Standards for the democratic regulation of big platforms to ensure freedom of expression online and an open and free Internet

A Latin American perspective for content moderation processes that are compatible with international human rights standards

July 2020

OBSERVACOM
Observatorio Latinoamericano
de Regulación de Medios y Convergencia

July 2020

Latin American organizations that sign the proposal

Experts who collaborated in the drafting of the proposal (in a personal capacity):

Javier Pallero
Policy Director, Access Now

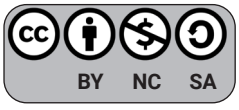
Joan Barata
Member of the Platform for the Defense of Freedom of Information (Spain)

Andrés Piazza
Consultant, ExLACNIC, ExLACTLD

Guillermo Mastrini
Research Professor at the National University of Quilmes (UnQ), University of Buenos Aires (UBA), Conicet

Martín Becerra
National University of Quilmes (UnQ), University of Buenos Aires (UBA), Conicet

Juan Ortiz Freuler
Berkman Klein Center Affiliated Researcher (2019-2020)



Index

Abstract

This document offers recommendations on specific principles, standards and measures designed to establish forms of public co-regulation and regulation to protect freedom of expression, information and opinion[1] of content platform users and to ensure a free and open Internet.

The proposal includes limitations to the powers of the large content platforms (such as social networks and search engines)[2] as well as protections to enable intermediaries to use adequate instruments to facilitate freedom of expression.

The proposal seeks to align with international human rights standards and takes into account existing asymmetries related to large Internet platforms without limiting innovation, competition or start-up development by small businesses or community, educational or nonprofit initiatives.

OBSERVACOM

1. *The notion of freedom of expression as used in this document includes the right to express and share ideas, information and opinions, as well as the right to search for and receive information, ideas and opinions of any kind.*

2. *"Content platforms" refers to online service providers that act as intermediaries or storage platforms, or provide services to search for or exchange information, opinions, expressions and other user-generated content and that perform some type of curation or moderation of such content. These include search engines, social networks and other platforms for exchanging text, images and videos. Large content platforms include Facebook, Twitter and YouTube, among others.*

Introduction

The growing intervention of Internet platforms in the contents of their users through the adoption of terms of service (ToS) and the application of business moderation policies has become an issue of concern throughout the world. Such forms of private regulation affect public spaces that are vital for democratic deliberation and the exercise of fundamental rights.

In fact, "private control" is considered by the international Rapporteurs on Freedom of Expression[3] as one of the three main challenges over the next decade and a "threat to freedom of expression". For the Rapporteurs, "a transformative feature of the digital communications environment is the power of private companies, and particularly social media, search platforms and other intermediaries, over communications, with enormous power concentrated in the hands of just a few companies."[4]

This concern is not new, given that on many occasions both international bodies and digital rights organizations have questioned such practices and made recommendations for corporations to make a change in policies and practices in order to align with international human rights standards[5].

For its part, the United Nations (UN) Office of the Special Rapporteur on Freedom of Opinion and Expression has published several reports on the issue[6], while the Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights (IACHR) has maintained for many years that "Intermediaries must thus keep their activities from provoking or helping to provoke negative consequences on the right to freedom of expression" in their voluntary measures for content moderation, which

3. *Joint Declaration: Challenges to freedom of expression in the next decade of the United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression; the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media; the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, 2019*

4. *Ibid.*

5. *Including the Santa Clara Principles*

6. *Internet Content Regulation, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2018*

"can only be considered legitimate when those restrictions do not arbitrarily hinder or impede a person's opportunity for expression on the Internet.[7]"

There is also a growing interest among governments and legislative congresses, including both authoritarian regimes and consolidated democracies, in regulating activities and the distribution of content, particularly through the regulation of content disseminated via social networks. However, most of these legal initiatives configure solutions that are illegitimate or disproportionate and increase the risk of violating the right to freedom of expression, assigning responsibilities and obligations that turn platforms into judges or even private police forces controlling the contents of third parties, for example.

The undersigned oppose such regulations and will continue to maintain this stance. However, we believe that the self-regulation model that has prevailed until now poses similar risks to the exercise of basic human rights.

A handful of corporations has centralized and concentrated the circulation, exchange and search for information and opinions and do so arbitrarily and without accountability mechanisms for duty bearers. This poses a risk to the exercise of rights. This risk becomes greater as these companies reinforce their dominant position in the market and develop non-transparent technologies for information governance. The natural path of this process increases the level of alarm and calls for urgent action to mitigate the risks to human rights and to the decentralized, free and open Internet that we have long struggled for.

