Accountability and Transparency in practice: necessary improvements to the 2630/2020 Bill

Following international trends in digital platform regulation – especially in online content moderation – legislative proposals have emerged in Brazil seeking to increase platforms' accountability and guarantee more rights for the users. The Bill No. 2,630/2020, Platform Accountability and Transparency Bill, more commonly known as "Fake News Bill", arises precisely with the intention of increasing transparency and defining duties for intermediaries, including the establishment of due process for users to question platform decisions. The purpose of this second article of the 2630/2020 Bill series is to present the main points of this piece of legislation on the accountability of platforms, its convergences with international literature and principles on the subject, as well as the possibilities for improvement towards achieving its high-minded and intended objective: to ensure that users' rights can be safeguarded and that transparency measures are implemented.

Digital platforms are private entities that play the role of intermediaries in public communication. Although they cannot be held accountable for the content posted online – as the article 19 of the Brazilian Internet Civil Rights Framework has so clearly and adeptly defined[1] –, it is through them that content is disseminated. As private actors, they have the right to make their own rules of conduct in the so-called terms of community services and policies (quasi-legislative power) and also to apply these rules in their environments (quasi-judicial and quasi-executive power). Thus, platforms are self-regulated entities, and this self-regulation is performed through techniques of online content moderation[2]. There is a consensus in the literature that this activity is inherent to the existence of platforms[3], and is desirable by users who seek conflict resolution based on these mechanisms[4].

Content moderation is performed both by humans and by automated systems of artificial intelligence, however, it is not free of error. Several cases of removal of lawful content

[1] The Article 19 of the Internet Civil Rights Framework attributes to intermediaries a conditional immunity in relation to content posted by third parties. I. e. they are not directly responsible – a correct legislative choice that strengthens freedom of expression –, only if notified in court and, based on what was stipulated in the legal sentencing, must remove the content specified in the court order.

[2] KLONICK, Kate. The New Governors*: The People, Rules, and Processes Governing Online Speech.* **Harvard Law Review**. N. 131, 2017. Disponível em: https://ssrn.com/abstract=2937985. Acesso em 3 jun. 2021.

[3] Grimmelmann, James, 'The Virtues of Moderation' (2015) 17 Yale Journal of Law & Technology 42.

[4] Citron, D. & Jurecic, Q (2018). 'Platform Justice: Content Moderation at an Inflection Point', Aegis Series Paper No. 1811, available at < https://www.lawfareblog.com/platform-justice-content-moderation-inflection-point>.

have been reported[5] - which has very serious implications for the freedom of expression and information of its users. And not only the removal of lawful content has been the target of criticism. The ordering of news feeds by the algorithms operating on the platforms, with the spotlighting of paid content, as well as the highlighting of content that has more potential to produce "engagement" and user permanence on the platform through data collection techniques for directed recommendation (or "algorithmic recommendation") has become an activity considered to be harmful.

Especially after public scandals such as *Brexit* and *Cambridge Analytica*, the significant role played by the platforms in the dissemination of harmful content – such as disinformation and hate speech – was recognized. As a result, principles documents and legislation were edited in different countries (such as the Network Enforcement Act , or NetzDG, in Germany) to demand greater transparency in the platforms' activities in discursive coordination, as well as demands for the creation of appeal mechanisms and the justification of decisions to ensure due process to users. As stated in a previous article, there are relevant recommendation initiatives by civil and academic institutions, such as the Manila Principles, Santa Clara Principles and Change the Terms – all with the objective of drawing attention to ways of achieving transparency and improvements in regards to users' rights.

The Internet Civil Rights Framework, drafted in 2012 and in force since 2014, does not establish clear provisions for the accountability duties of the platforms, since, at that time, this was not a regulatory priority – the priority was the guarantee of net neutrality and the preservation of freedom of communication. The Internet Civil Rights Framework is somewhat out of date in relation to the complexity of the platforms' activities in this new scenario - which is evidenced by the recognition that different pieces of legislation, such as the 2630/2020 Bill, need to be drawn up to define rules of transparency and accountability for companies.

In May 2021, a draft of a Decree prepared by the Federal Government was made public, aiming to amend Decree No. 8,771/2016, which regulates the Internet Civil Rights Framework[6]. The decree rightfully questions the lack of clear of users' rights regarding content moderation. However, its proposition – in addition to overrunning the discussion of such a relevant topic in the appropriate institution for public deliberation, that is, the Legislative

---

[5] COBBE, Jennifer. Algorithmic Censorship by Social Platforms: Power and Resistance. **Philosophy & Technology**, p. 1-28, 2020;

[6] Disponível em: <https://www.internetlab.org.br/wp-content/uploads/2021/05/SEI_53115.012742_2021_90.pdf>. Acesso 9 jun. 2021.

