



# Protecting online civic space

A review of current threats to freedom of expression on the internet

Asociación por los Derechos Civiles



With the support of:



**July 2021**

[adc.org.ar](http://adc.org.ar)

Written by: Eduardo Ferreyra

Layout: Matías Chamorro

Cover design: El Maizal



*Protecting Online Civic Space. A Review of Current Threats to Freedom of Expression on the Internet* is published under a Creative Commons Attribution-NonCommercialShareAlike 4.0 International (CC BY-NC-SA 4.0). To view a copy visit: [creativecommons.org/licenses/by-nc-sa/4.0/deed.en](https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en)

# Table of contents

Executive Summary | 4

Introduction | 5

The criminalization of online speech | 6

Monitoring of social media posts (cyber-patrolling) | 10

Private influence on content moderation decision-making | 13

Political speech and violence | 16

Preliminary recommendations | 19

Notes | 24

## Executive summary

The Asociación por los Derechos Civiles (Association for Civil Rights - ADC) has the mission to advocate for fundamental rights by contributing to the design of public policies within the state as well as the private sector. Government responses must promote a broad and robust civic space on the Internet. This duty implies ensuring users' freedom of expression, guaranteeing favorable conditions for the creation of online groups (freedom of association) and allowing the exercise of the right to protest on platforms (freedom of assembly).

Similarly, private companies must also commit to this purpose. Large platforms have extensive power to determine the conditions under which speech can circulate over the web. The influence of these companies is even greater than that of many countries and therefore, their action can seriously hinder the dissemination of ideas and opinions. Thus, the private sector is also accountable for the protection of human rights.

The health crisis caused by Covid-19 and the deepening of political polarization around the world - with the United States as a leading example - has drawn public attention to the role of states and private platforms in addressing phenomena such as disinformation and hate speech. Actions taken by governments and companies should be subject to the closest public scrutiny, as they will influence the way debates will be handled in the future.

The recommendations that will be provided here aim to contribute to an outcome that strengthens the rule of law, the prevalence of a democratic culture and the enjoyment of an inclusive online civic space.

## Introduction

Crises are often opportunities for the emergence of new regulations by states and large companies. This time has been no exception. On the one hand, the outbreak of the coronavirus pandemic has raised concerns about the effects that disinformation may have on the population's health and well-being. On the other hand, the institutional crisis undergone by the United States has drawn attention to the issue of hate speech, discrimination and violence, coming not only from ordinary citizens but political actors around the world as well. Both phenomena gave rise to a variety of responses from public authorities and platforms. As has been said, these measures were devised for an exceptional situation, but are more than likely to continue in force after the return to normality. And if not, the precedents they leave behind will serve as an anticipation of future regulatory trends. In any case, the underlying challenge remains. Large corporations have increased their power to such an extent that several aspects of their behavior are in collision with our understanding of democracy and how it should function. Likewise, governments take advantage of the context to put forward policies that foster surveillance and suppression of dissident voices.

This paper will focus on four aspects of the current scenario:

- The criminalization of online expression
- The monitoring of posts on social networks (cyber-patrolling)
- Decision-making on content moderation in social media platforms
- State officials and violent speech

First, we intend to examine these trends and then, offer some preliminary recommendations to assist public authorities and companies in their decision-making process, with an overall aim to address the above issues in a perspective which is respectful of human rights.

## The criminalization of online speech

The use of criminal law to repress illegal behavior should be a democratic state's last resort (ultima ratio principle of criminal law). This precept should be enhanced when it comes to freedom of expression. Pressing criminal charges for something said or posted is a typical trait of authoritarian regimes that stifle political debate. For this reason, it has been stated on several occasions that "Criminal Law is the most restrictive and severe means to establish responsibility for unlawful conduct".<sup>1</sup>

In Argentina, as well as in other countries around the region, the outbreak of the pandemic gave rise to several criminal proceedings against individuals for comments made on social networks. The reason argued was the commission of "public intimidation"<sup>2</sup> by making posts aimed at creating public fear or stirring up incidents.

