Emerging Challenges in Social Media Regulation: Resurgent States?

Edoardo Celeste
Ilton Norberto Robl Filho
Ingo Sarlet
Orgs.

Editora Fundação Fênix

DCU
Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

CDEA
UFRGS - PUCRS

idp
CENTRO DE ESTUDOS PETER HÄBERLE - IDP

idp

PUCRS | ESCOLA DE DIREITO

Editora Fundação Fênix

9 786554 603065

Emerging Challenges in Social Media Regulation: Resurgent States?

Série Direito

Conselho Editorial Internacional

Edoardo Celeste
Ilton Norberto Robl Filho
Ingo Sarlet
(eds)

Emerging Challenges in Social Media Regulation: Resurgent States?

Editora Fundação Fênix

Porto Alegre, 2026

Direção editorial: Ingo Wolfgang Sarlet
Diagramação: Editora Fundação Fênix
Concepção da Capa: Editora Fundação Fênix

O padrão ortográfico, o sistema de citações, as referências bibliográficas, o conteúdo e a revisão de cada capítulo são de inteira responsabilidade de seu respectivo autor.

Obra editada com apoio: Apoio: CAPES/PROEX Auxílio Nº 1605/2024, Processo Nº 88881.973765/2024-01.

Série Direito – 131

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Emerging challenges in social media regulation [livro eletrônico] : resurgent states? / [Edoardo Celeste, Ilton Norberto Robl Filho, Ingo Sarlet orgs.]. -- Porto Alegre, RS : Editora Fundação Fênix, 2026.
PDF

Vários autores.
Bibliografia
ISBN 978-65-5460-306-5

1. Direito digital 2. Governança 3. Mídia digital - Legislação - Brasil 4. Privacidade na Internet 5. Redes sociais 6. Regulação I. Celeste, Edoardo. II. Robl Filho, Ilton Norberto. III. Sarlet, Ingo.

26-356150.1 CDU-372.721

Índices para catálogo sistemático:

1. Regulação : Privacidade e proteção de dados pessoais : Direito 372.721
Maria Alice Ferreira - Bibliotecária - CRB-8/7964

DOI – https://doi.org/10.36592/9786554603065

SUMMARY

1. INTRODUCTION

Edoardo Celeste
Ilton Norberto Robl Filho
Ingo Sarlet

The relationship between the state and social media has undergone a rapid and profound transformation since the dawn of the platform era. To better understand the implications of this shift for democracy and constitutionalism, this volume explores the evolving dynamics of digital governance, with a particular focus on the Brazilian legal context. Following years of debate centered on self-regulation and co-regulation, our central inquiry asks: are states reasserting a primary role in regulating the social media environment, and what are the resulting opportunities and challenges?

This collection emerged from the course "Emerging Challenges in Social Media Regulation: Resurgent States?", conducted online between March and May 2025, thanks to a joint partnership between Dublin City University, IDP Brasilia and PUCRS. The curriculum featured insights from twelve experts spanning diverse jurisdictions and academic disciplines, providing the foundation for the student contributions found in these pages. We would like to thank all these colleagues for their time and availability as well as the administrative staff members from our universities who helped us set up the online module. Special thanks also go to Luis Felipe Alvarez Vega for his editorial assistance.

The present book collects the essays produced by the participating students at the end of the module and is organized into three thematic sections.

Part I: *Digital Constitutionalism and Fundamental Rights* critically examines the tension between platform economic models and the protection of human rights. It specifically addresses how the exercise of freedom of expression often clashes with the proliferation of disinformation and hate speech. In particular, in Chapter 2, titled "Digital Constitutionalism in Brazil: Balancing Fundamental Rights in the Digital Era," Euzébia Krusser Ferrari dissects the foundational role of digital constitutionalism, arguing for a balanced approach to protecting fundamental rights amidst rapid digitalization. This sets the stage for Chapter 3, "Digital Constitutionalism and the

Limits of Freedom of Expression in Brazil: From Social Media Platforms to the Federal Supreme Court," where Marcella de Pinho Pimenta Borges Ramos scrutinizes the judiciary's role in defining platform liability, a critical boundary in the Brazilian legal landscape.

