Settled Human Rights Standards as Building Blocks for Platform Accountability and Regulation

A CONTRIBUTION TO THE BRAZILIAN DEBATE

ELECTRONIC FRONTIER FOUNDATION

accessnow

Since at least a decade ago, human rights institutions have acknowledged the enabling potential of the internet to realize a range of human rights. Digital technologies have been incredibly transformative tools for allowing people to speak out against arbitrary acts of public and private powers, empowering the expression of historically vulnerable, marginalized and silenced groups, catalyzing civic organization and participation, and facilitating innovative ways to collectively build and share knowledge. Since then, the right to seek, receive, and impart information has enabled the exercise of other rights and strengthened the internet ecosystem, but not without backlashes and critical challenges.

The current discussion about platform regulation in Brazil, both in the draft bill known as "PL 2630" and in constitutional cases pending in the country's Supreme Court, demonstrates that much effort is going into addressing these challenges, but also shows that proper responses are not simple to craft. We should be able to tailor these responses safeguarding the positive potential of digital technologies and the essential role freedom of expression, including access to information, plays in preserving democratic societies.

Quick Background

The PL 2630, also known as the "Fake News Bill", was first introduced in the Brazilian Senate in 2020. The push from civil society organizations and coalitions, such as *Coalizão Direitos na Rede*, to improve the text and their work with the bill's rapporteur in the Chamber of Deputies were critical to neutralize threats like the traceability mandate of end-to-end encrypted messages. By then, Brazilian digital rights groups had also stressed that the regulation should focus on content moderation processes (e.g. transparency and due process rules) rather than restriction of certain types of content. After the release of a new draft text in early 2022, the bill remained halted in Brazil's Chamber of Deputies until the beginning of 2023.

Following the failed attempt earlier this year of the far right to overthrow the new administration of President Lula da Silva and a peak of violent attacks in Brazilian schools, PL 2630 has consolidated its position as the legislative path to address more comprehensive concerns on the use of digital technologies in contexts of social unrest. For that, the Executive branch proposed to the bill's rapporteur a new text that introduced several changes, looking at laws like the German *NetzDG*, the EU Digital Services Act (DSA), and draft legislations such as the controversial UK Online Safety Bill .

The latest published version of the bill incorporates some of these proposals, such as risk assessment rules, duty of care obligations, and new exceptions to Brazil's general rule to online intermediary liability. According to Article 2 of the bill, it applies to social networks, search mechanisms, and instant messaging services constituted as a legal entity and with more than ten million monthly users in Brazil. Although the DSA is often mentioned as an inspiration and democratic precedent grounding the new proposal, the revamped bill has important differences and still fails to ensure checks and balances considering the Brazilian context and institutional framework.

In parallel, the country's Supreme Court has pending cases about online intermediary liability (general repercussion issues 533 and 987) and the blocking of websites and applications by judicial authorities (ADI 5527 and ADPF 403). Currently, the general online intermediary liability regime in Brazil is set by Article 19 of Law n. 12.965/2014, known as Marco Civil da Internet. According to Article 19, internet applications can be held liable for user content only when they fail to comply with a judicial decision ordering the removal of infringing content. There are exceptions where an extrajudicial notice can make platforms liable for third-party content. They are copyright infringement, unauthorized disclosure of private images containing nudity or sexual activity, and content involving child sexual abuse.

Some Supreme Court's justices have expressed their opinion that Marco Civil's general regime needs an update to stiffen online intermediary rules, and the pending constitutional cases may be a way to do so if Congress does not address the issue in a timely manner. While the increasingly powerful role of major internet applications has prompted debates and initiatives to review current intermediary liability regimes across geographies, there are key questions we must ask, tools we should consider, and lessons learned to build on before introducing changes that can seriously impact protected expression and people's ability to strengthen their voices and rights using digital technologies.