The most famous case in our country was the one involving Kevin Guerra. On April 7, 2020, the 20-year-old youth posted on his Twitter account: "Hey, what's up with those of us who do not receive the 10-thousand-peso bonus. The looting plan is still afoot, isn't it?",<sup>3</sup> in reference to his not collecting the emergency economic assistance.<sup>4</sup> The comment was detected by the Technological Crimes Sub-Department of the National Gendarmerie together with the Cyber Patrol Area of the Buenos Aires Province Police.<sup>5</sup> The security forces searched for posts on Twitter under the parameter "looting/lockdown/argentina" and when coming across Guerra's tweet, they issued an "early warning" on the understanding that the publication accounted for a crime.<sup>6</sup> Guerra was thus notified that he was being prosecuted for an offense punishable by 2 to 6 years of imprisonment, although the federal district attorney hearing the case later considered that no felony had been committed. The Federal Court No. 3 of Mar del Plata agreed, and at the end of 2020, Guerra was acquitted.<sup>7</sup> At the same time, criminal investigations

for public intimidation were launched against people spreading information considered false about the pandemic. In different cities of Argentina, there were charges for sending WhatsApp voice messages<sup>8</sup> or posting allegedly fake news on websites.<sup>9</sup> Even journalists suffered administrative sanctions of this type.<sup>10</sup>

The situation led the Inter-American Commission on Human Rights (IACHR) and the Office of the Special Rapporteur for Freedom of Expression (RELE) to express their concern. Both organizations reaffirmed that “the introduction of criminal types could backlash the region into a situation where opinions about state officials or matters of public interest are punished and become a means of effective restraint on the publication of ideas, criticism and information”.<sup>11</sup> Undoubtedly, the use of definitions such as “public intimidation”, “public fear” or alike has always raised suspicion because of their vagueness. But their application to online speech creates an additional set of problems. To name but a few:

- **Wider restrictions on protest.** “Public intimidation” as a criminal type has always been used to curb citizens’ right to demonstrate against government measures,<sup>12</sup> apply for jobs<sup>13</sup> or file claims in court.<sup>14</sup> Thus, its disproportionate application has led to numerous instances in which social protest is criminalized. There is no reason to believe that its use in the online sphere will be any different. **The restrictions on public gatherings imposed due to the pandemic have forced people to engage more time on the Internet doing activities hitherto done in person, and the right to protest is among these activities.** Indeed, Kevin Guerra’s case can be seen as an example of sarcastic dissent against the shortcomings in the government social assistance programs. The subsequent action carried out by the security forces shows how the type of abuse previously mentioned recurs and even may be aggravated in the digital realm, considering the huge amount of content being passed through social networks.

• **Increased risk of misapplication of public intimidation charges.** Throughout history, governments have resorted to the idea of danger and threats of violence to suppress critical and dissident voices.<sup>15</sup> For this reason, the standard for labelling speech as unlawful must be very precise. Judicial doctrine states that such should be applied only to expressions capable of producing an “imminent illegal action”.<sup>16</sup> Aside from these exceptional cases, the state cannot legitimately assume the role of deciding what sort of speech is acceptable or not. Therefore, attempts to apply the “public intimidation” statute to social media posts are most likely to be legally flawed. Such expressions are mostly abstract invocations of violence with minimal chances of real repercussions. In Argentina, several people were investigated for tweeting threats against former president Mauricio Macri and former minister Patricia Bullrich, although they were later acquitted.<sup>17</sup> In other cases, they are mere jokes that may be offensive but do not amount to being capable of stirring up violence. This does not imply, however, that expressions on social media can never provoke violence at all, and it is frequently argued that Donald Trump’s posts are a good example of this. Nonetheless, he is a public figure of great relevance, with millions of followers and within a context of unprecedented institutional crisis. In order to justify the decision to suspend Trump’s account, Facebook claimed that he had emphatically approved of the people storming the Capitol when he wrote, “We love you. You’re very special, and when calling his supporters “great patriots” and encouraging them to “remember this day forever” he was also showing endorsement for the rioters.<sup>18</sup> **The bond between a political leader and his followers warrants a more careful consideration, due to the influence that the former exerts on the latter. This condition does not apply to an ordinary citizen making similar comments on the Internet. Therefore, the same reasons that lead us to view comments made by the former**



president of the United States with concern should prompt us to reject any government attempt to punish ordinary citizens on the grounds that their words may incite riots or similar actions.

• **Risk of private standards influencing the interpretation of criminal definitions.** All major platforms have content moderation policies on violent or hate speech,<sup>19</sup> determining what is to be understood as such and which sanctions are to be imposed. In addition, companies create their own procedures to enforce these guidelines. In fact, these policies are not necessarily in line with international standards of human rights. **Private platforms use broad definitions and criteria which allow content removal in a manner we would deem questionable if it were done by the government.** For example, Facebook's hate speech policy covers expressions containing "harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, or insults towards a protected characteristic."<sup>20</sup> In turn, Twitter bans statements that "glorify, celebrate, praise or condone violent crimes, violent events where people were targeted because of their membership in a protected group, or the perpetrators of such acts".<sup>21</sup> Thus, Twitter defines glorification "to include praising, celebrating or condoning statements, such as "I'm glad this happened", "This person is my hero", "I wish more people did things like this", or "I hope this inspires others to act".<sup>22</sup> If governments or courts begin to apply definitions such as "public intimidation" to online speech, it is possible that their criteria for delimiting the term will be influenced by the manner private companies resolve similar disputes. Considering the novelty of these situations, state and judicial authorities are yet to establish clear guidelines, hence, the likelihood to turn to the criteria used in the platforms to guide their decisions. Ultimately, public and democratic regulation should model the development of content moderation policies, never vice versa.