The conceptual tension within these rights is further explored in Chapter 4, "Freedom of Speech and its Limits: Ideal or Myth?" as Cyntia Melo Rosa questions whether the absolute ideal of free speech remains a functional reality or has become a myth requiring modern regulation. Tansitioning from speech to privacy, Chapter 5, "The Mandatory Disclosure of Geolocation Data by Internet Search Providers: A Constitutional Analysis of Geofence Warrants under Brazilian Law," authored by Yury Rufino Queiroz, offers a technical analysis of investigative efficiency versus the right to digital anonymity.

Part II: *State Functions vs Private Governance* tackles the friction between traditional constitutional state functions and the private entities that own and manage social media. Topics include the balance of efficiency and digitalization against the principles of public administration, and the struggle to disseminate trustworthy information during crises like the COVID-19 pandemic. The second part moves into the operational friction between public institutions and private tech entities. The vulnerability of the state is addressed in Chapter 6, "Judicial Ransomware and Democratic Continuity within Brazil's Digital Electoral Processes," where Celso Reic Urbieta explores how cyberattacks on judicial institutions threaten democratic stability. This theme of crisis management continues in Chapter 7, "Resurgent States, Disinformation and the Covid-19 Infodemic," by Thiago Lopes Cardoso Campos, which illustrates how the state's role as a content governor was accelerated by the need to combat life-threatening disinformation.

As the state adopts new technologies, Chapter 8, "Resurgent States and New Challenges on Algorithmic Decision-Making Regulation," by Pedro Nilson Moreira Viana, evaluates the ethical hurdles of integrating Artificial Intelligence into judicial drafting. The section then shifts toward models of oversight: Chapter 9, "A Delegated Geographical Regime for Social Media Governance," by João Pedro Barbosa Mota, proposes a principle-based geographical framework, while Chapter 10, "Content

Moderation in Brazil: Platform Self-Regulation and the 2025 Judicial Shift on Liability," by Larissa de Lima e Campos, analyzes the recent transition toward a more interventionist interpretation of the Brazilian Civil Rights Framework for the Internet. Finally, Chapter 11, "The Privatization of the Public and Efficiency as a Smokescreen: The Elon Musk Effect in Center-Right Municipal Administrations in Bahia," by Vitória Andréa De Almeida Nicolau, provides a localized critique, arguing that administrative efficiency must not bypass the principle of legality.

Part III: *National and Transnational Regulatory Frameworks* contextualizes social media regulation as a cross-border phenomenon. It examines various jurisdictional attempts to assert control over digital spaces, with a specialized focus on the evolving legal landscape in Brazil. In particular, Chapter 12, "Civil Liability of Social Networks for User-Generated Content in the Brazilian Legal System," by Vitória Monego Sommer Santos, provides a comprehensive grounding in current Brazilian legal doctrine.

The volume concludes with a comparative lens on global tech giants. In Chapter 13, "The Urgency of Social Media Regulation: Meta's Divergent Approaches to the EU and Brazil Regarding AI Training on Publicly Available User Content," Luisa Maciel Perez highlights the "regulatory gap" that allows platforms to operate under lighter constraints in Brazil compared to the EU. Supporting this view, Chapter 14, "Regulating the Attention Economy: Meta's Non-Compliance with the EU," by Mahon McCann, argues that the extractive "attention economy" model is fundamentally at odds with the EU's Digital Services Act (DSA). Together, these chapters illustrate the complex struggle to bring private digital power under the rule of law.

Part I - Digital Constitutionalism and Fundamental Rights

2. DIGITAL CONSTITUTIONALISM IN BRAZIL: BALANCING FUNDAMENTAL RIGHTS IN THE DIGITAL ERA

https://doi.org/10.36592/9786554603065-01

Euzébia Krusser Ferrari[1]

Abstract
The objective of this essay is to deal with the theme of digital constitutionalism and its role in the protection of fundamental rights in times of digitalization, and, at the same time, not to prohibit free expression. The work was divided into two parts: in the first, the impact of disinformation on fundamental rights is addressed, from the dissemination on social media, and in the second, the definition of digital constitutionalism is made, which is addressed as an instrument of protection for the new fundamental rights that emerge from the technological era, and also the need to impose limits on the right to freedom of expression, in the face of the manipulation of public opinion through the practice of disinformation, with the intention of weakening democracy.
Keywords: Digital constitutionalism; Freedom of expression; Regulation; Social Media.

1. Introduction

The purpose of this essay is to reflect on the moment we are living, with regard to the State's action to regulate social media[2], in the face of the challenges that have been arising in relation to digitalization, especially due to the advancement of technology by the use of artificial intelligence, which proves to be a great challenge for the protection of fundamental rights and human rights.