In response to this polarized scenario of corporate self-regulation vs. authoritarian regulation, several Latin American organizations believe that a third way is not only necessary but also possible. This third way implies developing a proposal for democratic, proportionate and intelligent regulation that can ensure appropriate regulatory environments that will protect human rights from the interventions of technological giants, while respecting international human rights standards.

The gatekeeper role of these companies requires that democratic societies set limits on such powers to guarantee historically recognized rights and freedoms, as well as the predominance of the general and public interest.

These proposals are not intended to cover all Internet intermediaries. Rather, they are limited to certain types of platforms and applications whose main purpose is to enable or facilitate access to information on the Internet

7. *Freedom of expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 111*

and/or provide support for expressions, communication and exchange of content among users, including social networks, search engines and video-sharing platforms, but not messaging services[8].

A principle of "progressive regulation" is proposed based on the impact that the measures taken by intermediaries have on the exercise of fundamental rights on the Internet, in particular on freedom of expression. Thus, for example, regulation should be stricter for those platforms that, because of their size, reach or market position, have become public spaces of deliberation, and whose massive nature makes them near monopolies, able to dominate deliberation options and/or the main routes for access to information in digital environments.

In view of these special characteristics, the aim is to create a regulatory environment that is appropriate for the functioning and characteristics of the Internet and that includes mechanisms of self-regulation, co-regulation[9] and public regulation. This should be done keeping in mind that the challenges presented by the new digital scenario (including the speed and volume of information) do not allow for the application of one-size-fits-all solutions.

This document does not propose legislation that determines which content can be disseminated on the Internet. Nor does it require that platforms moderate their content. However, if they decide to do so, a series of conditions should be established so that their users' fundamental rights are not violated in the private moderation process that these companies already carry out in a unilateral and non-transparent way.

Thus, proposals are included regarding the limits to content moderation that these platforms already implement, ensuring that their terms of services, criteria and procedures are compatible with international human rights standards, particularly as they concern the protection of minorities and vulnerable groups.

A democratic and balanced regulatory system should also protect platforms from the illegitimate pressures of governments and other stakeholders.

8. *It is known that, in some cases, messaging services work as mass communication services, surpassing their original function of interpersonal communication. However, due to their characteristics and primary functions, and the absence of content moderation by companies, these services are not included in this proposal*

9. *Defined as the institution, through a formal law passed by a democratically constituted Congress, of guidelines and results that must be achieved by companies, with their direct application and an oversight process conducted by an adequate body, with guarantees of autonomy and independence from governments and companies, with enforcement in cases of non-compliance, also defined by law*

As intermediaries, they are key to facilitating the exercise of these rights, and therefore the proposals include recommendations so that the regulatory frameworks allow such platforms to fulfill that role in an appropriate manner — no legal responsibility for third-party content or prohibition of obliging them to undertake generic monitoring or supervision of contents.

Private regulation of the Internet emerges from and is aggravated by a context of intense concentration of power in the hands of a few international corporations. Public regulation of the activities of these platforms should adopt antitrust measures to counter concentration and lack of competition. Such proposals are not included in this document, however, the simple fact that the main public spaces for the circulation of information and opinions can all be controlled by just one company should oblige antitrust bodies to take action.

This document is not intended to give a solution to all the challenges posed by online content governance, such as like disinformation, but we believe that the set of standards proposed herein – regarding transparency, due process, limits to terms of service, etc.– will have a positive effect on these issues, by limiting the conditions the encourage the spreading of disinformation, clarifying the responsibilities of large corporations in public debate, and defining a regulatory environment able to meet such challenges in a way that aligns with the right to freedom of expression. Complementary documents will include more specific solutions to such issues.

Neither have we included in this proposal issues such as mechanisms to guarantee pluralism and diversity on the Internet or to address tax issues. Rather, the document focuses on issues related to content moderation, offering principles that can be applied in general terms. At the same time, the characteristics of certain services may require specific approaches. For example, cultural services may require obligations for the protection and promotion of cultural diversity in line with the UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions.

Finally, any adopted norms and institutional designs must be adequately developed. This should take into account the needs of market regulations subject to continuous development, the specific characteristics of the digital environment in each country, and the unique requirements of Latin America within the context of international human rights standards.

