

Disinformation and freedom of opinion and expression

Institute for Technology and Society submission for the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression

The mission of the Institute for Technology and Society ([ITS Rio](#)) is to ensure that Brazil and the Global South respond creatively and appropriately to the opportunities provided by technology in the digital age, and that the potential benefits are broadly shared across society. Through its own research and in partnership with other institutions, ITS Rio analyzes the legal, social, economic and cultural dimensions of technology and advocates for public policies and private practices that protect privacy, freedom of expression and access to knowledge. We appreciate this opportunity to input into the Special Rapporteur's consultation on disinformation.

Disinformation global challenges from the Brazilian perspective

[Disinformation is not a new phenomenon](#), but in the digital age it has taken unprecedented proportions. The internet lowered the bar for circulation of information and eroded borders. Fake and misleading content can easily and fast circulate worldwide. The Global South may be particularly impacted and democratic institutions are especially at risk.

Brazil is one of such cases. It should be noted that it is the second [country](#) in terms of time spent on social media (an average 3.45 hours against less than 2 hours in most developed countries) and it is a [major source of news for most Brazilians](#). The 2018 elections are a significant example of the impact disinformation may have in the democratic system. The candidates relied heavily on digital services and social media was prominent among them. Researchers disclosed in a [NY Times article](#) that in the month prior to the election only 4 out of the 50 most shared messages on WhatsApp could be considered fully true. In general, the election was considered as well fraught with false and out-of-context news.

Additionally, the wide and fast spread of disinformation is often the result of coordinated behaviour and use of automation tools for digitally boosting false and political news in the country, as this study conducted by the [ITS](#) shows.

The pandemic is another factor that impacted on the online environment. It has been reported that in several instances false and misleading information about the health crisis was circulating in the country. In a [study](#) published in november 2020, Brazil stands as the only one in which drugs like Chloroquine and Ivermectin still lingered in the public debate, even after no scientific research supported them as having any significant positive effect against the disease.

The same study indicates that disinformation around the pandemic is being used in the context of local power disputes at the different levels of government. Supporters of the

President are reported to discredit the safety of vaccines produced in the country under the *aegis* of opposing governors.¹

The Report is structured in three sections providing an overview of the efforts implemented by (i) the government, (ii) tech companies and (ii) civil society in order to curb the problem in Brazil. They will highlight opportunities and risks to implement and guarantee human rights - particularly freedom of expression, access to information and privacy - in the Global South.

1. Brazilian Government efforts to fight disinformation

Disinformation was perceived as a major challenge by all levels of government in Brazil. This provided a scenario where both the Legislative and the Judiciary acted in order to tackle causes and effects of the phenomenon.

Chief amongst the initiatives were a series of bills proposed in Congress to discuss how to deal with the subject through several lenses from criminalization of conducts to proposing several obligations to internet service providers (intermediary enterprises mostly). Congress also established a bicameral parliamentary inquiry commission to investigate an alleged 'disinformation network' financed with public funds.

The Judiciary was called as well to deal with numerous complaints of 'fake news' dealing with defamation, attacks to personal honor, intimacy, bullying among others. Claims were both under civil and criminal procedures. The Supreme Court too opened a controversial inquiry under the allegation that both the institution itself (the Court) was the target of a disinformation campaign and its members. As the procedure was started *proprio motu* by the Justices themselves, there were allegations of violations of due process and bias, which is reported to have been [referred to the Inter-American Human Rights Commission](#).

2. Legislative Initiatives

2.1. Brazilian Legal Landscape

Brazil does not have a legal definition of disinformation, fake news, or informational disorder. Often, criminal law - libel, defamation, disturbing the electoral process, false communications of crimes, slanderous denunciations, and other infractions - is used as a way to deal with the phenomenon. This, however, raises a series of difficulties as such instruments are not well suited to keep up with the pace, volume, velocity and impact of disinformation online and may run counter to the human rights protected within the Constitution and international obligations.