The definition of digital constitutionalism is addressed, as an instrument for the protection of new fundamental rights that emerge from the technological era,

[1] PhD student in Law at the Pontifical Catholic University of Rio Grande do Sul (PPGD-PUCRS). Master in Human Rights from the Ritter dos Reis University Center (UniRitter). Lawyer (Porto Alegre, RS, Brazil). Lattes iD: http://lattes.cnpq.br/8135006177026905. Orcid iD: https://orcid.org/0009-0003-
4863-8863 E-mail: krusserferrari@krusserferrari.adv.br
[2] The term "social media" will be used in a broad aspect to refer to digital platforms that allow the creation and sharing of information, photos, videos, texts and other types of media, and allow interactions between users, communities and organizations, and also encompasses the definition of "social networks" that enable the construction of connections between people, such as Facebook and Instagram.

which advances in an accelerated movement and allows the development of new artificial intelligence systems with the potential to reach fundamental and human rights, and the imposition of limits on the right to freedom of expression is also addressed. in the face of the manipulation of information, or dissemination of disinformation, with the intention of manipulating public opinion, and weakening democracy.

2. Reflection on the impact of disinformation on fundamental rights, from social media

Society is being severely impacted by the influence of social media, which has been indelibly and inexorably changing the way we think and communicate. Not only that, it is transforming the way we will design our future, from the simplest preferences, such as the type of food we consume, to issues of extreme complexity, as the influence on the electoral process. This can be lethal for the exercise of citizenship and the maintenance of the Democratic Rule of Law.

We cannot deny that the development provided by technological advancement has a positive impact on important areas such as health, and allows knowledge to be shared more quickly, such as in the cases of improving protocols and making diagnoses faster and more efficiently.

However, from the emergence of the first platforms, in the 90s, which reached a limited number of people, to the current moment, when giants such as Facebook, Instagram, TikTok, operate globally, with billions of users, who "react" to content, sharing their preferences, information has become a valuable currency, which allows you to establish a profile of each user, and use that data to shape preferences.

In addition, the intention is the dissemination of wrong information, directed at a particular audience, with the intention to manipulate the opinions forming an immense wave of disinformation (which has at its core the intention of reaching a certain parcel of people and convincing them about "that truth").

The dissemination of dubious news has always permeated political campaigns, but to a certain extent they were admitted by opponents as part of the game, and when they caused damage they were solved in the judiciary. This cannot

be said of the current moment, because from the instant certain news is released on the internet, it is relentless, due to the speed with which it is shared, even if removed from the platform, the damage may have already become irreversible.

On this subject, in Brazil, Inquiry 4.781/DF has been underway[33] since March 14th, 2019 (Fake *News Inquiry),* which investigates the use of fake news to attack democracy and affect the honorability and security of the Supreme Court.

Another issue of great importance, with special attention to the Democratic Rule of Law, is to establish to what extent there is liability of digital platforms due to damages caused by the user of social media to a third party, and, to the same extent, what are the consequences for the omission of social media to act proactively, based on the identification of the violation of fundamental rights, and humans.

