder space for the debates on regulation, co-regulation and self-regulation of digital platform services. However, we would like to note the absence of meaningful regional representation and consultation. To our knowledge, there have been only a few selective regional consultation meetings with civil society organizations. In order to create

---

<sup>1</sup> Access Now, About Us, 2021, available at <https://www.accessnow.org/>. As a grassroots-to-global organization, we partner with local actors to bring a human rights agenda to the use, development, and governance of digital technologies, and to intervene where technologies adversely impact our human rights. By combining direct technical support, strategic advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age.

<sup>2</sup> See Draft 1 of the “Guidelines on Regulating Digital Platforms: a multistakeholder approach,” available at: <https://www.unesco.org/en/internet-conference/guidance-regulation>.

a shared and rights-respecting global regulatory guidance on this thorny and multi-faceted issue, it is imperative that UNESCO holds a series of meaningful and comprehensive consultations with various regional stakeholders, particularly from regions in the Global South which are often disproportionately impacted by the over-regulation, or inadequate governance, of digital platforms.

2. Firstly, we would like to highlight the lack of clarity and transparency of the consultation process which resulted in the first draft of the “Guidance for regulating digital platforms: a multistakeholder approach.” Since the start of the process, we have not observed sufficient publicly available information or guidelines on UNESCO’s stakeholder engagement and consultation. To date, it remains unclear which and how many stakeholders UNESCO has engaged with during this process and who contributed to the first draft of the Guidance (e.g. civil society, private companies, academia, and States), the number of participants, their respective countries or regions, and the comments they provided.
3. Language should not be a barrier to participation in the consultation process. The Guidance appears to be only available in English and French, despite promises for the Spanish. UNESCO should have sufficient capacity to incorporate comments in different languages and from different contexts. It remains unclear how UNESCO plans to evaluate and incorporate stakeholder comments and feedback. Comments should be analyzed by people who are fluent in the language in which they are submitted and are familiar with the regional context of those submitting them.
4. Secondly, it is important to highlight that during the consultation process, situations have been encountered that limited the participation and representation of civil society from the Global South. We participated in a consultation in Latin America, however it is unknown what criteria were taken into account to invite civil society organizations or academic scholars to participate in providing comments (oral and written), and if any of these were or will be incorporated.
5. During the Latin American regional consultation, for instance, participants questioned the short timeframe given to make any substantive contributions to this complex subject. They also raised concerns about the lack of clarity around the potential outcomes of the regional consultation, and whether their contributions will be incorporated into the UNESCO’s process and its upcoming in-person conference in Paris. It should be noted that in the Latin American context, there is strong criticism of the creation of regulations to determine the moderation of content. **Priority should be given to safeguarding the right to freedom of expression against any guideline that could be abused to violate this right.**
6. Furthermore, actors representing important positions on these issues in the Global South have raised many questions about the lack of regional and local representation in the Paris consultation, where it is unknown who is going and why, especially in view of the challenges faced by organizations and academia from the Global South to travel to UNESCO headquarters in Paris, where they may not have enough time and resources to make their voices heard in the framework of the face-to-face meeting.
7. Thirdly, the creation of guidance which aims to be global in scope and applicability requires a diligent assessment of online information ecosystems, the current regulatory models in various jurisdictions, and salient human rights risks associated with different content governance models including self-regulation, co-regulation, and state regulation. In authoritarian and non-democratic contexts, language about state regulation of digital platforms might open the door wide open for abuse, and encourage further restrictions on

the right to freedom of opinion and expression, the right to privacy, and a host of other fundamental rights and freedoms enabled and enjoyed through the use of social media platforms and other online services. The current Guidance draft fails to encompass these regional considerations and acknowledge the risks of suggesting state regulation in non-democratic governance systems that lack strong checks and balances, separation of powers, and independent oversight. Therefore, regional consultations will serve as a robust “reality check” for UNESCO’s proposed regulatory guidelines, their feasibility, and risks for potential abuse in fragile democracies or non-democratic contexts.

8. Given the serious implications of vague and imprecise guidance on regulatory frameworks on the enjoyment of human rights online, in addition to the restricted mandate of UNESCO to oversee state obligations under the International Covenant of Civil and Political Rights (ICCPR), we recommend that UNESCO closely consults and coordinates with the UN Human Rights Office (OHCHR) and engages with existing expert guidance and initiatives including the B-Tech Project, the Rabat Plan of Action, and Istanbul Process.
9. Finally, we welcome the reference made by UNESCO to the Santa Clara Principles on Transparency and Accountability in Content Moderation,<sup>3</sup> and invite UNESCO to further consult civil society recommendations on content moderation including the recently published joint document, “Declaration of principles of content and platform governance in times of crisis.”<sup>4</sup>

## **2. Substantive comments on the text of the Guidance for regulating digital platforms for public good**

### **Guidance for regulating digital platforms: a multistakeholder approach Draft 1.1 (December 2022)<sup>5</sup>**

#### **Access Now Direct Edits to Text and Accompanying Commentary**

*Disclaimer: This document presents the overall UNESCO initiative and an initial draft proposal of the guidance document for regulating digital platforms. A further developed draft of this document will be circulated by the end of January 2023 ahead of the Global Conference Internet for Trust, which will provide a space for debate about the broader issues behind the paper, the proposals themselves, and future actions. A glossary of key concepts, based on international human rights law, UN system resolutions and declarations, and other soft law documents produced by the Universal System of Human Rights, is being developed and will be added in the next version of this draft.*

All comments should be sent to the email: [internetconference@unesco.org](mailto:internetconference@unesco.org), mentioning the specific paragraph number the comment refers to.

#### **COMMENTARY**

- a. Despite the disclaimer that a glossary of key concepts is being developed (to be added in the next version of this project), it is still important that the next draft be discussed in light of the glossary, as the lack of definitions has generated too many vague definitions, a situation that has raised a large number of comments.

<sup>3</sup> The Santa Clara Principles on Transparency and Accountability in Content Moderation, 2018, available at: <https://santaclaraprinciples.org/>

<sup>4</sup> Access Now, Declaration of principles for content and platform governance in times of crisis, November 2022, available at: <https://www.accessnow.org/cms/assets/uploads/2022/11/Declaration-of-principles-for-content-and-platform-governance-in-times-of-crisis.pdf>

<sup>5</sup> Retrieved from: <https://unesdoc.unesco.org/ark:/48223/pf0000384031.locale=en>

- b.** Additionally, the next consultation should have sufficient representation from civil society and academia at the regional and local levels, as well as provide sufficient time and transparency to ensure clarity in the procedure.

