



Written Submission for the Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights (IACHR/RFOE), the OAS Department of Electoral Cooperation and Observation (DECO) and the Department of International Law (DIL)

Public Consultation:

Disinformation in electoral contexts

January 31, 2019

Submitted by Electronic Frontier Foundation

Introduction

The Electronic Frontier Foundation (EFF) is the leading international non-governmental organization dedicated to protecting individual's fundamental rights through impact litigation, policy analysis, grassroots activism, and technology development. With nearly 40,000 dues-paying members around the world and a social media reach of well over 1 million followers across different social networks, EFF has been active since 1990, engaging directly with digital users worldwide and providing leadership on cutting-edge issues of free expression, privacy, and human rights.

The public consultation seeks to receive general comments or work in progress that address one or more of the following issues: (i) empirical studies on mass dissemination of false information, especially in electoral contexts; (ii) human rights principles or standards applicable to the problem; and (iii) possible actions and actors involved. In response, this submission focuses mainly on the normative dimension, offering also recommendations regarding actions and operational measures.

1. The inter-American Legal Framework on freedom of expression in the face of disinformation in electoral contexts

The public consultation calls stakeholders to submit their positions on a crucial question—whether current international standards on freedom of expression are adequate to tackle the disinformation phenomenon that is occurring during elections. EFF presents this submission to express that the inter-American Legal Framework remains a comprehensive and sound baseline to address the challenges posed by the dissemination of disinformation. Inter-American case law, opinions, the Declaration of Principles on Freedom of Expression alongside their interpretation, as well as developments on this right performed by the Office of the Special Rapporteur for Freedom of Expression provide substantial guidelines to understand and deal with the “fake news” phenomenon.

Although the spread of "fake news" raises serious concerns, especially regarding its potential impact on elections, it is not a new problem. The Americas¹ have long

¹ As stressed in the Open Letter from Latin American and Caribbean Civil Society Representatives on the concerns around the discourse about fake news and elections. Available at: <https://direitosnarede.org.br/c/openletter-latinamericacivilsociety-ifg2017/>

dealt with the challenges of disinformation campaigns coordinated by State and non-State actors. Understanding how errors and inconsistencies in news reporting affect individual human rights is an ongoing challenge. Fortunately, the inter-American System has analyzed issues concerning free expression and misinformation, and therefore can provide a set of comprehensive guidelines for addressing the spread of disinformation to governments and society.

As outlined in the 2017 Vienna’s Joint Declaration on Freedom of Expression and “Fake News,”² the Internet and other digital technologies played a transformative role in supporting an individual’s ability to access and disseminate information and ideas. The way in which the Internet fostered the ability to seek, receive, and impart information, and enabled the collaboration, creation, and sharing of knowledge, represents a “democratization” of freedom of expression.³ The increase or abundance of information should not be deemed, itself, a problem. The responses and solutions to the “fake news” phenomenon—if they’re unable to adhere to proper human rights standards—could be. Responses to disinformation must be precisely calibrated to their target and balance the rights involved.

To that end, we highlight three reasons why the inter-American framework regarding free expression is an adequate baseline to meet this challenge, providing:

- (i) Substantive parameters for circumscribing disinformation;
- (ii) Adequate balance between protected rights; and
- (iii) Sound elements for addressing electoral contexts.

The explanation for each one is detailed in the following sections.

1.1 Substantive parameters for circumscribing disinformation

Any limitation to freedom of expression due to disinformation in electoral contexts should encompass an effort to clearly and precisely identify what the purported *disinformation* is about. This is an instrumental starting point in order to address

² Joint Declaration on Freedom of Expression and “Fake News,” Disinformation and Propaganda, adopted by the four special rapporteurs on freedom of expression (UN, OAS, OSCE, and ACHPR). Available at: <https://www.oas.org/en/iachr/expression/showarticle.asp?artID=1056&IID=1>

³ IACHR. Office of the Special Rapporteur for Freedom of Expression. *Standards for a free, open, and inclusive Internet*. OEA/Ser.L/V/II CIDH/RELE/INF.17/17. 15 March 2017. Para. 81.

the legitimate concerns involved without exerting a chilling effect on freedom of expression. In that sense, some important parameters are offered by the inter-American framework.

First, only facts, and not opinions, are susceptible to judgments of truthfulness or untruthfulness.⁴ An individual should not be held liable when the information challenged is a value judgement rather than a factual assertion, since subjective opinion cannot be proven true or false.⁵ Value judgments also include humorous and satirical speech.⁶

Second, “erroneous” speech should only be punishable when found to be produced with “actual malice.”⁷ According to the *standard of actual malice*, there must be proof that the person expressing the opinion did so with the intent to cause harm and the knowledge that she was disseminating false information, or that she did so with a reckless disregard for the truth of the facts.⁸ Hence, the demonstration of actual subjective knowledge of falsity (or actual subjective serious doubt as to the truth) stands out as a key prerequisite for establishing subsequent liability.

Building upon that, the determination of intent should also take into account relevant aspects of current configurations of the disinformation phenomenon. The concerns that it raises are intimately related to its potential ability to manipulate public opinion and meddle with public interest decision-making. For this reason, for speech to be punishable, the intention to deceive and the knowledge of falsehood must be proven. Further, the extent to which the information was disseminated is also an important defining element.

When proposing a delimitation of the phenomenon, the former Special Rapporteur Catalina Botero underscores the conjunction of three elements: *material* (massive diffusion of false information), *cognitive* (knowledge of the falsehood), and

⁴ IA Court H. R., Case Kimel v. Argentina. Merits, Reparations and Costs. Judgment of May 3, 2008. Series C No. 177. para. 93; IA Court H. R., Case of Tristán Donoso Vs. Panama. Preliminary Objection, Merits, Reparations and Costs. Judgment of January 27, 2009. Series C No. 193. para. 124.