---

[3] Inquiry 4781/STF (reference) is being processed in secrecy of justice, as well as the other related processes, however, it is possible to verify the object of investigation, through the summary of Inq 4781-AgR (twelfth), judged on July 03, 2023, which provides as follows: [...] STRONG INDICATIONS OF THE PARTICIPATION OF THE INVESTIGATED IN A CRIMINAL ORGANIZATION ("DIGITAL MILITIAS"). USE OF PROFILES ON SOCIAL
NETWORKS TO PROPAGATE HATE SPEECH, SUBVERSION OF ORDER AND ENCOURAGEMENT TO BREAK INSTITUTIONAL AND DEMOCRATIC NORMALITY. ABUSE OF THE RIGHT TO FREEDOM OF EXPRESSION. NECESSITY AND ADEQUACY IN THE BLOCKING OF PROFILE TO STOP CRIMINAL ACTIVITY. INTERLOCUTORY APPEAL DISMISSED. 1. The object of this inquiry is the investigation of fraudulent news (fake news), false reports of crimes, slanderous denunciations, threats and other infractions coated with *animus caluniandi, diffamandi* or *injuriandi*, which affect the honorability and security of the FEDERAL SUPREME COURT, its members; as well as their families, when there is a relationship with the dignity of the Justices, including the leakage of confidential information and documents, with the intention of attributing and/or insinuating the practice of unlawful acts by members of the SUPREME COURT by those who have the legal duty to preserve confidentiality; and the verification of the existence of financing schemes and mass dissemination on social networks, with the intention of harming or exposing to danger of injury the independence of the Judiciary and the Rule of Law. 2. The initial steps, described in the records, especially in the decision dated 5/26/2020, indicate the existence of organized use of computer tools, notably social media accounts, to create, disseminate and disseminate false information or capable of harming the institutions of the Rule of Law, notably the FEDERAL SUPREME COURT. 3. Necessity, adequacy and urgency in the interruption of speech with hate content, subversion of order and encouragement to break institutional and democratic normality by blocking accounts on social networks, such as Facebook, Twitter and Instagram, of the investigated, with the aim of interrupting the injury or threat to the right (art. 5, XXXV, Federal Constitution). 4. The investigated persons would have, in theory, a direct or indirect connection with the criminal association and its financing, since, evaluating the content of their pronouncements and the procedure for dissemination on social networks, there are indications of alignment of their illicit messages with the alleged scheme narrated by the parliamentarians heard in these records. 5. Appeal dismissed. Available at:
https://jurisprudencia.stf.jus.br/pages/search?classeNumeroIncidente=%22Inq%204781%22&base=acordaos&pe
squisa_inteiro_teor=false&sinonimo=true&plural=true&radicais=false&buscaExata=true&page=1&pageSize=10&s ort=date&sortBy=asc&isAdvanced=true. Accessed on: 26 Jun 2025.

In this sense, recently, on June 26th, 2025, Extraordinary Appeal N. 1037396, on Topic 987, was judged[4], which partially declared unconstitutional article 19, Law N. 12.965/2014 of the Law (Civil Rights Framework for the Internet), and established the liability of social media, including objectively and jointly and severally in relation to certain content[5].

This decision is of fundamental importance to ensure fundamental rights and guarantees, as it clearly establishes the responsibility of platforms for the maintenance of content that causes harm to users, as well as establishes the duty to indemnify, as financial compensation is crucial for social media to adopt stricter measures to avoid being held liable.

---

[4] Topic 987 - Discussion on the constitutionality of article 19 of Law No. 12,965/2014 (Civil Rights Framework for the Internet) which determines the need for a prior and specific court order to delete content for the civil liability of internet providers, websites and social network application managers for damages resulting from unlawful acts practiced by third parties. Available at: &sort=_score&sortBy=desc&isAdvance=true&classeNumeroIncidente=RE%201037396. Accessed on: 25 Jun 2025.
https://jurisprudencia.stf.jus.br/pages/search?base=acordaos&sinonimo=true&plural=true&page=1&pageSize=10

[5] With regard to strict liability, it was established in RExt No. 1037396: "[...] *3. The internet application provider is civilly liable objectively* ***and regardless of notification****, for damages arising from content generated by third parties, in the following cases: 3.1. when* ***they recommend, boost (paid or not) or moderate such content****, with joint and several liability with the respective advertiser or sponsor, in the case of advertisements or sponsored material; 3.2. when it is an* ***inauthentic account*** *(also called a '****fake profile****'), or a* ***de-identified and/or automated account****; 3.3. when it comes to copyright and related rights, jointly and severally with the third party responsible for the effective publication/posting of the content, in accordance with arts. 102 to 104 of Law No. 9,610, of 1998; 3.4. when they constitute practices provided for in the following* ***exhaustive list****: (a) crimes against the Democratic Rule of Law (CP, art. 296, sole paragraph; art. 359-L, art. 359-M, art. 359-N, art. 359-P, art. 359-R); (b) acts of terrorism or preparatory to terrorism, typified by Law No. 13,260, of 2016; (c) crime of inducement, instigation or assistance to suicide or self-mutilation (CP, art. 122); (d) crime of racism (Law No. 7,716, of 1989, articles 20, 20-A, 20-B and 20C); (e) any kind of violence against children, adolescents and vulnerable persons, including the crimes provided for in arts. 217-A to 218-C of the Penal Code, as amended by Laws No. 12,015, of 2009, and No. 13,718, of 2018, and in Law No. 8,069, of 1990, and in compliance with Law No. 13,257, of 2016, and CONANDA Res. No. 245, of 2024; (f) any kind of violence against women, including the crimes of Law No. 14,192, of 2021; (g) sanitary infraction, for failing to execute, hinder or oppose the execution of sanitary measures in a situation of Public Health Emergency of National Importance, under the terms of article 10 of Law No. 6,437, of 1977; (h) trafficking in persons (CP, art. 149-A); (i) incitement or threat to commit acts of physical or sexual violence (CP, art. 29 c/c arts. 121, 129, 213, 215, 215-A, 216-A, 250 and 251c/c art. 147); (j) dissemination of facts that are notoriously untrue or seriously decontextualized that lead to incitement to physical violence, threats against life, or acts of violence against socially vulnerable groups or members; (k) disclosure of facts that are notoriously untrue or out of context with the potential to cause damage to the balance of the election or the integrity of the electoral process (Res. No. 23,610/2019, articles 9-C and 9-D); 3.5. If there is* ***reasonable doubt*** *about the configuration of one of the conducts mentioned in item 3.4, the regime of* ***article 21 applies,*** *in the form of item 2 of this thesis; [...]".* Available at: https://digital.stf.jus.br/publico/publicacoes. Accessed on: 30 Jun 2025.