## Monitoring of social media posts (cyber-patrolling)

The investigation and prosecution of individuals for comments on networks cannot take place without the security forces conducting social media intelligence (SOCMINT) to find expressions deemed as criminal.<sup>23</sup> This activity poses a high risk, as it can severely affect the online civic space, thus changing it into a place where critical opinions are discouraged.

Although Argentina has been using this technique for several years, the issue acquired greater repercussions in April 2020 when the National Minister of Security declared that they were resorting to cyber policing to monitor social mood.<sup>24</sup> The government was particularly concerned about possible calls for looting and riots, and drew up a protocol for action in response,<sup>25</sup> allegedly in line with international human rights standards.<sup>26</sup> However, the data protection authority recommended suspending its enforcement as it failed to adequately protect individual privacy.<sup>27</sup>

The danger that using SOCMINT poses to the creation of a free Internet environment is obvious for several reasons, among which we can mention:

- **It legitimizes police-led deterrence of speech, not behavior.** This has been informally called “cyber patrolling”, since it is likened to the prevention tasks that police normally carry out in public spaces (streets, squares, parks, etc.). This comparison, however, omits a substantial difference: the main purpose of police action is to thwart illicit behavior that may direct and immediately affect the life, physical integrity or property of a person. Law enforcers are meant to act on the street in case of assault, gunpoint robbery, vandalism, burglary, etc. In contrast, **cyber-patrolling focuses**

mainly on speech, i.e., comments, messages, posts and other forms of online interaction. Thus, the object of surveillance are expressions that, in general, do not immediately put people's property and interests at risk. Of course, this does not imply that the Internet cannot be used as a means of exchanging information to carry out illegal activities, whether online or in person. But the perils of SOCMINT leads to it eventually being used for social control or to restrain legitimate speech.

- **Police cannot be held accountable for their actions with ease:** citizen control of law enforcement agents is essential for people to maintain trust in the system. Several aspects help in this task when the patrolling is done in physical spaces, Firstly, officers are clearly identifiable by their uniforms and badges, and move around in vehicles which are identifiable as well. In this way, individuals can monitor their actions. Secondly, when patrolling happens in physical places, people can witness their activity, and finally, eventual situations of police abuse can be recorded by photo or video. For instance, journalism has played an important role in revealing how police may violate human rights.<sup>28</sup> These circumstances are absent in cyber patrolling. We have no means of knowing if police intelligence is currently viewing our comments or posts. We cannot access the offices where SOCMINT occurs. We have no information about which pages are visited and if we do, we can only rely on what the authorities say. Chances of citizens or journalists recording a possible case of abuse are slim. **There is a secrecy which is inherent to the way cyber-patrolling is conducted. Thus, we must be very cautious in introducing its use.**

- **It assumes that content appearing in publicly available sources is not subject to a high degree of data protection.**The main argument for SOCMINT is that social media posts are freely accessible information and preventing law enforcement from monitoring them would be analogous to blindfolding police officers

walking on the streets. The comparison, however, is inaccurate. The uniqueness of social media platforms is that the information hosted is publicly available in principle, but in fact, the sites are operated by private corporations imposing their own rules. Hence, there is a hybrid nature which calls for considering the issue differently. Protection of personal data is of major importance in the digital era, since it allows people to control the use and recipients of their own data, according to the principle of informational self-determination. **People express their political opinions, sexual preferences or health status on social networks. The sensitivity of this data requires that its processing should not be subject to covert surveillance or screening by the authorities.** In its statement on the protocol for cyber-patrolling, the data protection authority stressed that sensitive information enjoys special protection and therefore, any processing is subject to the strictest scrutiny<sup>29</sup> In that sense, any policing activity is more like an officer standing on the sidewalk and looking through a house window all day to see what is going on inside. And likewise, the equivalent of “drawing the curtains” would be users closing accounts or refraining from expressing their opinions on controversial issues.