Branch – prevents platforms from moderating any content without a court order, which violates the commercial freedom of the platforms and could engender even more ungovernable online environments, with an increase not only in disinformation and hate speech, but also in in commercial fraud and the spread of inappropriate content.

The discussion about accountability rules for platforms is complex and needs to be informed by the civil society. Recently, there have been motions in Congress to open a new public consultation for discussion of the 2630/2020 Bill, which highlights the urgent need to collect more information in order to improve legislative technique. As stated in the previous article, the version presented to Congress by the Senate already included important innovations, among them the requirement to establish due process for content moderation, and specific obligations to increase transparency and the justification of decisions. We intend, in this article, to raise certain relevant points in order to raise public awareness regarding this debate.

(i) Transparency

The 2630/2020 Bill, in its current wording, requires the production of quarterly transparency reports, made public within 30 days after the end of the quarter, which must contain information such as: (a) total number of Brazilian users with access to services the platforms (art. 13, I); (b) total number of accounts and content moderation measures adopted due to compliance with the terms of private use of social media providers, due to compliance with the Law and due to compliance with court orders (art. 13, II, III and IV); with (c) specification of the reasons, the methodology used to detect the irregularity and the measures adopted. Documents should also indicate the total number of automated accounts detected and the methodology for detection, distribution networks, sponsored content and advertising.

Due to the lack of specification about how the report should present the information required, it is possible to criticize the bill, considering the space it opens for platforms to produce reports with generic information. This procedural duty seems to be directly inspired by NetzDG, which also requires reporting obligations. As research based on the results from German legislation observes, the result of the requirements of the law was the drafting of generic and non-uniform transparency reports.[7] The bill mainly requires the publication of numerical data on the total amount of content moderation techniques applied. Such data tends to show only the measures taken by the platforms, without identifying possible biases or errors

[7] HELDT, Amélie Pia. Reading between the lines and the numbers: an analysis of the first NetzDG reports. **Internet Policy Review**, v. 8, n. 2, 2019.

in moderation – and without encouraging such curation to be carried out. The disclosure of numerical data, therefore, does not necessarily make the platforms more transparent, since they are not enough to explain to the receiver of that information how the decision-making process of the content moderation was made. This means that, in practice, there is no increase in accountability of the platforms. It is necessary to ask to what purpose each specific information serves[8]. Furthermore, there is no plausible explanation for the requirement of reports on a quarterly basis – without even considering the structure of the platform and its technical capacity to do so. Such a requirement may have very negative effects of barring new players from entering the market, with even more perverse effects on competition.

Another problem with the lack of definition of a methodology for organizing and presenting data is that it can make it difficult to compare reports (1) from different platforms that have similar functions, such as Twitter and Facebook, or (2) of the same platform for different periods. This also makes it difficult to analyze the consequences of legislation, or even the consequences of the platforms' terms of use in the content moderation decision-making process. A small but necessary correction to improve this rule would be making the regulatory authority – Internet Transparency and Accountability Council (created by art. 26 of the 2630/2020 Bill) – responsible for preparing models of forms and reports to be followed by platforms.

The bill also seemed not to take in consideration the effect that such a requirement (i.e. to produce reports) might have on the functioning of different platforms, such as encrypted platforms. How will it be possible for them to submit the required data? Also, there is no clarification concerning how the platforms must disclosure these data without violating the Brazilian General Law of Personal Data Protection. In this sense, the bill must incorporate the above legislation to ensure that data processing is adequately carried out by the platforms when the reports are published.

Other transparency predictions in the bill deal with sponsored content. Advertising is defined in art. 5, VI, as "advertising messages conveyed in exchange for pecuniary payment or estimated value in cash for companies covered by this Law". Articles 14 to 17 bring requirements on the tagging of sponsored content, so that it is identified as paid or promoted

[8] ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20:3, 2016. DOI: 10.1177/1461444816676645 .

content, and the sponsor must also be identified, in addition to specific demands regarding sponsored content during election time.

