



June 2023

## **Contributions to the Special Rapporteur on contemporary forms of racism**

*Enabling practices for the propagation of hate speech in the digital  
environment.*

In preparation for the 78th session of the United Nations General Assembly, the Special Rapporteur on contemporary forms of racism has called for contributions<sup>1</sup> to elaborate her periodic report, which will focus on online hate speech.

Taking into consideration the main points of analysis proposed in the call, *Asociación por los Derechos Civiles* (Association for Civil Rights-ADC) offers the Special Rapporteurship this contribution, oriented to chart the current situation and provide certain guidelines in order to address this issue.

ADC is a civil society organization, founded in 1995, based in Buenos Aires, Argentina, that works in the promotion and defense of fundamental rights, with regional and international scope. Among the most important aspects the organization works with are freedom of speech, access to justice, inclusion and diversity, privacy and personal data protection.

In recent years, ADC has focused on the interaction between human rights,

technological development and digital environments. As a result, the organization has identified the relationship between the growing activity developed in social networks and content platforms and the expansion of the phenomenon of online hate speech.

Taking the research in four Latin American countries (Argentina, Brazil, Panama and Costa Rica) as a starting point, this contribution supports its analysis on three fundamental axes: digital platforms, hate speech and the role of intermediaries.

### **Digital platforms: the consolidation of the online civic space**

In recent years, the Internet has become one of the spaces with the greatest impact on the lives of citizens. The United Nations has recognized it as a human right, which also enables the exercise of other rights, in light of the new dimensions that the digital environment proposes for their deployment.<sup>2</sup>

In this sense, freedom of speech is one of the fundamental rights that has found significant diversification in the digital sphere, given that a large percentage of citizens use this space to express opinions, publish content and access information.<sup>3</sup>

The digital environment is increasingly strengthening interaction on content platforms which, unlike face-to-face conversation, uses anonymity as a means of protection for users in the context of exercising their freedom of speech. However, the right to preserve one's identity on the Internet has raised many questions, based on the prevention of crime and the punishment of conducts such as harassment and the dissemination of discriminatory messages.<sup>4</sup>

### **Online hate speech**

In terms of regulation of the digital environment, one of the biggest

controversies revolves around the legitimacy of action against the dissemination of online hate speech. As an initial approach, there is concern on how digital technologies have enabled the spread of denigrating and harmful messages, affecting the dignity of members of vulnerable groups, exposed to multiple attacks and humiliations. From a different perspective, excessive regulation of activity on platforms increases the risk of confronting freedom of speech which, among other aspects, includes dissent and certain forms of offensive speech as a logical consequence of the circulation of plural and diverse discourses.<sup>5</sup>

In order to separate "tolerable" messages from those that cannot be shielded by the scope of this right, it is essential to unify criteria in order to clarify what is understood by hate speech. However, the definition of this phenomenon has been problematic from the beginning and, to this day, the discussion has not been settled. With the emergence of the digital dimension as a means for the proliferation of these speeches, the need to find answers has increased.<sup>6</sup>

The main problem with this conceptual void is that legal and judicial systems have been forced to build criteria around it. In a context of expanding decision-making power in favor of digital platforms, these divergences among the public sector have facilitated discretionality in the hands of the private sector, leading to two possible outcomes. The first one is the confrontation of freedom of speech in the digital environment, as a result of policies that excessively limit public debate. The second one, on which this contribution will place particular emphasis, is the enabling of an unrestricted circulation of online hate speech, as a consequence of excessive permissiveness.<sup>7</sup>

### The legal dimension

Among the international treaties that outline a definition of hate speech is the American Convention on Human Rights (ACHR), which categorically excludes it

from the protection of freedom of speech, as long as a connection can be established between the expression in question and the promotion of violence.<sup>8</sup>

In this regard, it would be extremely enlightening to have parameters delimited by the Inter-American Court of Human Rights, as the final interpreter of the scope of the ACHR. However, its existing jurisprudence has not yet provided sufficiently detailed guidelines on the categorization of speeches that, depending on their effects, should be totally or partially excluded from the sphere of freedom of speech.<sup>9</sup>

The International Covenant on Civil and Political Rights (ICCPR) is aligned with the ACHR in excluding hate speech from the scope of protection provided by freedom of speech. However, following the definition proposed by this instrument, it is not the connection with incitement to violence that must be proven, but with the promotion of hostility or discrimination.<sup>10</sup>

The United Nations Human Rights Council, in its Resolution on the promotion, protection and enjoyment of human rights on the Internet, has reinforced the provisions of the ICCPR on the scope of freedom of speech. Specifically, it stressed the "importance of combating the promotion of hate, which constitutes an incitement to discrimination and violence on the Internet, by encouraging tolerance and dialogue instead".<sup>11</sup>