In turn, the Supreme Court's ruling on blocking of websites and applications has been halted since 2020, when Justice Alexandre de Moraes requested the file for review, returning it only in March this year. These cases refer to WhatsApp blockings in Brazil in 2015 and 2016, involving the issue of whether authorities could require an internet application to undermine its privacy and security features by design, *i.e.*, end-to-end encryption, to disclose user communications data within a criminal investigation. The ruling started in 2020 with key Justices' votes supporting privacy and security protections inscribed in digital systems' architecture and rejecting interpretation of Brazilian law to allow state-ordered blocking aimed at impairing such protections. Unfortunately, the possible outcomes of resuming this ruling in the current context is unpredictable. Following its pioneering role of recognizing personal data protection as a fundamental right in Brazil's Constitution, it's crucial that the Supreme Court endorses Justices Rosa Weber and Edson Fachin votes in favor of robust privacy and security by design.

Despite moves from the Executive branch and the Supreme Court for changes in Brazil's current legal framework, political actors agreed, at least for now, that Congress is the proper venue for a democratic debate on platform regulation. We agree. It's relevant, then, to look into the draft law under discussion. While it contains positive elements, we must also highlight points yet to be improved.

Important Points of Concern

The PL 2630 purports to strengthen users' rights in face of the power of large internet applications, like Facebook, Youtube, and Twitter. Yet, there are crucial points of concern that Brazil's regulation debate and PL 2630 should carefully tackle. Other

groups in the region, like Derechos Digitales, have raised points of attention. As we further elaborate on this piece, there is a set of issues that stakeholders must consider and address before passing a new law. The most relevant are:

- Neutralize risks of abuse of content-based regulations, dropping duty of care obligations, focusing on systemic impact assessments, and making it explicit that platform accountability doesn't mean general monitoring and filtering of user content.
- Ensure robust checks, balances, and due process safeguards for the application of specific rules to situations of conflict and imminent risk.
- Carefully design and ensure adequate means to establish a proper independent, autonomous, participative, and multi-stakeholder oversight structure for the upcoming regulation.
- Establish clear safeguards against increasing surveillance and related security risks.
- Refrain from giving special speech protections to government officials, who bear special responsibilities under human rights standards.
- Ensure sanctions in accordance with human rights standards and due process guarantees, particularly when it involves blocking online applications.

The last point refers to the administrative penalties that may apply in case internet applications within the scope of the bill fail to comply with its rules. The "temporary suspension of activities" is among this list of penalties. In practice, this means that a government administrative authority would have the power to block an entire website or app. Website blocking in Brazil generally happens following a judicial order, although the Ministry of Justice has recently stated that consumer administrative bodies would have this authority as per traditional suspension penalties set in consumer law. Human rights standards indicate that blocking of entire websites and applications is an extreme measure with technical challenges, great risks of abuse, and significant impacts on fundamental rights. In 2021, the UN Human Rights Council reiterated the adoption of a resolution unequivocally condemning the use of internet shutdowns and online censorship, which includes social-media shutdowns, to arbitrarily prevent or disrupt access to or dissemination of information online. We highlighted such concerns in the context of PL 2630. And while in previous versions of the bill only an absolute majority of a judicial collegiate body could apply this blocking penalty, the current draft gives this power to an unspecified administrative authority. Brazilian lawmakers should acknowledge the dangers of the arbitrary use of online blockings and step back.

Additionally, the legitimate enforcement of possible sanctions is closely tied to the bill's set of rules and oversight structure. The other points of concern we mention above highlight relevant remaining gaps on this front. We elaborate on them in the next section.

From 2011 to 2023: Address Current Challenges Building on Existing Principles and Safeguards

Since the 2011 Joint Declaration on Freedom of Expression and the Internet of Special Rapporteurs for Freedom of Expression, human rights institutions have underscored that government initiatives seeking to regulate online communications should preserve and adapt to the unique characteristics of the internet. This is for these initiatives to be both effective and respect internet features enabling fundamental rights and freedoms. Any restrictions must follow the "three-part test", that is, they must be clearly set by law, strictly necessary and proportionate to achieve a legitimate aim in a democratic society. Important concerns around internet fragmentation, collateral censorship, over-removal of legitimate expression, and more recently, inherent intricacies of content moderation at scale, have led experts throughout the years to avoid content-specific regulations. The risks of arbitrary application and interpretation of content restrict rules in nondemocratic or conflictive settings add other layers to this set of concerns.

