Derechos Digitales' comments to the UNESCO *Guidance for regulating digital platforms: a multistakeholder approach*

draft 1.1[1]

20 January, 2023

1. About Derechos Digitales

Derechos Digitales[2] 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.

2. General comments on the process

For Derechos Digitales, it is important to comment not only on this text, but also the process around its drafting and discussion. A first concern is around the transparency with which the process has been conducted. The selection of regulators, engineers, social media platforms, Internet Service Providers (ISPs), technology companies, civil society representatives, individual experts from different parts of the world, good faith whistleblowers and technology-focused journalists, to formulate input through of bilateral consultations, is relevant to understand the current content of the proposal.

What can be expected from a process of this nature, spearheaded by an international organisation such as UNESCO, would be to have the details of the individuals and stakeholders that have had the opportunity to influence the design of the proposal in its current and previous iterations, as its contents reveal specific, non-neutral visions about the identified problems and the role of the different actors involved in content moderation.

Briefly after being introduced to the Draft 1 version of this Guidance (in Spanish), Derechos Digitales participated in a closed regional consultation, in which we expressed much of the criticisms made in this document, as well as subsequently remitted writing recommendations

[1] Comments drafted by M. P. Canales, M. R. de Souza, and J. Carlos Lara. Please direct comments to info@derechosdigitales.org.
[2] https://www.derechosdigitales.org/

in Spanish to the consultation[3]. Despite this, we consider that the consultation process as a whole suffers from problems that give us great concern.

Consistent with the principle of transparency, the stages of this process would ideally have been informed to the public at the beginning of the process and not so far into its phases. Timely information could have guaranteed the possibility of a broader and more diverse participation in the initial contributions to the design of the document that are essential to define its structure. This is compounded by other limitations: the draft under discussion is presented in English and comments are expected from stakeholders, largely from the global south and including Derechos Digitales, for which English and French are not national or majority languages.[4] The timing for the comments and the UNESCO Global Conference also prevent from dedicating resources to this process.

We hope that future versions of this text take into account the possible interventions of different stakeholders prior to advancing to other consultation instances, allowing them to know in advance the aspects that will be part of an open debate. A true "multistakeholder approach" is not only one that leads to an open process to produce a guidance document, but also one that sets up participatory processes for the implementation of its recommendations too.

3. General comments on content and structure

Regarding the text and its structure, we wish to highlight that this proposal focuses on processes and structures as a positive thing, although it would be convenient for this approach to be captured by the concrete recommendations for implementation of the identified principles. We believe that this approach can allow solving the issue of the diversity of legal traditions, local regulatory frameworks, and institutional development that a proposal with a universal character faces for implementation.

A subject as impactful as content management in internet platforms, in contexts that present great variation from one country or region to another, the key to a proposal for a "regulatory framework" lies in providing enough room for local nuances or adaptations that make it possible to avert the dangers to freedom of expression and other human rights of well-intentioned regulations with unwanted side effects. We are concerned about the lack of definition of the problems identified and that the draft seeks to solve through regulation proposals, with a more global perspective in its diagnosis and the perceived usefulness and feasibility of implementation of its recommendations. An already set structure entails serious methodological challenges to the provision of feedback, restricting the substance and richness of the contributions.

---

[3] The document is available here: https://www.derechosdigitales.org/wp-content/uploads/UNESCO-Regulacion-de-plataformas-digitales-Comentarios-DD.pdf

[4] Though a Spanish version was announced in UNESCO's website, it was not available at the time of drafting these comments.

We are very concerned that a vision of outsized trust in the role of States predominates, overlooking adequate consideration for the institutional fragility in many countries, particularly in the global south. In contrast, while the focus of many recommendations is on platform regulation and self-regulation, governmental actions are not considered in similar detail or degree. In particular, with regards to governmental requests for content removal or blocking, as well as governmental requests for information from platform users, there needs to be a stronger

The draft under consultation presents a series of decisions on structure that we believe should be revised. Our general and particular comments are based on the experience of Derechos Digitales in Latin America, as we have participated in regulatory debates and evidenced the dangers in content moderation regulations that have been proposed in several countries of our region.

4. Substantive comments to the draft Guidance

Introduction

On **paragraph 1**, the draft describes the scope of the proposal, and its purpose as a "guidance document", developed "through multistakeholder consultations". We wish to emphasise again that this is not reflected in the process for this consultation, and not reflected on the shape of the recommendations and their implementation. It is also concerning that there is a specific focus on platforms of "the largest size and reach" before mentioning minimum safety requirements that might be necessary for platforms of potentially infinite diverse shape and size.