## Private influence on content moderation decisions

Both from civil society<sup>30</sup> as well as from academia,<sup>31</sup> the need to adopt international human rights standards for decisions on content moderation has been emphasized. While it is obvious that no single solution will magically solve all the problems of online speech governance, the human rights perspective has certain advantages. In terms of legitimacy, it provides us with a global language on which we can rely to start a discussion on an equal basis. Disputes about post removals, account suspensions and so on should be resolved on standards that the international community as such could recognize as its own. On the other hand, it would be a foundation for the design of regulatory policies that keep governments from falling into authoritarian temptations. Although this aspiration should be conceived as an indispensable first step, it is insufficient in itself. **Any implementation of norms -including those concerning human rights- will be thwarted if there is no decentralization of the power that companies hold on deciding about content moderation.** A change in the rules without altering the current state of affairs regarding the bodies and the procedures by which these rules are enforced, may lead to the following situation:

- **Increased power of social media platforms.** human rights standards need to be interpreted before their application. The question of who is in charge of such enforcement is of crucial importance. If the same companies were to continue deciding unilaterally and exclusively, changing their private guidelines for the International Covenant on Civil and Political Rights will not have a significant effect. Discretionary decisions would continue, now in the name of human rights. The power of these private corporations would actually be augmented, due to the new source of legitimacy.

- **Increased likelihood that external content moderation mechanisms influence the resolutions of public authorities.**

. Companies deal with thousands of cases on a daily basis. This intensive work has prompted them to develop expeditious procedures for resolving disputes over content. In addition, some of them have adopted external overseeing mechanisms for their decisions. A prime example of this is Facebook's Oversight Board (OB), which has applied community norms but resorting to international human rights standards as a reference. Judicial authorities - especially those in our country and region - do not have equivalent experience, inclining them to be influenced by such interpretations had they to intervene. Consequently, judgements of official authorities could be a mere replica of previous decisions made by private mechanisms already established.

- **It may prevent users from playing a central role in decision-making.** Platforms have admitted being aware of the problem caused by online content governance being exclusively in their hands. For that reason, some of them have created mechanisms so that external bodies can oversee those resolutions with binding powers. The most outstanding example so far is Facebook setting up its Oversight Board, which issues the final ruling in certain controversies over content removal.<sup>32</sup> This is a positive step, insofar as the platform recognizes its lack of legitimacy to make such sensitive decisions. The solution, however, does have its flaws: although the board members have sufficient background and expertise to deal with these issues and there is no evidence that their independence may be impaired, the task continues to be in the hands of a small number of people. **Responses to such controversial and global problems should aim to involve citizens significantly in decision making.** The Internet already has alternative models of content moderation where individuals have a good deal of influence. For example, Wikipedia's dispute resolution system relies fundamentally on the discussion and decision of

users and volunteers. Of course, this model is not suitable for exact replication on platforms with very different characteristics. But it is relevant to point out that there are no insurmountable barriers to broadening citizen engagement in the resolution of content moderation conflicts.

## Political speech and violence

The decision of the major companies to expel former U.S. President Donald Trump from their platforms has led the discussion over content moderation to focus on the relationship between politicians' discourse and violence. On the one hand, we are told that citizens should have the easiest possible access to whatever is said by public authorities of such relevance. Thus, banning a president from communicating through the most popular media amounts to a significant restriction. On the other hand, such leaders have enough power over a large number of people – unlike average citizens – to stir up violence through inflammatory comments. There are several possible responses to this dilemma, but it is essential to point out a number of concerns underlying the process of its resolution, since it is those that will come into play on creating a healthy online civic space. To mention a few:

- **Ability of social media platforms to assess the context.**

According to the companies, Trump's expulsion came forth after a study of the U.S. situation showed that his statements had a clear potential to provoke acts of violence.<sup>33</sup> This action demonstrates the importance of a proper evaluation before deciding on content moderation. Nonetheless, this is something that can be done effectively – to some extent – by major companies in the United States, since it is there where most of them are based. In contrast, their ability to understand the effects of speech in given social environments or political climates of other countries is diminished. Consequently, decisions may be made without the proper information of such scenarios, especially those with deep differences from that of the U.S.