¹ This was the case for instance, the president of the Palmares Foundation (a Brazilian federal foundation) wrote on [Twitter](#): 'I am from São Paulo and I appeal to my family members, to whom I wish all the good in the world: do not take the Chinese vaccine from Doria!'. The statement until February 10 (date of writing of this input) had almost 20,000 likes and was retweeted more than 3.6k times and was still available on the social media platform. According to [Yahoo news](#), the following message from Roberto Jefferson, Brazilian politician close to Bolsonaro, has been replicated more than 4,000 times on Twitter: "Chinese laboratory creating vaccine against Chinese virus and with research sponsored by a governor who is a great partner of China? I don't want this vaccine, how about you?".

Several bills have been put forward in Congress to change this scenario. Their strategies refer to: (i) revisiting the liability of internet intermediaries - it should be noted that under the current Brazilian legislation (Internet Bill of Rights) intermediaries are *not* liable for content generated by third parties, unless they do not comply with a judicial decision of content removal; (ii) enhancing obligations to intermediaries to collect and store data (despite data protection minimization obligations); (iii) limiting cryptography; (iv) adding to the penal code and other criminal legislations crimes connected to information disorder; and (v) enhancing investigative powers.

In the following sections positive and negative aspects referred to such proposals will be highlighted. The relevance of this exercise is to showcase the plethora of options that may impact the enjoyment of human rights, particularly freedom of expression both in Brazil and in the world. Many of the proposals pushed forward in the country echo other international initiatives.

2.2 Internet Bill of Rights - the *juste milieu* may bow under pressure

One of the ways pursued to deal with online disinformation is through regulation of content moderation and mechanisms in order to hold platforms liable (and accountable) for the content available in their services. In this sense, the Brazilian regulatory framework is considered one of the most balanced and a great achievement. The [Brazilian Internet Bill of Rights](#) (Law No. 12.965/2014) is the result of a long collaborative process with a broad multistakeholder support. It establishes the guiding principles for internet governance in the country which include freedom of expression, access to information and privacy.

Article 19 of the aforementioned bill created a liability regime for internet providers - social media, video sites, encyclopedias, messaging apps, and any platform (including news) that contains comments and contributions from its users - known as **judicial notice and takedown**.

Pursuant to Article 19, technology companies are not liable for the acts of third parties (their users) until a court order states that the content is illegal. From that moment on, if they do not remove the content (text, photos or video) they become responsible for it. The exception is due to content that justifies quick removal mechanisms, such as nudity or unauthorized pornography (“revenge porn”), which are regulated under article 21.

On the other hand, platforms may be liable for their own actions - removing or reducing the availability of content. Several users, feeling harmed by the removal of pages, videos and photos, have already sued providers in Brazilian courts. Please note that Brazil does not have an immunity for actions of content moderation, such as Section 230 provides for in the US.

Hence, this highlights a compromise achieved under the Internet Bill of Rights, a *juste milieu*, where as a principle individuals are allowed free speech online and platforms are incentivised to moderate content under a responsible manner, knowing that their actions should respect individual rights. It is incumbent on the Courts - and not technology companies - to take a definitive decision on whether a specific content is legal or illegal, holding service providers liable only if they do not comply with a court order.

However, there is an outcry that this regime may not be effective particularly against disinformation campaigns. Thus, several initiatives propose amending the Internet Bill of Rights with a varying degree of risks for enjoyment of human rights in the country.

2.3 'Disinformation' Draft Bills

Today, more than 50 proposals about disinformation are under analysis, which, in summary, would:

- address the criminalization of the dissemination or sharing of false or incomplete information on the internet,
- hold social media liable when untrue materials are posted on the internet and the content is not removed within twenty-four hours,
- force social media platforms to provide filters and tools to prevent the dissemination of harmful information, and
- hold content providers and providers of internet services liable for damages caused by the dissemination of fake news on the internet.