## Introduction

- 1.** UNESCO is developing, through multistakeholder consultations, a guidance document for actors seeking to regulate, co-regulate and self-regulate digital platforms, with the aim of supporting freedom of expression and the availability of accurate and reliable information in the public sphere, while dealing with content that potentially damages human rights and democracy. The scope of this guidance covers digital platform services that can disseminate users' content to the wider public, including social media networks, search engines and content sharing platforms. While this guidance is developed for those platforms whose services have the largest size and reach, minimum safety requirements should be applied to all platform service companies regardless of size.
- 2.** This document aims to provide high-level guidance for those Member States and other relevant stakeholders that are considering how to regulate [user-generated] online content online. It sets standards to help them in the development of legislation and policies that are consistent with international human rights standards, and which enhance the availability of accurate and reliable information in the public sphere. It also intends to serve as a guidance for co-regulatory and self-regulatory processes, as well as a concrete tool for a process of checks-and-balances, through which ~~companies~~, civil society organizations, academics, the technical community, and journalists can hold platforms accountable ~~the players in charge of regulating, co-regulating and self-regulating this space~~.

## COMMENTARY

- a.** Para 2. The term “*minimum safety requirements*” generates uncertainty due to the absence of a notion of its scope and what it consists of. We understand that this term could be included in the glossary.
- b.** Para 1. “*The availability of accurate and reliable information in the public sphere*” is a vague definition and lacks clarity: the scope is poorly designed. The guidance seeks to regulate very large platforms but it also seeks to establish minimum safety requirements to other “platform services”. While this distinction exists in the DSA legal framework, it is not clear what services the guidance wants to address.
- c.** Para 2. “*The players in charge of regulating, co-regulating and self-regulating this space*” is an overly broad and confusing phrase and lacks clarification as to who the players referenced are. As written, the text seems to call on private actors to regulate potentially legal but harmful content. It also implies companies should hold regulators accountable. Private actors should not be in charge of regulating potentially legal but harmful and illegal content online.

## Why UNESCO?

3. *UNESCO has a global mandate to promote the free flow of ideas by word and image. As part of the Organization's Medium-Term Strategy for 2022-2029 (41 C/4), Strategic Objective 3 is to build inclusive, just and peaceful societies by promoting freedom of expression, cultural diversity, education for global citizenship and protecting heritage. Strategic Objective 4 is to foster a technological environment in the service of humankind through the development and dissemination of knowledge and skills and the development of ethical standards.*
4. *The development of guidance for Members States (including a diverse range of public entities, among which may be different types of independent regulators) and digital platform services themselves, to secure information as a public good, contributes to all five of UNESCO's functions as a laboratory of ideas, a clearing house, a standard setter, a catalyst and motor for international cooperation, and a capacity-builder.*
5. *More specifically, the development of guidance for the regulation of digital platform services builds on the Organization's work in the domain of broadcast regulation developed over several decades.*
6. *This guidance for regulation of digital platform services **focuses on the structures and processes** to help users have a safer, critical, and self-determined interaction with online content, dealing with content that is potentially damaging democracy and human rights, while supporting freedom of expression and the availability of accurate and reliable information in the public sphere.*
7. *This guidance will:*
  - 7.1 *Take forward the Windhoek+30 Declaration on Information as a Public Good, as it calls on all parties to mainstream media and information literacy, as well as promoting increased transparency of relevant technology companies and media viability, principles unanimously endorsed by UNESCO's Members States during its 41<sup>st</sup> session of its General Conference.*
  - 7.2 *Create a multistakeholder global shared space for the debates on regulation, co-regulation and self-regulation of digital platform services, through an inclusive consultative process and research ahead of the conference.*
  - 7.3 *Enable a network of regulators and regulatory systems to draw upon this guidance, and facilitate the creation of an international community of practice, capable of exchanging good practices on how to approach regulation of digital platform services to secure information as a public good while while providing adequate safeguards for human rights protection, in particular respect for the inherent dignity, right to privacy, right to freedom of expression, non-discrimination and equality (...) protecting freedom of expression and other human rights.*

**7.4** *Serve as an advocacy and accountability tool for all the relevant stakeholders, who will be able to advocate for user-centric smart regulation aligned with human rights, where it is missing, and to hold relevant players (parliaments, regulators, companies) accountable, guaranteeing that any regulatory, coregulatory and self-regulatory measures discussed and implemented are in line with international human rights standards.*

**7.5** *Offer inputs to “Our Common Agenda”, including the Global Digital Compact and the UN Summit of the Future to be held in September 2024.*

**7.6** *Feed into discussions about the upcoming 20-year review in 2025 of the World Summit on the Information Society (WSIS) and the review of the Internet Governance Forum (IGF).*

**7.7** *Build on and gain insights from the work linked to the development and implementation of the UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted by the UNESCO General Conference in November 2021, and particularly regarding its guidance on digital platforms and online media under the Policy Area for Communication and Information.*

#### **COMMENTARY**

- a.** Para 6. “*content that is potentially damaging democracy*” is terminology that is vague and unclear. It leads to confusion as to whether UNESCO intends to address potentially harmful but legal content or whether the Guidelines also include illegal content online. In this regard, it is essential to clarify the scope of the Guidance. in order to prevent its use as a justification of abuse and illegitimate restrictions of human rights by undemocratic governments Moreover Different regulatory approaches should apply to illegal content in contrast to potentially harmful but legal content.
- b.** Para 7.1. “*relevant technology companies*” inconsistent terminology - who are relevant technology companies? How are they defined or selected by the Guidance? This text would strongly benefit from a clearly defined scope that is rather superficially described and incoherent. We encourage UNESCO to clarify the scope of applicability, particularly the criteria applied to determine what constitutes “relevant technology companies.” For example, will this apply to platforms primarily used by individuals for user-generated content? Is there a threshold for annual revenue and monthly active users? Further, we understand the [Windhoek+30 Declaration on Information as a Public Good](#) uses the terms “technology companies” and “media viability.” However, this document does not provide a definition of the term. Furthermore, the term is inconsistent with the terminology used throughout the Guidelines and does not have a definition. We encourage UNESCO to standardize the terminology used throughout the document and to define the terms in the glossary.
- c.** Para 7.2 “*a network of regulators and regulatory systems*” There is a lack of understanding of what regulators and regulatory systems UNESCO has in mind. Regulatory frameworks tackling internet intermediaries are very diverse, depending on jurisdictions. By

regulatory systems, do they mean intermediary liability laws? This also underlines the lack of data protection and competition law elements in the scope of the guidance. Intermediary liability law, adequate data protection legal standards as well as competition law are mutually enforceable and indispensable for meaningful accountability of platforms and their surveillance-based advertising business models.

- d. Para 7.3 “*protecting freedom of expression and other human rights language*” the human rights language must be reinforced.

