THE LEGAL UNCERTAINTY ASPECT OF THE REGULATORY APPROACH REGARDING ONLINE CONTENT MODERATION

Eduardo Helfer de Farias and Gabriel Rached
Fluminense Federal University, Niterói/RJ, Brazil

ABSTRACT

Since the creation of the internet, online content moderation has been facing uncertainty regarding its lawful application even in situations where there is a specific law addressing the issue. Internet users around the world have been dealing with the struggle to define the boundaries of lawful disagreement and unlawful speech. In this sense, this research aims to identify how the State's regulatory approach may influence the legal uncertainty regarding online content moderation. For this purpose, there were selected specific countries representing three different regulatory models for the internet: the USA (self-regulation); Brazil (command and control regulation); and Germany (standard-based regulation). The research intends to compare the laws and precedents of each selected legal system and contrast how they address: (i) the moderator's identity; (ii) the criteria for moderation; (iii) the moderation tools; (iv) the "the checks and balances" over the decision-making of the moderator; and (v) the liability to be applied against wrongful decision. The level of legal uncertainty in each of these five aspects shall be measured exclusively by the extent of the law's ambiguity and the discretion of the decision-makers in interpreting it. So, it will not address the "subjective uncertainty". In this discretion, given the legislation may vary according to the subject matter presented for content moderation, three main topics were taken into consideration: privacy, fake news, and hate speech. These topics have been selected for being sensitive issues in all countries, including the three selected for this study. The preliminary results indicate that the command and control regulation as presented in the Brazilian version might be the one with the highest legal uncertainty while the self-regulation as presented in the US version appears as the one with the lower legal uncertainty. In the conclusion, these results would be summarized indicating how the law's ambiguity and decision-maker's discretionary reflects on the cyber environment.

KEYWORDS

Online Content Moderation, Legal Uncertainty, Regulation, Decision-Making, Liability, Cyberspace

1. INTRODUCTION

This research intends to identify how the legal uncertainty regarding online content moderation is influenced (lower or higher) by the regulatory approach of the internet adopted by the State agents, such as self-regulation, command and control regulation; and standard-based regulation. Specifically, the intent is to find out to which degree the regulatory approach adopted for online content moderation increases or decreases the legal uncertainty in the cyber environment, reflecting on the decision-maker's choice.

By "legal uncertainty" refers to the law's ambiguity and the decision-makers discretion in interpreting it (CALFEE & CRASWELL, 1986; SEGAL & STEIN, 2006). It is not considered in this study the so-called "subjective uncertainty", which means people's perception of the law (DAVIS, 2011), because it would require a different methodological approach.

US Supreme Court Judge Oliver Wendell Holmes understood the law as a "prediction of how the courts behave" (HOLMES, 1897), which means the law offers a degree of certainty about public administration behavior. Therefore, most scholars sustain that the uncertainty of the law may frustrate the law's goal of deterring antisocial behavior and even undermine the rule of law (KERHUEL & RAYNOUARD 2010). Also, it cannot be presumed that legal uncertainty reduces naturally over time giving the bias of judges and legal scholars in "upsetting expectations" and the "economic incentives" in favoring those "disadvantaged by legal rules" (D'AMATO, 1983). So, the issue is a problem requiring action to be solved. Assuming these premises,

the regulatory approach which succeeds in minimizing uncertainty contributes also to minimizing antisocial behavior and preserving the rule of law.

Unfortunately, most academic research about the legal uncertainty arousing from content moderation restricts themselves to the social media environment (HELDT & DREYER, 2021; CHEN, 2007; FAUST. 2017). These references give insights into how content moderation should be handled, but they are restricted to an environment controlled by a single agent. Besides, platforms such as social media are just the most visible level of the internet. There are other four internet categories capable of content moderation: internet service providers, domain registers, payment services, antivirus providers, cloud storage services, and platforms.

To narrow the scope, the analysis is restricted to the online content moderation exercised through the five levels of the internet under the labels of "privacy", "fake news" and "hate speech" in countries representatives of the three different internet regulatory approaches: USA (self-regulation); Brazil (command and control regulation); and Germany (standard-based regulation). These countries have been selected as representatives for their regulatory approach because of the maturity of their legal systems. For all purposes, "Content moderation" refers to all actions taken to suppress, restrict or promote a specific speech.

The methodological approach consists of analyzing the law and precedents of these three countries regarding the answers they provide to these five questions: 1) who is the moderator (the one who decides about the content to be promoted or restricted); 2) which contents are moderated; 3) which tools are applied for this purpose; 4) which "checks and balances" should be applied to the moderation decisions, and 5) which liability process should be enforced to correct or minimize the consequences of wrongful decisions.

Considering this is part of a thesis ongoing research, the content to be shown in the final remarks could be considered preliminary results.

2. A THEORY OF CONTENT MODERATION: CASE STUDIES

The legal uncertainty regarding online content moderation is detrimental to the exchange of ideas necessary for preserving basic human rights, such as free speech. To evaluate how the State's regulatory approach to the internet influences this uncertainty, this research shall analyze the sources of objective legal uncertainty (law's ambiguity and decision-makers discretionary) in the regulatory approaches of the USA (self-regulation), Brazil (command and control regulation) and Germany (standard-based regulation): (i) establish the moderator's identity; (ii) chose the criteria for moderation; (iii) addresses the moderation tools; (iv) impose "checks and balances" over the moderator's decision-making; and (v) take measure to correct wrongful decisions.

For the first question, it is hypothesized three possible moderators: government, internet infrastructure owner, or the general public. "Government" means any state structure, may it be the executive, legislative, or judiciary. "Internet infrastructure owner" means any holder of the technologies that sustain the internet. "General public" means any internet user that does not fit the two first categories.