⁵ IACHR. Background and Interpretation of the Declaration of Principles. Para. 47.

⁶ Id. Para. 48.

⁷ Id., para 35.

⁸ IACHR. Office of the Special Rapporteur on Freedom of Expression. *The Inter-American legal framework regarding the right to freedom of expression*. OEA/Ser.L/V/II CIDH/RELE/INF. 2/09. 30 December 2009. Para. 109.

volitional (intent to deceive the public or one part of it).⁹ This combination should be considered when framing *what* should be tackled.

Regarding *how* to best address the phenomenon, we can also rely on the guidelines outlined in the inter-American framework. The last two reasons will be analyzed together.

1.2. Adequate balance between protected rights and sound elements for addressing electoral contexts

The Inter-American Court of Human Rights (IA Court) and the Inter-American Commission of Human Rights (IACHR) have recognized that the right to freedom of expression is instrumental to democratic institutions. It is an indispensable requirement for the very existence of a democratic society,¹⁰ since open and free discussions prepare society for the conflicts that could otherwise lead to its destruction.¹¹ In this sense, the IACHR has underscored the triple function of freedom of expression in democratic systems:

*a) as an individual right that reflects the human capacity to think about the world from our own perspective and communicate with one another; b) as a means of open and uninhibited deliberation about matters of public interest; c) as an essential instrument for the guarantee of other human rights, including political participation, religious freedom, education, culture, equality, and others.*¹²

This is why Article 13 of the American Convention sets a general presumption of coverage. By that, in principle, no persons, groups, ideas or means of expression are excluded *a priori* from public debate. The right to freedom of expression must be upheld, not only in instances involving favorable or indifferent ideas and information, but also in cases of offensive, shocking, and disturbing speech. This is required for the diversity of opinions, tolerance, and a spirit of openness--all of

⁹ The delimitation proposed is that disinformation consists of “the massive publication or diffusion of false information of public interest the publication or mass dissemination of false information of public interest, knowing its falsity and with the intent of deceiving or confusing the public or a fraction thereof”. Botero Marino. La regulación estatal de las llamadas “noticias falsas” desde la perspectiva del derecho a la libertad de expresión. In: OAS. *Libertad de expresión: a 30 años de la Opinión Consultiva sobre la colegiación obligatoria de periodistas*, p. 69.

¹⁰ IACHR. Declaration of Principles on Freedom of Expression.

¹¹ IACHR. Background and Interpretation of the Declaration of Principles. Para. 24.

¹² IACHR. Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights. *Standards for a Free, Open and Inclusive Internet*. OEA/Ser.L/V/II CIDH/RELE/INF.17/17. 15 March 2017. Para.70.

which are required in a democratic society.¹³ The same is valid for prior conditioning of expressions in the name of truthfulness or correction.

In a landmark advisory opinion regarding free expression, the IACHR stated:

*A system that controls the right of expression in the name of a supposed guarantee of the correctness and truthfulness of the information that society receives can be the source of great abuse and, ultimately, violates the right to information that this same society has.*¹⁴

Principle 7 of the Declaration of Principles on Freedom of Expression approved by the IACHR enshrines such understanding. The Office of the Special Rapporteur, when interpreting it, has asserted that requiring truthfulness as a prerequisite for speech could lead to virtually automatic censorship of all information that cannot be proven as true. All public debate based primarily on ideas and opinions could fall under this overreaching restraint. Such a restraint would fail to consider that even information regarding concrete events, that may be factually proven, is subject to various interpretations.¹⁵ And the Rapporteur goes further:

*Moreover, even assuming that it is possible to determine the truth about everything, the debate and exchange of ideas clearly is the best method to uncover this truth and to strengthen democratic systems based on plurality of ideas, opinions and information. Prior imposition of a requirement to report only the truth expressly precludes the possibility of engaging in the debate necessary to reach it. The prospect of penalties for reporting on a subject that free debate later shows to be incorrect creates the potential that informants will engage in self-censorship to avoid penalties, with the attendant harm to citizens who are unable to benefit from the exchange of ideas.*¹⁶

The arguments and concerns presented are exacerbated during elections. Not only do the intensity of debates and controversies increase, but so does the need for a society to openly and vigorously discuss itself and its representatives. Curbing this process by imposing prior constraints in speech would severely impair political participation and, ultimately, the democratic system.

¹³ IACHR. Office of the Special Rapporteur on Freedom of Expression. *The Inter-American legal framework regarding the right to freedom of expression*, supra note 8, para. 30 and 31.

¹⁴ IACHR, Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism, Advisory Opinion OC-5/85 Series A, No. 5. Para. 77.

¹⁵ IACHR. Background and Interpretation of the Declaration of Principles. Para. 32.

¹⁶ Id., para 33.

In addition to violating the right of each individual to express himself, prior censorship impairs the right of each person to be well informed.¹⁷ Under this protection, prior censorship cannot be applied to counter abuses of free expression, but such cases should be subject to subsequent imposition of liability.¹⁸ Such restrictions must be expressly established by law, where the ends sought to be achieved are legitimate, and the means for establishing liability are necessary to achieve those ends.¹⁹

The preservation of a democratic environment in which dissenting opinions and minority voices can still be heard inspires art. 13(5) of the American Convention. Without prejudice to the presumption of coverage of all forms of human speech by freedom of expression, propaganda for war and advocacy of hatred that constitute incitements of lawless violence do not enjoy protection under Article 13.²⁰ Incitement of violence is a requirement for such restriction and imposed sanctions under this charge must be backed by actual, objective, and strong proof that it constitutes an incitement to violence with the intent and ability to cause such violence.²¹

Therefore, *disinformation* on its own cannot be considered unlawful speech under the American Convention. If, in specific cases, it implicates a real and proven incitement of violence, then, regardless of its veracity, the expression may be subjected to proper sanction.²² This is a key standard for addressing disinformation, and should be stressed in electoral processes.