In **paragraph 2**, the document sets a difference between setting "standards" for the development of legislation and policies, and providing guidelines, oscillating between different objectives. There is an emphasis on "the availability of accurate and reliable information in the public sphere" which is concerning: all types of expression and communication happen online, in general under the protection of the fundamental right to freedom of expression, regardless of their status as "accurate and reliable information". There is a risk that emphasising that as a desired goal can have the unwanted effect of empowering actions that in the name of combating misinformation end up restricting freedom of expression and several other fundamental rights.

Why UNESCO

Paragraphs 3-5 refer to UNESCO's mandate and its purpose to provide "guidance" to States, all based on previous experience with broadcasting services.

Paragraphs 6-7 emphasise the guidance's focus on structures and processes, which is a positive element, notwithstanding its contradiction with the aforementioned emphasis on "accurate and reliable information".

What is striking is the phrase that sets the objectives in **Paragraph 6**: "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". The wording presents three different values (safety, critical thinking, self-determination), while at the same time placing them in a manner in which they risk contradicting each other, as we pointed out in our previous comments (especially between critical and self-determined).

In **paragraph 7**, the "multistakeholder" element is emphasised in relation to the Draft 1, which is a welcome emphasis but inconsistent with the lack of commitment to what it entails in practice, both for this process as for the implementation of the recommendation. Finally, there is reference to several instances of previous and current processes, including the recent Recommendation on the Ethics of AI, all added as an input for this process, yet the mentions of ongoing UN processes appear without any clear linkage anywhere between this effort and those UN initiatives. It is astonishing that, while this effort acknowledges the ongoing work of UN institutions and experts, the Guidance seems disconnected from all of them. We are deeply concerned that this isolation can impact negatively in the implementation of the Guidance.

Paragraphs 8-13 focus on "Independent Regulation" for internet platforms. The need to consider the differences between different systems and the role of different authorities in different jurisdictions is highlighted, which is welcome.

Although the "independence of the regulatory systems" highlighted by the proposal is essential, that value must be complemented with technical capacities and resources. It must also be made compatible with the necessary coordination with other institutions with jurisdiction over topics as diverse as the protection of personal data, consumer protection, access to public information, electoral regulation, telecommunications regulation, antitrust and market regulation authorities, and the protection of human rights. Setting out the regulatory power in matters of digital content services as an isolated function threatens to deprive it of the necessary coordination for the adequate protection of freedom of expression in consistency with various legal frameworks in force at the local and regional level, constitutional provisions and international obligations arising from human rights treaties.

Paragraph 13 mentions there are elements which the platforms should address when reporting and that could be specified by the regulator or the regulatory system, leaving it open for the system to determine what they would be. We are also aware that those issues were not necessarily part of "reporting" from platforms but general issues to address (in version 1 of the Spanish version), and would like to understand how the change came to be.

The benefits of the guidance

Paragraphs 14-15 describe the benefits of this guidance, in a way that differs from a previous approach in Draft 1 that mentioned a regulatory framework rather than a guidance. The target audience of this guidance audience is focused not on States alone but multiple stakeholders.

Proposed Guidance

Section One: the goal of the regulation

Paragraphs 16-21 discuss the overall goal of this Guidance, yet they do not answer basic questions about the definition of the categories of platforms on which it is promoted to apply the proposed regulatory framework, related to their differences in terms of the type of services provided, size, reach, dominance, business model, et cetera, as we have mentioned before.

In **paragraph 16**, the protection of freedom of expression is set as the first objective, and then the availability of reliable information in the public sphere, once again emphasising "information" of certain characteristics. It says nothing about ensuring plurality of views and opinions.

Paragraph 18 sets the standard according to which the regulator must evaluate the actions of the States that require actions from the platforms, referring to "human rights standards" as the basis and adding in the final sentence the scrutiny over the scope of requests between "illegality and freedom of expression". It is unclear what this "scope of requests" is referring to. The paragraph, while promoting transparency from governments, still provides a large room for opacity through recognition of "sensitivities" around governmental requests of content removal, which can provide cover for abusive requests from governments, and can prevent scrutiny on the compliance of international human rights standards as well as the Guidance's own recommendations in those requests.