- **Local solutions with global impacts.** Although Trump's deplatforming was on grounds of avoiding resurgence of violence in the domestic scene, the consequences reached users all over the



world. In other words, the citizens of Argentina, Denmark, India, Australia and elsewhere were deprived of reading the outgoing president's posts due to a particular situation in the United States. There was obviously no chance that those countries could be exposed to any sort of violence emanating from Trump's tweets. Therefore, in a certain perspective, the measure could be reckoned as disproportionate. Indeed, platforms implement their decisions with uniformity in every country that they operate. Aside from economic reasons, there are technical and operational factors that justify this policy. Moreover, a segmentation of the norms – e.g., Trump banned on Twitter in the US but allowed anywhere else – is not a solution that can be easily implemented and may pose bigger problems than those it intends to solve. Nonetheless, it is a phenomenon that deserves consideration. **Global court-ordered removals are often questioned because they are said to lead to a “race to the bottom” in the application of norms restricting free speech. The conclusion should be no different on decisions made by a private company.**

- **Consistency of criteria among companies.** Increasingly, large platforms are making similar decisions on content moderation issues. The era in which different strategies and guidelines were adopted to deal with disinformation and determine what content can remain on their networks is giving way to a coincidence – either spontaneous or deliberate – of the actions to be taken, a phenomenon that has been described by some authors as the emergence of “content cartels”.<sup>34</sup> If this trend towards uniformity deepens, the ensuing lack of diversity might lead to a number of problems. First of all, it would expand the effects of a wrong decision, e.g., the arbitrary elimination of content from one platform cannot be mitigated by another's decision to maintain it. Secondly, it reduces our chances to experiment, compare, and evaluate the benefits and drawbacks of each moderation policy. Thus, the diversity which is necessary to evolve in the

understanding of a subject such as Internet governance, which is only in its early stages, is impoverished. Finally, the aforementioned “cartelization” increasingly reveals the high degree of concentration in the hands of few players. Therefore, uniform action by the major platforms can shut down an individual from any meaningful interaction on the most popular sites.<sup>35</sup>

## Preliminary recommendations

The challenges are complex, and no magic solution is available. But there are actions that, if implemented, can contribute to the design of an online civic space respectful of people's rights. Below are provided some preliminary recommendations to address the problems mentioned in previous sections.

### Criminalization of online speech

- **Do not resort to concepts provided for in the Criminal Code such as “public intimidation” or similar to prosecute posts made on social networks.** Prosecutors and criminal courts must implement the standards of freedom of expression when deciding to investigate or press charges against individuals for their activity on social networks. These principles indicate that only speech that is aimed at producing imminent illegal behavior and has a certain likelihood of doing so can be indicted. Under this standard, the vast majority of comments made on the Internet should not be criminalized even if they are unpleasant or irritating. On the other hand, it would be advisable to enact a law that clearly establishes that expressions referring to matters of public interest are not included in the public intimidation definition. As in the case of slander and libel, this clarification would be useful to avoid criminal persecution of persons using the Internet to spread information.
- **Consolidate and expand the scope of the right to protest on the Internet.** In a country with deep inequalities like Argentina, less favored social groups have few means to make their voices heard to political authorities, and online communication could serve as a remedy to this shortcoming, however minimal. This should be a reason for courts and the authorities to especially protect those expressions of dissent. Examples such as Kevin

Guerra's tweet could have been used to address the underlying problem of social assistance to people affected by lockdown instead of launching criminal investigations for jokes that had no chance of generating any serious impact.

## Cyber-patrolling

- **Internet policing should be subject to strict scrutiny: it cannot be standardized the same way as street patrolling for one simple reason.** our law forbids gathering information or storing data on people for their legitimate activities.<sup>36</sup> This is precisely what happens when mass online monitoring is not conducted within a particular criminal investigation. It is the government who must prove that less costly options are not available to its duty to ensure public order. Similarly, it must provide that its actions are limited and specific. In terms of protection of the civic space, this implies that cyber patrolling should not be used to actively track citizens for alleged speech-related crimes.
- **Create independent accountability mechanisms with citizen participation.** Internet policing must have an external system of checks and balances to oversee that law enforcement agencies guarantee individual rights are protected. This implies a number of conditions, but as a start, the activities and policies related to social media should be publicly disclosed, so as to be impartially scrutinized. On the other hand, it requires the creation of oversight committees formed by citizens to ensure a democratic control of police procedures. Finally, it calls for human rights impact assessments and audits by independent entities with a solid record and expertise.

## Decision-making processes on content moderation

- **Deepen and develop the existing standards on freedom of expression.** Governments, judiciaries, and human rights protection systems – each within their scope – must continue their work in specifying and delimitating existing principles on free speech. Otherwise, the vagueness of terms can be exploited by large platforms to justify decisions made in view of their own interests while proclaiming to be in line with the international human rights framework. Indeed, issues such as hate speech or disinformation generate legitimate disagreements on how to deal with them. However, this does not prevent public authorities from promoting discussion and formulating proposals that can be debated open and democratically.