The definition of sponsored content and transparency requirements also include content that is unrelated to political advertising, as in art. 17, which defines the obligation to identify all users who broadcast advertisement, covering posts from digital influencers and small businesses. Such stipulation may generate barriers to the establishment of small businesses in the platforms' marketplaces due to the extensive bureaucratic process necessary to obtain the required licenses, resulting in very negative economic impacts.

The point of greatest criticism to the articles establishing transparency rules for advertisement lies in the fact that the targeting and recommendation of content by platforms, whether sponsored or not, is done by automated algorithmic systems. **The bill does not address the lack of transparency about algorithmic systems** - nor even the difficulty of justifying decisions made by algorithms - when it makes the demands of exposing the motivation of moderation in the preparation of transparency reports.

(ii) Due Process

An important step forward in relation to users' rights in the current draft of the 2630/2020 Bill is the creation of a mechanism for user appeal regarding the platform's moderation decisions (art. 12, §3˚), acknowledging that content moderation decisions are not free of error. In face of the huge amount of content posted daily by its users, artificial intelligence systems are increasingly being deployed for automated moderation. Such systems, however, are still quite rudimentary and precarious when it comes to assessing the context of social interactions. Moreover, most automated decision systems are developed and trained on English, causing a large drop in performance when they are used in non-English speaking countries. In this sense, such systems often do not identify contents that would be considered harmful according to platforms' policies or legislation (false negatives), or, on the flip side, categorize unharmful content as harmful (false positives). Internal review mechanisms are, therefore, essential for providing possible clarification to users.

The wording given to art. 12 in the version of the 2630/2020 Bill approved by the Senate foresaw the need for "an accessible and clearly visible mechanism, available for at least three months after the decision, so that the creator or sharer, as well as the complainant, can appeal the decision". In the version under discussion in Congress, the provision, however, was removed, leaving the generic provision of §3˚ in which "the right of the user to appeal against

the removal of contents and accounts must be guaranteed by the provider". While the previous wording established in more detail the requirement of a complaint mechanism – of own content removed, or of content reported by the user –, the current version does not entail a direct obligation to create a mechanism to fulfill this function, but only the provision of a "right to appeal" when the user has their own account or content removed. There is, thus, the loss of (1) the guarantee of an accessible and prominent mechanism; (2) the provision of a time frame for the review to be carried out; (3) the possibility of transparency regarding reports of illegal content by third parties – even those that may directly affect the user who filed the complaint.

It is important to emphasize the need for the development of specific rules regarding the establishment of these appeal mechanisms, besides the justification of the decisions, so that users can effectively exercise its rights to appeal, not becoming stuck in a limbo of information asymmetry. In order to allow the monitoring of this duty, procedural rules need to be provided, enforcing the principles of due legal process[9].

In this sense, the bill could go beyond general provisions and mandate that content removals be carried out in a way knowable to users – following the advances brought by the Consumer Protection Code, that aimed to correct the asymmetries between service providers and consumers, and the Brazilian General Law of Personal Data Protection[10], that established the possibility of human review of decisions made by automated systems. Even further, the legislation could mandate that means of reporting illegal content and appealing to decisions should be of easy access to users, as well as requiring the platforms to establish clear deadlines in their terms of service. In such a way, despite the automation of moderation decisions, legitimacy may still be improved vis-à-vis the community, reducing the generalized feeling of arbitrary decision-making – which is also in the interest of the platforms.

Rule-makers must understand what they are trying to regulate, so that the regulation can be effective. The 2630/2020 Bill shows that regulators are concerned with important topics, such as increasing platforms accountability and the provision of users' rights. These concerns are being manifested through requirements for greater transparency and for the establishment

---

[9] VAN LOO, Rory. Federal Rules of Platform Procedure. **University of Chicago Law Review, Forthcoming**, 2020.

[10] The Brazilian General Law of Personal Data Protection, acknowledging that decisions made by algorithmic systems can be unjust, establishes, in its art. 20, the right of the data owner "to request the review of decisions that have been made exclusively by automated data processing, if such decision affects their interests. "The provision for human review of automated decisions has been gaining strength and would also fit in the section of the bill which refers to automated moderation decisions..

of appealing mechanisms. However, the possible consequences of these requirements over the existing reality must always be taken in account. We conclude that more specific provisions are needed regarding the duties of platforms, so that the rights of users are actually guaranteed and exercised, platforms truly comply to the obligations established by the law, and Public Administration can have clear criteria to audit and supervise the platforms.

This series of articles is an initiative organized by the Research Group on Online Content Moderation of the Center for Technology and Society at Getulio Vargas Foundation.