## Independent Regulation

8. **Online User-generated content online** represents a new regulatory challenge that many actors, including states, are struggling to deal with. Existing regulatory systems vary from country to country. In some jurisdictions, there may be an existing broadcast regulator which is being granted new powers over digital platform regulation. In other states a new regulator may be established to regulate online content. There are other cases in which more than one regulatory body or institution oversees these issues. taking into account the wider implications of digital content for our societies. For instance, there are contexts where we have audio-visual, electoral, telecom, data protection regulators that deal with different aspects of the digital platform’s services. This is why this text is using the concept of a regulatory system. Whichever is the case, this guidance outlines the importance of establishing the independence of the regulatory system, however constituted, as well as ensuring that regulators have the necessary skills.
9. In 2006, the World Bank published its *Handbook for Evaluating Infrastructure Regulatory Systems*<sup>6</sup> in which it says the following about independent regulation:
  - 9.1. “The key characteristic of the independent regulator model is decision-making independence. This means that 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).”
10. In a guiding document commissioned by UNESCO (2016), the expert on broadcasting independent regulatory systems, Eve Salomon, highlighted:
  - 10.1 “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

---

<sup>6</sup> World Bank Handbook for Evaluating Infrastructure Regulatory Systems, p.50  
<http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-6579-3>

*unwarranted dismissal) is better placed to act impartially in the public interest and to avoid undue influence from political or industry interests. This ability to operate impartially is vital to protect freedom of expression, which is necessary in a functioning democracy. Independence is also required for the proper operation of all of the major functions of broadcasting regulation, including licensing, applying content standards and positive content obligations, and ownership and competition regulation.”<sup>7</sup>*

- 11.** *This guidance is principle-based – with the regulator (or the regulatory system when there is more than one regulatory entity or body), setting the overall goal for regulation which the digital platform services must fulfill.*
- 12.** *Importantly, this guidance recommends that any regulatory system focuses on the structures and processes that services use to make content available, rather than seeking to intervene in actual content decisions.*
- 13.** *In addition to setting out the primary regulatory goal which is to support freedom of expression and the availability of accurate and reliable information in the public sphere while dealing with content that damages human rights and democracy, the guidance suggests that the regulator or the regulatory system could specify a number of issues that the digital platform services should address when reporting. These are currently set out in several separate sub-items<sup>8</sup> –. The guidance goes on to set out the constitution, powers, and scope of the regulatory system.*

#### **COMMENTARY**

- a.** Para 8. *“online content.”* Online content can also be sponsored content, for instance for the purpose of political advertising.
- b.** Para 8. *“there are contexts where we have audio-visual, electoral, telecom, data protection regulators that deal with different aspects of the digital platform’s services.”* The push for regulating online content can be problematic in non-democratic contexts and might open the door to further digital repression and state-sanctioned abuse. In particular contexts, such as in parts of the Middle East and North Africa, Latin America, and elsewhere, the lack of a regulatory framework for online platforms has actually allowed users to more robustly exercise their rights online, whereas other sectors such as media and audio-visuals have been heavily regulated, and consequently restricted, by the state.
- c.** Para 13. *“primary regulatory goal, which is to support freedom of expression and the availability of accurate and reliable information in the public sphere while dealing with content that damages human rights and democracy.”* This requires further clarification and refinement. As written, the text implies the guidance does not go beyond freedom of expression. But then in Paragraph 15.1 the guidance exceeds this scope to include “other human rights.” We encourage UNESCO to include the right to privacy, freedom of peaceful assembly and of association and the right to non-discrimination within the scope of these guidelines.

<sup>7</sup> <https://unesdoc.unesco.org/ark:/48223/pf0000246055>, see also See also Principle 17 of Declaration of Principles on Freedom of Expression and Access to Information in Africa <https://www.achpr.org/legalinstruments/detail?id=69>

<sup>8</sup> 10 in total, although this may change as the result of further consultations



## **The benefits of this guidance**

**14.** *The guidance should be helpful to a range of stakeholders: for policymakers in identifying some of the stakeholders, objectives, principles, processes and procedures that might be considered in legislation; for regulators in implementation; for companies in their policies and practices; and for other stakeholders in their advocacy and accountability efforts.*

**15.** *Specifically, this guidance will:*

**15.1** *provide guidance in developing regulation that can help Member States address moderation and curation processes of content that potentially damages democratic discourses and structures and human rights, while protecting freedom of expression and other human rights;*

**15.2** *provide guidance on the strengthening and constitution of an independent regulator(s) or regulatory system;*

**15.3** *facilitate the ability of digital platforms to align to a common framework and develop coherent systems across regions to minimise internet fragmentation to protect freedom of expression and enhance the availability of accurate and reliable information in the public sphere*

**15.4** *help develop regulatory stability and a more coherent global governance (which would benefit the development of a more diverse range of companies and local economies) and help Member States be “future ready” anticipating new challenges;*

**15.5** *support platforms by providing practical and implementable guidance with a view to realizing the regulatory goal.*

### **COMMENTARY**

- a. Para 15.1. “while protecting freedom of expression and other human rights” We encourage UNESCO to introduce the concept and expand the scope of the guidance to include “other human rights” consistently throughout the guidance. The phrase “other human rights” does not appear consistently throughout the guidance and is first mentioned in Para 15.1. Please see the comment above related to Paragraph 13.

## **The proposed Guidance for regulating digital platforms: a multistakeholder approach**

### **Section One - The goal of regulation**

**16.** *The objective of this guidance is to protect freedom of expression and enhance the availability of accurate and reliable information in the public sphere, while dealing with content that potentially damages human rights and democracy.*

**17.** *The guidance sets out how the regulatory system can oversee the conduct of digital platform services in respect to content issues.*

- 18.** *The guidance also outlines government [obligations] to be transparent and accountable about the requirements they place upon digital platform services, particularly regarding the alignment with international human rights standards. For example, governments should be open, clear, and specific about the type and volume of requests they make to companies to remove and block content. In the case of sensitivities about publicising these requests – for instance, content relevant to national security or the prevention of serious crime – then the regulator or regulatory system should be able to examine, based on human rights standards, the validity of such requests and be able to report publicly on their findings and actions. The regulator should also be able to scrutinize the scope of requests to ensure adequate balance between illegality and freedom of expression.*
- 19.** *This will also require finding a means to deal with the potentially harmful content that may damage democracy and human rights – current examples include hatred of defined groups; incitement to violence; harassment; mis- and disinformation; and hostility directed at women, racial and minorities, human rights defenders or vulnerable groups - while protecting international standards of freedom of expression. But we should recognise that new dangers may arise that are not foreseen now, and that any regulation to protect human rights must be flexible enough to adapt to new or changing circumstances.*
- 20.** *Finally, this guidance shows that this goal can be achieved only if there is a cooperation among the companies providing services and the regulatory systems, while being effective and implementable and providing real accountability.*
- 21.** *This guidance for regulation will be based on five key principles: platform, policies and operations need to be (1) human rights-based, (2) transparent, (3) empowering, (4) accountable and (5) verifiable, to help ensure:*
- 21.1 Platforms have content governance policies and practices consistent with human rights standards,** *implemented algorithmically or through human means(with adequate protection for the well-being of human moderators);*
- 22.2 Platforms are transparent,** *being open about how they operate ~~(taking into account commercial confidentiality)~~ with policies being explainable;*
- 22.3 Platforms empower users** *to use digital services in a self-determined and empowering manner, including being able to assess the quality of information received and how their data is used to recommend content;*
- 22.4 Platforms are accountable** *to users, the public, and regulators in implementing terms of service and content policies, including ~~giving recognizing and facilitating~~ users rights to effective ~~of~~ redress against content-related decisions and other safeguards of procedural fairness;*

**22.5 There is independent oversight** and assessment of the impact that regulation has on companies' rules and practices, with a view to adjusting regulation to more effectively protect human rights and information as a public good.