We detail our points of concern below.

Concerning Duty of Care Obligations

The progression of PL 2630's versions was an expression of opting for a process-based approach instead of a content-focused one within a regulation initiative aiming to advance platform accountability. However, after amendments earlier this year, the bill now contains a list of illicit practices, connected to illicit content, that internet applications "must act diligently to prevent and mitigate (...) making efforts to improve the fight against the dissemination of illegal content generated by third parties." This relates to a duty of care obligation that the bill doesn't define, but nevertheless operationalizes its application, mainly in Article 11. The list of such illicit practices in Article 11 points to provisions in six different laws that amount to around 40 criminal offenses, each one containing a set of elements that must be present for the conduct to be illegal. Some offenses also have causes that exclude certain conduct from being the basis of a crime. For instance, both Brazil's Antiterrorism Law (Law n. 13.260/2016) and the crimes against the democratic state set in the Penal Code don't apply to critical political demonstrations based on constitutional rights. As per Article 11 of the bill, it would be up to the internet application to consider all these elements and assess whether conduct or content visible through their platforms constitute a criminal activity.

In some cases, it's even harder to understand what exactly the provider should check, or whether it is something the provider should check at all, despite its inclusion in the list of Article 11 criminal offenses. For example, Article 11 generically refers to the crimes

against children and adolescents of Law n. 8.069/1990. Among these criminal offenses, there's the failure of a doctor, nurse, or the head of a healthcare facility to correctly identify a newborn and its birth mother at the time of delivery (Article 229 of Law n. 8069/1990). What's the duty of care expected from internet platforms here? This rule is an example of a provision encompassed by Article 11 which doesn't seem to have any clear relationship with online platforms. Article 11 is also not very clear about how and which institution(s) will assess the compliance of duty of care obligations by internet applications. It states evaluations will not focus on isolated cases and will include information internet applications provide to authorities on their efforts to prevent and mitigate the practices listed, as well as analysis of platform's reports and on how they respond to notices and complaints.

Within the same bill, Article 45 stipulates that "when the provider becomes aware of information that raises suspicion that a crime involving threat to life has occurred or may occur, it must immediately report their suspicion to the competent authorities." While a crime involving a threat to life is definitely an emergency and a dire situation, Article 45 establishes a new policing role for internet applications that, even within this strict scope, may give rise to controversial outcomes, potentially affecting, for example, women in Brazil seeking information online about safe abortion.

Duty of care obligations as set in PL 2630 rely on a regulatory approach that reinforces digital platforms as points of control over people's online expression and actions. They require internet applications to be judges of whether acts or content are lawful based on a list of complex criminal offenses, as if it were simple for content moderation tools and processes to be programmed to recognize every element that constitutes each offense. But, to the contrary, these are often close calls that even judges and juries may have difficulty with. In many cases, users disseminate sensitive content precisely to call out institutional violence, human rights violations, and the perpetration of crimes in conflict situations. Sharing videos on social networks that expose cases of discrimination contribute to holding the perpetrators accountable. During the wave of protests in Chile, internet platforms wrongfully restricted content reporting the police's harsh repression of demonstrations, having deemed it violent content. In Brazil, we saw similar concerns, for example, when Instagram censored images of Jacarezinho's community's massacre in 2021, which was the most lethal police operation in Rio de Janeiro's history. In other geographies, the quest to restrict extremist content has already removed videos documenting human rights violations in conflicts in countries like Syria and Ukraine.

As the Office of the Inter-American Commission on Human Rights (IACHR) Special Rapporteur for Freedom of Expression pointed out, as private actors, internet applications *"lack the ability to weigh rights and to interpret the law in accordance with freedom of speech and other human rights standards,"* particularly when the failure to restrict specific contents can lead to administrative penalties or legal liability.

It's not that internet applications shouldn't make efforts to prevent the prevalence of pernicious content in their platforms, or that we don't want them to do a better job when dealing with content capable of causing serious collective harms. We agree they can do better, especially by considering local cultures and realities. We also agree that

their policies should align with human rights standards and that they should consider the potential impacts of their decisions to human rights, preventing and mitigating possible harms.