Paragraph 19 highlights potentially harmful content that can affect democracy and human rights, but lists a series of examples that may constitute illegal or restricted forms of expression, in the same category as protected forms of expression. The end of the paragraph has a reference to flexible regulations to ensure the effective protection of human rights, without emphasising a co-regulation approach as in the previous draft. Though the latter revision is welcome, there is still a focus on online content as a source of harm and subject to restriction.

Paragraph 21 presents a series of principles for "platform policies and operations". The paragraph is confusing, as it is unclear whether the principles apply to the guidance itself, or

to the implementation of its recommendations, or to all regulatory efforts by either states or companies.

In any case, the enumeration in **Paragraph 21.1** now starts with "governance policies consistent with human rights standards", where implementation is allowed “algorithmically or through human means”, which also is linked to the idea of automated systems to process and register complaints (**P. 32.3**). Recent experience of ambiguous results for the protection of human rights from the use of automated systems for content moderation has been prominent in the discussions on the regulation of artificial intelligence, for example those that are reflected in the UNESCO Recommendation on Ethics of Artificial Intelligence. In this sense, the problems related to the use of artificial intelligence and with the use of algorithms are related to several factors, such as a lack of transparency about what data is used to generate content segmentation and on the definition of these criteria. Furthermore, there is little or no broad discussion about how the criteria for moderation of driven content are constructed, for example when it comes to selecting the advertising content that is displayed to female users. Another problem is related to its limitations in interpreting the context, considering that in order to make moderation decisions it is necessary to have a vision of the culture, language, and local specificities.

This makes the automated implementation of policies supposedly aligned with human rights standards highly risky of not occurring in practice. To avoid undesirable effects of the use of automated technologies, the document should reinforce: (i) the need for human supervision of automated processes, (ii) the opportunity for users to request human review of automated moderation actions , and (iii) the lack of full reliance on automated systems to remove content or suspend accounts (except in cases of in the absence of an emergency, or for the purpose of detecting spam or malware, for example).

Paragraph 22.3, in relation to empowerment mechanisms, is not clear whether this would translate into control tools that favour decision-making by users regarding how the content can be viewed or accessed, or if the reference seeks to reinforce transparency information regarding the origin of the content in order to provide elements for the development of a critical vision of the plurality of content that is accessed. It could be a combination of different mechanisms, but it must be taken into account that self-management tools can also have the effect of reinforcing echo chambers of the users, amplifying narratives that confirm unsupported beliefs, or grant more visibility to abusive behaviour. This would not necessarily contribute to the objective of information as a common good.

Paragraph 22.4. in relation to accountability, lacks a specific reference to the requirements of due process, expeditious access and in accessible language to review mechanisms regarding content moderation, as well as the obligation of platforms to substantiate and notify their content moderation decisions to the users impacted with them. It focuses on the implementation of terms of service, although their enforcement and its revision mechanisms should be accountable as well.

One last issue that is not covered by the principle in its current form and that is essential for the responsibility of the platforms in content moderation is the duty of due diligence that must be constantly developed in the form of evaluation and mitigation of risks for different groups of users, as well as conducting human rights impact assessment studies when the potential risks identified show levels of relevance and particularity that require a detailed study and mitigation recommendations adapted to the specific situation.

This entire area of accountability responsibility has been developed for technology companies in recent years on the occasion of the application of the United Nations Guiding Principles on business and human rights, and has been the particular subject of the work of the B-Tech Project led by the Office of the High Commissioner for Human Rights[5]. It would be convenient to integrate this experience into the proposed framework, as well as the issue of the obligation of proactive risk assessment that has been addressed in recent regulatory models such as the *Digital Services Act* (DSA) recently issued by the European Union.

Paragraph 22.5, regarding "independent oversight", is mostly confusing because here it refers to how regulation impacts the rules and actions of the platforms, but not the impact of platform policies and processes like the rest of the principles.

Section Two - Fulfilling the goal

Paragraphs 22-38 present a much different structure than in Draft 1, which was focused on providing detail to the implementation of the aforementioned principles. It begins by establishing in **Paragraph 23** the responsibilities of governments according to international standards with direct reference to the ICCPR, and a list of topics that partly coincide with several standards and recommendation documents from the past decade, but which are not referenced. **Paragraph 24** then launches into providing detail about the “structures and processes” that digital platform services should implement.