This document contains the following sections:

1. SCOPE AND NATURE OF REGULATION
2. SERVICE TERMS AND CONDITIONS
3. TRANSPARENCY
4. DUE PROCESS
5. RIGHT TO DEFENSE AND APPEAL
6. ACCOUNTABILITY
7. REGULATION AND CO-REGULATION

1 SCOPE AND NATURE OF REGULATION

1.1 This regulation proposal concerns online service providers when they act as intermediaries or storage platforms, or provide services to search for or exchange information, opinions, expressions and other content generated by their users and that perform some type of curation or moderation of such content (referred herein as "content platforms"). These include search engines, social networks and other platforms for exchanging text, images and videos. These proposals do not cover messaging services.[10]

1.2 Limits to the power of large content platforms should be structured based on a co-regulation model, where self-regulation and public regulation structures act together[11] to create legal, contractual and technical solutions that guarantee freedom of expression online, in line with other fundamental rights. Regulatory and co-regulatory instruments[12] should be the result of a multistakeholder governance process that takes into account local and regional contexts.

1.3 Platforms should directly incorporate into their conditions of service and their community standards the relevant human rights principles that ensure that the measures related to the content are governed by the same criteria of the protection of expression through any media[13]. These principles include transparency, accountability, due process, necessity, proportionality, non-discrimination and the right to defense. All platforms must ensure full respect for consumer rights.

1.4 Content platforms providing significant access to information and opinions of public interest, having an influence on public debate or who define themselves as such, while having significant market power ("large content platforms") should be subject to asymmetric regulation with respect to other providers, in view of the importance and impact that their business decisions may have on the exchange of information, opinions and cultural property, as well as the exercise of freedom of expression and public debate with political, social and cultural effects. Definition of the influence or marketing power of the various content platforms is the responsibility of the

10. *Messaging services work as mass communication services in some cases, surpassing their original function of interpersonal communication. However, due to their characteristics and primary functions, and the absence of content moderation by companies, these services are not included in this proposal*

11. *The institutional design and division of responsibilities is developed in section 7 of this proposal*

12. *Co-regulation refers to a system in which the general guidelines and expected results of platform policies are defined in a legal instrument, with input from multiple sectors, which must be applied directly by platforms taking into consideration local and regional context, and in line with human rights principles. An appropriate body, with guarantees of independence and autonomy, should oversee the companies' application of these standards*

13. *Regulation of content on the Internet, Special Rapporteurship on the Promotion and Protection of the Right to Freedom of Opinion and Expression, 2018*

legitimate regulatory body and based on the specific situation of each country[14].

1.5 This asymmetrical treatment does not imply violating the principle of equality, given that all content platforms must comply with the minimum human rights standards and principles. Smart regulation should not impose excessive burdens on actors that, due to their characteristics and development, could not live up to them. Thus, it should treat big content platforms differently in comparison to those that are smaller or have more specific ends[15].

14. *Previous versions of this document suggested using concepts of economic competition, such as Significant Market Power, relevant market and substitutability, but we prefer a perspective in which their importance is based on the access to information, the exercise of freedom of expression and their influence on public debate. The goal is not to impose disproportionate burdens on low economic capacity actors, start-ups or even actors in services of specific interest. Such obligations may become obstacles to their entry and permanence, affecting the diversity of services available to Internet users. Monopolistic or oligopolistic positions could intensify requirements, given their more severe effect on public debate. The definition of significant market penetration may consider, for example, the percentage of penetration as regards the total number of users or the amount of users, as used by the German NetzDG.*

15. *Equality is sought under similar conditions. Specific obligations should be imposed only on certain actors when they have a predominant role in public discourse and the capacity to comply with such specific requirements. This does not imply denying the existence of minimum requirements (non-discrimination, transparency, etc.) for all content platforms to ensure the protection of human rights. This includes non-profit, scientific, or educational platforms with a*

2 SERVICE TERMS AND CONDITIONS

2.1 The terms of service (TOS) of all content platforms[16], as well as other complementary documents (such as guides or content application guidelines) should be written in a clear, precise, intelligible and accessible[17] way to all users. Big platforms should also present them in the user's national language. Services for children and teenagers, regardless of their legal limitations to hire such services[18], need to establish term and conditions that can be easily understood by this group of people.

2.2 All content platforms should establish and implement TOS that are transparent, clear, accessible and in line with international human rights norms and principles, including the conditions under which interference with the right to freedom of expression or user privacy[19]. In particular, the user should be aware of the conditions that may lead to the termination of the contract (account drop, for example) as well as the removal, de-indexing or significant reduction of the scope of their expressions and contents from unilateral modifications made by curation algorithms[20].

2.3 No content platform should be able to unilaterally modify the terms of service and conditions of use, or apply new terms, without clearly informing the users of the reason and without giving, with reasonable notice, the possibility of canceling the contract[21]. Terms of service should not contain abusive, unfair or disproportionate clauses.