It should be noted that recent attempts to curb the spread of disinformation have not respected human rights. According to the 2018 Freedom on the Net report, 13 countries prosecuted citizens for spreading "fake news," and enforced criminal penalties in these instances.²³ Internet shutdowns have also been implemented in

¹⁷ IACHR, Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism, Advisory Opinion OC-5/85 Series A, No. 5. Para. 54.

¹⁸ IACHR. Background and Interpretation of the Declaration of Principles. Para. 22.

¹⁹ *Id.*, para 24.

²⁰ IACHR. Office of the Special Rapporteur on Freedom of Expression. *The Inter-American legal framework regarding the right to freedom of expression*, supra note 8, para. 57 and 58.

²¹ *Id.*

²² Botero Marino. La regulación estatal de las llamadas "noticias falsas" desde la perspectiva del derecho a la libertad de expresión, supra note 9, p. 72.

²³ Freedom House. *Freedom on the Net 2018 - The Rise of Digital Authoritarianism*. October 2018, p. 12. "Rwandan blogger Joseph Nkusi was sentenced in March 2018 to 10 years in prison for incitement to civil disobedience and the spreading of rumors, having questioned the state's narrative on the 1994 genocide and criticized the lack of political freedom in the country. Police in Bangladesh arrested media activist Shahidul Alam only hours after he live-streamed a video on Facebook in which he decried a disproportionate crackdown on protesters in August."

an effort to stop the spread of disinformation. Authorities in India and Sri Lanka temporarily shut down mobile networks or blocked social media apps during protests to allegedly halt the flow of disinformation and incitement of violence. Regarding legislative measures, 17 governments approved or proposed laws restricting online media in the name of countering "fake news" and online manipulation.²⁴

The 2018 elections in Brazil spurred a significant number of draft bills related to "fake news."²⁵ The vast majority of the 15+ proposals treat disinformation as a criminal offense. One bill proposes penalties of up to eight years in prison for "creating, disclosing or sharing false news that may modify or misrepresent the truth about a natural or legal person that affects the relevant public interest" on the Internet.²⁶ Another aims to hold social media companies accountable for "disseminating false, illegal or harmfully incomplete information to the detriment of individuals or companies." It establishes a R\$50 million (about USD 12.8 million) fine on companies that do not delete such posts within 24 hours and proposes that companies create filters and tools to prevent the spread of false news.²⁷ Two new bills that were introduced in the Brazilian Senate after the elections (PLS 471/2018 and 533/2018) offer exceptions for speech that is classified as opinion, artistic or literary expression or humorous content. Nevertheless, the bills still insist upon a criminal response for the spread of any false information that is *not* protected by these exceptions and proposes to hold online platforms liable if they fail to remove or block reported content within 24 hours.

The Germany's Network Enforcement Act was the first European law to compel social media companies to promptly take down false content related to criminal offenses outlined in the country's criminal code or face fines of up to €50 million. It came into force in 2017 and stirred ongoing discussions of a similar regulation across the European Union as a whole.²⁸

In the face of these initiatives, EFF reaffirms the important safeguards and guidelines already provided by the inter-American framework that should be observed in any efforts aimed at countering the disinformation phenomenon.

²⁴ Id., p. 2.

²⁵ We can mention the following draft bills: (Chamber of Deputies) PL 6812/17, PL 9647/18, PL 9761/18, PL 9838/18, PL 9884/18, PL 9931/18, PL 9554/18, PL 8592/17, PL 7604/17, PL 9532/18, PL 9626/18, PL 7072/17, PL 5742/05, PL 1589/15, PL 1589/15, PL 1589/15—many of the them were attached to bills 6812/17 or 3453/2004; (Senate) PLS 473/17, PLS 471/2018, PL 533/2018.

²⁶ This is PL 9884/2018, currently in the Brazilian Chamber of Deputies.

²⁷ This is PL 7604/2017, currently in the Brazilian Chamber of Deputies.

²⁸ Freedom House. Freedom on the Net 2018, *supra* note 23, p. 13.

Some have already been underscored therein: (i) parameters and elements to properly determine what this phenomenon is; (ii) rejection of measures implicating prior censorship or conditioning of “false speech.”

Others refer to the subsequent imposition of liability and the role of Internet intermediaries in enabling third-party expression.

2.1.1. Subsequent imposition of liability

Article 13(2) of the American Convention sets forth the requirements by which the exercise of free expression may be subjected to subsequent imposition of liability—the restriction shall be expressly established by law to the extent necessary to ensure: a. respect for the rights or reputations of others; or b. the protection of national security, public order, or public health or morals.

The application of this provision should take into account the following standards and interpretations:

Democratic principle and three-part test – based on the provision above, the case law of the inter-American System has developed a “three-part test” which requires that the limitation must be: 1) clearly and precisely defined in a law, both substantively and procedurally, and must serve compelling objectives authorized by the Convention; 2) necessary and appropriate in a democratic society to accomplish the compelling objectives pursued; and 3) strictly proportionate to the objective pursued.²⁹

With regard to the *first* part, we should note that “falsity” is an excessively vague and ambiguous criterion for restricting free expression and threatens a whole range of legitimate speech.³⁰ In this sense, EFF endorses the Joint Declaration on Freedom of Expression and “Fake News” which states that “[g]eneral prohibitions on the dissemination of information based on vague and ambiguous ideas, including ‘false news’ or ‘non-objective information,’ are incompatible with international standards for restrictions on freedom of expression, (...), and should be abolished.” Examinations on the requirements of clarity and precision should consider the parameters and elements discussed in the first section.

²⁹ IAHCR. Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights. *Standards for a Free, Open and Inclusive Internet*, supra note 12, para. 74

³⁰ Botero Marino. La regulación estatal de las llamadas “noticias falsas” desde la perspectiva del derecho a la libertad de expresión, supra note 9, p. 79.