Paragraph 25 presents recommendations on Transparency of process. Though there is a list of requirements here, the draft lacks a detailed and systematic guide of indicators regarding what must be delivered in compliance with the principle of transparency, to whom, and for what purpose. The proposal could make a contribution by helping to standardise the requirements in this matter at a global level.

It is very different to discuss transparency about an individual case, where it would be ideal to have a notification of a content restriction decision with its justification, than to build up a public registry with that information, which can affect the privacy of the user and generate more damage to the exercise of freedom of expression and limit the right to anonymity. If what is sought is transparency for auditing or independent investigation, this can be resolved through more appropriate mechanisms than a public file. This seems to be the proposal raised in the section on "access to data" that proposes the opening of APIs for the development of

[5] Ver https://www.ohchr.org/en/business-and-human-rights/b-tech-project

research based on duly anonymized information. It would be convenient to include in the proposal a more systematic treatment of the different hypotheses within this section on transparency, and thus clarify when transparency requirements are being proposed with respect to individual cases, and when the opening is proposed for aggregate analysis processes that have other research purposes or performance evaluation of the platforms in the work of content moderation.

On the other hand, progress can also be made in the requirements of specific indicators in the transparency reports that contain the aggregate set of cases of content restriction as a way of operationalizing the principle of transparency, to make the information more granular, comparable and useful for evaluating and comparing the performance of different platforms. Today, each platform has their own reporting practices, with different criteria and limited possibilities for comparative analyses. We suggest reviewing some of the standards and criteria that have been proposed by initiatives such as the Santa Clara Principles or the documentation from the Internet & Jurisdiction Policy Network.

One last aspect of transparency is advancing in the requirements for substantiation and communication of application criteria of the moderation rules by the platforms themselves to ensure consistency in their application from the perspective of the users.

A structural revision of the section would be advisable to make it more systematic and detailed in the proposal of indicators, and to separate it from aspects that are not really conducive to meaningful transparency, such as the information necessary to present complaints that is part of the due process of the accountability mechanism accounts which are discussed later.

Paragraphs 26-30, on Content management policies, cites at the beginning important documents such as the UNGPs (in addition to ICCPR and regional treaties) and “best practices”, though without citing specific sources for those practices aside from one principles document. The Manila Principles are missing from these references.

Paragraph 27 covers basic requirements that try to provide certainty to the application of platform rules. However, a previous version not only required previsibility but adequacy to three-part test from article 19 ICCPR, a requirement that now resides in a footnote to paragraph 26.

Paragraph 27.1 refers to resources needed and the required action against certain types of content. Given the very different natures and customs of very different platforms and user communities, such mentions of what content can be restricted are contrary to freedom of expression and the determination of the purposes of different digital spaces. Yet several of the categories are much different than others, for instance, where undefined concepts like “promotion of terrorism” can be used to punish political dissent in local law. Additionally, it requires “quick and decisive action” as “a minimum” on the listed types of content, imposing the need for enforcement measures of a certain character regardless of context. Expecting the same standard for very different types of content creates the risk to restrict legitimate

expression and is not fully consistent with international human rights law standards. We recommend the Guidance focuses efforts on procedural aspects of content moderation, and recognise different needs and expectations with relation to different types of content and contexts.

Paragraph 27.2 explicitly references the possibility of a contradiction between national regulation of illegal content and human rights standards, establishing that the standard in such a case for the platform is "report on how it responds to such requests", a valuable but oddly specific requirement given the breadth of other recommendations.

Paragraph 28 has changed from Draft 1 from requiring systems to requiring reporting on systems in place, to identify a series of contents and behaviours linked to information disorders. Language in Draft 1.1 seems not to be final, and though it is valuable that the paragraph states that these systems must still protect privacy and anonymity, we must stress that these acts and behaviours, if poorly defined, risk becoming problematic due to the impact they can have on the exercise of the right to anonymity, essential for expression in contexts of repression or of traditionally marginalised groups. No restriction on anonymous or pseudonymous speech, for the mere fact of being so, conforms to international standards on freedom of expression. Value judgements regarding legitimate forms of expression or identity, that make the platform act as an arbiter of veracity, contradict the essence of the diversity and plurality of expressions protected by freedom of expression.