#### COMMENTARY

- a. Para 16. *“protect freedom of expression.”* The guidance should go beyond freedom of expression and places more emphasis on the human rights legal framework as a whole. There are other significant rights negatively impacted by platforms' conduct and their business model.
- b. Para 16. *“content that potentially damages human rights and democracy.”* This is a vague notion that may include various categories of user-generated content. We recommend to UNESCO to reassess their approach of singling out concrete categories of user-generated content and instead, focus on platforms' systems and processes and systemic societal risks they impose to peoples' human rights. Para 17 *“to content issues.”* Again, there is a significant problem with the scope of the guidance. Is it transparency measures, interface design, content curation and content moderation (or both), due diligence etc.? This needs to be properly classified.
- c. Para 18. *“government.”* First and foremost, the governments have positive and negative obligations to protect human rights. This should be underlined in the guidance from the very start.
- d. Para 18. *“adequate balance.”* It seems that the guidance suggests that the balancing of rights exercise should be done solely by online platforms themselves, including the assessment of content's legality, which constitutes an unacceptable delegation of governmental functions that is at odds with fundamental rights. However, the guidance fails to place emphasis on the need for public scrutiny.
- e. Para 19. *“to deal with the potentially harmful content.”* The guidance promotes the content based approach to regulation, i.e. combating concrete categories of user-generated content instead of focusing on systems and processes deployed by private actors, safeguards of procedural justices, meaningful transparency and due diligence. This is shortsighted approach that will have a detrimental impact on human rights.
- f. Para 19. *“current examples include hatred of defined groups; incitement to violence; harassment; mis- and disinformation; and hostility directed at women, racial and minorities, human rights defenders or vulnerable groups - while protecting international standards of freedom of expression.”* This is a mix of a wide range of user-generated content, some of it illegal and some of it potentially harmful but legal categories. Each of them requires a different regulatory approach and from an international human rights law perspective, they should not be approached from the same regulatory angle.
- g. Para 19. *“must be flexible enough”* Therefore, the guidance should promote regulation of platforms' systems and processes, focusing on due diligence and other procedural safeguards. This way, UNESCO could build on existing in-house expertise, such as the UN B-tech project and other work of the OHCHR in this area.
- h. Para 20. *“among the companies providing services and the regulatory systems.”* This notion

is too vague. There are instances of non-transparent cooperation and ongoing regulatory pressure by governments on platforms, including the non-transparent pressure and removal requests by Internet Referral Units and corporate overcompliance with state sanctions. The criteria for human rights compliant and transparent cooperation with public authorities are currently missing in the text.

- i. Para 22.2. “(taking into account commercial confidentiality).” Legally mandated transparency criteria should not be subject to any exemptions based on commercial confidentiality or trade secrets. Such exemption would make those measures weaker. For instance, the EU DSA framework does not include exemption based on trade and commercial secrets elsewhere in transparency measures. The only time this notion appears in the text is in the data access framework (article 40) that goes beyond transparency requirements.
- j. Para 22.5. The creation of an “independent oversight” is only applicable in democratic governance systems. Any similar body created in authoritarian contexts could be dedicated to social media monitoring and exerting pressure on social media platforms to comply with government requests for accessing users’ data, account suspensions, and content removal.

## Section Two - Fulfilling the goal

**22.** *Before setting out the responsibilities of digital platform services in respect to the regulator, it is helpful to set out the responsibilities of governments that are considering legislation to regulate processes impacting content moderation and/or content curation, so that such legislation fulfils the regulatory goals of providing and ensuring information as a public good, while protecting people’s human rights freedom of expression.*

**23.** *Governments should:*

**23.1** *Protect and respect users’ rights to freedom of expression, the right to privacy, information, equality and non-discrimination;*

**23.2** *Respect the requirements of Article 19(3) of the International Covenant on Civil and Political Rights (ICCPR), in that any restrictions applied to content should have a basis in law, have a legitimate aim, and be necessary and proportional to the harm that is being restricted;*

**23.3** *Ensure that any restrictions are also consistent with Article 20 of the ICCPR;<sup>9</sup>*

**23.4** *Be transparent about the requests they make to companies to remove or restrict content or accounts, and be able to demonstrate how this is consistent with Article 19 of the ICCPR;*

**23.5** *Guarantee that any content removals are subjected to the adequate due process of law, including independent judicial review;*

---

<sup>9</sup> “1. Any propaganda for war shall be prohibited by law. 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law.”

**23.6** *Not impose indirect restrictions to companies (for example, internet shutdowns) for alleged or potential breaches of regulations;*

**23.7** *Not subject staff of companies to criminal penalties for an alleged or potential breach of regulations, as this will have a chilling effect on freedom of expression;*

**23.8** *Regulators with responsibilities in this area should be structured as independent regulators, with the proper accountability systems in place.*

**24.** *In turn, this guidance recommends that the regulatory system expect digital platform services to have in place structures and processes and to report to them on the following issues:*

**25. Transparency of process.** *How digital platform services fulfill the principles of transparency, explicability, and reporting against what they say they do in their terms and conditions (T&Cs) and community standards. This should include:*

**25.1** *Information about the reasons behind any restrictions imposed in relation to the use of their service being publicly available in an easily accessible format in their terms and conditions;*

**25.2** *How content is managed, including through algorithmic decision making and human review, as well as content that is being removed or blocked under either T&Cs or pursuant to government demands/requests, and information relevant to complaints about the removal, blocking, or refusal to block content;*

**25.3** *Any information about processes used by the platform to enforce their T&Cs and sanction users, as well as government demands/requests for content removal, restriction, or promotion;*

**25.4** *Any safeguards applied in relation to any content moderation that are put in place to safeguard freedom of expression and the right to information, including in response to government demands/requests, particularly in relation to matters of public interest, so as to ensure a plurality of views and opinions;*

**25.5** *How users can access the complaints process;*

**25.6** *Any use made of automated means for the purpose of content moderation, 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.*

#### **COMMENTARY**

- a.** Para 22. Platforms systems and operations have a far reaching impact on human rights far beyond freedom of expression, including right to equal treatment and right to privacy. This language should be changed across the whole text of the guidance. In general, human rights language in the recommendation 23 is very limited. As already mentioned above, States have positive and negative obligations to protect human rights of groups and individuals, including positive obligation to protect against interference by private

actors.