The decision is quite comprehensive, and it will not be possible to address all its contents in this article However, the portion that establishes a period of 180 (one hundred and eighty) days for internet application providers to indicate or create a private entity that, among other attributions, develop artificial intelligence mechanisms for the removal of illegal content,and user education, deserves to be highlighted. Thus, from the presentation of these two cases, the inquiry that investigates the attack on the institutions of the Rule of Law, in the investigation phase, and the decision that declared the partial unconstitutionality of article 19 of Law N. 12.965/2014 , the intention is to address the issue of digital constitutionalism as an instrument for the resolution of issues brought by the context we are experiencing, of intense digital transformation, without, however, entering into the similarity of the current moment with other historical moments, due to the need for in-depth development of the theme, and also due to its transversality, which demands study in the field of political philosophy, international law, not exhausted in these areas.

It is a fact that complex issues such as those mentioned here overflow in the confrontation by the constitutional courts, which are required to decide on the protection of fundamental rights, such as freedom of expression, and, at the same time, these institutions are attacked through social media, through the use of disinformation and with the clear intention of dividing opinions and weakening democracy, a study that must be faced by digital constitutionalism.

3. Addressing the meaning of digital constitutionalism and its importance for the protection of fundamental rights and limits to freedom of expression

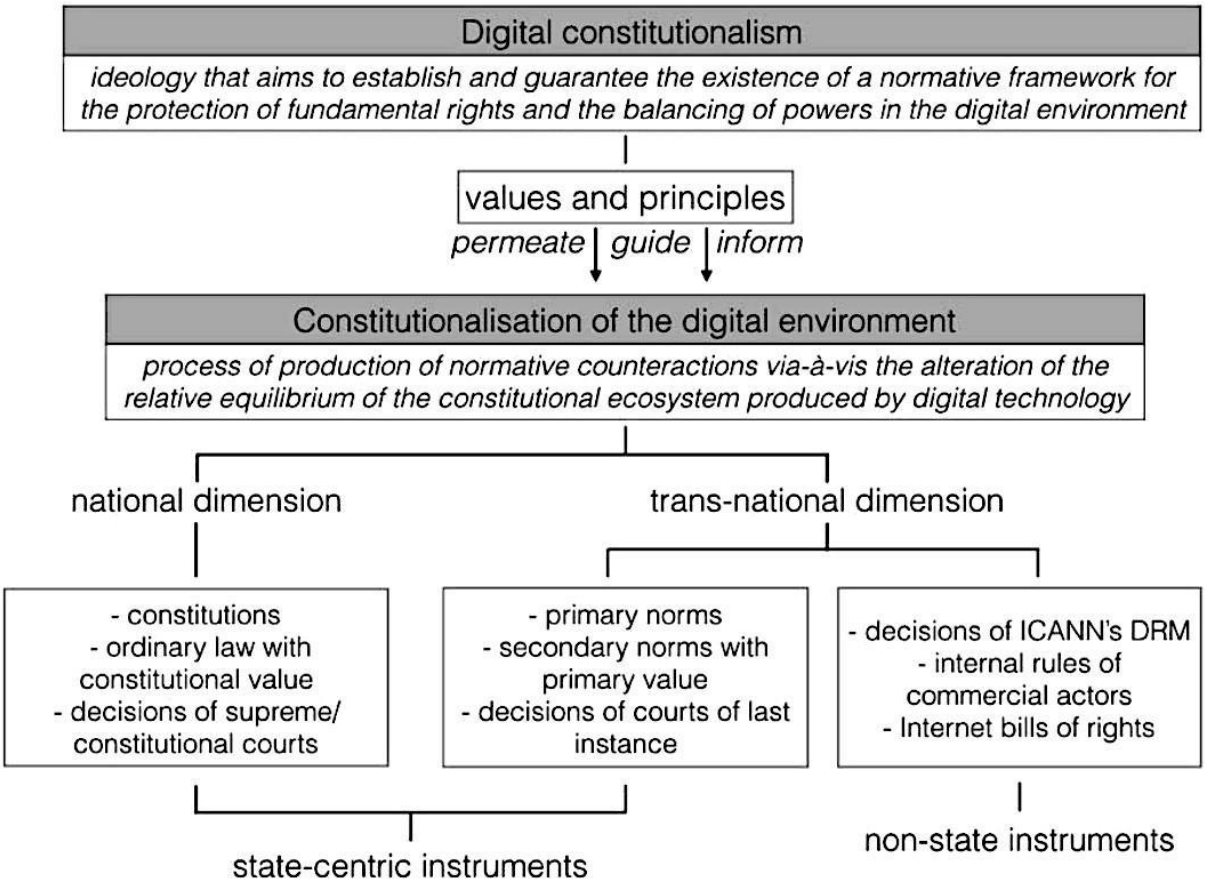
We have traveled a path of technological development, which has evolved over the years, from the first communication networks, with limited reach to the student public of a university, networks that have been interconnected, breaking down geographical limits, to surpassing national borders and transforming the world into a connected universe.