- **Draft Bill # 2630/2020 - The 'Fake News' Bill²**

One of the bills that most reverberated in recent times and was approved in one of the legislative houses was Draft Bill # 2630/2020, also known as 'Fake News' Bill. The processing of the project involved a series of problems, largely because of its fast pace allowing less than optimal social participation. This is a huge contrast with the collaborative history of the Internet Bill of Rights.

The bill has undergone notable changes, with different versions being published every few days. In its various versions, the bill raised several controversies about the control of platforms and users. In the version approved by the Federal Senate, some of these criticisms were absorbed and many of the provisions in question were removed from the text. Many controversial clauses, however, were kept, such as an obligation to mass collect and store information regarding message exchanges in order to be able to trace origin and spread of messages - disregarding privacy concerns and circumventing encryption protocols.

In this scenario, the 'Fake News' Bill accumulates national and international criticisms: from [Human Rights Watch](#), the [Global Network Initiative](#), to the [National Human Rights Council](#) and this UN rapporteur for freedom of expression and opinion. The [Direitos na Rede](#) Coalition, articulation between 39 civil, academic and digital rights activists in Brazil, including ITS, has repeatedly pointed out the project's risks and failures besides its lack of social support and participation on its construction.

ITS highlighted in this [technical note](#) the problems in rushing the processing and conceptual inaccuracies in its text. Among which key takeaways are:

² There's an unofficial [english version](#) made available by CTS/FGV.

- **The text may encourage censorship and undermine freedom of expression on the Internet.** It reverses one of the greatest advances brought about by Brazil's Internet Bill of Rights as the bill may lead platforms to monitor all content published by users and intent as they may be liable for third party publications on their platforms. Thus, providers may aggressively remove legal content that has the slight hint of not being in accordance with the law; reversing a presumption in favor of free speech to one against Platforms will be prong to police content and become in fact gatekeepers. The 'chilling effect' this obligation may have on speech can hardly be ignored.
- **The text brings imprecise definitions:** imprecise or even conflicting definitions with other legal texts can lead to confusion and may give rise to abuses of power and the attribution of police functions to platforms, in order to define "disinformation" in freedom of speech violation cases. For instance, there are no reservations about misinformation, which can sometimes be the result of the action of an ill-informed user or a journalistic error. These cases of bad information do not have the expected intentionality of the misinformation and deserve different treatment.
- Users suspected of creating bots or anonymous accounts may be required to **present IDs to technology companies (mass identification), without a court order:** accounts could be reported as suspicious for a wide variety of reasons — including, in theory, political disputes. Also, conditioning the registration on social media to the presentation of documents in a country where millions of people still lack legal proof of ID could exacerbate an already present digital divide.
- **Draft Bill #246/2021**

The [Bill #246/2021](#) establishes a civil liability regime for internet application providers in cases of content moderation, not only for removal, yet for content labeling as well. The Draft Bill #246 seems to transplant a discussion very much alive in the United States regarding an immunity for content moderation under Section 230 of the Communications Decency Act (known as "[good Samaritan](#)"). This legal measure is at best superfluous, as, as mentioned earlier, this immunity never existed in Brazil. Platform actions directly on the content are "acts of the platform" and as such not covered by the immunity in article 19 of the Brazilian Internet Bill of Rights.

If internet companies end up acting abusively and remove or even label content wrongly they may be liable. Brazil already has several cases in which the author of posts or the holder of improperly removed accounts won their cases in the Judiciary. Brazilian courts have already [condemned](#) Google, for example, for mistakenly removing parodies of songs claiming it was a copyright infringement. A congressman has already won a lawsuit against Facebook because the company removed his account alleging that he was spreading fake news and the company was [forced](#) to republish the posts. Likewise, it has already [condemned Facebook](#) for mistakenly removing a deputy's fanpage. That is, the tech companies responsibility for moderation activity already occurs without the need for a new law.

The danger for human rights may not be the obligations present in the proposed bill, but the environment it creates. It seems to embolden an environment against freedom of expression online, focusing on reshaping the equilibrium set in the Internet Bill of Rights and pushing

internet companies to shield speech that may be discriminatory, offensive or may even incite violence.