- b.** Para 23.6. In reference to “*indirect restrictions to companies (for example, internet shutdowns)*,” the wording and the example provided are unclear. What constitutes “indirect restrictions” of companies? And why are internet shutdowns considered as indirect restrictions of companies' operations? Throttling or shutting down digital platform services is used as a sanction against social media companies for non-compliance (see for instance Turkey’s 2020 amendments of the internet law,<sup>10</sup> and the newly adopted amendments of October 2022 known as the “censorship law”<sup>11</sup>) which we think is easily prone to abuse. It is also unclear what companies fall under this category.
- c.** Para 25. “*Transparency of process.*” We welcome that the Guidance refers to Santa Clara Principles. However, it is rather disappointing that concrete recommendations on transparency and platform processes are so vague and not granular enough. Precisely due diligence obligations and to some extent, transparency requirements are areas where international standards can be developed and promoted at international scale. The Guidance should be more ambitious here and perhaps, depart from an attempt to combat concrete categories of user-generated content. Transparency should also include information as to which sanctions regimes apply to the company, and how the platform complies with and implements the relevant sanctions, with regular disclosures of enforcement actions.
- d.** Para 25.1. “*Information about the reasons behind any restrictions imposed in relation to the use of their service being publicly available in an easily accessible format in their terms and conditions;*” This is a vague recommendation. Platforms should notify users when they take any action on their content or account(s), and provide full, clear, and accessible information on reasons behind any content moderation decisions or actions taken including the terms and conditions they violated. Users should also know if this action is taken as a result of a government request (whether through voluntary reporting mechanisms such as internet referral units (IRUs) or via legal orders). However, as worded, it is unclear how user notice on restrictions can be made public as part of the platform's terms and conditions. If the intended reference here is to platforms’ rules and guidelines on impermissible or prohibited content, then we suggest a clearer and more specific wording. Platforms should provide clear, precise, and publicly accessible information on their rules and policies, including types of prohibited content and actions that could be taken against it (such as content removal, account suspension, algorithmic demotion, break-glass measures, etc).
- e.** Para 25.2. “*How content is managed.*”The Guidance should use consistent language and clear language. “Managing content” does not specify whether it is a reference to both content moderation or content curation. Furthermore, there are hardly any recommendations specifically addressing content recommender systems, how they are optimized, what data is being used for their optimisation or direct tools that would empower users in understanding how the content is being personalized and ranked. Finally, in recent years civil society organizations and academia have put forward proposals for unbundling content recommender systems as well as requirement of

<sup>10</sup> Human Rights Watch, “Turkey: Social Media Law Will Increase Censorship,” July 27, 2020. Available online: <https://www.hrw.org/news/2020/07/27/turkey-social-media-law-will-increase-censorship>

<sup>11</sup> Human Rights Watch, “Turkey: Dangerous, Dystopian New Legal Amendments,” October 14, 2022. Available online: <https://www.hrw.org/news/2022/10/14/turkey-dangerous-dystopian-new-legal-amendments>

interoperability of large online platforms. These proposals are being reflected in the guidance in an extremely limited manner.

- f.** Para 25.3. “Any information about processes used by the platform to enforce their T&Cs and sanction users,” This point is redundant, see Para 25.1.
- g.** Para 25.4. “particularly in relation to matters of public interest,” There are no clear criteria here on what constitutes content of public interest. Who decides here what issues/content are of public interest? Does this refer to specific types of content such as journalistic content and media reporting, specific users such as HRDs and journalists, or could this be content tied to specific periods of time (such as elections, uprisings, etc)?
- h.** Para 25.6. “Any indicators.” This should be more specific to provide meaningful transparency. There are many different testing metrics to assess algorithmic content moderation models such as accuracy, precision and recall, etc. Platforms should publish data according to specific metrics that provide clear insight into their accuracy such as error rates and numbers or percentages of false positives and false negatives, including in other non-English languages. Transparency criteria should also contain unambiguous information whether the decision was made by automated decision making or by human content moderators.

### **Content management policies**

- 26.** *The content management policies of digital platform services should be consistent with the obligations of corporations under the UN Guiding Principles for Business and Human Rights, the International Covenant for Civil and Political Rights and relevant regional treaties. They should also follow best practices as expressed, for example, in the Santa Clara Principles.<sup>12</sup>*
- 27.** *Any restriction upon content posted should be clearly set out in the platform rules, which should be implemented consistently, without arbitrary distinctions made between types of content or between users.*
  - 27.1** *Platforms should, in policy and practice, through adequately trained and staffed personnel, ensure that, at a minimum, there is quick and decisive action against child sexual abuse materials, promotion of terrorism, promotion of genocide, clear threats of violence, gender-based violence and incitement to hatred based on protected characteristics.*
  - 27.2** *There is often a tension between national laws and international human rights standards, which poses a challenge for any attempt to define global guidance on regulation. Should illegal content be defined in a jurisdiction that may violate international human rights law, the platform will be expected to report on how it responds to such requests.*

---

<sup>12</sup> Any restrictions upon content being posted, or content removed, should be defined in law, have a legitimate purpose and be necessary in a democratic society and be applied proportionally.

**28.** *In addition, platforms should report on systems they have in place that would help enable them to identify the following, while protecting the right to privacy and anonymity:*

**28.1** *multiple accounts created by the same source;*

**28.2** *[false or inauthentic behaviours that promote mis / disinformation or other damaging content]<sup>13</sup>;*

**28.3** *[synthetic content designed to mislead or create a false impression (unless clearly identified as such for artistic or creative purposes)]<sup>14</sup>;*

**28.4** *the use of automated programmes designed to mimic users (bots);*

**28.5** *content created by accounts registered by state actors or otherwise credibly determined to be state-affiliated.*

**29.** *Platforms should then have explicit processes to deal with these phenomena, whether it is to label and identify such content or accounts, while protecting the right to privacy and anonymity, to restrict the virality of content arising from such accounts, or to flag with a warning that the nature of this content could be misleading or otherwise problematic (similar to splash page warnings now provided by banks before allowing transactions). The purpose for these provisions is to allow users to understand the nature and origin of questionable content and accounts and allow them to make their own judgment as to their provenance.*

**30.** *Finally, platforms should notify users when their content is removed or subject to content moderation. This will 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.*

**31. An enabling environment.** *This guidance recognises the difficulties of identifying content that is potentially damaging to democracy and human rights. For example, separating misinformation) from disinformation (is complex, bringing with it dangers to free expression of suppressing content legitimately protected under international human rights law.*

**31.1** *Platforms should show what they do to provide an enabling environment that facilitates expression, that challenges false or misleading information, warns of offline consequences to speech that might be dangerous (e.g., hate speech) or simply flags different perspectives or opinions.*

**31.2** *Where possible, users should be given the ability to control the content that is suggested to them - platforms should consider ways to encourage users' control over the selection of the content to be displayed because of a search and/or in news feeds. The availability to choose between content recommendation systems that display different sources and different*

---

<sup>13</sup> This item was highlighted by different stakeholders consulted as requiring extra safeguards for the balancing of rights, therefore we would appreciate detailed comments on how to address this point.