However, we should not mix platform accountability with reinforcing digital platforms as points of control over people's online expression and actions. This is a dangerous path considering the power big platforms already have and the increasing intermediation of digital technologies in everything we do. Article 11's approach is also problematic in that it establishes such control based on a list of potentially unlawful practices that political forces can change and expand at any time or lead to opportunistic or abusive enforcement to restrict access to information and silent criticism or dissident voices.

On the contrary, platform accountability prioritizes a process-based and systemic approach by which the provider assesses and addresses, to prevent and mitigate, the negative impacts of its activities to human rights. This is consistent with the UN Guiding Principles on Business and Human Rights. The PL 2630 itself has provisions on systemic risk analysis and mitigation measures related to companies' activities. Brazilian lawmakers should prioritize this approach over the concerning "duty of care" obligations.

Moreover, the concept of duty of care, as we currently see in the Brazilian debate, has yet another risk. It allows for interpretations that internet applications should engage in general monitoring of the user content they host. Such interpretations are not explicitly denied in the text of PL 2630, as they are, for example, in the EU DSA.

Repel Rules and Interpretations That Can Lead to Content Monitoring Obligations

The Special Rapporteurs for Freedom of Expression have also stated: *"At a minimum, intermediaries should not be required to monitor user-generated content."* And that: *"Content filtering systems which are imposed by a government and which are not end-user controlled are not justifiable as a restriction on freedom of expression."*

There are at least two main reasons why general monitoring obligations are a very bad idea. First, such obligations are perhaps the ultimate expression of treating internet applications as a policing force of everything we do and say online, with pernicious consequences to free expression, access to information, and overriding privacy expectations. If applications' commercial practices often raise similar concerns, societal pushback to corporate surveillance has driven data privacy regulations and changes in companies' policies to better protect user privacy. Second, general monitoring and related pervasive filtering constantly fail, and the fact it performs poorly poses even more concerns to human rights. Given the sheer volume of new content that people post and share on internet applications every minute, content moderation increasingly relies on automated tools, reflecting their limitations and flaws. Regulations or interpretations

mandating the adoption of these tools and tying such an obligation to sanctions or liability of internet applications amplify the potential for errors and problematic enforcement.

Speaking just in terms of probability, when a system that's already prone to making mistakes is scaled up to moderate content that churns out at a rate of many millions to billions of entries per day, more mistakes will occur. And when learning models are employed to educate the artificial intelligence (AI) inside these methods, there are rarely chances for the learning models to recognize and self-correct those mistakes. More often than not, such technologies reproduce discrimination and biases. They are prone to censoring legal, non-offending, and relevant speech. While we advocate, and will continue advocating, for human review in content moderation processes, having enough human moderators working in adequate conditions to prevent undue content restrictions will be a continuous challenge.

AI systems usually employed in content moderation include image recognition algorithms and natural language processing models. As for the intricacies of training AI language models, experts underscore that language is highly dependent on cultural and social contexts, and varies considerably across demographic groups, topics of conversation, and types of platforms. Moreover, training language processing algorithms demand clear and precise definitions of targeted content, which is very hard to achieve with complex terms normally implicated in characterizing a criminal or illicit practice. Even if we generally consider that the current stage of available natural language processing tools perform effectively in English, they vary significantly in quality and accuracy for other languages. They can also reproduce discrimination in data, disproportionately affecting marginalized communities, like LGBTQIA+ people and women. Multilingual language models also have their limitations, as they may not reflect well the day-to-day language used by native speakers and fail to account for specific contexts.

In turn, despite current advances in technology, image recognition tools also have their limitations. A good example relates to sexual imagery recognition. Since even people can't agree on where the line is drawn regarding offending and non-offending sexual imagery, the systems we build to automatically recognize it and remove it from online platforms will naturally tend towards the more conservative estimates to minimize legal risks. Without value judgment, that means expression that is otherwise protected, legal, and often coming from sexual minorities, will be deemed inappropriate. A landmark case of platform censorship in Brazil precisely reflects this problem. In 2015, Facebook blocked a picture from the early 20th century of an indigenous couple partially dressed, posted by the Brazilian Ministry of Culture to release the launch of the digital archive Portal Brasiliana Fotográfica right before Brazil's Indigenous Day.