Some of these situations (multiple accounts, *bots*, state-owned or state-controlled accounts) may be better addressed through transparency obligations, rather than reference to action against these categories of content in content moderation policies. Thus, more than platform decisions regarding this content, what would be fostered would be a critical consumption of information by users with additional contextual information on the way and source of content generation through transparency actions. **Paragraph 29**, takes these ideas to recommend labelling or identifying those types of contents or their provenance, and it is valuable that the need to respect privacy and anonymity is reinforced. Special emphasis must be given to discussing precise language that properly covers aspects such as restricting “virality” of content.

With regards to **paragraph 30**, from a formal perspective, to favour clarity in the structure of the document, it would be convenient to separate from this section the issue of notification to the user, which is part of the due process regarding the accountability mechanisms.

Finally, it would be convenient to integrate **paragraph 37** “*language and accessibility*” in this section, since the accessibility of the content and its moderation rules for users are part of what is required under the international human rights framework.

On **Paragraph 31** on “enabling environment”, the first thing to note is that the paragraph lacks systematic references to the principles enunciated in the previous paragraphs. It seems

to refer to the principle of empowerment of the users, but it broadens it considerably with respect to what was stated before.

Additionally, and contrary to the principle of the empowerment of the users, it seems to emphasise the responsibilities of the platforms in taking proactive actions to promote a space for the circulation of information aseptically. Such a definition seems problematic in its compatibility with international standards on freedom of expression. Additionally, it is paternalistic to house that responsibility on the platform, as it diminishes the decision-making agency in the information consumption of each user. We believe that the difficulties described in the paragraph 31 of the section are better addressed with more transparency on the part of the platform regarding the context and the provenance of certain content, thus providing elements to users to judge their information consumption and to promote a regime of responsibility.

As we commented in Draft 1, the control of the users of the selection of the content that will be displayed as a result of a search and/or in the news or of the recommender systems, can generate undesired echo chamber effects that can be detrimental to the objective of information as a common good raised by the document. Such measures in isolation from the issues of economic incentives and business models will not have a significant impact on improving access to diverse and plural information for users.

Paragraph 32, on user reporting, refers to a series of characteristics that the accountability system of digital platforms that develop content moderation should be endowed with. The structure of the document would benefit if the section made clear reference to the development of the principle identified above, and grouped the elements that have been erroneously integrated into other sections that we have identified above. There are obvious typographic errors in Draft 1.1.

The requirement for the availability of information on moderation policies accessible in an digestible format and in all relevant languages should be moved to the section on content moderation policies compatible with human rights principles and standards, for greater consistency in the structure of the document.

On the other hand, the reference to the fact that the system of complaints gives preference to “threatening or intimidating” content, requires definitions in this regard that require more guidance for the platforms that would implement the system. Otherwise, a recommendation could be formulated to encourage companies to have general complaint mechanisms and emergency procedures, with respect to which characteristics of precise severity are indicated so that they are triggered.

A different issue, and one that has also been raised by civil society in situations of recent social upheaval in Latin America, is the need for complaint processes to include special procedures for cases in which a large number of simultaneous requests must be processed, as it happens in contexts of social unrest or massive violations of human rights. This matter seems to be

addressed in **paragraph 36** referring to "major events" which includes the obligation of platforms to have risk assessments and crisis mitigation policies for events such as conflicts, natural disasters, health emergencies and migrations, where information disorders are likely to increase. The draft would benefit from a more detailed proposal on the characteristics of the mitigation measures required, including what minimum guarantees of due process must be respected in these cases, since many times they will refer to the same content or categories of content that are particularly relevant for their public interest in the particular context.

Paragraph 33, on Content that potentially harms democracy and human rights, covers some of the most problematic subjects in generating standards. The section recognizes the lack of a universal definition of what constitutes potentially harmful content, however, it insists on assigning responsibility to the platforms for a consistent application of their policies regarding an undefined concept. The transparency measures mentioned in other sections of the proposal seem more specific and clearer to apply than the vague recommendation formulated here. The document addresses as desirable the regulation of potentially harmful content, which is an ambiguous concept and on which there is no local or global consensus. International human rights standards, including the Inter-American human rights system, focus on the need that all restrictions on expression must satisfy criteria of legitimate purpose, legality, necessity and proportionality, so that that the regulatory approach should focus on illegal content as the lowest common denominator of a regulation that has an ambition of universality. The paragraph is adjusted from Draft 1 to include the challenges in identifying this type of content for both platforms and regulators. Reference to the problem of scale is then added and it is considered "acceptable" that "in principle" this is mainly handled automatically, but it is not said what it should be supplemented with.