<sup>14</sup> This item was highlighted by different stakeholders consulted as requiring extra safeguards for the balancing of rights, therefore we would appreciate detailed comments on how to address this point.



viewpoints around trending topics should be made available to users in online platforms.

**32. User reporting.** *In supporting freedom of expression and the availability of accurate and reliable information in the public sphere, it is critical to empower users of digital platform services. In this regard, all companies, governments, civil society organisations and academic institutions have a role to play. Companies in particular, in addition to the platform providing information about its policies accessible in a unambiguous and machine readable digestible format and in all relevant languages, it should show how it allows users to report potential abuses of the policies, whether that be the unnecessary removal of content, the presence of violent or threatening content, or of any other content which is in breach of the policies. Where possible, users should have access to a platform representative in their own country.*

**32.1** *The user reporting system should give high priority to content that is threatening or intimidating, particularly to vulnerable groups ~~groups with protected characteristics~~, ensuring a rapid response<sup>15</sup> and, if necessary, by providing specific means of filing the report. This is particularly important when it comes ~~it to comes~~ to gendered online violence and harassment. A pre-set template would allow the aggregation of similar complaints that would help identify systemic failings on the platform. At the same time, this guidance recognises that much of this will depend upon local and regional contexts.*

**32.2** *There should also be an effective user complaints mechanism to allow users meaningful opportunities to raise issues of concern. This should include a clear, easily accessible and understandable reporting channel for complaints and users should be notified about the result of their appeal.*

**32.3** *There will clearly be issues of scale for platforms with large numbers of users. In such circumstances, platforms may need to deploy automated decision making processes ~~systems to process~~ and record complaints and the regulatory system will review the operation of these systems.*

#### COMMENTARY

- a. Para 27.1. “child sexual abuse materials, promotion of terrorism, promotion of genocide, clear threats of violence, gender-based violence and incitement to hatred based on protected characteristics.” Most of these listed categories of content are considered illegal by international as well as national legal order. The guidance delegates the responsibility and enforcement directly to private actors without putting forward any processes oriented oversight and public scrutiny safeguards. This paragraph also says “as a minimum, act swiftly on...” this means that more than swift action (i.e. automated filtering) might be good or acceptable in all those cases. We have argued that such extreme measure is only acceptable in cases where illegality is such that there is no possible protected use of that content (e.g. CSAM). For all other cases, upload filtering is excessive and should be clearly rejected.
- b. Para 30. “platforms should notify users when their content is removed” Based on our positioning, whether 26 Recommendations DSA position papers, we have been

<sup>15</sup> One mean could be through an escalation channel for the most egregious threats.

advocating for users' notifications prior any action against content is taken by platforms (of course, this does not necessarily include illegal content irrespective of its context, although opinions in the community differ).

- i. *“subject to content moderation.”* We do not understand this - all user-generated content is subject to content moderation.
  - ii. *“Also they should have processes in place that permit users to appeal such decisions.”* This should be a stand alone recommendation with more granular and precise demands.
- c. Para 31.1. *“show what they do to”* What does it legally mean for a company to “show what they do”?
  - d. Para 32.1. *“gendered online violence and harassment.”* It is unclear why these two specific categories of content are singled out. This requires more clarity and explanation.
  - e. Para 32.1 *“pre-set template”* It is unclear how such a pre-set template corresponds with notice and action procedures required by at least some intermediaries liability laws around the world.
  - f. Para 32.3. *“The regulatory system.”* Does this refer to the national regulatory system? If so, there are serious data protection and safety red flags with this proposal. If regulatory bodies are entrusted with overseeing the operation of the complaint mechanisms, users will be less tempted to report in fear of reprisal, especially if they believe that the platform action was taken in response to a government request or interference. Similarly, it is unclear what responsibilities and types of access regulatory powers will have in overseeing platforms' complaint mechanisms. For instance, will they have access to the complainant's personal information such as name and location? Furthermore, the guidance does not properly articulate safeguards against overboard government powers and so-called enforcement overreach. Taking down user communication is a highly intrusive act that interferes with the right to privacy and threatens the foundation of a democratic society. Ideally, only judicial authorities should be authorized to issue such orders. In many jurisdictions around the world, non-independent administrative authorities are often under the supervision of executive political power and don't necessarily consider the legitimate interests of all parties involved, including the protection of their human rights.

### **33. Content that potentially damages democracy and human rights, including mis- and disinformation and hate speech**

**33.1** *For platforms and independent regulators, attempts to identify potentially damaging content that is not manifestly illegal can be a significant challenge, as most freedom of expression legal standards emphasise the importance of context and intent – saying the same words in different contexts and in different ways can have very different legal implications. And sometimes apparently legal speech which constitutes disinformation can be deployed with the intent of causing severe harm. Different opinions and viewpoints on the potential damage posed by content will arrive at very different solutions. This is made even*

more difficult by the sheer volume of content uploaded continually across all platforms, which can feasibly be managed, at least in the first instance, mainly automatically.

**33.2** Platforms should say how they define and respond to a wider set of damaging content through a systematic risk assessment. The regulatory system should assess if platforms are consistently applying their systems and processes to effectively enforce their own standards (including the protection of legitimate speech) which should be aligned to international human rights standards.

**33.3** This guidance recognises the formidable challenge of identifying damaging speech which may be legal in one context but damaging in another. For example, it is important to distinguish between content promoting hatred directed at women, children, youth, LGBTTIQ, indigenous groups, people with disability and vulnerable communities and content that is simply offensive to a particular group of people in a particular context.

**33.4** Platforms should demonstrate how they would respond to potentially damaging speech – either by providing alternative reliable information,<sup>16</sup> flagging concerns about the quality of this information, curbing its virality or any other means. Content removal or de-platforming of users should be considered only when the intensity<sup>17</sup>, and severity<sup>18</sup> of content that has the intention to harm a group or individual occurs. The platform should also be explicit about whether it partners with outside organizations or experts to help it make these kinds of decisions particularly in countries or regions where the platform itself has little local knowledge.

**33.5** Platforms should show whether they apply specific protection measures to particular groups. If they do, these measures might include risk assessment and mitigation processes or the creation of specific products that enable these specific groups to actively participate online.

**34. Media and information literacy.** Platforms should set out the resources they make available to improve media and information literacy, including digital literacy about their own products and services, for their users. There should be a specific focus inside the company on how to improve the digital literacy of its users with thought given to this in all product development teams. The platform should be reflecting on how any product or service impacts upon user behaviour and not just on the aim of user acquisition or engagement.