The second *part* establishes an important condition for the restriction of free expression—it must be necessary and appropriate in a *democratic society*. This indicates the general rule that precedes the three-prong test and by which the interpretation of the Article 13(2) restrictions must be judged by reference to the legitimate needs of democratic societies and institutions.³¹ The interpretation of *compelling objectives* falls under these grounds. According to the IA Court, in general terms, “public order” cannot be invoked to suppress a right guaranteed by the Convention, to change its nature or to deprive it of its real content.³² In order to impose any kind of penalty in the name of the defense of public order (understood as security, public health, or morals), it is necessary to show that the “order” that is being defended is not an authoritarian one, but a democratic one understood as the existence of the structural conditions that enable all people to exercise their fundamental rights.³³ Recognizing that a functional democracy is the highest guarantee of public order, any impairment of such order as a justification to limit free expression must be based on real and objectively verifiable causes that present the certain and credible threat of a potentially serious disturbance of the basic conditions for the functioning of democratic institutions.³⁴ Thus, mere conjectures or hypothetical circumstances based on interpretations of authorities do not meet these requirements.

The democratic principle also guides the application of the *necessary, appropriate, and proportionate* test. Any limitation to free expression must be a measure that is effectively conducive to attaining the legitimate objectives pursued—it must be suitable to contribute to the achievement of the aims compatible with the American Convention, or be capable of aiding in their accomplishment.³⁵ In addition to being *appropriate*, the free expression restriction must be *necessary*—meaning that the compelling objective cannot reasonably be accomplished by any other means less restrictive to human rights. Finally, the restriction must be *strictly proportionate* to the legitimate aims that justify them. One must determine whether the sacrifice of freedom of expression this limitation entails is exaggerated or excessive in relation to the advantages obtained through such measure.³⁶ According to the IA Court, three factors must be examined: (i) the degree to which the competing right is affected (serious, intermediate, moderate);

³¹ IACHR. Office of the Special Rapporteur on Freedom of Expression. *The Inter-American legal framework regarding the right to freedom of expression*, supra note 8, para. 66.

³² *Id.*, para. 80.

³³ *Id.*, para. 58.

³⁴ *Id.*, para. 82.

³⁵ *Id.*, para. 87.

³⁶ *Id.*, para. 88.

(ii) the importance of satisfying the competing right; and (iii) whether the satisfaction of the competing right justifies the restriction to freedom of expression. There are no *a priori* answers or formulas in this field, but if the subsequent imposition of liability in a specific case is disproportionate, it means there is a violation of Article 13.2 of the American Convention.³⁷

The examples mentioned above demonstrate different cases that fall short of following these standards. In truth, generic provisions that forbid the publication of false information in order to protect indeterminate legal concepts, such as public order, or public interest, do not meet requirements set by the inter-American framework.³⁸ Current debates on the disinformation phenomenon also require that considerations about national security take these standards into account as a baseline for any measure or regulation.

Rights and reputations of others – the respect for the rights or reputations of others is also a compelling objective that may authorize subsequent imposition of liability by following the three-part test. Honor, dignity, and reputation are human rights enshrined in Article 11 of the American Convention. Nevertheless, the exercise of these rights must be reconciled with free expression, as it is not a right with a higher level or authority.³⁹ The simultaneous exercise of the rights to honor and to freedom of expression, including access to information, must be guaranteed through a balancing exercise in each specific case, which weighs each right in each individual case.⁴⁰ It is undertaken within the rules-based framework mentioned above, which includes the requirement for provable falsity. Moreover, the case law of the inter-American system specifies that the limitation of freedom of expression in such cases requires a clear harm or threat to the rights of others, which should be proven by the authority imposing the limitation.⁴¹

In that sense, the spread of false information that does not harm the rights of others protected by the American Convention cannot be punished under the inter-American framework.⁴² The standard of "actual malice" should also be stressed and considered as indicated in the first section. In addition, the disinformation

³⁷ *Id.*, para 89.

³⁸ Botero Marino. La regulación estatal de las llamadas "noticias falsas" desde la perspectiva del derecho a la libertad de expresión, *supra* note 9, p. 74.

³⁹ IACHR. Office of the Special Rapporteur on Freedom of Expression. *The Inter-American legal framework regarding the right to freedom of expression*, *supra* note 8, para. 103.

⁴⁰ *Id.*, para 103 e 104.

⁴¹ *Id.*, para. 77.

⁴² *Supra* note 38.

phenomenon in electoral contexts bring attention to other crucial inter-American standards:

(a) Specially protected speech – political speech and speech involving matters of public interest are recognized as specially protected speech by the inter-American System. The same treatment is applied to speech regarding public officials in the exercise of their duties and candidates for public office. Although all forms of expression are protected in principle by the American Convention, there are certain types of speech that receive special protection due to their relevance to the exercise of other human rights, or to the consolidation, proper functioning and preservation of democracy.⁴³ This means that the State must be even more cautious of placing limitations on these forms of speech, and that State entities and officials, as well as those who aspire to hold government positions, must have a higher threshold of tolerance in the face of criticism because of the public nature of their duties.⁴⁴ Consequently, the IA Court has established that, in applying the proportionality test, speech about the practices of State institutions enjoys greater protections, in the interest of furthering democratic debate within society.⁴⁵ The disinformation phenomenon as outlined in the first section refers primarily to these types of protected speech. States regulations and measures as the ones mentioned above also primarily target at least “matters of public interest.” The political or public interest dimension of the speech is even greater during electoral processes, which highlights the need to consider this special regime when assessing the subsequent imposition of liability.