Relatedly, and as we edge closer to sophisticated AI systems able to accurately determine sexual imagery from other material, we stumble onto the age-old problem of art versus porn. Classical art that depicts the nude form continues to be flagged as improper by moderation algorithms, despite overwhelming consensus that it is firmly in the "art" category, and not illegal or contrary to community standards. Contemporary art further blurs those boundaries, often intentionally. Our capabilities for expression as

humans are ever-changing, and this will continue to be a challenge for developers of computer systems built to recognize and categorize user-generated content, which at scale will produce even more mistakes.

A considerable rate of mistakes can also happen in image recognition systems based on hashes. Common errors faced by this type of technology, such as the so-called "collisions," occur because two different imagens can have the same hash value, leading to false positives, where an image is incorrectly identified as something it is not. This can occur for various reasons, such as if the images are very similar, if the hash function is not very good at distinguishing between different images, or if the image has been corrupted or manipulated. The opposite can also occur, that is, to manipulate infringing images so the hash function does not recognize and flag them. Beyond efficiency issues, these systems undermine protections in the architecture of digital platforms that, by design, ensure the inviolability of communications, privacy, security, and data protection, as is the case with end-to-end encryption.

When moderation systems are scaled up to disproportionately large sizes, the reach of their attached monitoring and reporting obligations, if existent, are scaled the same way. And these things can, and have been, tooled as the eyes and ears of arbitrary, nondemocratic forces.

Platform regulation should not incentivize interpretations or further regulation demanding general content monitoring and filtering. PL 2630 should be more explicit to repel such interpretations, and Brazil's regulatory debate over platform accountability should reject such mandates as not being necessary and proportionate responses.

Robust Checks, Balances, and Due Process Safeguards for Exceptional Measures in Crisis Situations

PL 2630 also establishes special obligations for when there is an imminent risk of damage or negligence of an application provider (Articles 12-15). In assessing this section of the bill, it's crucial to recall the 2015 Joint Declaration about crisis situations. Among other recommendations, it highlights that *"[s]tates should not respond to crisis situations by adopting additional restrictions on freedom of expression, except as strictly justified by the situation and international human rights law. Administrative measures restricting freedom of expression should be imposed only where they can be justified pursuant to the three-part test for such restrictions."*

While this section of the bill purports to act as a legal basis for restricting fundamental freedoms during crisis situations, its current language fails to provide enough precision and clarity, as well as proper checks and balances to substantiate an intervention that is necessary and proportionate.

According to PL 2630, the decision implementing the security protocol will specify, among others things, the impacted providers, the protocol's deadline (up to 30 days, which can be extended), and a list of relevant issues or requirements that providers must address through effective and proportionate mitigation measures during the protocol's period. While the protocol is in force, and for the types of content specified in the implementation decision, the impacted providers are subject to joint and several liability for user-generated content as long as providers have prior knowledge of such content. A simple user notification, using the notice mechanism Article 16 requires internet applications to provide, is enough to constitute such prior knowledge. The bill, thus, creates an exceptional notice-and-takedown mechanism to be applied while the protocol is in effect and relating to certain types of contents (as per the protocol's "thematic delimitation").

Notice-and-takedown mechanisms raise many concerns. They can fuel the weaponization of notice systems to censor critical reporting, political criticism, and voices from marginalized groups. They too often lead to over-removals. The Office of the IACHR Special Rapporteur for Freedom of Expression has noted that they create incentives for private censorship as they put *"private intermediaries in the position of having to make decisions about the lawfulness or unlawfulness"* of user-generated content. Such *"intermediaries are not necessarily going to consider the value of freedom of expression when making decisions about third-party produced content for which they might be held liable."* Brazil's own experience in courts shows how tricky the issue can be. InternetLab's research based on rulings involving free expression online, released five years after Marco Civil's approval, indicated that Brazilian courts of appeals denied content removal requests in more than 60% of cases. In the public hearing that Brazil's Supreme Court held to receive inputs on its cases about online intermediary liability, the Brazilian Association of Investigative Journalism (ABRAJI) presented data about takedown requests filed in courts from 2014 to 2022. According to ABRAJI, at some point of the judicial proceedings, judges agreed with content removal requests in around half of the cases, and some were reversed later on.