- **Explore options to increase the democratic legitimacy of the bodies and procedures by which content moderation policies are designed and applied.** However fair they may be, the content moderation decisions of large digital platforms are lacking in democratic legitimacy. Civil society organizations, academia and individuals in general should be given substantial participation in the internal processes that each platform performs to moderate content. “Substantial”, in this case, means participation beyond the role of mere advisory boards, enjoying real and binding power of decision over the most controversial situations. Equally, the democratization of online discourse governance models demands that platforms create novel mechanisms to incorporate the citizenry in their decision-making. While being too early to deem this as the most effective solution, it does appear as a suitable path to follow for tackling – along with other measures – the aforementioned democratic deficit.

## Political speech and violence

• **In contexts of high conflict, public officials have a special duty to prevent their statements from affecting fundamental rights.** Freedom of expression is a right for the entire population and therefore, includes public officials. However, leaders have the obligation to protect citizens' fundamental rights not only through their actions but also through their words. As the Inter-American System has held, in situations of high political and/or social polarization, state officials should avoid speech that put certain social groups in a situation of vulnerability.<sup>37</sup> For example, the Inter-American Court of Human Rights has indicated that public officials should refrain from making statements that, in a context of social polarization, increase the risk of journalists and media outlets to suffer aggressions from third parties. However, these decisions must be backed up by an understanding of the local context. Thus, regular interaction with national actors, with the possibility of influencing companies' decisions to a certain degree, is imperative.

• **Measures to curb expressions of violence by public officials must be sufficiently adequate so as not to affect the principle of proportionality.** The indefinite suspension of an account implies the most severe sanction in terms of guaranteeing freedom of expression. Therefore, platforms must explore other less harmful alternatives that can achieve the same end. Donald Trump's expulsion is a clear example of this. If the argument for shutting down his accounts was the extraordinary political context in the United States, the period of suspension could have been limited to the period of transition of power. And had it been estimated that the danger lingered, the ban could have been extended for another period and so on. An indefinite exclusion leaves companies with absolute discretion to assess possible new

situations. In contrast, a time-bound decision may force platforms to analyze changes in the social context and be in the need to publicly justify the extension of a ban.

\* \* \*

## Notes

1 Inter-American Court of Human Rights, Palamara Iribarne Case v. Chile. Merits, Reparations and Costs. Judgment of November 22, 2005. Series C No. 135, para. 79, [https://www.corteidh.or.cr/docs/casos/articulos/Seriec\\_135\\_esp.pdf](https://www.corteidh.or.cr/docs/casos/articulos/Seriec_135_esp.pdf)

2 Article 211 of the Argentine Criminal Code states that “anyone signaling, raising alarm, threatening to commit an offense which endangers public order or resorting to any suitable means to instill public fear or incite riots or disorder, , shall be punished with imprisonment of two to six years” (art. 211).

3 <https://twitter.com/KevinGuerra99/status/1247709948554903554>

4 *La Nación*, “Ciberpatrullaje: Made a joke on Twitter and now is under prosecution”, April 15, 2020, <https://www.lanacion.com.ar/sociedad/ciberpatrullaje-hizo-chiste-saqueos-twitter-le-abrieron-nid2354402/>

*Clarín*, “Cyber-patrolling - ‘The looting plan is still afoot, isn’t it?’, the tweet which earned a Balcarce youth criminal prosecution”, April 14, 2020, [https://www.clarin.com/policiales/-sigue-pie-saqueo-tuit-valio-causa-judicial-joven-balcarce\\_0\\_Q2GKqAle1.html](https://www.clarin.com/policiales/-sigue-pie-saqueo-tuit-valio-causa-judicial-joven-balcarce_0_Q2GKqAle1.html)

5 *La Izquierda Diario*, “The Cyber-patrolling effect. Kevin Guerra: ‘The Investigations Department told me that the president didn’t like my tweet and now I am charged’”, April 15, 2020, <https://www.laizquierdadiario.com/Kevin-Guerra-La-DDI-me-dijo-al-Presidente-no-le-gusto-tu-tuit-y-ahora-estoy-procesado>

6 CELS, “Federal Courts dismiss Kevin Guerra for expressions on Twitter”, January 14, 2021, <https://www.cels.org.ar/web/2021/01/la-justicia-federal-sobreseyo-a-kevin-guerra-por-sus-expresiones-en-twitter/>

7 Ibid.



8 *Diario El Tiempo*, "Another criminal case being investigated for 'public intimidation'", March 29, 2020, <https://www.diarioeltiempo.com.ar/nota-se-esta-instruyendo-otro-sumario---penal-por-intimidacion-publica-163794>