(b) Civil liability instead of criminal – the IACHR discourages the use of the criminal law to criminalize speech. The special rapporteurs for freedom of expression asserted in their Joint Declaration on “Fake News” that “[c]riminal defamation laws are unduly restrictive and should be abolished. Civil law rules on liability for false and defamatory statements are legitimate only if defendants are given a full opportunity and fail to prove the truth of those statements and also benefit from other defenses, such as fair comment.” This assertion is emphasized during elections as the speech in question refers to matters of public interest like the conduct of political candidates, for example. The IACHR has considered that the use of criminal law mechanisms to punish this type of speech *in and of itself* violates Article 13 of the American Convention, since it is unnecessary, disproportionate, and can also constitute an indirect means of censorship given its

⁴³ IACHR. Office of the Special Rapporteur on Freedom of Expression. *The Inter-American legal framework regarding the right to freedom of expression*, supra note 8, para. 32.

⁴⁴ *Id.*, para 35.

⁴⁵ *Id.*, para 105.

intimidating effect on debate concerning matters of public interest.⁴⁶ Additionally, the subsequent liability that arises from the abuse of freedom of expression must always be ordered by an independent and impartial judge or court authority, and respectful of due process guarantees.⁴⁷

This understanding does not overlook the damaging impacts that disinformation campaigns may have in electoral processes and, ultimately, in democratic societies. Yet, by taking these impacts into account, the reasoning points to a much greater harm if criminal responses prevail. Resorting to criminal law to punish speech related to State institutions, or the exercise or abuse of power by public officials, severely impair this same democracy and the means society has to defend it. As mentioned above, even regarding factual assertions the existence of one indisputable truth is, in itself, controversial.

In addition to subsequent civil liability, other measures and sanctions may apply to specific agents engaged in disinformation campaigns when their actions entail violations of other legitimate and well-grounded legislation. Investigations concerning the Brazilian elections, for example, indicate that political campaigns or companies directly supporting such campaigns may have acted in disagreement with the electoral law.⁴⁸ Potential unlawful behavior during elections should be investigated and subject to the review of an independent and impartial judge in accordance with due process safeguards.

2.2.2. The role of Internet intermediaries in enabling third-party expression

Article 13(3) of the American Convention establishes that the right of expression may not be restricted by indirect methods or means. An indirect restriction would compromise not only the right to impart information, but also the rights to seek

⁴⁶ Id., para. 114.

⁴⁷ IAHCR. Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights. *Standards for a Free, Open and Inclusive Internet*, supra note 12, para. 74.

⁴⁸ The accusations include: (i) companies' funding of the spread of election material through WhatsApp that may qualify as their donation to political campaigns, which the electoral legislation forbids; (ii) unlawful use of third-party databases to send WhatsApp messages; and (iii) massive activation of chips with ID numbers of elderly people without their knowledge to circumvent WhatsApp's anti-spam limits. See Pablo Ortellado. O que sabemos sobre o uso do WhatsApp nas eleições. Folha de São Paulo, 04 December 2018. (><https://www1.folha.uol.com.br/colunas/pablo-ortellado/2018/12/o-que-sabemos-sobre-o-uso-do-whatsapp-nas-eleicoes.shtml><)

Patrícia Campos Mello. Empresários bancam campanha contra o PT pelo WhatsApp. Folha de São Paulo, 18 October 2018 (><https://www1.folha.uol.com.br/poder/2018/10/empresarios-bancam-campanha-contra-o-pt-pelo-whatsapp.shtml><)

and receive content. Since most communication on the Internet is facilitated by intermediaries, such as Internet service providers (ISPs) and social media platforms, unnecessary and disproportionate measures expected of them result in undue limitation of the rights of freedom of expression and access to information. A government's order to shut down mobile networks or block entire social media platforms in the name of combatting disinformation clearly constitutes an excessive and dangerous overreach of power.

Any restriction on the right to free expression must follow the standards in the three-part test and should consider, in its application, the different impacts it may have on each type of intermediary. Targeting specific content on a social media platform is less pervasive than, for example, resorting to Domain Name System (DNS) blocking, which can prevent users from accessing lawful and unlawful speech alike. EFF has pointed out that problems with censorship by direct hosts of speech are tremendously magnified when core infrastructure providers are pushed to block or filter content.⁴⁹ The risk of powerful voices silencing marginalized ones is greater, as are the risks of collateral damage. Takedowns by infrastructural intermediaries—such as certificate authorities, DNS, or content delivery networks—are far more likely to cause collateral censorship. For that reason, EFF has called these parts of the Internet free speech's weakest links⁵⁰ and believes that the most consistent defense these links can take is to decline attempts to use them as a control point. Conduits such as ISPs should also not be treated as publishers. Their legitimate scope to limit content is critically restricted by network neutrality principles that require them not to discriminate online content.⁵¹ Such principle, as indicated by the Office of the Special Rapporteur, is a necessary condition for exercising freedom of expression on the Internet pursuant to the Article 13 of the American Convention.⁵²

The Office of the Special Rapporteur has also asserted that restrictive measures should at all times include safeguards to prevent abuse. Such safeguards might include transparency about the content whose removal has been ordered, as well as detailed information about the need and justification of the takedown. At the same time, a measure of this kind should be adopted only when it is the only

⁴⁹ See McSherry, York, and Cohn. Private Censorship Is Not the Best Way to Fight Hate or Defend Democracy: Here Are Some Better Ideas. 30 January 2018. Available at <https://www.eff.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some>.

⁵⁰ See the EFF's "Free Speech Weak Links" at <https://www.eff.org/free-speech-weak-link#home>

⁵¹ *The Manila Principles on Intermediary Liability Background Paper*. Version 1.0. 30 May 2015, p. 35.

⁵² IACHR. Office of the Special Rapporteur for Freedom of Expression. *Freedom of expression and the Internet*.