Yet, PL 2630's notice-and-takedown mechanism attached to a security protocol seems to play a moderating role amidst an increasing push from the Executive branch and the Supreme Court to expand the exceptions to Marco Civil's general rule on online intermediary liability. The fact this mechanism would be limited in time and in scope could help with some of the concerns above, as well as Article 18's rules, which include users' right to appeal content moderation decisions. However, the overall dynamic of the security protocol still poses serious problems. A paramount concern is that crisis situations don't become permanent by extending the duration or reiterating the occurrence of measures that, by definition, are restricted to exceptional circumstances. Clear and effective controls are required so that a legal discipline for crisis situations doesn't turn into the standard regulation.

Here are the main issues and possible mitigations Brazilian lawmakers should consider:

- Article 12 defines a crisis situation in an extremely broad way. The imminence of risks set in Article 7, which includes a range of issues (e.g., the dissemination of illicit contents listed in Article 11 and risks to freedom of expression, public

health, and the democratic State), **or** the "negligence or insufficiency of a provider's action" is enough to trigger the implementation of the security protocol. The criteria to typify what constitutes such insufficiency or negligence depend on regulation that is yet to exist. However, the provision doesn't relate the application's negligent action to the risks set in Article 7. An insufficiency or negligence of a provider related to any matter or an imminent risk set in Article 7 is enough to configure a crisis situation. This also means that even if providers are taking important steps in good faith to address Article 7's imminent risks , they can still be subject to the security protocol's exceptional measures. At a minimum, **the provision should combine both requirements, using *and* instead of *or* in its language**. But there are still other critical concerns.

- Previous versions of the bill qualified the protocol's situation of imminent risk. It used to refer to "imminent risks of harm to the collective dimension of fundamental rights." This is a critical qualifier, especially because Article 7 is still quite broad in the risks it lists. While its checklist may work to guide big provider's impact assessments, it raises concerns about possible abusive interpretations and malicious uses in the context of a security protocol that sets exceptional obligations to internet applications. Hence, there should be **a risk of harm to the collective dimension of fundamental rights** to allow an authority to put this security protocol in place. Furthermore, the bill should be explicit **that the authority's assessment must follow strict necessary and proportionate standards when making such a decision**.
- The bill is silent about which authority has the power to declare a crisis situation and establish the security protocol's terms. We address the bill's oversight design in the next section, and the fact it currently lacks a proper democratic oversight structure is a major concern within the application of a security protocol. The 2015 Joint Declaration states that *"[a]dministrative measures which directly limit freedom of expression, including regulatory systems for the media, should always be applied by an independent body. It should also be possible to appeal against the application of administrative measures to an independent court or other adjudicatory body."* **In this regard, and building on important related safeguards, the security protocol mechanism should count on robust checks and balances, including:** *(i)* an independent government entity or oversight structure that assesses the crisis situation based on clear, transparent criteria and determines the implementation, or extension, of the security protocol by a reasoned decision within a public administrative proceeding abiding by due process safeguards; *(ii)* a referendum or prior consultation of a multistakeholder, participative council as part of the decision proceeding (both for implementing or extending the protocol); *(iii)* just like the administrative proceeding, not only a summary, but the resolution itself implementing or extending the security protocol is public; *(iv)* the right to a judicial review; *(v)* proper ongoing transparency over providers' measures deriving from the security protocol and government-related oversight activities.
- Finally, Article 16 setting the notice mechanism leaves crucial definitions to further regulation. It should at least clarify that user notices must specifically indicate the location of the allegedly unlawful material and explain why the user deems it unlawful. The bill should also make it explicit that due process safeguards that Article 18 ensures for users who have their content restricted

remain applicable in the context of a security protocol, covering the providers and types of content affected, and the entire period the protocol is in effect.