- **Draft Bill # 291/2021**

The [Bill #291](#) was presented earlier February this year. The Draft Bill aims to prohibit the removal of messages from users by an application provider in disagreement with the constitutional guarantees of freedom of expression, communication and expression of thought. However, it should be noted that, as it stands, the bill may have a limited reach.

The Bill adds to the Brazilian Internet Bill of Rights providing for the nullity of any contractual clauses and the terms of use related to the provision of internet application services that foresees the take down of user content due to political orientation or expression of an opinion. Today, in theory, there is no term of use that allows deletion of content for “political orientation or expression of opinion”.

This Bill, if approved, may as well have a ‘chilling effect’ on tech companies using their terms of use as justification to remove content as they may be held subject to high penalties. The vague nature of the terms used in the proposed bill increases such risk.

2.4 ‘Fake News’ Parliamentary Inquiry Commission (‘Fake News CPMI’)

The Parliamentary Inquiry Commission on ‘fake news’ (‘Fake News CPMI’) is closely associated with the increase in proposed bills on disinformation. The Commission was motivated by allegations of [“cyber attacks against democracy and the public debate, online harassment, the use of fake profiles to influence the 2018 elections, enticement of children in hate crimes and suicide, and attacks against authorities”](#).

In December of 2019, house representative Joice Hasselmann testified exposing a group called [“office of hate”](#) within President Jair Bolsonaro’s government. [According to Hasselmann](#), almost BRL 500,000 have already been proven to have been spent to pursue political opponents, by daily global and national news reports as well as sentiment analysis, which are used to guide decision-making, as well as extensive use of bots to spread misinformation online. ‘I suggest to follow the money trail as we are talking about millions’, Hasselmann argued.

As it became clear, one of CPMI's main contributions was the identification of financing practices for disinformation campaigns, which motivates bills by first identifying the fake news industry's "follow the money" strategy. Fake News CPMI was extended for an additional 180 days and would include the spread of disinformation about the COVID-19 pandemic, since as the coronavirus spread through Brazil, [ominous stories began filtering out through social media](#), often via lawmakers with thousands of followers. The risks of spread of disinformation gained new urgency because of the seriousness of the health issue.

One of the positive points of the work carried out by the Commission was the fact that a [wide range of actors were heard](#). Journalists, experts in combating disinformation and tech companies provided testimony. [According to the rapporteur of the Commission](#), these inputs

will inform the legislative house to “establish legislation that can punish culprits without jeopardizing freedom of expression”.

3. Judicial initiatives - ‘blocking orders’

In March 2019, Justice Dias Toffoli, Federal Supreme Court’s President at that time, ordered the [opening of an inquiry](#) which became known the ‘Fake News inquiry’ ([Inquiry N° 4781](#)) to investigate the alleged existence of slander, threats and fraudulent reports against the Court, its members and relatives.³

Justice Alexandre de Moraes, inquiry’s rapporteur, issued several controversial decisions during its proceedings. The magistrate ordered social networks Facebook and Twitter to block access to the accounts of 16 individuals being investigated for allegedly spreading disinformation and hate speech online. However, after national blocking, users outside of Brazil, or who use a VPN, continue to be able to access the accounts. Justice Moraes is reported to have requested [global suspension of the profiles](#) after realizing that the blockade he ordered earlier had a reduced impact. The event had a further repercussion, as one of the 16 who had their accounts globally blocked was able to use another account to make a very controversial statement.

Specialists and academics criticized the controversial decision. [which at first was objected to by Facebook](#) and Twitter. The company stressed that any order with global effect should be treated with caution as not to impact the rights of individuals in other countries. Later higher [fines and the risk of criminal liability to one of their employees](#) lead the companies to compliance, pending an appeal. [It has been reported](#) that the procedure was brought to the Inter-American Court of Human Rights attention alleging violations of the rights of the investigated, including freedom of expression, information and the press.