An interesting proposal in paragraph 33.2 is the reference to the "systematic risk assessment" being carried out to identify this harmful content, of which characteristics are not stated. Reference is added to transparency when specific protection measures are offered to specific groups, indicating that in such cases risk assessments and mitigation processes are developed or the creation of specific products to facilitate their online participation. There is no reference to which risks the assessment should focus on, leaving out a focus on human rights, leaving out the process to identify and define those risks, and leaving out whether platforms themselves or independent assessors can conduct those assessments.

Paragraph 33.4 recommends demonstration on potential responses to speech, though the recommendation is vague. The paragraph limits when content removal or “de-platforming” of users should be considered, depending on “intensity” and “severity” of content along with intention of harm, yet this leaves out acts of moderation that are solely related to the purpose of each platform or groups and communities within platforms, severely limiting platforms’ and groups capacity to define their own standards of permissible behaviour and expression.

Paragraph 34 on Media and information literacy, recognises that this constitutes a primary duty of States that is not appropriate to outsource to private actors, even when their collaborative actions are welcome. However, more than a literacy directed to specific groups

identified as vulnerable, the effort should be to promote an inclusive and participatory design of different sectors of society, so that the policies and products of the companies can have the opportunity to be influenced by citizen participation, beyond the economic incentives of product development.

Paragraph 35, on Election integrity, recommends a risk assessment process "for any election event", including engagement with public authorities and civil society groups in paragraph. As with other recommendations, this is one where large platforms (a focus of this Guidance) might be the few services with the capacity to implement these measures. The recommendation can be refined to more specifically address the issue of transparency in election advertising.[6] The recommendation on a library of political advertisement in paragraph 35.3 lacks sufficient justification for its suggestion of use of public funds, and assumes a higher level of risk by recommending higher scrutiny to what can be legitimate political expression in the case of issue advertisement. Recommendations of transparency on automated tools in paragraph 35.4 belong elsewhere in the document, although it is valuable to strengthen them in relation with political advertising.

Paragraph 36 on Major events once again recommends risk assessments and mitigation policies, for which the same comments to paragraph 34 apply, as it is unclear which risks are the target of the recommendation. The recommendation also is unclear whether the likelihood of higher effects of misinformation and disinformation are a factor that justifies the recommendation or a criterion for the platform to define events that justify the risk assessments and mitigation policies.

Paragraph 38 on Access to data, describes conditions necessary to conduct research on content that is potentially damaging to democracy and human rights, including reference to the protection of privacy and trade secrets. There need to be transparency and accountability mechanisms to ensure this access is provided. Additionally, the safeguards should ensure data integrity and security, including protocols on the process to obtain access.

Section Three- The independent regulatory system

Paragraphs 39-40, on the independent regulatory system, explicitly acknowledge institutional diversity in different jurisdictions. It is important to maintain openness to that diversity, so that the Guidance will account for the different institutional, social, political, cultural and legal realities when establishing the regulatory system proposals, particularly with respect to countries in the Global South where the independence of technical bodies in many cases is affected, in the form of inappropriate influence on the functioning of regulatory systems by some government authorities. We recommend more specific guidance on transparency requirements imposed by the regulatory system that have a lower risk of manipulation for political purposes than other types of measures.

[6] For instance, see the OAS' guide to protect freedom of expression in the face of deliberate misinformation in electoral contexts, available at: https://www.oas.org/es/cidh/expresion/publicaciones/Guia_Desinformacion_VF.pdf

Paragraphs 41-45 on the Constitution for the regulatory system, references the "Regulatory System" and the "control body" as if they were two subjects within the proposal but without clarifying their differences. Up until this point, the draft speaks of "independent regulators" and "independent oversight", but in this paragraph these are attributed to the "Regulation System" which seems confusing since the system includes a set of rules that must be observed by the obligated subjects (companies, States, users, media, etc.), not a unitary entity that is in a position to render accounts to the control body. In this sense, the draft would benefit from greater clarity regarding the fact that the constitution requirements described refer to the control or regulatory body, whatever the specific form it adopts according to the different local institutional realities. For instance, paragraph 42 includes a reference to the grounds for dismissal of members of the "system" and the need for a process established to ensure their independence. From a substantive perspective, regulatory independence must be complemented with technical independence and made compatible with coordination with other institutions that are already in charge of various aspects that are involved with the issues dealt with by the regulatory system. Finally, in addition to carrying out "*regular multi-stakeholder consultations on its operation*" (p. 45), it is important that the regulator's design include participatory processes with a diversity of stakeholders, both for the election of its members and for the increase the transparency of their own operation and to feed proposals for reform of the regulatory system according to the evolution of the needs of society and the development of technology.