OEA/Ser.L/V/II CIDH/RELE/INF.11/13. 31 December 2013. Para. 25.

measure available for achieving an imperative end and is strictly tailored to achieve it.⁵³ Importantly, the Office of the Special Rapporteur for Freedom of Expression has stressed that subsequent liability should be imposed upon the authors of the speech in question rather than on the intermediaries.⁵⁴

This is in line with the Manilla Principles, which indicate that intermediaries shouldn't be liable for third-party content in circumstances where they have not been involved in modifying that content.⁵⁵ Put forward by EFF and other civil society organizations from around the world, the Manilla Principles are a framework of baseline safeguards and best practices based on international human rights instruments for States with respect to intermediary liability.⁵⁶ According to the Principles, States should limit the liability of intermediaries for third-party content (Principle 1), not require the restriction or removal of content without a court order issued in accordance with due process rights and guarantees (Principles 2 and 3), ensure that the laws meet the three-part test on freedom of expression and include the principles of transparency and accountability (Principles 5 and 6).

However, current proposed or approved regulations related to disinformation hold intermediaries accountable for third-party content without a prior court order. In most cases the demand is to promptly remove the content, within 24 hours after being reported. This flawed solution disregards its negative impacts on innovation as well as the chilling effects on free expression by incentivizing self-censorship by platforms—which the Office of the Special Rapporteur for Freedom of Expression has underscored.⁵⁷

Further, legislative initiatives mentioned above misguidedly propose that companies create automatic filters and tools to prevent the spread of

⁵³ Id., para. 87.

⁵⁴ Id., para. 102.

⁵⁵ The modification which may give rise to liability should not include measures such as downranking or filtering, since this is not related to the creation of content, and should not comprise strictly technical modifications: "It is not intended that technical modifications, such as the addition of HTTP headers by an caching intermediary, would give rise to liability for the cached content." *The Manila Principles on Intermediary Liability Background Paper*. supra note 51, p. 21.

⁵⁶ See the principles at <https://www.manilaprinciples.org>.

⁵⁷ "In view of the uncertainty about potential liability, intermediaries can be expected to end up suppressing all of the information that they think, from any point of view, could potentially result in a judgment against them. A system of this kind would seriously affect small and medium-sized intermediaries, as well as those who operate under authoritarian or repressive regimes. It would also jeopardize the right of all persons to use the media they deem appropriate for the transmission of ideas and opinions". IACHR. Office of the Special Rapporteur for Freedom of Expression. *Freedom of expression and the Internet*, supra note 52. Para. 99.

disinformation. Once more, such proposals disregard declarations from the Office of the Special Rapporteur for Freedom of Expression, that recognize, among other faults, “that [automated] systems for blocking and filtering Internet content frequently block legitimate websites and content. Some governments have used them to prevent their populations from accessing information that is fundamentally in the public’s interest but that governments are interested in hiding.”⁵⁸ In this sense, the Joint Declaration on Freedom of Expression and “Fake News” states that content filtering systems which are imposed by a government and which are not end-user controlled are not justifiable as a restriction on free expression.

Governments should also refrain from pushing intermediaries to indirectly use their terms of service or community rules to expand the legally established grounds for restriction.⁵⁹ Over the past decade we have seen the emergence of secretive agreements between companies that seek to control users’ behavior online, and governments that want a shortcut and workaround to regulation.⁶⁰ EFF has called this “shadow regulation” and stressed its dangerous and undemocratic nature.⁶¹ Regulation should take place through open doors and involve meaningful participation from all affected stakeholders.

Any State’s attempt to tackle disinformation in electoral contexts must not imply circumventing or undercutting the deep connection between democracy and freedom of expression. It is during elections that the fiercest debates over society and a government’s direction take place, and during which public engagement is often maximized. Controversial, exaggerated, or biased speech therein often involve matters of public interest and refer to candidates or public officials in the exercise of their duties. While abuses of free expression by the person responsible for the content will and should be addressed by subsequent civil liability, companies should not be turned into a sort of speech police.

Private platforms are prone to error and can disproportionately affect the less powerful.⁶² Internet intermediaries when establishing terms and rules for their platforms should do so by following standards of transparency, due process, and

⁵⁸ Id., para 90.

⁵⁹ IAHCR. Office of the Special Rapporteur for Freedom of Expression of the Inter-American Commission on Human Rights. *Standards for a Free, Open and Inclusive Internet*, supra note 12, para. 124.

⁶⁰ McSherry, York, Cohn. Private Censorship Is Not the Best Way to Fight Hate or Defend Democracy, supra note 49.

⁶¹ See more at <https://www.eff.org/issues/shadow-regulation>.

⁶² McSherry, York, Cohn. Private Censorship Is Not the Best Way to Fight Hate or Defend Democracy, supra note 49

accountability. The Joint Declaration on Freedom of Expression and “Fake News” points out some recommendations for intermediaries to that end and further considerations on this matter are presented below. As the Office of the Special Rapporteur has previously highlighted, intermediaries have become points through which it is technically possible to exercise control over online speech.⁶³ Solutions that pretend to bolster democracy cannot exploit these control points to silence dissent and stifle political debate; any technically possible solution must be, above all, legitimate, necessary, and proportionate.

2. Suggested guidelines on how to address actions and operational measures

The following guidelines for tackling disinformation during elections are not intended to be exhaustive, nor are they intended to cover the range of issues that may arise from potential or implemented initiatives. They consist of focus points regarding EFF’s current work on the matter and seek to contribute to a broader debate.

(a) Advancing Transparency and Accountability in Content Moderation

The Santa Clara Principles on Transparency and Accountability in Content Moderation offer an important baseline framework for advancing better content moderation practices among platforms.⁶⁴ Put forward by EFF alongside other civil society groups and academic experts, the principles underpinned an open letter to Facebook’s CEO Mark Zuckerberg signed by over 100 organizations around the world.⁶⁵ According to these principles, there are *three minimum steps* that companies engaged in content moderation should take to provide meaningful due process for impacted speakers and better ensure that the enforcement of their content guidelines is respectful of users’ rights:

1. Notice – companies should provide notice to each user whose content is taken down or account is suspended. It should include the specific clause of the company’s guidelines that the content was found to violate, sufficient information to identify the specific content that was restricted, and detail about how the content was detected, evaluated, and removed. Individuals must also have clear

⁶³ IACHR. Office of the Special Rapporteur for Freedom of Expression. *Freedom of expression and the Internet*, supra note 52. Para. 92.