2.4 Users shall retain copyright —moral and proprietary— recognized under the law of their country of origin as regards the content they post. Platforms should be assigned proprietary rights by individual users only through explicit consent, without imposing abusive conditions or taking advantage of the asymmetry between the parties.

2.5 In the content moderation process, restrictions arising from copyright protection of protected content should consider the limi-

reduced, closed group of users.

16. *Ideally, all big platforms should use the same vocabulary in their TOS, in order to make things easier for users, digital rights organizations and regulators.*

17. *The Web Content Accessibility Guidelines is a good accessibility reference*

18. *The use of platforms by minors, their legal capacity or incapacity to understand the terms of service and the situations of real parental control when accepting contract conditions are not in the scope of this proposal. However, the language used in the terms of use must be clear enough so that children and teenagers can understand them*

19. *Freedom of expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 112*

20. *These are algorithms of content prioritization and moderation*

21. *EU agreement with Facebook, Google and Twitter in 2018 "Better social media for European consumers"*

tations and exceptions recognized in international treaties and national laws, such as fair use and use of short fragments, especially with the purposes of criticism, social critique, or educational purposes. In particular, upload filters that inhibit fair use are inconsistent with the prohibition of prior censorship established in the American Convention on Human Rights. Users must be notified about who made the complaint. The treatment of allegedly infringing content must follow the notification and counter-notification process, as provided for in 4.8, especially considering that platforms must not be held responsible for third-party content (see section 4.9).

2.6 The terms of services should not grant unlimited and discretionary power for the platforms to determine the appropriateness of user-generated content[22]. In particular, terms of service that imply limitations in the exercise of the right to freedom of expression and access to the information of its users should not be formulated in a vague or broad way that allows for arbitrary interpretation and application.

2.7 Regarding the curation/prioritization of the visualization of user-generated content (in news feeds, search results, news access services, etc.), big platforms should:

A. Make transparent the criteria used by algorithms to prioritize, reduce or redirect the reach of content, and explain the effects on the user[23].

B. Not use discriminatory criteria that create unfair differentiation[24] that could illegitimately affect the freedom of expression and the right to information of their users.

C. Provide customized filtering mechanisms in a clear, transparent, explicit, revocable/modifiable manner and under user control, so that they decide what content they want to prioritize and how they want to do it (e.g., chronological order).

D. Respect the users' right to know and control which of their personal data are collected and stored, and how they are used in the distribution of content, respecting the principle of informational self-determination.

2.8 If large platforms decide, of their own accord, to incorporate in their ToS certain restrictions and even prohibitions to the publication of contents generated by their users they may only do so with the following limitations, so that they are compatible with the international human rights standards:

A. They may prohibit, even by automatic

22. *Last sentence taken from the EU Agreement with Facebook, Google and Twitter in 2018 "Better social media for European consumers"*

23. *The transparency obligation regarding algorithms should not imply the infringement of trade secrets or intellectual property rights. However, it should be possible to verify the effects of algorithms in the moderation process in order to assess their compliance with human rights standards. In this regard, algorithms can be audited by a defined group of individuals, including machine learning and/or artificial intelligence, ensuring total access for those responsible while maintaining confidentiality and protecting trade secrets*

24. *The concept of unfair differentiation has been debated by the Council of Europe in its declarations on artificial intelligence and algorithmic decision-making, but it has not been clearly defined yet. One of the challenges posed by this is the consistent definition of the concept of "fairness" and "unfairness"*

filtering[25], contents that are clearly and manifestly illegal and that, at the same time, are recognized as legitimate restrictions to freedom of expression in international human rights declarations or treaties, such as sexual abuse or exploitation of minors, or propaganda in favor of war and any advocacy of national, racial or religious hatred that constitutes incitement to violence or any other similar illegal action against any person or group of people, for any reason, including those of race, color, religion, language or national origin[26].

B. They may restrict, as a non-definitive precautionary measure, contents that, even if they are not recognized as illegal, cause serious, imminent and irreparable damage or difficult reparation, to other persons such as unauthorized dissemination of sexual content, gender-based and sexual-orientation-based violence, explicit and excessive cruelty, or even imminent and irreparable harm to public or individual health. In these cases, the list and definitions of restricted content should be included in the ToS in a restrictive, clear, precise manner and should consider, in the analysis of the measure to be taken, the context of the expression published, ensuring they are not part of legitimate expressions (educational, informative content or content to make complaints or express criticism).

C. Content such as cyberbullying or explicit and abusive drug use may be restricted to specific audiences, such as children and teenagers.