**34.1** Platforms should implement specific media and information literacy measures for

---

<sup>16</sup> Several media platforms have instituted “disputed news” tags that warn readers and viewers about contentious content

<sup>17</sup> The frequency of occurrence over a given period and their range.

<sup>18</sup> Its scale, scope or irremediability. Scale means the gravity of the impact on the human right(s). Scope means the number of individuals that are or could be affected. Irremediability means the ease or otherwise with which those impacted could be restored to their prior enjoyment of the right(s).

women, children, youth and indigenous groups<sup>19</sup>.

### **35. Election integrity**

**35.1** Digital platform services should have a specific risk assessment process for any election event and should engage with the election's administrator/regulator (and relevant civil society groups), if one such exists, prior to and during an election to establish a means of communication if concerns are raised by the administrator or by users/voters. Within the assessment, they should review whether political advertising products, policies, or practices arbitrarily limit the ability of candidates or parties to deliver their messages.

**35.2** Digital platform services that accept political advertising should ensure in their terms of service that to accept the advert, the funding and the political entity are identified by those that place the adverts.

**35.3** The platform should retain these advertisements and all the relevant information on funding in a publicly accessible library online. Political advertisements which refer to issues rather than parties or candidates should be scrutinised to ensure they are consistent with the overarching policies of the platform in relation to hate speech or speech targeting people with protected characteristics.

**35.4** Digital platform services should adopt transparency measures regarding the use and impact that the automated tools they use, although not necessarily the specific codes with which they operate, may have in practice, including the extent to which such tools affect the data collection, targeted advertising, and the disclosure, classification, and/or removal of content, especially election-related content.

**36. Major events** – Digital platform services should have risk assessments and mitigation policies in place for “major events” crises such as conflicts, wars, natural disasters, health emergencies, and sudden world events where mis- or disinformation and hate speech are likely to increase and where their impact is likely to be rapid and severe.

**37. Language and accessibility.** Digital platform services operate globally, and the main language of many such platforms is English. There are over 7,000 languages spoken in the world today, though many are spoken only by small groups of people.<sup>20</sup> It is critical if regulation is to be effective that users can operate in a language that they understand. Setting a reasonable expectation for which languages platforms should be able to operate in will depend upon the scale, reach, and sensitivity of the service. For global platforms, it would be reasonable to suggest that users can contact them either in one of the six UN languages<sup>21</sup> or in one of the 10 languages spoken by more than 200 million people.<sup>22</sup> Automated language translators, while they have their limitations, can be deployed to

---

<sup>19</sup> Although this guidance is aiming to highlight the asks of the regulatory systems towards the platforms, it is important to underline that UNESCO has a series of recommendations for governments regarding media and information literacy policies, which must be implemented by relevant authorities, particularly in the education sector.

<sup>20</sup> <https://www.ethnologue.com/guides/how-many-languages>. Around 40% of languages are spoken by fewer than 1,000 speakers.

<sup>21</sup> Arabic, Chinese, English, French, Russian and Spanish

<sup>22</sup> English, Chinese, Hindi, Spanish, French, Arabic, Bengali, Russian, Portuguese and Urdu.

increase the number of languages available. Platforms may wish to ensure the provision of information in additional languages during election events, perhaps by increasing the capacity of moderation in local languages during such events. Consideration should also be given to persons with disabilities, and the ways in which they can interact with, and make complaints in relation to, the platform.

**37.1** It is recognised that this signifies an important shift in the way platforms operate, as English is a predominant international language. Nevertheless, as the major force in global communication, platforms must recognise their responsibility to allow people to communicate effectively if they are to be accountable to them.

**38. Data access.** Platforms should provide stable access, wherever safe and practicable, to non-personal data and anonymised data. Access should be provided to data that is aggregated, or manifestly made public data for research purposes through automated means such as application programming interfaces (APIs) or other open and accessible technical solutions allowing its analysis. They should provide access to data necessary to undertake research on content that is potentially damaging to democracy and human rights and support good faith research that involve their services. There need to be safeguards with providing data access that ensures the protection of privacy and respect of commercial confidentiality. For platforms to build reliable interfaces for data access, there will need to be alignment among regulators that can determine what is useful, proportionate and reasonable for research and regulatory purposes.

## COMMENTARY

- a. Para 33. “*Disinformation and hate speech.*” Placing disinformation and hate speech next to each other is not compliant with international human rights standards. They are very distinctive categories of content.
- b. Para 33.2. “*Damaging content*” is yet another vague term used in the guidance. It is essential that the guidance’s drafters ensure the consistent terminology throughout the text of the guidance. If these recommendations are addressed to regulators in numerous jurisdictions around the world, the number one priority would be to avoid vague defined terminology in legislation, including criminal liability laws, that will create the space for abuse as well as the environment of legal uncertainty.
- c. Para 33.3 and Para 33.4. “*damaging to another.*” Adding to the comment above, what constitutes “damaging speech”? The inconsistent use of this vague terminology is dangerous. States use such overly-broad and vaguely-defined terms to criminalize and restrict speech online. For instance, some wording can be found in a number of draconian cybercrime laws that criminalize freedom of expression under elastically-defined crimes such as “damaging the state’s reputation.” We recommend being extremely cautious, precise, and consistent when referring to categories of speech.
- d. Para 34. “*Content removal or de-platforming of users should be considered only when the intensity, and severity of content that has the intention to harm a group or individual occurs.*” Any restrictions to the right to freedom of opinion and expression must meet the three-part test as prescribed by international human rights law (requirements of proportionality and necessity etc.).
- e. Para 34. “*The platform should also be explicit about whether it partners with outside organizations or experts to help it make these kinds of decisions particularly in countries or regions where the platform itself has little local knowledge.*” This is a security flag. Any exposure of this sort could put human rights experts, human rights defenders, and civil society actors at serious risk of reprisal, particularly those operating in situations of armed conflicts or under authoritarian regimes.
- f. Para 34. “*or their users.*” This paragraph could emphasize here the need to localize such sources and make it available in different and accessible spoken languages.
- g. Para 35.3. “*The platform should retain these advertisements and all the relevant information on funding in a publicly accessible library online.*” Singling out political advertising instead of focusing on human rights abuse caused surveillance based advertisement as a whole will provide an incomplete response to ongoing human rights violations by private actors. We recommend that the guidance addresses data harvesting business models and pervasive targeting techniques.
- h. Para 36. “*Major events.*” Please consult the recently launched joint [Declaration of principles for content and platform governance in times of crisis](#).
- i. Para 38. “*Data access.*” This paragraph could be further clarified by underlying the importance of an adequate data protection framework. There are national jurisdictions that have no adequate data protection laws in place which are compliant with

international human rights legal standards. It raises questions and concerns, who will be vetting researchers in authoritarian regimes; what safeguards can we expect to be put in place by such governments where there is no pseudonymisation principle, purpose limitation and data minimization standards in existence. Data protection elements and especially safeguards against abuse of personal data should be included in the guidance, also in the context of data harvesting business models and its impact on content curation (content recommender systems) and ad delivery techniques.