*El Popular*, "Two policemen and local of Azul charged for spreading false information", March 25, 2020, <http://www.elpopular.com.ar/142512>

9 *Infobae*, "Woman's home raided for spreading fake news about coronavirus. Secretary of Energy found sleeping there", April 4, 2020, <https://www.infobae.com/politica/2020/04/04/allanaron-a-una-mujer-por-difundir-noticias-falsas-sobre-el-coronavirus-y-encontraron-al-secretario-de-energia-durmiendo-en-su-casa/>

*Misiones Online*, "Coronavirus: house in Oberá raided for fake news and National Secretary of Energy Sergio Lanziani found there", April 4, 2020, <https://misionesonline.net/2020/04/04/coronavirus-por-una-causa-judicial-de-fake-news-allanan-una-vivienda-en-obera-y-encuentran-a-sergio-lanziani-secretario-de-energia-de-la-nacion-violando-la-cuarentena1/>

10 *Página 12*, "Mayor of Pichanal fines social communicator", <https://www.pagina12.com.ar/274042-el-intendente-de-pichanal-multo-a-un-comunicador>

*Diario Clarin*, "Controversy in Salta: mayor fines journalist 40.000 pesos for reporting coronavirus cases in his town", 24 de Junio de 2020, [https://www.clarin.com/politica/salta-intendente-multo-40-mil-periodista-reportar-casos-coronavirus-ciudad\\_0\\_bev8G3nMu.html](https://www.clarin.com/politica/salta-intendente-multo-40-mil-periodista-reportar-casos-coronavirus-ciudad_0_bev8G3nMu.html)

11 Press Release R78/20, "IACHR and its SR FOE express concern about restrictions on freedom of expression and access to information in States' response to the COVID-19 pandemic", April 18, 2020, <https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=1173&IID=2>

12 CELS, "The Right to Protest in danger", December 15, 2017, <https://www.cels.org.ar/web/2017/12/el-derecho-a-la-protesta-en-peligro/>

- 13 *Diario Judicial*, "Social protest won the battle", December 9, 2005, <https://www.diariojudicial.com/nota/51879>
- 14 *Página/12*, "A statute for criminalizing social protest", July 5, 2017, <https://www.pagina12.com.ar/48389-una-figura-para-criminalizar-la-protesta>
- 15 In the United States, socialist and communist activists were criminally convicted for distributing pamphlets opposing compulsory military service to fight in World War I. The convictions were upheld by the U.S. Supreme Court under the "clear and present danger" standard, cf. The convictions were upheld by the U.S. Supreme Court under the "clear and present danger" standard, see *Schenck v. United States*, 249 U.S. 47 (1919), <https://www.law.cornell.edu/supremecourt/text/249/47> y *Abrams v. United States*, 250 U.S. 616 (1919), <https://www.law.cornell.edu/supremecourt/text/250/616>
- 16 The "imminent legal action" standard is enshrined by the U.S. Supreme Court in *Brandenburg v. Ohio*, <https://www.law.cornell.edu/supremecourt/text/395/444>. In Argentina it has been adopted to analyze cases of freedom of expression on the Internet such as "Vita, Leonardo G. Y González Eggers, Matías s/procesamiento" of the 1st Chamber of the Federal Court, <http://www.hfernandezdelpech.com.ar/JurisprudenciaArgLiberExpreFalloCamaraFed.htm>
- 17 *La Vaca*, "To Tweet against Macri is not a crime", May 22, 2019, <https://lavaca.org/notas/tuitear-contramacri-no-es-delito/>
- Perfil*, "Youth imprisoned for tweeting against Macri is declared innocent", March 23, 2019, <https://www.perfil.com/noticias/sociedad/declaran-inocente-al-joven-que-escribio-un-tuit-contramacri.phtml>
- 18 For a summary of Facebook's explanations, see the Oversight Board's ruling on the case, <https://www.oversightboard.com/sr/decision/2021/001/pdf-english>
- 19 Website <https://letrachica.digital/> monitors and analyzes content moderation norms of Facebook, Twitter and YouTube.