And here we are, in the digital age, in which major transformations occur in a short period of time, and, in this sense, the scope and the way in which new technologies can be used demands a specific area of law that understands it, and, at the same time, that protects fundamental rights.

In this sense, the importance of understanding the meaning of digital constitutionalism, a topic that has been faced by jurists in order to define this very complex area, which speaks intimately with the technological area.

Based on Edoardo Celeste (2025), the definition adopted is that 'Digital constitutionalism refers to the concept of establishing a set of principles, norms, and rules that govern the use, protection, and regulation of digital technologies within a society.

Also, in a didactic way, Celeste (2025) provides a logical reasoning that facilitates the understanding of the concept of digital constitutionalism and the application of the theory on the environment of digital constitutionalisation, so that it allows the reader to follow the path it has taken (although the difference between the themes is not the object of development at this time, it is important to note, for a future study):



(Source: Celeste, 2025, p. 25)

Thus, digital constitutionalism has the role of outlining how governments will face the absence of borders and the speed with which artificial intelligence breaks paradigms, and in particular, how to protect fundamental rights.

Edoardo Celeste *et. al* (2023) also brought to the debate the complexity of the decision on the issue of which rules should prevail in the exercise of moderation of content disseminated on social media, considering that international law is limited to providing general principles, and, in contrast, as large technology companies have developed, they have been creating their own rules.

Therefore, there seems to be a need, in the same way that a traditional constitution defines fundamental rights, safeguards numerous guarantees, establishes duties for the State itself, form of government, etc., that digital constitutionalism also studies technologies and their transformations, and proposes rules that can meet social demands, and in particular, face the challenge of protecting against the manipulation of information disseminated on the network,

and against the effects of disinformation, which leads us to address the limits of freedom of expression.

Freedom of expression is presented, at this moment, as an essential pillar of democracy, which often conflicts with other fundamental rights (Robl and Sarlet, 2016), and requires the judge to decide which fundamental right should prevail in the specific case.

In Brazil there is a clear tendency for freedom of expression to prevail over other fundamental rights (Robl and Sarlet, 2016), but, obviously, there are restrictions that must be considered, so as not to attribute a hierarchy that does not exist among constitutional norms.

And in this sense, the authors point to the express restriction on anonymity, in article 5, IV, and against censorship, in the article 220, paragraph 2, of the Federal Constitution (Brazil), in addition to providing for the right of reply and compensation for violation of personality rights.