### **Section Three- The independent regulatory system**

- 39. There are vastly different types of bodies involved in online regulation throughout the world. These 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 general 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 may exist independently of the government and others might be constituted as government agencies. It is therefore difficult to set out detailed guidance when regulation can take so many varying forms and potentially involve so many agencies.*
- 40. Nevertheless, in whatever form regulation operates, it will constitute what this guidance calls a “regulatory system” of some kind. The guidance below is therefore meant to be generally applicable to any system of regulation, however established, however varied. Of course, this guidance recognises that this approach, if adopted, could imply significant changes to the way regulation operates in some Member States.*

#### **Constitution**

- 41. Any regulatory system, whether a single body or multiple overlapping bodies, charged with managing online content (overseeing systems and processes) needs to be independent and free from economic or political pressures or any external influences. Its members should be appointed through an independent merit-based process of appointment, overseen by an oversight body (which could be the legislature or an independent board/boards). Its members should not seek or take instructions from any external body, whether public authority or private actor.*
- 42. The dismissal of members of the regulatory system should be based on clear criteria and must also be subjected to a thorough process, guaranteeing that their dismissal is not a result of any political or economic pressures.*
- 43. The members of the regulatory body/bodies should make public any possible conflict of interest.*
- 44. The regulatory system must have sufficient funding to carry out its responsibilities effectively.*

**45.** *The regulatory system should make a regular report to an oversight body on its findings and will be accountable to it. The regulatory system should also hold periodic multi-stakeholder consultations on their operation.*

### **Powers**

**46.** *Regulation will set the overarching goals for platforms to safeguard information as a public good, empowering and protecting users (particularly vulnerable users such as children or minorities) and specifying expectations as to how the stated goals should be fulfilled. It should not make judgements about individual pieces of content but will focus upon the systems and processes used by the platforms.*

**47.** *While the guidance is developed for those platforms whose services have the largest size and reach, minimum safety requirements should be applied to all platform service companies regardless of size.*

**48.** *In-scope digital platform services will be required to report regularly on how they are achieving the goals, though regulators may commission off-cycle reports if there are exigent circumstances, such as a sudden information crisis (such as that brought about by the COVID-19 pandemic) or a specific event which creates vulnerabilities (e.g., elections, protests, etc.).*

**49.** *The regulatory system will have the power to call in any digital platform service deemed not to be complying with its own policies or failing to protect users and, after discussion, may recommend a specific set of measures to address the identified failings. Any such judgment 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. Failing to comply with this stage could lead to penalties which are proportionate, dissuasive, and effective (but excluding personal criminal liability).*

**50.** *It will have the power to 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.*

**51.** *It is expected that illegal content will be removed solely in the jurisdiction where it is illegal.<sup>23</sup>*

**52.** *One option as an additional protection for users is for there to be an ombudsman for*

---

<sup>23</sup> However, it is important to recognise that no systems and processes will be 100% precise in identifying illegal content (at least not without disproportionate intrusion and monitoring). Therefore, it should not automatically be a breach of the regulations if illegal content is found on the service, unless it can be shown that the platform knew of it and failed to report it, or if the relevant systems and processes can be shown to be inadequate. Moreover, identification of illegal content should be interpreted consistently with international human rights law to avoid unjustified restrictions on freedom of expression.



complaints about platforms.<sup>24</sup> While in the first instance, complaints should be made directly to the digital platform service itself, in the event of no or an inadequate response, the user could go directly to the ombudsman. This may result in an unmanageable workload, and an alternative for digital platform services with large volumes of content is for them to have independent complaints/appeals/redress processes, which the regulatory system can then evaluate.

- 53.** Given the likely volume of complaints, the regulatory system will be expected to prioritise those complaints that demonstrate importance and relevance, systemic failings and/or substantial user harm. In this event the relevant regulator will have the power to intervene and require action, including on an interim/urgent basis if necessary.

### **Review of the regulatory system**

- 54.** There will be provision for a periodic independent review of the regulatory system, conducted by a respected third-party reporting directly to the legislature and subsequent consideration by the legislature.

- 55.** 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 will be subject to review in the courts if it were believed it had exceeded its powers or acted in a biased, irrational or disproportionate way.

- 56.** Decisions on eventual limitations of specific content should be taken by an independent judicial system, following a due process of law.

#### **COMMENTARY**

- a.** Para 40. “Nevertheless, in whatever form regulation operates, it will constitute what this guidance calls a “regulatory system” of some kind.” Given that the guidance acknowledges the varying regulatory systems around the world and the difficulty in creating a harmonised model, it is counterproductive to engage in the exercise of creating or offering guidance on a “regulatory system,” which UNESCO has no mandate or the means to ensure its (correct) implementation.
- b.** Para 41. “managing online content (overseeing systems and processes).” This is problematic and also unclear. Will regulatory bodies manage content or oversee systems? These are two different things.
- c.** Para 45. “The regulatory system should make a regular report to an oversight body.” The terminologies used here are confusing. As proposed so far in the guidance, a “regularly system” is a form of an oversight body that oversees platforms’ compliance. Does UNESCO propose setting up an additional oversight body that oversees the work of regulatory authorities. If so, what are the different roles, mandate, and governance structures of these two entities?

<sup>24</sup> The scope of complaints would be limited to a failure to comply with its regulatory duties, rather than as an additional appeal mechanism where users are unhappy with specific decisions.

- d. Para 46. “*Regulation.*” Regulatory bodies follow local and national laws of the jurisdictions they operate in. What should regulatory bodies do when the state have repressive laws that support censorship and removal of types of content deemed illegal by the state?
- e. Para 46. “*the systems and processes used by the platforms.*” While we welcome the focus on systems and processes deployed by online platforms, very few recommendations in this guidance address these systems more specifically, despite the fact that there are numerous international standards and regulatory frameworks developed in this area.
- f. Para 47. “*minimum safety requirements*” The term “*minimum safety requirements*” generates uncertainty due to the absence of a notion of its scope and what it consists of. We understand that this term could be included in the glossary.
- g. Para 48. “*In-scope digital platform services.*” The scope of digital platform services have not been defined earlier in text, which makes in-scope and out-scope references here unclear.
- h. Para 49. “*Any such judgment should be evidence-based.*” Is the regulatory system substituting courts here?
- i. Para 49. “*Failing to comply with this stage could lead to penalties which are proportionate, dissuasive, and effective (but excluding personal criminal liability).*” It should be obvious here, but any penalties should be provided by law and not only at the discretion of the so-called "regulatory system.”



**Access Now** (<https://www.accessnow.org>) defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, legal interventions, and convenings such as RightsCon, we fight for human rights in the digital age.

For more information, please contact: [un@accessnow.org](mailto:un@accessnow.org)