The current state of the Draft Bill to implement the "Law on Liberty, Responsibility and Transparency for the Internet" in Brazil: It is essential to improve the procedural duties of digital platforms, while respecting the liability regime adopted by the Brazilian Civil Rights Framework for the Internet (Marco Civil da Internet)

More than a year ago, a legislative proposal popularly known as "*PL das Fake News*" (Fake News Draft Bill) aiming to implement the "Brazilian Law on Liberty, Responsibility and Transparency for the Internet" (*Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet*) was filed in the National Congress. In a previous article on this matter, we already discussed that, despite the existence of dozens of other draft bills proposing the regulation of social media, attention on this one bill in particular is justified, beyond other factors, because **it innovates by providing a due process for content moderation**. The proposal establishes procedural duties of transparency, justification and mechanisms for contesting decisions taken by digital platforms.

However, the draft bill has been attracting attention from brazilian society and academic sectors not due to its positive aspects, but, mainly, because of its pitfalls. Various issues contained in the original wording were modified by the senators. For instance, the criminal approaches were removed and the concept of "inauthentic account" was reformulated. Nonetheless, the text approved by the Senate is still far from what we want for the Brazilian internet. It's, therefore, not surprising that the bill was criticized by dozens of research centers, besides national and international bodies, such as the United Nations and the Organization of American States (OAS) in a letter sent to the Ministry of Foreign Affairs of Brazil.

An informal working group was formed at the Chamber of Deputies to discuss the matter, resulting in the presentation of a modified version of the bill by Representative Orlando Silva, which wasn't supported by the other parliamentarians of the team. Requests for public hearings were also approved, with a schedule yet to be set. Another Clean Bill is expected, but, for the time being, the version of the text approved by the Senate still invests in the massive identification of users, including through the presentation of official identity documents (art. 5º, I; art. 7º) and on the traceability of conversations in message applications (art. 10), measures that may foster online censorship and surveillance. There is an urgent need to redress these issues as well as to enhance the positive aspects of the bill, especially the procedural duties.

The golden rule is to preserve legitimate anonymity on the internet and the liability regime centered on judicial notice, adopted by the Brazilian Civil Rights Framework for the Internet (Marco Civil da Internet - MCI)

The official wording approved by the Senate reformulates the concept of "inauthentic account" in an attempt to protect legitimate forms of anonymity, the use of social names and pseudonyms. However, its effectiveness is questionable, since the actual wording ends by stimulating platforms to indiscriminately over-remove accounts. **How does that happen?**

Platforms will keep having the duty to *"ban inauthentic accounts"* (art. 6º, inc. I) under penalty of being fined (art. 31, II). This means that they will need to develop their own means of identifying the "*inauthentic accounts"*. These, in turn, are those accounts that aim to *"assume the identity of third-parties to deceive the public"* (art. 5º, inc. II), an imprecise definition which makes it hard to comprehend what exactly the law aimed to curb.

Even without a clear definition of the goal pursued by the bill, one thing is certain: platforms may be held responsible regardless of a judicial notification in case they fail this monitoring duty. The practical consequence of this dynamic is a scenario of great legal uncertainty both for the users (that may have their account suspended) and for the enterprises (which will have to "wage war" against inauthentic accounts, without knowing how to properly identify the "enemy"). How does one discover the purpose behind an account? What does it mean to *"deceive the public"*? And how to distinguish "deceiving" accounts from those that, rightfully, use pseudonyms, social names or are anonymous, just to mention the legal categories that are, at least in theory, protected by the bill?

In order to avoid responsibility due to neglect of this duty, it is likely that platforms will adopt a posture of over-removal, by banning all accounts that present any glimpse of "inauthenticity". And this is how this bill overturns the liability regime provided by article 19 of the Civil Rights Framework for the Internet, currently in force in Brazil. In reality, platforms will be obliged to deploy a subjective analysis on the merits of contents, being held responsible by third-parties' actions, without judicial notice, if they fail in their assessment.

Platforms should not be held responsible for the content created by users, but they should be responsible for the communicative architecture that they manage

Aiming to protect freedom of expression and prevent censorship, the Civil Rights Framework for the Internet ensures that, as a rule, platforms cannot be made liable for third-party content. But the law says little about the duties that intermediaries should assume as managers of the communicative architecture and, especially, what are their responsibilities when dealing with systemic issues, such as disinformation, hate speech and electoral manipulation, issues that are much more pronounced today than it could ever be imagined at the time that MCI came into force.