D. For any other measure of prioritization or restriction to expressions and other user-generated content that platforms may consider —for commercial or other reasons— "offensive", "inappropriate", "indecent" and similar vague or broad definitions, which could illegitimately affect freedom of expression, big platforms should provide mechanisms and notices for other users —voluntarily and based on their moral, religious, cultural, political or other preferences— to decide whether they want to have access to it[27]. Such con-

25. *Currently, platforms must comply with automatic and upload filtering legal obligations. Nevertheless, in our opinion, these rules or directives often lack legitimacy, since they go against international standards, so they should be modified*

26. *American Convention on Human Rights, art. 13, section 5. With the scopes and interpretations made by different bodies of the Inter-American System, this provision includes gender aspects, sexual orientation, and so on (i.e. "no reason")*

27. *Restrictions available for users should be carefully established to prevent divergent thoughts and different expressions from being erased from their field of view*

3 TRANSPARENCY

tent should not be prohibited, removed or reduced in scope by default if it passes the test of legality, necessity and proportionality, since doing so would disproportionately affect users' right to freedom of expression.

2.9 Services that are not intended or designed for children and teenagers should have effective measures to prevent this group from using them, providing active and widespread information on this condition to ensure children those responsible for them are aware.

3.1 Platforms should publish their content restriction and prioritization policies online, in clear language and in accessible formats, keeping them updated as they evolve, and notifying users about changes as appropriate[28]. Services accessible to children and teenagers must consider the use of the appropriate language for this audience.

3.2 When content is restricted in a product or service of the intermediary that allows the display of a notice when trying to access it, the notice displayed should clearly explain what content was removed and why[29].

3.3 In the prioritization of online content accessible to the user (feeds, search results and so forth), the commercial nature of the communication, sponsored content and electoral or political advertising should be clearly defined, identifying the contracting party without raising doubts about its meaning[30], while being transparent about the content metadata (prices, etc.). Platforms should also be transparent when it comes to their relationships with companies that recommend their products or contents through their services.

3.4 Large platforms should notify their users in a clear, explicit and accessible[31] way, at least about:

A. What types of content and activities are prohibited in their services.
B. What are the criteria and mechanisms for the curation and moderation of content. Which are directly controlled by the user and which are not. How does content visibility affect the curation algorithm used.[32]32
C. In what cases, when and how does auto-

28. *Manila Principles*

29. *Manila Principles*

30. *Based on EU agreement with Facebook, Google and Twitter in 2018 "Better social media for European consumers"*

31. *To "allow users to predict with reasonable certainty what content places them on the dangerous side of the line" (Regulation of content on the Internet, Special Rapporteurship on the Promotion and Protection of the Right to Freedom of Opinion and Expression, 2018)*

32. *For example, through real-time transparency panels allowing users to compare the content to which they have been exposed with the whole universe of published stories during a certain period of time, in order to enhance public understanding of how the algorithm works*

matic content removal apply.[33]

D. In what cases, when and how does human review of content apply. This question makes particular reference to the criteria for decision-making, taking into account the context, the wide variation of idiomatic nuances and the meaning and the linguistic and cultural peculiarities of the contents subject to possible restriction[34].

E. How many moderators do they have, describing in detail their professional profile (experience, specialization or knowledge), their spatial location and their distribution of tasks (in terms of themes, geographical areas, etc.).[35]

F. What are the rights of users regarding the content generated and published by them and the policies applied by the company in this regard.

G. How is the personal information of users used and processed, including personal and sensitive data, in algorithmic decision-making that has an impact on their rights.[36]

3.5 Governments and authorities with regulatory powers must have the obligation to report about their relationship with companies as regards content moderation, including:

A. The number of requests made for user data.

B. The reasons that justified such requests.

C. The legal frameworks that underpin the requests made.

D. The cases in which specific content moderation measures were requested, such as content removal.

E. The response that companies gave to each case.

F. The number of posts that were removed or limited in scope and the number of posts that were reinstated.

33. *Regulation of content on the Internet, Special Rapporteurship on the Promotion and Protection of the Right to Freedom of Opinion and Expression, 2018*

34. *Ibid.*

35. *Without detriment to the respect for the right to privacy and anonymity of moderators*

36. *It is a personal right to have access to the information about the use of their personal data, not only to organize and prioritize content but also to be aware of how decisions that have an impact on their quality of life and fundamental rights are made*

4 DUE PROCESS

4.1 In the design and application of their community content management policies, platforms should ensure that any restriction arising from the application of the terms of service does not unlawfully or disproportionately restrict the right to freedom of expression[37]. To do so, they must respect the requirements of searching for an imperative purpose, as well as the need, suitability and proportionality of the measure to achieve the intended purpose[38].