20 Facebook Community Standards. Part III Objectionable Content. Hate Speech, [https://www.facebook.com/communitystandards/objectionable\\_content](https://www.facebook.com/communitystandards/objectionable_content)

21 The Twitter Rules. Glorification of Violence Policies, <https://help.twitter.com/es/rules-and-policies/hateful-conduct-policy>

22 Ibid.

23 Cfr. ADC, *Unseen followers. A first approach to government use of Open-source intelligence (OSINT) and social media intelligence (SOCMINT)*, October 2018. <https://adc.org.ar/en/reports/unseen-followers-a-first-approach-to-government-use-of-open-source-intelligence-osint-and-social-media-intelligence-socmint/>

24 *Infobae*, "Controversial revelation: the Minister of Security admitted that the forces under her charge carry out cyber patrols on social networks to 'detect social mood'", April 8, 2020, <https://www.infobae.com/politica/2020/04/09/polemica-revelacion-la-ministra-de-seguridad-admitio-que-las-fuerzas-a-su-cargo-realizan-ciberpatrullaje-en-redes-sociales-para-detectar-el-humor-social/>

*Perfil*, "Frederic habló de 'ciberpatrullaje para medir humor social' y desató otra polémica", 8 de Abril de 2020 <https://www.perfil.com/noticias/actualidad/sabina-frederic-ciberpatrullaje-redes-sociales-medir-humor-social-no-me-van-a-callar.phtml>

25 Resolution 144/2020: Open-Source Cyber Patrolling Protocol. National Ministry of Security, May 31, 2020, <https://www.boletinoficial.gob.ar/detalleAviso/primera/230060/20200602>

26 Télam, "Security created protocol to detect cybercrimes related to coronavirus pandemic", June 2, 2020, <https://www.telam.com.ar/notas/202006/471150-crean-un-protocolo-para-detectar-delitos-relacionados-con-la-pandemia-del-covid-19-en-internet.html>

*Página/12*, "Coronavirus: A protocol for cyber-patrolling", June 2, 2020, <https://www.pagina12.com.ar/269916-coronavirus-un-protocolo-para-el-ciberpatrullaje>

27 Public Access to Information Agency, Reply to Note NO-2020-41096462-APN-UGA-MSG – Consultative Board for the evaluation and monitoring of the General Protocol for Police Crime Prevention with the use of Open Digital Sources, July 23, 2020, <https://www.argentina.gob.ar/sites/default/files/no-2020-47326285-apn-aaip.pdf>

28 *La Izquierda Diario*, "Maxi Kosteki y Dario Santillán. Hostile Avellaneda: the story of Sergio Kowalewski's photo", June 26, 2020, <https://www.laizquierdadiario.com/Avellaneda-hostil-la-historia-de-la-foto-de-Sergio-Kowalewski>

*Página/12*, "Images proving who fired the shots", June 26, 2005, <https://www.pagina12.com.ar/diario/elpais/1-52917-2005-06-26.html>

29 Public Access to Information Agency Reply to Note NO-2020-41096462-APN-UGA-MSG - Consultative Board for the evaluation and monitoring of the General Protocol for Police Crime Prevention with the use of Open Digital Sources, <https://www.argentina.gob.ar/sites/default/files/no-2020-47326285-apn-aaip.pdf>

30 Observacom et al. "Standards for democratic regulation of large platforms to ensure freedom of expression online and a free and open Internet", July 2020, <https://www.observacom.org/wp-content/uploads/2020/09/Estandares-para-una-regulacion-democratica-de-las-grandes-plataformas.pdf>

31 David Kaye, *Speech Police: The Global Struggle to Govern the Internet*, June 2019

32 Content Advisory Board, "We guarantee respect for freedom of expression on the basis of independent judgment". <https://oversightboard.com/>

33 Cfr. Mark Zuckerberg's post of January 7, 2021, [https://www.facebook.com/story.php?story\\_fbid=10112681480907401&id=4](https://www.facebook.com/story.php?story_fbid=10112681480907401&id=4)

34 Cfr. Evelyn Douek, "The rise of content cartels", February 11, 2020, <https://knightcolumbia.org/content/the-rise-of-content-cartels>

35 Cfr. CELE, "The shunning of Trump", January 30, 2021, <https://observatoriolegislativocele.com/el-ostracismo-de-trump/>

36 Art. 4.2 Law 25520, <http://servicios.infoleg.gob.ar/infolegInternet/anexos/70000-74999/70496/texact.htm>

37 Cf. Office of the Special Rapporteur for Freedom of Expression, Inter-American Commission on Human Rights, "Inter-American Legal Framework on the Right to Freedom of Expression", 2010, paragraphs 201 to 212. [http://www.oas.org/es/cidh/expresion/docs/cd/sistema\\_interamericano\\_de\\_derechos\\_humanos/index\\_MJIAS.html](http://www.oas.org/es/cidh/expresion/docs/cd/sistema_interamericano_de_derechos_humanos/index_MJIAS.html)



por los Derechos Civiles

[adc.org.ar](http://adc.org.ar)