⁶⁴ See at <https://santaclaraprinciples.org>.

⁶⁵ Available at <https://santaclaraprinciples.org/open-letter/>

information about how to appeal the decision. In general, platforms should provide detailed guidance to the community about what content is prohibited, including examples of permissible and impermissible content and the guidelines used by reviewers. Companies should also provide an explanation of how automated detection is used across each category of content.

2. *Appeal* – companies should provide a meaningful opportunity for timely appeal of any content removal or account suspension. It should include a human review by a person or panel of persons that was not involved in the initial decision as well as the opportunity to present additional information to be considered in such review. The results of the review should be notified with a statement of the reasoning sufficient to allow the user to understand the decision. In the long term, independent external review processes may also be an important component for users to be able to seek redress.

3. *Numbers* – companies should publish the numbers of posts removed and accounts permanently or temporarily suspended due to violations of their content guidelines.⁶⁶ This data should be provided in a regular report in an openly licensed, machine-readable format.

(b) Deploying further measures or assessing ones in course in dialogue with civil society

In addition to content moderation, platforms may or have already put in place other measures to counter disinformation. It is important that companies do so in collaboration with civil society, and consider local nuances and concerns. The following items include comments and proposals of solutions platforms could adopt:

1. *Collaboration with fact-checkers* – the Joint Declaration on Freedom of Expression and “Fake News” indicates that intermediaries should cooperate with initiatives that offer fact-checking. Fact-checking is a good practice and has played a role in highlighting false content as well as raising public awareness of the need to verify the information we receive. Still, fact-checking initiatives must be transparent and fair about how, where, and by whom the verification is made. They should also provide means by which their verification may be objected to. Bearing in mind previous sections of this submission, judgments about what is true or false are inherently complex, especially in the electoral context. Using fact-

⁶⁶ See in the Santa Clara Principles’ statement the minimum dimensions along each the numbers should be broken down, *supra* note, 64.

checking to guide content downranking or removals could reinforce errors and biases in content moderation. This is why placing fact-checker assessments and alternative sources in controversial posts seem more suitable practices. Further, the negative rating or downranking of content by the mere fact it refers to anonymous and pseudonymous journalism is a usual fact-checking practice that should be reviewed. From a human rights perspective, the exercise of privacy protections should not be punished.

2. Users' customization of feed and search algorithms – users should be in control of their own online experience. Smart filters help users to choose content they want to see and filter out content they prefer not to receive. It should continue and evolve. There is a big opportunity to address online disinformation challenges through innovation. This could include user interface designs and user controls that encourage productive and informative conversations that would temper the spread of wildly fabricated material while giving readers transparency and control over that process. At the same time, media and information literacy initiatives can play a key and complementary role in encouraging users to avoid bias and to seek more diverse sources and opinions.

3. Greater transparency for electoral propaganda – political parties and candidates using online platforms to target users with electoral propaganda should do so transparently. Labels, specific user interfaces or other technological solutions should be implemented by companies in order to enable users to identify election material and to know what campaign is targeting which content. Third-party tools should also be fostered by using API's, allowing scraping, and ensuring interoperability. Considering the legislation in place, this is part of users' right to consent to the collection and processing of their data by platforms as well as to know with whom such data is being shared.⁶⁷

4. Other measures to consider – recent changes in WhatsApp's platform due to disinformation concerns in countries like India and Brazil deserve attention. Identifying forwarded messages in WhatsApp, for example, increases transparency and encourages users to double check content sources. However, the app also limits message forwarding to five recipients, which may not be the best way to address current concerns. The viral spread of contents therein is not specifically serving disinformation. It can enhance activism, broader political debate, and can also play a role in spreading messages of public interest (e.g. to provide support in

⁶⁷ Schwartz, McSherry, McKinney, and Tien. New Rules to Protect Data Privacy: Where to Focus, What to Avoid. 2 July 2018. Available at <https://www.eff.org/deeplinks/2018/07/new-rules-protect-data-privacy-where-focus-what-avoid>

situations of public disasters). How effective restrictive measures can be in the face of malicious actions is another relevant question. Malicious agents may exploit ways to circumvent the new limitations while regular users will be subjected to them.⁶⁸ Companies can certainly decide to change features on their platforms; but when deploying them or being pushed to do so, it is important for all stakeholders involved to consider both the intended and unintended consequences.

(c) Undermining encryption is not an answer

The spread of disinformation through WhatsApp during Brazil's elections resurfaced debates about end-to-end (E2E) encryption in communications systems. Unlike platforms like Facebook and Twitter, where companies' moderators and algorithms can actually read the posts, WhatsApp's encryption would allegedly have been a hindrance to countering the spread of disinformation over the elections.⁶⁹ Although further investigation is needed and will probably occur, there are already some key points to highlight: *First*, E2E encryption has not prevented researchers to follow disinformation patterns throughout WhatsApp groups by joining those groups.⁷⁰ *Second*, E2E encryption has not prevented participants of those groups to share controversial content with fact-checking initiatives and/or with the campaign the false news targeted; in both cases efforts were made to respond the false posts.⁷¹ *Third*, most of the information disseminated through WhatsApp was also shared over platforms such as Facebook and Twitter and/or checked through initiatives available.⁷² *Forth*, there is evidence that much of the disinformation that was disseminated is related to legally

⁶⁸ See supra note 48. In addition to evidences of massive activation of chips with ID numbers of elderly people without their knowledge, Brazilian digital marketing companies may also have used tools to automatically generate phone numbers with different country codes to circumvent anti-spam limits that WhatsApp's had set to Brazil. Although the company has recently extended worldwide the new message forwarding limit, this was not the situation during Brazil's elections—by then, the text forwards were limited up to 20 in that country.