A third instrument mandated to define hate speech is the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD). This treaty requires States to consider the issue, but proposes a broader definition than the previous ones. According to this Convention, there is no need to identify incitement to violence or discrimination, since the mere dissemination of expressions based on notions of hatred or racial superiority is enough to take action.<sup>12</sup>

In light of these conceptual divergences, the adaptation of international treaties to national legal systems has not been a simple task. In addition, interpretative conflicts may arise between two international instruments that have the same hierarchy in a given country, or between an international treaty and a national law, if the State accepts the former and the latter on an equal level of obligation.<sup>13</sup>

Regarding the internal treatment of the concept of hate speech in the Latin American region, the examples of Argentina, Brazil, Costa Rica and Panama can be mentioned as countries that directly or indirectly address the concept in their laws, but do not define it concretely.<sup>14</sup> The States coincide in the possible targets of these speeches, based on aspects such as race, ethnicity, gender, religion and national origin.<sup>15</sup>

As for the behaviors encouraged by what constitutes hate speech, the dissent is similar to that of the international community, and even more atomized. Argentina refers to the persecution of groups or individuals<sup>16</sup>, while Costa Rica refers more precisely to violence.<sup>17</sup> Finally, Brazil<sup>18</sup> and Panama<sup>19</sup> consider that the determining consequence is discrimination.

The absence of a precise definition is also reflected in the jurisprudential decisions of the aforementioned countries. Either because there are no judicial decisions on the subject, as in the cases of Panama and Costa Rica, or because the few precedents that exist are ambiguous and make very limited contributions to the consolidation of stable criteria, as in the cases of Brazil and Argentina.<sup>20</sup>

In the latter two countries, case law has set as standards for identifying hate speech the incitement to persecution, hatred or discrimination. This derives into a detachment from the Inter-American system, which explicitly determines the

incitement to violence as a parameter.<sup>21</sup>

In reference to the level of correlation required between the hate speech and the act encouraged by it, it is worth mentioning the example of the courts in Argentina, which have applied the test of certain and current or imminent risk when delimiting the scope of a discriminatory act. Judicial resolutions highlighted the importance of analyzing the expressions in the corresponding context, in order to determine whether they are capable of creating a current and evident risk for a person to be the object of persecution.<sup>22</sup>

Brazil, in contrast, has resorted to a broader conception, focusing on the harm that discrimination causes to those who suffer it. In this sense, it has referred to racist discourses as manifestations of power and instruments of political domination, which reproduce historical inequalities and keep vulnerable groups on the margins of the rights protection system.<sup>23</sup>

### The social dimension

Outside the strictly legal sphere, the task of defining the concept of hate speech is also limited by the particularities of the contexts in which this question is addressed. In general terms, it can be noted that these discourses are linked to predominant power relations in the social environment, which condition and determine the opinions and behaviors of citizens. For an adequate identification and prevention of this phenomenon, it is essential to take the analysis of these origins as a starting point.<sup>24</sup>

The Office of the Special Rapporteur on freedom of religion or belief made particular reference to hate speech as a consequence of a public discourse that has been co-opted by political interests, which stigmatize and dehumanize minority and historically vulnerable groups: migrants, refugees, women and religious communities, among others.<sup>25</sup>

In terms of the experiences of vulnerable groups, it can be pointed out that these discourses reinforce and consolidate sociocultural hierarchies that ultimately lead to unequal power relations, exacerbating the vulnerability of oppressed groups through discriminatory practices and acts of symbolic violence, embedded in these discourses or at least enabled by them.<sup>26</sup> Regarding the materialization of discriminatory expressions, it is important to note their intersectional origin, which implies the simultaneous reflection and reinforcement of a multiplicity of oppressions suffered by members of vulnerable groups, marginalized by various features of their identity.<sup>27</sup>

The analysis shared so far is equally applicable to the traditional forms of hate speech dissemination and to the more contemporary mechanisms enabled by digital technologies. However, the particularities of the latter demand a specific approach to the new challenges that arise due to the massiveness of the Internet.