4.2 The criteria for making decisions, so as not to affect human rights, should take into account the context, idiomatic nuances and the linguistic and cultural peculiarities of the content subject to possible restriction[39].

4.3 In addition, in the analysis of the content restriction measures applicable in each case, the principles of proportionality and progressivity should be respected, weighing the severity and reach of the damage, the recurrence of the violations and the impact that such restrictions could have on the Internet capacity to guarantee and promote freedom of expression against the benefits that the restriction would bring to the protection of other rights[40].

4.4 Users should always have the right to have the content restriction decisions made by large platforms respect due process[41], particularly when it comes to measures that could affect their right to freedom of expression. As a general principle, and except for duly justified exceptional cases[42], people affected by a restriction or interference measured by the platforms and, where appropriate, the general public, must be notified in advance[43] about the restriction measures that affect them[44]. There should also be a possibility for them to submit

37. *Freedom of expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 112*

38. *Freedom of expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 55*

39. *Retrieved from David Kaye's 2018 report (the original gives transparency recommendations for platforms)*

40. *Freedom of expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 54*

41. *Due process implies, at least, the guarantee of equal treatment, the justification of the decisions made, and the possibility for users to seek effective defense, appeal decisions and have procedures completed in a reasonable time.*

42. *See 4.6*

43. *Freedom of expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 115*

44. *The previous RFOE (IACHR) clearly states that prior notification must be interpreted as an absolute requirement and applied "where appropriate". Regardless of this, there might be exceptions that require swift action (content moderation, content removal, etc.) and then there should be an explanation*

counterarguments or voluntarily remove the published content before a measure is unilaterally taken by the platform.

4.5 In view of the aforementioned principles of necessity and proportionality, in the case of possible breaches of the ToS, platforms should adopt less burdensome measures than the removal of content or others of similar effects, opting for warning or notification mechanisms, flagging, or linking with opposing information, etc.

4.6 More drastic unilateral measures taken without notice or due to prior process, such as the removal of accounts, profiles or content, or other measures that have a similar impact of exclusion from the possibilities of participating in the platform should be taken by large platforms only under the following conditions:

A. When dealing with non-arbitrary or discriminatory technical management interventions (such as spam, fake accounts[45], malicious bots, among others);
B. When dealing with duplicates or unmodified reiterations (not commented on or edited for journalistic or informative purposes or other legitimate purposes) of other content and expressions of manifest illegality that were already restricted after human evaluation following the aforementioned standards;
C. In following situations:
   a. The grounds set forth in section 2.8 A;
   b. The observance of orders from competent authorities of immediate withdrawal and the commission of common crimes already recognized in national legislation;
   c. Serious, imminent and irreparable damage or of difficult reparation to the rights of other persons as in the cases listed in 2.8, sections B and C.

In all these cases, except in the case of orders from competent authorities[46], the platform should proceed to the immediate subsequent notification, with the possibility to appeal for a possible revision of the measure under the terms of section 5 of this document. The decision should also be communicated to the general public, reaching at least the users who interacted with the content in question.

4.7 Upload filtering and blocking is only legitimate and compatible with international human rights standards when it comes to child and adolescent protection[47] or in the first two situations described in the previous point. Otherwise, it should be considered as an act of prior censorship, under the terms established by the American Convention on Human Rights[48].

4.8 Any other measure of content restriction that the platform intends to adopt in the event of a possible breach of the ToS or the complaint of third parties (for example with regard to an impact on copyright), the content in question should be kept on the platform until a final decision arising from due process is made where, after the user is notified:

a. the voluntary withdrawal of the content in question is promoted

45. *Parody or satirical accounts shall not be considered as fake accounts*

46. *As it is not the platform's responsibility*

47. *American Convention on Human Rights, art. 13, section 4*

48. *As stated before, it should be taken into account that there are certain legal automatic filtering obligations that platforms must comply with. However, our opinion is that these are illegitimate under international standards and should thus be modified*

b. the exercise of the right to defense[49] of the user is guaranteed allowing a justified counter-notification, before taking a decision.

4.9 No content platform should be held liable for content generated by third parties, as long as they do not modify or edit such content, or refuse to execute judicial orders or orders from competent and independent official authorities, provided such orders comply with appropriate due process guarantees and clearly identify the content that needs to be restricted and the reasons why such content is illegal.