The Rapporteur is requested to dwell on the proper extent of judicial order. Taking into consideration that countries differ on how to implement certain rights - such as freedom of expression - and how to balance them against each other, it is to expect that **global orders that may impact human rights should be an exception and only applicable in situations of global consensus or to protect *jus cogens*.**

4. Tech Companies’ Initiatives

4.1 Introducing Friction to Information Ecosystems

A promising way to curb disinformation on social media platforms is by introducing friction to the information ecosystems they have helped create and now inhabit. This practice is relatively new - it gained steam during the 2020 presidential election in the US - but its impacts have been mostly positive for two reasons: (1) by decreasing the speed with which information is shared on social media, friction-inducing measures buy more time for users, fact-checkers and platforms to act and tackle disinformation campaigns, and (2) by relying

³ This measure raised controversy regarding the legality of the processing, as it was initiated *ex officio*, that is, without request of another body, such as the Public Ministry, the Attorney General's Office or a police authority.

on a nudge-like approach, these measures have only a minor (and at times non-existent) impact on freedom of expression.

Friction-inducing initiatives are basically small design tweaks introduced to the platform with the objective of making information flagged as potentially false or misleading less protuberant and harder to find or share. As the [New York Times](#) noted in a piece describing how Facebook and Twitter behaved on election day in the US, 'it's a telling sign of self-awareness, as if Ferrari had realized that it could only stop its cars from crashing by replacing the engines with go-kart motors'. Since their emergence in the 2000s and early 2010s, digital platforms fostered spaces where information could be shared seamlessly and serious journalistic investigations by the Times would appear side by side with bogus articles by the weeks-old Denver Guardian.

Another way to look at it is by analyzing the evolution of Facebook's (in)famous motto: 'move fast and break things'. The problem is that by moving way too fast and failing to notice the downsides associated with the speed with which information was being shared throughout its frictionless platform, Facebook came very close to breaking a very important thing: the truth. The company later [updated its motto](#) to 'move fast with stable infrastructure'. This is another telling sign of self-awareness. The company now understands that, in order to move fast *safely*, its infrastructure should be built over a sturdy foundation capable of addressing the downsides of high-speed information beforehand. No more 'moving fast, breaking things, and fixing it later'.

In this input, we would like to highlight some of these friction-inducing measures. Please note that we are not aiming for a comprehensive list. The first noteworthy example comes from Twitter when the company introduced a 'read before you retweet' prompt to its users in September of 2020. The idea is to nudge people into reading the articles they intend to share before they actually click 'retweet'. As the company [noted](#) in a tweet of its own, 'sharing an article can spark a conversation, so you may want to read it before you Tweet it'. If a user tries to share an article he or she did not read, Twitter would show a screen saying that 'headlines don't tell the full story' and that the user 'can read the article on Twitter before Retweeting'.

This is an important innovation because sensational headlines are often used by bad actors to spread false or misleading information on social media. According to the Washington Post, a 2016 study by scientists at Columbia University and the French National Institute found that '59% of links shared on social media have never actually been clicked: In other words, most people appear to retweet news without ever reading it'. By introducing friction to the retweet function, Twitter took a huge step towards fixing the problem. According to the company's [own metrics](#), 'people open articles 40% more often after seeing the prompt' and 'some people didn't end up RTing after opening the article', which points to a more informed conversation around the articles and decreases the chances that people will end up boosting false or misleading headlines.

Another interesting innovation was also introduced by Twitter during the presidential election in the US. The company decided to temporarily change the retweet function to nudge people into adding their own "quotes" before sharing a piece of content. As [The Verge](#) noted, 'Twitter is hoping that by introducing some friction into the process, people might better consider exactly what they're retweeting or take the opportunity to add their own perspective'. Although prompting users into adding their unique perspectives to the debate promotes a more informed debate and curbs disinformation, Twitter [rolled back](#) the changes after noting that 45% of Quote Tweets were 'single-word affirmations and 70% had less than 25 characters'. It remains to be seen if a 'Quote Tweet Prompt 2.0' can learn from this

experience and efficiently nudge people into joining the conversation in a more meaningful way.