One of the main problems is the growing presence of fake accounts and trolls in social networks, classified as haters and in many cases in charge of the rapid multiplication of harmful messages. Far from questioning anonymity as a safeguard for free and safe participation on the Internet, it is indisputable that it generates a problem for those affected by these messages, when it comes to identifying the speakers. This has drawn attention to the role played by the platforms.<sup>28</sup>

### **Content moderation: the role of the platforms**

In the Internet sphere, the conceptual void on hate speech has increased the decision-making power of digital platforms, which have adopted the removal of publications classified as harmful as part of their content moderation policies.<sup>29</sup> Thus, this concern about the extremes when regulating from the public sector

acquired a new edge: the discretion of private companies to regulate the online civic space according to their own rules, criticized both for extreme permissiveness and for the tendency to censorship.<sup>30</sup>

The proper identification of hate speech is not an easy task, especially considering the complexity of telling the difference between messages shared as a deliberate attempt to harm, and those replicated with a reduced awareness of their potential impact.<sup>31</sup> However, this aspect can't lead to overlooking the responsibility of the platforms regarding the impact of their business model on the enjoyment and exercise of citizens' rights.

Within the framework of Business and Human Rights, the United Nations Guiding Principles on Business and Human Rights (UNGPs) combine the role of States in terms of guaranteeing the exercise of human rights with the duty of companies to respect them. In relation to the latter, the commitment with due diligence does not end with the reparation of damages caused by their activity. They must previously assess the risks of its development in order to make their business model compatible with the adequate exercise of rights.<sup>32</sup>

### **Concluding remarks**

Dissent is an unavoidable component when dealing with online hate speech. This reflects an acceptable level of disagreement within societies, especially when dealing with such complex issues. However, without the proper approach, these conceptual divergences can operate as enablers of harmful speeches. To avoid this, it is essential to articulate their coexistence on the basis of some consensus, bearing in mind that the different proposals of approach arise from particular contexts, and in this sense it is not appropriate to classify them as right or wrong.



The first point of consensus is the incorporation of the human rights perspective, as it proposes a universal language that guarantees a balanced discussion.<sup>33</sup> In this sense, the UN Rabat Plan of Action proposes an analysis of freedom of speech in its interaction with other fundamental rights, establishing the importance of proving the connection between the speech and the harm attributed to it.<sup>34</sup>

The second point is the promotion of a plural dialogue that brings together all the parties involved to ensure legitimacy in decision-making. For this purpose, the role of human rights organizations, which often bring together the experiences of groups affected by hate speech, is particularly relevant.<sup>35</sup> But it is also essential that the platforms are invited by the States to this collective discussion. They cannot be disqualified as agents in the incorporation of the human rights framework. On the contrary, their commitment is fundamental to fulfill this objective, and this implies replacing antagonism with a proposal for collaborative work.<sup>36</sup>

## Notes

1. Call for contributions by the Special Rapporteur on Contemporary Forms of Racism <https://www.ohchr.org/en/calls-for-input/2023/call-input-online-hate-speech>
2. UN adopts Resolution on the promotion, protection and enjoyment of human rights on the Internet [https://ap.ohchr.org/documents/dpage\\_e.aspx?si=a/hrc/res/32/13](https://ap.ohchr.org/documents/dpage_e.aspx?si=a/hrc/res/32/13)
3. Association for Civil Rights (ADC) "[Freedom of speech in the digital realm](#)" In Spanish, (2016).
4. Association for Civil Rights (ADC) "[Freedom of speech in the digital realm](#)" In Spanish, (2016).
5. Association for Civil Rights (ADC) "[More Than Words: Seeking Consensus to Characterize Hate Speech](#)" In Spanish, (2020)
6. Association for Civil Rights (ADC) "[More Than Words: Seeking Consensus to Characterize Hate Speech](#)" In Spanish, (2020)
7. Association for Civil Rights (ADC) "[More Than Words: Seeking Consensus to Characterize Hate Speech](#)" In Spanish, (2020)
8. Article 13.5 of the ACHR stipulates that "Any propaganda for war and any advocacy of national, racial or religious hatred that constitutes incitement to violence or any other similar unlawful action against any person or group of persons on any grounds including those of race, color, religion, language or national origin shall be prohibited by law."
9. Association for Civil Rights (ADC) "[More Than Words: Seeking Consensus to Characterize Hate Speech](#)" In Spanish, (2020)