⁶⁹ See Cristina Tardáguila. Criptografia ou capacidade de viralização? O WhatsApp precisa enxergar esse paradoxo. 03 December 2018. (><https://piaui.folha.uol.com.br/lupa/2018/12/03/artigo-epoca-criptografia-whatsapp/><)

⁷⁰ See WhatsApp Monitor (><http://www.monitor-de-whatsapp.dcc.ufmg.br/><) and Mariana Simões. Grupos pró-Bolsonaro no WhatsApp orquestram fake news e ataques pessoais na internet, diz pesquisa. Agência Pública, 22 October 2018 (><https://apublica.org/2018/10/grupos-pro-bolsonaro-no-whatsapp-orquestram-fake-news-e-ataques-pessoais-na-internet-diz-pesquisa/><)

⁷¹ See What WhatsApp "API access" meant for Comprova. First Draft, 29 November 2018. (><https://medium.com/1st-draft/what-whatsapp-api-access-meant-for-comprova-b0e2944607ab><) and Em 12 horas, Haddad recebe 5 mil denúncias de fake news. 03 October 2018. (><http://www.pt.org.br/em-12-horas-haddad-recebe-5-mil-denuncias-de-fake-news/><).

⁷² Juliana Gragnani. Eleições com fake news?: Uma semana dentro de 272 grupos políticos no WhatsApp mostra um Brasil dividido e movido a notícias falsas. BBC Brasil. 05 October 2018 (><https://www.bbc.com/portuguese/brasil-45666742><).

controversial practices that demanded human and financial resources; E2E encryption does not prevent those practices from being investigated and fairly sanctioned.⁷³ *Fifth*, the degree in which such messages influenced voters is still being determined, especially regarding the disinformation shared through the big groups that populated WhatsApp during elections and to which users had been massively added without their authorization.⁷⁴ *Sixth*, and most importantly, the inability for the platform to read users' private messages is definitely not a problem we should tackle, but a standard we should continuously foster.

Encryption protects civil liberties and individual rights. It grants less powerful people crucial safeguards that allow them to communicate without fear of being harassed and surveilled. It empowers activism and civil organization. It is not only about one-to-one communication, but also about the exchange of ideas, information, and actions in and among groups. It is about democracy; and about preserving means that are necessary to resist authoritarian regimes or political persecution. As previously noticed by the UN Special Rapporteur for Freedom of Expression, encryption provides individuals and groups with a zone of privacy online to hold opinions and exercise freedom of expression without arbitrary and unlawful interference or attacks. Privacy and freedom of expression are interlinked and encryption play a critical role in ensuring those rights.⁷⁵

Likewise, EFF has continuously emphasized that attempts to weaken encryption jeopardize everyone's security and privacy, unduly burden innovation, and disregard the fact that malicious actors and criminals will always have an alternative tool available.⁷⁶ Regarding what was pointed out above, there is a fair amount of other pathways to understand and counter the disinformation phenomenon that does not entail shattering the integrity and security of users' communications—alternative pathways that are truly consistent with legal,

⁷³ See supra notes 48 and 68. See also Anthony Bodle. Facebook's WhatsApp flooded with fake news in Brazil election. Reuters. 20 October 2018. (><https://www.reuters.com/article/us-brazil-election-whatsapp-explainer/facebooks-whatsapp-flooded-with-fake-news-in-brazil-election-idUSKCN1MU0UP><)

⁷⁴ Survey commissioned by Avaaz and carried out by IDEA Big Data found that nearly 90 percent of Bolsonaro's voters believed in false news posted on Facebook and Twitter; levels of credibility differed depending on the news. (><https://www1.folha.uol.com.br/poder/2018/11/90-dos-eleitores-de-bolsonaro-acreditaram-em-fake-news-diz-estudo.shtml><). In turn, in the survey made by Ibope Intelligence only 5% said they rely on information shared on Facebook and 4% on WhatsApp. (><https://www1.folha.uol.com.br/poder/2018/11/pesquisa-ibope-aponta-que-66-confiam-mais-em-sites-de-noticias.shtml><).

⁷⁵ UN, General Assembly. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye. A/HRC/29/32. 22 May 2015. Para 16. See also UN, General Assembly. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue. A/HRC/23/40.

⁷⁶ See supra note 48 and 68.

necessary and proportionate requirements: that is to say, genuinely compliant with international human rights guarantees. Responses to the disinformation phenomenon must not override these guarantees.

(d) Paying attention to network neutrality and platform competition

Brazil's elections also raised concerns about the role of zero-rating agreements and their effect on the spread of disinformation.⁷⁷ Major telecom operators in the country usually do not discount users' mobile data caps when they use Facebook, WhatsApp, and Twitter. Since on WhatsApp and Twitter one must access non-zero-rated applications in order to open links shared on those platforms, users could be disincentivized to read the actual piece instead of simply reading the headline. Further online research to check additional sources could equally be compromised. In this sense, further studies about zero-rating impacts on disinformation would be welcomed. In addition to promoting free expression, network neutrality bolsters innovation and competition. Zero-rating reinforces the online dominance of a few players by forging a dynamic that keeps users coming back to this very small group of providers. Network neutrality and competition are also important factors in fostering a healthier informational environment. In promoting platform competition, data portability and interoperability are two key measures that could substantially improve the way people seek, receive, and impart information online through and out electoral processes.

⁷⁷ See Juliana Gragnani. Como planos de celular com Facebook e WhatsApp ilimitados podem potencializar propagação de notícias falsas. BBC Brasil. 16 April 2018. (><https://www.bbc.com/portuguese/brasil-43715049><).