Proper Independent and Participative Oversight Structure

The bill stipulates obligations to internet applications and powers to an unspecified administrative authority to oversee compliance with PL 2630's rules. The bill's enforcement without a genuinely independent and democratic oversight structure jeopardizes its purported goals. So far, the proposal's text fails to ensure the basis for such a structure, giving a greater margin to arbitrary enforcement of PL 2630 rather than setting the grounds for preventing such abuses. Although Legislative branch-proposed bills have limits in creating new entities within the federal administration, this is a political equation that Brazil's Congress and federal government must sort out, in debate with civil society, before passing PL 2630.

Anatel, the Brazilian telecommunications regulatory agency, has been working to fit as the answer. The agency already exists and counts on essential attributes ensured by law, such as administrative independence, absence of hierarchical subordination, stability of its directors, and financial autonomy. Yet, its expertise and legal mandate pertain to telecommunications services and infrastructures, not to internet applications and content moderation activities. Moreover, Anatel has a bad track record in fulfilling its mandate as a telecommunications oversight agency and ensuring meaningful civil society participation in its decisions.

Coalizão Direitos na Rede emphasized a set of Anatel's shortcomings in a public statement released earlier this year. Among them, the digital rights coalition criticizes Anatel's favoring of large telecom operators in the auction of 5G spectrum bands. It also points out flaws regarding Anatel's oversight efficiency and transparency, based on reports from Brazil's Federal Court of Auditors (TCU). Conversely, *Coalizão Direitos na Rede* advocates for a new autonomous oversight agency backed by a participative and multi-stakeholder council.

This is in line with the Special Rapporteurs for Freedom of Expression's 2019 Joint Declaration, upholding "independent and multi-stakeholder oversight, transparency and accountability mechanisms to address private content rules that may be inconsistent with international human rights and interfere with individuals' right to enjoy freedom of expression."

The Special Commission on Digital Rights of the Brazilian Bar Association (OAB) has also proposed a more elaborate oversight structure. It would involve three fronts: *(i)* an oversight and deliberative entity formed by representatives of the government's three branches (Legislative, Executive, Judiciary), Brazil's competition and data protection authorities, Anatel, and OAB; *(ii)* a self-regulatory entity responsible for addressing specific cases of content moderation, and *(iii)* Brazil's Internet Steering Committee

(CGI.br), which already plays a key role issuing studies, guidelines, and recommendations for the development of the internet in Brazil. One crucial point is that any design must uphold CGI.br's current role and nature.

Coalizão Direitos na Rede's and OAB's Special Commission's proposals reflect the need for robust checks and balances, including meaningful civil society participation, in PL 2630's oversight design. This is still missing, and filling this fundamental gap demands a committed and participative debate.

Clear Safeguards Against Incrementing Surveillance and Related Security Risks

Given the new obligations PL 2630 sets to providers, including specific rules for crisis situations, it's important to make it explicit that none of its provisions will imply changes in platforms' systems to introduce security vulnerabilities or undermine privacy protections by design. This is particularly crucial to preserve the features of end-to-end encrypted applications and avoid intents to weaken encryption's fundamental principles and protections.

In this sense, the 2016 Joint Declaration of Freedom of Expression Special Rapporteurs addressing government efforts to combat violent extremism underlines that States should not adopt, and should review, laws and policies that involve measures weakening existing digital security tools. Article 8 of PL 2630 already stipulates that measures providers implement in compliance with the bill should preserve information security and personal data protection. This is good, but the provision should go further to explicitly repel applications of the law seeking to introduce vulnerabilities in platforms' systems or make internet applications adopt any other measures that can systematically increase the risk of security incidents.

Moreover, the bill contains rules that expand existing data retention obligations. On this point, the 2015 Joint Declaration about crisis situations states that *"requirements to retain or practices of retaining personal data on an indiscriminate basis for law enforcement or security purposes are not legitimate. Instead, personal data should be retained for law enforcement or security purposes only on a limited and targeted basis and in a manner which represents an appropriate balance between law enforcement and security needs and the rights to freedom of expression and privacy."*

The most problematic language related to data storage obligations is found in Article 46 of PL 2630. The text requires internet applications to preserve metadata associated with all content that was removed or disabled in compliance with PL 2630 rules or judicial orders. Although it may seem, at a first glance, a "targeted" measure related to potentially offensive content, the volume of restricted content will likely be massive by the very nature and dynamic of user content creation on big platforms. If it makes sense

to store the restricted content for a specific period, the bill's language is overbroad on the related metadata that applications would have to store along with such content.