In all countries selected for study, one of the three possible moderators assumes as a protagonist over the other two, without eliminating them. The USA's First Amendment and Section 230 of the Communications Decency Act turns protagonist the general public. In opposition, Brazil nominates its Judiciary as the main moderator regarding speeches (arts. 9 § 1, 18 and 19 of Internet Civil Regulation). Germany delegated the moderation responsibility to the internet infrastructure owners, setting boundaries for them to act under government supervision (Netzwerkdurchsetzungsgesetz of 2017). Therefore, the study of each one of them gives a glimpse of a different approach to dealing with the issue.

For the second question, it has been selected three categories under which a speech may deem to be vetoed or restricted: privacy, misinformation, and hate speech. "Privacy" refers to any identifiable information regarding a natural person, such as name, social security number, health status, and behavior. "Misinformation" means any false statement, intentional or not. "Hate speech" means any speech that intends for the audience to take violent actions against an identifiable group of persons. These choices were made to have a comparison standard between these three countries, given that all of them are dealing with these issues.

The internet infrastructure owners – all based in the USA – have vowed to moderate the online content to prevent the spread of misinformation. On the other hand, the Brazilian top Judiciary Court started a confidential investigation by itself (Inquérito 4.781/2020) to arrest and reprimand those whom the Court deemed to be part of an online criminal organization to discredit the Judiciary. In Germany, the Gesetz Zur Verbesserung der

Rechtsdurchsetzung in sozialen Netzwerken imposed on the internet infrastructure owners the duty to remove "clearly illegal content" on their behalf. Therefore, all three countries are taking action on the issue.

In the third question, the main moderator's tools are identified as fear, friction, and flooding (ROBERTS, 2017). "Fear" is the result of widespread awareness that one specific speech will be sanctioned by moderators if publicized. "Friction" means any imposition of restrictions to access or promote a specific speech, such as filtering or demanding conditions for accessing, such as confirming age or payment. "Flooding" means the widespread promotion of a specific speech through the internet to make it harder for internet users to be aware of the existence of opposing speech.

Internet infrastructure owners are taking the responsibility to moderate online content by themselves or being forced to take it by the government in all three jurisdictions. Moderation is always as visible as restriction of accesses. It may also be as subtle as promoting one view disproportionality disregarding the other or imposing barriers for one specific content to be known. This research shall see which techniques are being deployed by each jurisdiction.

For the fourth question, it is evaluated if there are clear boundaries for a moderator to make his/her decision, if these boundaries are enforced by a third party beyond the moderator and if the subject target of the moderation may present defense on his/her behalf.

This question investigates how each country set boundaries to avoid bias or mistaken decisions among moderators. Assuming there is no foolproof moderation system, it's important to check how this risk is minimized.

For the fifth question, we evaluate if there is any legal remedy the subject targeted by moderators may seek reparation in case it is proven it was not according to law for his speech to be restricted or banned. Even with fail safes in place, it is inevitable for "false positives" to happen. The final question would be how each of these three jurisdictions shall deal with the aftermath of a failed moderation.

3. CONCLUSION

Online content moderation is a sensitive issue for any regulatory approach on the internet.

In the current research status, the observed preliminary results point out that the self-regulation of the internet in its USA approach seems to be the one with the least legal uncertainty, and the command and control regulation in the Brazilian approach seems to be the one with the highest legal uncertainty, standard-based regulation of Germany being in between regarding its certainty.

It's important to note that certainty does not refer to concepts of fairness or equity, being a topic related to predictability. Therefore, this research does not address the morality surrounding the regulatory approaches, which is a topic that may be explored in future research.

It may be impossible to eliminate the legal uncertainty surrounding online content moderation, but it's reasonable to search for ways of minimizing it. This research's goal is to contribute to settling some degree of predictability over law enforcements and business behavior regarding the boundaries of speech to avoid the escalation of conflicts.

REFERENCES

Calfee, J. E. & Craswell, R. (1984). Some Effects of Uncertainty on Compliance with Legal Standards. In Virginia Law Review, Vol. 70 pp. 965-1003.

Chen, P. J. (2013). *Australian Politics in the Digital Age.* Anu Press, Canberra, Australia.

D'Amato, A. (1983). Legal Uncertainty. In California Law Review, Vol. 71 pp.1-53

Davis, K. E. 2011. The Concept of Legal Uncertainty: Definition and Measurement. Available at SSRN: https://ssrn.com/abstract=1884664 or http://dx.doi.org/10.2139/ssrn.1884664.

Faust, G. (2017). Hair, Blood, and the Nipple: Instagram Censorship and the Female Body. In: Urte Undine Frömming et al. Digital Environments: Ethnographic Perspectives Across Global Online and Offline Spaces. Bielefeld, Germany, pp. 159–170.

Heldt, A., & Dreyer, S. (2021). Competent third parties and content moderation on platforms: Potentials of Independent Decision-Making Bodies from a Governance Structure Perspective. In: Journal of Information Policy, Vol. 11, pp. 266-300.

Holmes, O. W. (1897). The path of the law. In: Harvard Law Review, Vol. 10, pp. 457-478.

Kerhuel, A. et. al. (2011). Measuring the Law: Legal Certainty as a Watermark (In French). International Journal of Disclosure and Governance, Vol. 8, 4, pp. 360–379.

Roberts, M. E. (2018). Censored: distraction and diversion inside China's Great Firewall. Princeton University Press, New Jersey, USA.

Segal, U. & Stein, A. (2006). Ambiguity Aversion and the Criminal Process. In Notre Dame Law Review Vol. 86, pp. 1495-1551