Facebook also experimented with a number of temporary, friction-inducing measures during the 2020 election in the US. In a summary of the most important changes, the [New York Times](#) noted that the company has ‘put in place a new, cumbersome approval process for political advertisers, and blocked new political ads in the period after Election-day. It throttled false claims, and put in place a “virality circuit-breaker” to give fact-checkers time to evaluate suspicious stories. And it temporarily shut off its recommendation algorithm for certain types of private groups, to lessen the possibility of violent unrest’. In other words, Facebook temporarily replaced the engine of its Ferrari to allow fact-checkers and engineers to keep track of the vehicle and, if necessary, change tracks.

Furthermore, Facebook also introduced friction to its share function to help fight COVID-19 disinformation. The company calls the approach ‘Informative Sharing’ and it consists of offering people more background information on a given content before they click ‘share’. According to OneZero, ‘Facebook added notification screens making people pause, just for a moment, and consider some context before they share COVID-19 related posts, and outdated articles as well’. This is consistent with MIT Professor David Rand’s research on [‘scalable accuracy nudge intervention’](#). According to Rand, nudging people into thinking about the accuracy of a COVID-19 information before they share it significantly reduces the likelihood that they will pass along false or misleading information.

From the point of view of human rights law, **design tweaks that introduce friction to information ecosystems may be a solution that encroaches less upon freedom of expression in a significant way.** After all, most of the initiatives described above are nudges or, to put it differently, attempts to push human behavior in a certain direction without directly restricting people’s options or overly patronizing them. For example, when Twitter shows its users a screen asking if they would like to read the article before sharing, the company is not restricting user’s rights, only nudging the user to contribute to a more informed conversation online, making them more aware of what they are sharing. If the user decides that it is not in his or her best interest to actually read the content before retweeting it, he or she can simply ignore the prompt and move forward with the post.

This strategy is not without its risks. The decision on what to ‘nudge’ and in which direction may impact content and may have an effect on people’s choices. **Transparency concerning architectural changes is paramount in order to be compatible with human rights.**

4.2 Focusing on Big and Prominent Profiles

Focusing on large accounts has shown to be effective in curbing disinformation on social media. These companies often stress how difficult it is to monitor content on the platform, identify instances of misinformation or disinformation that violate their community standards and, finally, act accordingly. However, recent developments in the field point to a somewhat different direction. The experience with the deplatforming of former US President Donald J. Trump shows that platforms can go a long way in the fight against false or misleading information by monitoring large accounts that are responsible for a disproportionate share of bogus content. Nevertheless, unlike friction-inducing measures, **this practice raises serious concerns about human rights violations**, especially the right to free expression and the right to equality.

In a groundbreaking research on mail-in voter fraud in the US, Harvard Law Professor Yochai Benkler and his coauthors [found](#) that ‘contrary to the focus of most

contemporary work on disinformation, [the mail-in voter fraud disinformation campaign] was an elite-driven, mass-media led process. Social media played only a secondary and supportive role.’ According to the study, Donald J. Trump gamed the information ecosystem and turned mass-media outlets into purveyors of his elaborate lie of electoral fraud. Trump achieved that ‘by using three core standard practices of professional journalism [...]: elite institutional focus (if the President says it, it’s news); headline seeking (if it bleeds, it leads); and balance, neutrality, or the avoidance of the appearance of taking a side.’

The results of this study point to a difficult situation, they suggest that the problem of elite-driven, mass-media disinformation cannot be solved by fact-checking or algorithm-tweaking by social media platforms. Although the authors of the study do not venture this far off, one plausible conclusion is that, in order to effectively curb widespread disinformation campaigns like Trump’s voter fraud allegations, social media platforms should find ways to neutralize the source. Albeit for different reasons, Facebook, Twitter and other platforms suspended Trump for an indefinite period of time after the former US President incited a violent mob against Congress on January 6, 2021. Following ‘the great deplatforming’, the [Washington Post](#) noted that several independent research institutes came to a similar conclusion: online misinformation about the election dropped drastically.