4.10 Large content platforms should only be held responsible for their own actions when censoring public interest content protected by the right to freedom of expression[50] and for the active promotion of expressions that could affect the rights of third parties if they collide with the principles established in 2.8. They should also be held responsible when they fail to comply with due diligence in relation to content that has been judicially questioned or to avoid or limit coordinated malicious actions.

49. *See section 5*

50. *These do not include commercial ads, which should be considered as an economic action conducted by platforms*

5 RIGH TO DEFENSE AND APPEAL

5.1 All content platforms should clearly explain to users why their content has been restricted, limited or removed; or why the account or profile has been suspended, blocked or deleted:

A. Notifications should include, at least, the specific clause of the community rules that the user allegedly violated.
B. Notifications should be detailed enough to allow the user to specifically identify the restricted content and should include information on how the content or account was detected, evaluated and deleted or restricted.
C. Users should be provided with clear information on how to appeal the decision[51].[52]

5.2 Content platforms should not delete publications or other user-generated content without being notified, without providing clear justification and without giving users the possibility to appeal[53], so that they can exercise their right to defense and prevent abuse. In this regard, platforms must provide users with the opportunity to appeal content moderation decisions, under the following conditions:

A. Appeal mechanisms should be very accessible and easy to use.
B. Appeals should be subject to review by a person or panel of people who were not involved in the initial decision and are not a party.
C. Users should have the right to propose new evidence or material to be considered.
D. Appeals should result in prompt determination and response to the user.
E. Any exception to the principle of universal appeals recognized in due process[54]should be clearly disclosed and compatible with international human rights principles[55].

5.3 Users affected by any measure of restriction of their freedom of expression as a result of the decisions of platforms, depending on the specific regulations of domestic law, must have the right to access legal resources to dispute said decision and reparation mechanisms in relation to the possible violation of

51. *Santa Clara Principles*

52. *Manila Principles – "The notification on content restriction decisions adopted by a platform must, at a minimum, have the following information: The reasons why the content in question violates the intermediaries' restriction policies. The Internet identifier and a description of the alleged violation of the content restriction policies. The contact details of the issuing party or its representative, unless this is prohibited by law. A statement in good faith that the information provided is accurate"*

53. *EU agreement with Facebook, Google and Twitter in 2018 "Better social media for European consumers"*

54. *American Convention on Human Rights (San José de Costa Rica), Article 8.2, section H*

55. *Santa Clara Principles*

their rights[56].

5.4 In this regard, content platforms may not prevent their users from taking legal action against them in their country of residence, which would imply a denial of their right to access justice[57]57 as a complementary or separate way to claims through the internal appeal mechanisms. For this purpose, the contract executed between the user and a content platform must expressly include that the disputes will be governed by the law and the justice system of the country where the user has their habitual residence and not by the place where the offices[58] of the platform are located[59].

56. *Freedom of expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 115*

57. *EU agreement with Facebook, Google and Twitter in 2018 "Better social media for European consumers"*

58. *EU agreement with Facebook, Google and Twitter in 2018 "Better social media for European consumers"*

59. *The issue of jurisdiction is complex and has been subject to several reviews and suggestions in public consultation, the definition of the scope of damages and remedies and the juxtaposition of conflicting rules on different topics. This shall be analyzed thoroughly in future documents*

6 ACCOUNTABILITY

6.1 Content platforms should publish transparency reports that provide specific and disaggregated information about all content restrictions adopted by the intermediary, including actions taken before government requests, court orders, private requirements, and on the implementation of their policies[60] on content restriction[61].

6.2 Content platforms should issue periodic transparency reports on the application of their community rules that include at least:

A. Full data describing the categories of user content that are restricted (text, photo or video; violence, nudes, copyright violations, etc.), as well as the number of pieces of content that were restricted or removed in each category, by country[62].
B. Data on how many content moderation actions were initiated by a user's report (flag), a trusted flagger program or by the proactive application of community standards (for example, through the use of a machine learning algorithm)[63].
C. Data on the number of decisions that were effectively appealed and the number of decisions determined to have been made mistakenly[64].
D. Data reflecting whether the company performs a proactive audit of its non-appealed moderation decisions, as well as its error rates in moderation content decisions[65].
E. Aggregated data that illustrate trends in the compliance with standards and examples of real cases or detailed hypothetical cases that clarify the nuances of the interpretation and application of specific standards[66].

6.3 Apart from a periodic transparency report, big platforms should also create alerts for specific cases, such as service disruption and unusual behavior in the requests for content or account removals.