10. Article 20.2 of the ICCPR provides that "Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."
11. UN adopts Resolution on the promotion, protection and enjoyment of human rights on the Internet [https://ap.ohchr.org/documents/dpage\\_e.aspx?si=a/hrc/res/32/13](https://ap.ohchr.org/documents/dpage_e.aspx?si=a/hrc/res/32/13)
12. Article 4 of ICERD determines that "State Parties condemn all propaganda and all organizations which are inspired by ideas or theories based on the superiority of one race or group of persons of one color or ethnic origin, or which attempt to justify or promote racial hatred and racial discrimination in any form whatsoever, and undertake to take immediate and positive measures designed to eliminate all incitement to, or acts of, such discrimination, (...)"
13. Association for Civil Rights (ADC) ["More Than Words: Seeking Consensus to Characterize Hate Speech"](#) In Spanish, (2020)
14. The Association for Civil Rights (ADC) analyzed the legal and jurisprudential situation of these four States in the framework of the research "More than words: seeking consensus to characterize hate speech". The information provided in this contribution on legislative and judicial developments on hate speech in these countries is updated until 05/22/2020, the publication date of the report.
15. Association for Civil Rights (ADC) ["More Than Words: Seeking Consensus to Characterize Hate Speech"](#) In Spanish, 2020)
16. Argentina's [Law No. 23,592](#) on Discriminatory Acts punishes with imprisonment from one month to three years those who "encourage or incite persecution or hatred against a person or groups of persons because of their race, religion, nationality or political ideas".
17. [Law No. 9,145](#) for the prevention and punishment of violence at sporting events in Costa Rica punishes anyone who "...utters racist insults or insults that constitute any other form of discrimination contrary to human dignity or that incite hatred and violence against other human beings..." .
18. Brazil's [Law No. 7,716](#) on Anti-Racism punishes as a criminal offense the action of inducing or inciting discrimination based on race, ethnicity, religion and national origin.
19. [Law No. 82](#) of Panama criminalizes femicide and violence against women, reaffirming that "women have the right to a life of dignity and freedom from physical, sexual, psychological and patrimonial violence, to privacy, to not be subjected to torture or cruel and degrading treatment or any form of discrimination".
20. Association for Civil Rights (ADC) ["More Than Words: Seeking Consensus to Characterize Hate Speech"](#) In Spanish, (2020)
21. Association for Civil Rights (ADC) ["More Than Words: Seeking Consensus to Characterize Hate Speech"](#) In Spanish, (2020)
22. Federal Criminal and Correctional Chamber, room I, "Cherashny, G s/ procesamiento", in Rivera, Julio César "The limits of the constitutional protection of freedom of speech", Commented Constitution of the Argentine Nation, 2019.
23. ADO 26/DF, rel. Min. Celso de Mello, julgamento em 13.6.2019. - MI 4733/DF, rel. Min. Edson Fachin. Summary of the judgment in Portuguese available at <http://www.stf.jus.br/portal/informativo/verInformativo.asp?s1=racismo+e+crime&pagina=1&base=INFO>.
24. Association for Civil Rights (ADC) ["Contributions to the Special Rapporteur on Freedom of Expression on Hate Speech and Criminalization of Women in the Region"](#) In Spanish, (2022)
25. Hate speech and incitement to hatred or violence <https://www.ohchr.org/en/special-procedures/sr-religion-or-belief/hate-speech-and-incitement-hatred-or-violence>
26. Association for Civil Rights (ADC) ["Contributions to the Special Rapporteur on Freedom of Expression on Hate Speech and Criminalization of Women in the Region"](#) In Spanish, (2022)

27. Association for Civil Rights (ADC) ["Contributions to the Special Rapporteur on Freedom of Expression on Hate Speech and Criminalization of Women in the Region"](#) In Spanish, (2022)
28. Association for Civil Rights (ADC) ["Contributions to the Special Rapporteur on Freedom of Expression on Hate Speech and Criminalization of Women in the Region"](#) In Spanish, (2022)
29. Association for Civil Rights (ADC) ["More Than Words: Seeking Consensus to Characterize Hate Speech"](#) In Spanish, (2020)
30. Association for Civil Rights (ADC) ["Protecting civic space online"](#) In Spanish, (2021)
31. Association for Civil Rights (ADC) ["Contributions to the Special Rapporteur on Freedom of Expression on Hate Speech and Criminalization of Women in the Region"](#) In Spanish, (2022)
32. Association for Civil Rights (ADC) ["How to implement human rights due diligence in technology development"](#) In Spanish, (2020).
33. Association for Civil Rights (ADC) ["Protecting civic space online"](#) In Spanish, 2021)
34. The [Plan](#) recommends a threshold test consisting of six parameters, which take into account (1) the social and political context, (2) the category of the speaker, (3) the intent to incite the audience against a particular group, (4) the content and form of the speech, (5) the extent of its dissemination, and (6) the likelihood of causing harm, even imminently.
35. Association for Civil Rights (ADC) ["Contributions to the Special Rapporteur on Freedom of Expression on Hate Speech and Criminalization of Women in the Region"](#) In Spanish, (2022)
36. Association for Civil Rights (ADC) ["Contributions to the Special Rapporteur on Freedom of Expression on Hate Speech and Criminalization of Women in the Region"](#) In Spanish, (2022)