A study by Signal Labs, for example, found that election misinformation dropped 75% after Trump lost access to his social media accounts. As the Washington Post article states, ‘the research by Signal and other groups suggests that a powerful, integrated disinformation ecosystem - composed of high-profile influencers, rank-and-file followers and Trump himself - was central to pushing millions of Americans to reject the election results and may have trouble surviving without his social media accounts’. To be sure, this raises an important question: how far social media platforms should go to curb disinformation campaigns? It seems clear that deplatforming Trump for good is an effective measure against false or misleading information about electoral fraud, but is it proportionate *vis-à-vis* the right to free expression? Although we do not have an answer yet, we believe that the Special Rapporteur’s annual thematic report should acknowledge this tension.

Moreover, the focus on large accounts also **raises a question of equality**. On the one side, the most obvious point is that social media platforms would need to apply their community standards in a non-congruent way, restricting the speech of prominent figures on the platform while smaller accounts may get a pass. But, on the other side, the deplatforming of Donald J. Trump also shows how platforms may be willing to act in the US - where, if they fail, their reputation is more vulnerable - but not necessarily in other countries where they may have less to lose keeping accounts up. Although other world leaders have been accused of using social media accounts to sow division and even promote violence, not all had their ‘digital megaphones’ restricted. For example, as Tom Phillips noted in his piece for [The Guardian](#), ‘calls for action have been particularly loud in Brazil, which has been led since 2019 by Jair Bolsonaro, a far-right tweeter-in-chief who basks in portrayals as the “tropical Trump”. Nonetheless, despite the accusations of repeatedly using social media to undermine democracy and incite violence and [spread misinformation](#), the Brazilian president’s social media accounts remain active.’

There is an important tension upholding freedom of speech and finding ways to restrict significant vocal accounts that are accused of or are actually spreading disinformation. Yet, it is important to highlight that it may

5. Civil society initiatives

Civil society organizations have a prominent role in curbing disinformation while enhancing human rights. Media literacy initiatives are of significant importance in the efforts to deal with disinformation, particularly when they congregate different stakeholders.

5.1 The Brazilian Superior Electoral Court's Confronting Disinformation Program

One promising initiative was launched by the Brazilian Superior Electoral Court (TSE in its Portuguese acronym). TSE's [Confronting Disinformation Program](#) aiming at the 2020 Brazilian elections adopted a multistakeholder approach. The Program was [supported by 57 institutions](#), including political parties and public and private entities.

The program was organized under different strategies, of relevance are: (i) Media and Information Literacy; (ii) Containment of Disinformation; (iii) Identification and Disinformation Check; (iv) Improvement of the Legal Order; and (v) Improvement of Technological Resources. The full description of the program and its strategic plan can be found in Portuguese [here](#).

A fact-checking coalition was formed for the elections. The news from a group of nine checking agencies was published on the ['Fact or Rumor' page](#), available on the Electoral Justice Portal.

Additionally, the program has partnered with four of the world's largest social media and messaging platforms - Google, Facebook, Twitter and WhatsApp. For example, in the case of [Google](#), voters seeking information about the elections found at the top of the search results reliable and publicly useful content prepared by the Electoral Justice.

The partnership between the [Electoral Court and WhatsApp](#) to fight the mass firing of messages in the 2020 Elections gave citizens the power to denounce these illegal practices that can unbalance the electoral process.

During the electoral period, from September 27 to November 29, a platform was created for receiving reports of accounts suspected of sending massive amounts of disinformation messages. It received 5,180 reports, 199 of which were dismissed as unrelated to the elections. After this first filter, the Electoral Court sent 4,981 reports to WhatsApp, to check for possible violations of the app's Terms of Service.