60. *Manila Principles*

61. *The information disclosed must be competitively neutral and cautious about trade secrets and procedures guaranteed by intellectual property*

62. *Santa Clara Principles*

63. *Santa Clara Principles*

64. *Santa Clara Principles*

65. *Santa Clara Principles*

66. *Regulation of content on the Internet, Special Rapporteurship on the Promotion and Protection of the Right to Freedom of Opinion and Expression, 2018*

7 REGULATION AND CO-REGULATION

7.1 In so far as these are measures that could affect fundamental rights, the substantive aspects of the regulation proposed in this document should be adopted beforehand and by formal law, that is, a law approved by the legislative body (Congress, Parliament, National Assembly or similar), after public and open consultation. When necessary, regulatory delegations of enforcement agencies should be carefully established by law.

7.2 Content platforms should not depend on licenses for their operation in a given country, but there must be an obligation to identify legal officers and effective forms of communication and response for users and the respective authorities such as an email account, an electronic form, or equivalent means.

7.3 Content platforms should not be obliged to monitor or supervise content generated by third parties in a generic way, in order to detect alleged current violations of the law or to prevent future ones.

7.4 The operation of content platforms should be framed in an environment of co-regulation in accordance with the characteristics of the digital environment:

A. The principles and standards in this proposal should be included by content platforms in their terms of service and other complementary documents (such as guidelines);
B. Platforms should apply these principles and standards without prior intervention by state agencies;
C. The implementation of policies should be overseen by a public authority with a special understanding on the protection of freedom of expression, operating with sufficient guarantees of independence, technical and decision-making autonomy and impartiality, while having the capacity to evaluate the rights at risk and provide users with the necessary safeguards[67]. It should also be able to identify their adequacy[68] in relation to the compliance of A and B;
D. The regulatory body can be a dedicated authority or it can be part of an existing authority in the country operating in this area, provided that it complies with the guarantees in C. In all cases, it should be established by law with ordinary proceedings and it should be given its own resources to operate properly. It should have a multistakeholder advisory council, made up of representatives from all sectors involved, including civil society.
E. In case of non-compliance with the obligations of transparency, due process, right to defense and others, the agency must have sufficient enforcement capacity, being able to apply sanctions, if necessary. However, while it can assess individual cases that might be emblematic for the analysis of platform policies, it should not impose sanctions or have a binding decision in such cases, except when it comes to

67. *As stated in Freedom of Expression and Internet, Office of the Special Rapporteur for Freedom of Expression IACHR, 2013, para. 56*

68. *This does not imply the imposition of policies or standards about the treatment of specific contents on platforms*

independent and specialized authorities that serve quasi-judicial functions, such as electoral bodies.

F. The regulatory body should have the power to request any type of granular information necessary to fulfill its supervisory role and to impose fines or other remediating actions when platforms are unable to provide information in a timely manner.

G. Its attributions may include the elaboration of national and comparative studies, the promotion of citizen rights, and the cooperation with independent regulatory authorities in other countries, as well as cooperation efforts with self-regulation bodies of the companies.

H. As a rule, the regulator should have national jurisdiction, but regional solutions could be adopted with the approval of the involved parliaments, provided the regional legislation and practices are sufficiently consistent and coherent.

7.5 Individual cases where there is a violation of a user's rights that is not satisfactorily resolved within the internal scopes and mechanisms for dispute resolution should be resolved by judicial bodies or similar independent and specialized public bodies —in the country where the user has their habitual residence— by means of an abbreviated procedure, digital procedure and electronic notification (fast track) with guarantees of subsequent revision. Other authorities or state agencies that do not meet the previous characteristics should not be able to enforce platforms to remove or process specific content.

7.6 The creation or strengthening of offices of the Public Defender, Ombudsman's Offices or similar bodies should be encouraged to defend and advance the rights of users in platforms. They should have the power to receive and process claims in cases where there has been an infringement of fundamental rights in platforms and state bodies, including individual cases.

7.7 Without prejudice of the above, platforms should have internal and effective appeal mechanisms, as well as independent external stages for the revision of cases and adopted policies.


OBSERVACOM
Observatorio Latinoamericano
de Regulación de Medios y Convergencia


intervozes
coletivo brasil de comunicação social


idec
Instituto Brasileiro de
Defesa do Consumidor


CAinfo


TE
DIC.


PROLEDI
Libertad de expresión, derecho a la
información y opinión pública


USUARIOS
DIGITALES
conectados por nuestros derechos


desarrollo
digital


IPANDETEC
CENTROAMÉRICA


Datos
Protegidos


FLIP
FUNDACIÓN PARA
LA LIBERTAD
DE PRENSA


Datysoc
LABORATORIO DE
DATOS Y SOCIEDAD


HIPER
DERECHO

July 2020