We already know today that, if on the one hand platforms strive to remove harmful content (including to guarantee that the users remain connected and to preserve their economic interests), on the other, their own business models may trigger the systemic issues that they are trying to repair. Techniques of profiling and algorithmic recommendation are typically deployed to direct content, which may provoke "filter bubbles", "echo chambers" and political polarization, distorting the user's perception of reality.

The Stop Hate for Profit campaign denounced how Facebook's and Instagram's advertising policies would be promoting racist, misogynist and white supremacist speech; the Cambridge Analytica scandal showed how user's data were used for political manipulation,

while, currently, organized networks of disinformation try to distort scientific data related to the pandemic.

Episodes like this illustrate that, although platforms are not the creators of content, their policies and processes of moderation exerts decisive influence on the kind of information that will be prioritized, as well as the information that will be censored or covertly downranked (as is the case of shadowbans), which may have impacts ranging from businesses profitability to even public health issues and the repercussion of political manifests. These are processes that go far beyond a simple removal of a specific content and stay out of any kind of fiscalization, be it by the Judiciary Branch or any other institution in Brazil.

Towards a due process for content moderation

The establishment of a due process for content moderation is already adopted by foreign legislation, such as the NetzDG in Germany, and is recommended by a set of international initiatives, such as the Santa Clara Principles, the Manilla Principles and Change the Terms. Among the main points covered in these documents are requirements for: i) transparency reports produced by the platforms, informing the number of posts removed and suspended accounts due to violation of terms of use; ii) clear guidelines on community rules, offering justification to users in case of content removal, account suspension, or other techniques of moderation deployed; iii) mechanisms of appeal for users to contest the decisions taken by platforms.

It is worth mentioning that the implementation of procedural duties must add an extra layer of accountability to the platforms, but not subvert the judicial notice liability regime already in force. This is a warning made by the Manilla Principles, when stating that *"intermediaries must never be made strictly liable for hosting unlawful third-party content"* and that the laws referring to intermediary liability must be *"precise, clear, and accessible"*.

This is one of the mistakes of the "*PL das Fake News*", when it holds platforms liable if they fail to remove "inauthentic accounts" and a mistake also committed by the German law, when it holds platforms liable by not removing *"manifestly unlawful content"* within 24 hours. Various critics worry about the consequences of the law for freedom of expression, stating that it may encourage over-removal of rightful content by the platforms, especially taking in consideration the short time to assess the complaints and the high fines. It is essential to learn from the foreign experience and not repeat the same mistakes.

Great power, great responsibilities: we need to talk about procedural duties

Some research shows that, as a rule, platforms fail systematically in justifying content moderation decisions. The users are frequently in doubt about how the rules are created and how possible violations to the terms of use are identified: the complaints are made by other users, by a government body, by a human moderator or by an algorithm? In addition to not receiving sufficient explanation about the origin of the complaint, many users also do not understand what

exactly was the behavior that triggered a sanction by the platform: which rule was violated? Why was this post considered as a violation of the rule?

If platforms have become the "new governors of online speech", it is urgent that their decision-making processes be submitted to limits consistent with constitutionalism. They are free to elaborate their own terms of use, perform content moderation decisions (not only by humans, but also by automated systems) and boost ads, but now they need to render account of how they carry out these processes, why certain decisions were taken in each specific case and to offer opportunities for users to question their decisions. The draft bill is right in seeking the initial parameters for the fulfillment of these obligations, which, however, still need to be greatly improved.

Duties of Transparency, Appeal and Justification of decisions are a consensus that should be improved

The demand for transparency and justification of decisions cannot be an end in itself. Rather, it needs to be a mechanism that enhances accountability by platforms and not just a showcase for the sole purpose of exposing information. Some research evaluating the implementation of NetzDG in Germany has already shown the "low informative value" of transparency reports, which, in the way they were implemented, didn't lead to a better understanding of the decisions made by platforms.

Procedural duties are commonly reported as points of consensus by those who evaluate the "*PL das Fake News*", but deserve to be further explored, in compliance with academic recommendations and international documents on the matter. Are the duties provided by the draft bill sufficient? To what extent can they be improved to enable effective inspection? The next articles in this series will bring some considerations on this matter.

* 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.