After a preliminary review to remove duplicate or invalid numbers, WhatsApp identified 3,527 valid accounts and banned 1,042 numbers (29.5%) for violating its Terms of Service. Of the total banned accounts, more than 64% were blocked proactively and automatically by the WhatsApp integrity system, even before being reported.

The creation of the "Electoral Dispatch on WhatsApp", a chatbot developed free of charge by Infobip to facilitate access to reliable data on the elections, was also important for the Electoral Court to disseminate relevant information to Brazilian voters. During the electoral period, this channel had more than one million unique users and more than 18 million messages were exchanged with the robot within the application.

Such **multistakeholder partnerships** may have an impact on protecting rights as they **expedite action and provide an opportunity for different actors to do their part and be heard.**

5.2. Civil Society media literacy initiatives:

As highlighted by the [Inter-American Human Rights Commission](#), there is an ‘urgent need to promote digital literacy programs aimed at developing civic skills in a perspective of democratic coexistence and with a human rights approach.’ Initiatives from civil society may address this very *lacunae*. Several of them focus on different aspects of the phenomenon of disinformation.

[‘Exposing bot disinformation in Brazil’](#), for instance, is an effort to ensure that people are aware of the use of robots in social media, particularly those that spread disinformation. Through an algorithm that promotes transparency on bot behavior in social media, the project strengthens ownership of internet users, civil society organizations, media organizations and policy makers to identify and contextualize disinformation campaigns. So far the project’s algorithm (‘Pegabot’, bot catcher in English) has been used more than 100k times to check the probability of Twitter profiles being automated. Besides this, 750 members of the electoral justice system participated in trainings on disinformation and automation so they could incorporate this knowledge into dealing with electoral processes affected by disinformation.

Hence, initiatives by third parties based on media literacy and transparency may prove to be important ways to tackle the challenges posed by disinformation. Not only they may raise awareness about the issue, they also create resilience amongst those affected.

5.3. The role played by ‘digital influencers’

‘Digital influencers’⁴ are central to today’s information ecosystem. In social media, they are significant information producers and distributors. Spontaneously, influencers move trends and themes, they frame debates. In the Global South where a majority of the population consume their news through social media, influencers are one of the most important sources.

Initiatives aimed at raising awareness of digital influencers’ role in spreading information may improve the overall information environment.

Projects such as the Brazilian [‘Redes Cordiais’](#) which trains communicators from different niches to fight misinformation and hate speech on social media help forge a path to the future. They do not restrict rights, just the opposite, they empower not only influencers, but also the audience in a dialog on how to fight disinformation. As an example, Redes Cordiais has in 16 months brought together 109 influencers - who in total have more than 68.3 million followers - to train and discuss how to identify and translate to their public what is disinformation and hate speech online. Thus, the influencers become responsible

⁴ In general, "digital influencer" is understood as a new type of celebrity on the Internet, a person who accumulates many followers and is able to modify the opinion or behavior of his audience.

propagators aiming at finding reliable information whenever they discuss issues with their public.

Hence, the Special Rapporteur may recognize that initiatives aimed at raising awareness of digital influencers' role in spreading information help in the fight against disinformation, supporting the spread of reliable information and raising the level of media literacy of the population.

Conclusion

A disinformation global crisis persists. It is particularly challenging to effectively navigate the problem while protecting freedom of expression and access to information. Regulation and safeguards may play a role, but as the experiences so far demonstrate, they bring challenges of their own and may not be enough.

The solution should be beyond the government and beyond imposing burdensome obligations to internet intermediaries - changing liability regimes. Rights at stake are of extremely importance, they are at the very foundation of democracy and it is not up to the private sector to regulate them. What is necessary is to recognize that it is a multifaceted problem that needs a holistic approach. Different actors under the guidance of the government should engage in multistakeholder, multi-prong initiatives with each having an important role to play.

We recommend that the Special Rapporteur deeply consider the complexity of the scenario and to seek a holistic approach, recognizing that variety of contributions that can be made in order to create a better environment more conducive to the protection of people's rights.