As per Article 46, the storage obligation includes "any related data and metadata removed" along with the content, as well as the respective IP address, access logs, networking ports, subscriber information (e.g. name and address), "telematic data," and "other records and user information that can be used as probative material, including those related to the form or means of payment, if any." The storage period is 6 months, which can be extended.

Brazil's Data Protection Authority (ANPD) issued a statement criticizing the vague nature of provisions in the bill establishing the collection of personal data for criminal investigation purposes, with specific references to the language of Article 46. According to ANPD, "PL 2630/20 establishes data storage obligations for criminal investigation purposes using vague and imprecise expressions, which can lead to a disproportionate expansion of personal data collection, or even to abusive tracking and surveillance of personal data subjects." The Brazilian data protection authority highlights that government authorities must observe the need for setting the specific purposes for the processing of personal data, limit such processing to what is strictly necessary to achieve these purposes, adopt security measures proportionate to the risks involved, and ensure wide transparency of personal data processing operations. In this sense, ANPD recommends lawmakers review the bill's text to expressly and explicitly indicate which data may be collected.

Building on the principles of purpose, necessity ("data minimization"), and prevention of Brazil's Data Protection Law, the standard storage of metadata related to restricted content in the bill should not go beyond Marco Civil's data retention rules. With Marco Civil's retention of "access to application logs," which includes the user IP address, authorities can start an investigation and, within its proceedings, may request additional information or conduct further examinations as needed and depending on each case.

Review Problematic Immunity for Public Officials

Article 33, paragraph 6 of the bill extends the immunity that Brazil's Constitution ensures for members of Parliament for their opinions, words, and votes in the exercise of their mandates to content published by "political agents" on social networks and private messaging platforms. The term "political agents" in the provision seems to encompass any elected officials in the Executive and Legislative branches at the federal, state, and municipal levels, as well as ministers of state, state and municipal secretaries, and the heads of government entities in general. If this provision is approved, this large set of public officials would be immune to civil and criminal liability for the content they publish online.

The bill gives special speech protections to public officials, while inter-American freedom of expression standards acknowledge that these officials, on the contrary, bear

special duties for their statements. These include the duty to ensure that their statements do not constitute arbitrary interference, direct or indirect, with the rights of those who contribute to the public discourse through the expression and distribution of their thoughts, the duty to ensure that their statements do not amount to human rights violations, and the duty to reasonably verify the facts on which their statements are based.

In view of these duties, the 2021 Joint Declaration of Freedom of Expression Special Rapporteurs addressing increased concerns with the spread of disinformation stressed that States should *"a) [a]dopt policies which provide for disciplinary measures to be imposed on public officials who, when acting or perceived to be acting in an official capacity, make, sponsor, encourage or further disseminate statements which they know or should reasonably know to be false. b) [e]nsure that public authorities make every effort to disseminate accurate and reliable information, including about their activities and matters of public interest."*

The bill, whose roots rely on similar concerns, contains such a provision that seems to neglect the role that prominent public officials play in creating, funding, and disseminating harmful content online. This provision contradicts PL 2360's purported goals, and Brazilian lawmakers should reject its text.

Conclusion

Any laws seeking to strengthen users' rights in the face of dominant internet applications should build on these principles and safeguards instead of ruling them out. We will not be able to offer responses to challenges arising from the constant but ever-changing interrelation between digital technologies and society if we disregard relevant settled bases, grounded in human rights standards, at each step of this way. Empowering users before dominant internet platforms' huge corporate power also involves more structural and economic measures that are mainly missing from the current debate, such as fostering interoperability of social networks. We hope the concerns and principles we articulate here can contribute to the debate currently underway in Brazil.