Paragraphs 46-53 on the Powers of the system, requires a structural revision. It is unclear what the section seeks to define, since some things that are mixed include: the objectives of the regulatory system, the scope of application of the proposed regulation, the powers of the regulatory system, due diligence obligations that are sought to be imposed on companies, persecution policies, sanctioning procedure, among others. In summary, "Powers" does not represent the content of these paragraphs.

Paragraph 46 states that "Regulation will set the overarching goals for platforms to safeguard information as a public good", but as we have pointed out, there is a practical problem due to the lack of clarity in the proposal regarding what is meant by "information as a common good" in the context of platform regulation. We understand that to focus the proposed Guidance that is being advanced in the draft, there are two possibilities: (i) either it is already understood that information is a common good and the regulatory system will have, as a prerequisite for its existence, the application of the framework from the Guidance with its aims and objectives, anchored in respect for human rights; or (ii) on the contrary, it is necessary to have a regulatory system in each country that starts by recognising that information is a common good and, in accordance with that definition, in each case, define the objectives of the platforms. The document should provide considerably more guidance to States parties on what is meant by "information as a common good".

The investigative and sanctioning powers of the regulator must be clear and pre-established by the regulatory system. The document would benefit from a more specific and systematic guide to the minimum elements of a sanctioning process. Some matters to be covered would

be: investigative powers, rules of due process, reception of expert opinion, catalogue of sanctions, principle of opportunity in the selection of cases, intervention of third parties interested in the procedures, emergency measures, appeal mechanisms (coordination with local judicial system), among others.

Lastly, another point where the proposal could advance is in promoting that users have legal standing to challenge decisions of the platforms before the regulatory system, when dealing with cases involving systemic failures or collective damage to users. Here, the proposal could go further to declare that organised civil society will have legal standing to collectively challenge such decisions, as a measure to facilitate access to justice.

Paragraphs 54-56, on the Review of the regulatory system, refers to an unnamed "respected" third party to report to the legislature, assuming legislation at the centre of the system. The obligation to act within the law and respect human rights, as well as the possibility that the courts review the decisions of the regulator, are extremely important to us. However, such provisions should be within the "Constitution" section.

A different topic is the need to review and formulate proposals to improve the regulatory system as a whole, which could well be assigned to the regulator or legislator, with the support of external experts. Periodic consultations referenced in paragraph 45 seem to go in the same direction.

Paragraph 56 closes the section by acknowledging that the exclusive competence for deciding cases regarding specific content corresponds to the justice system. Once again, it is a structural limitation to the scope of application that we believe should be incorporated in the "Constitution" section of the regulatory system, since it is not a review of it or of the actions of its regulator.

5. Conclusion

UNESCO has made important contributions to the regulation and promotion of human rights in the digital environment, for example, through the ROAM and ROAM-X Universality Indicators, or processes such as the Recommendation on the Ethics of Artificial Intelligence. In such instances, the construction of these documents was accompanied by broad processes of participation to varying degrees by diverse interested stakeholders. However, as we have expressed in this document, we consider that this consultation process as a whole suffers from very concerning problems. As verified throughout this document, there are still many structural points that need to be improved. We believe that the questionable aspects of this draft Guidance are directly related to the limited opportunity for more open discussion and engagement with the process.

Derechos Digitales recommends that UNESCO uses the Conference as a space to receive inputs and feedback, but more importantly, that UNESCO takes advantage of this dialogue to

open up discussion before finalising the draft Guidance, so that the process effectively has a multistakeholder approach, with meaningful participation from all interested parties in all the world. Without meaningful, wide participation in a highly debated and contested topic as platform regulation, the opportunity for UNESCO to provide guidance for new regulatory efforts risks getting lost. We expect that the leadership shown by UNESCO in the digital environment leads to the best possible outcomes for human rights for all.

Derechos Digitales remains available to discuss the points expressed here in this contribution and to support UNESCO and its Member States to carry out a plural and meaningful discussion in order to protect human rights.