There are also implicit restrictions, according to Robl and Sarlet (2016), which refer to the need to harmonize freedom of expression with other fundamental rights, and mention hate speech as an exception to the protection of freedom of expression, especially when it involves the denial of notorious facts such as the Holocaust or incites hatred and violence against social groups (ethnic, religious or racial).

The way that has been found to resolve conflicts between fundamental rights is to carry out a judgment based on reasonableness and proportionality, on the recommendation of the authors (Robl and Sarlet), with the observance of criteria such as the truth of the facts, the lawfulness of the means of obtaining information, and the analysis of the seriousness of the offense to personality rights.

In a brief approach to the performance of the Federal Supreme Court, which has faced unprecedented attacks with reference to attempts to destabilize the rule of law, under the cunning argument that demonstrations are free and cannot be curtailed.

In this sense, disinformation has been the greatest political weapon, which uses platforms to divide public opinion and weaken the Democratic Rule of Law.

And, to conclude, as Mendes and Fernandes (2020) predicted, with the performance of the constitutional court, it shed light on the issue related to the regulation of various spheres of the use of digital spaces with the judgment of Extraordinary Appeal 1.037.396, representative of Topic 987 of the General Repercussion system, which has not yet become final, but which declared Article 19 of the Civil Rights Framework for the Internet partially unconstitutional, and in addition to establishing rules, established the duty to indemnify, and, in some cases, joint and several liability.

4. Conclusions

The theme brought to this essay is of very high relevance for the current moment and should be continuously studied, in order to bring light to the demands that have been arising from the advancement of new technologies, driven by artificial intelligence.

The confrontation with digital constitutionalism, issues such as freedom of expression, and due to the pending regulation regarding the responsibility of social media, strengthens the role of the constitutional court in the protection of users' rights, and reinforces its role in the protection of fundamental rights in times of digitalization, and, at the same time, in the protection of the right to free demonstration.

Therefore, by addressing the impact of disinformation on fundamental rights, the definition of digital constitutionalism, and the need to impose limits on freedom of expression, it is verified that the action of constitutional courts is indispensable for the protection of situations that arise from the evolution driven by new technologies, in order to prevent the practice of disinformation from weakening democracy, to the extent that citizens stop exercising citizenship through manipulation, which directly impacts the Democratic Rule of Law.

REFERENCES

CELESTE, Edoardo. Digital constitutionalism: a new systematic Theorization. **International Review of Law, Computers & Technology.** V.33, n.1, p. 76-99, 2019. DOI: https://doi.org/10.1080/13600869.2019.1562604. Available at: https://www.tandfonline.com/doi/epdf/10.1080/13600869.2019.1562604?needAccess =true. Accessed on: June 27, 2025.

CELESTE, Edoardo. Conceptual Approaches to Digital Constitutionalism: a CounterCritique. *In* **Digital Constitutionalism**. (Eds.) Indra Spiecker gen. Döhmann, Laura Schertel Mendes & Ricardo R. Campos. [Online Electronic Resource] (2025). DOI: https://doi.org/10.5771/9783748938644.

CELESTE, E.; PALLADINO, N., REDEKER, D.; YILMA, K. The Content Governance Dilemma. Information Technology and Global Governance and the Search for a Global Standard. Ed. Palgrave Macmillan, (eBook) (2023). https://doi.org/10.1007/978-3-031-32924-1_1

ROBL FILHO, Ilton; SARLET, I. W. Freedom of Speech in the Federal Constitution of Brazil and the Problem of its Collision with other Fundamental Rights, particularly personality rights. **Przegląd prawa konstytucyjnego.** [Online electronic resource], DOI: https://10.15804/ppk.2016.06.07. Available at: https://czasopisma.marszalek.com.pl/en/10-15804/ppk/2414-ppk2016607. Accessed on: 26 Jun 2025.

MENDES, Gilmar Ferreira; FERNANDES, Victor Oliveira. Bridging Legislation and Jurisprudence in Democratic Digital Constitutionalism: a look at the brazilian supreme court's approach. *In* Digital Constitutionalism. [Online electronic resource]. https://www.nomos-elibrary.de/de/10.5771/9783748938644/digitalconstitutionalism?page=1

3. DIGITAL CONSTITUTIONALISM AND THE LIMITS OF FREEDOM OF EXPRESSION IN BRAZIL: FROM SOCIAL MEDIA PLATFORMS TO THE FEDERAL SUPREME COURT

https://doi.org/10.36592/9786554603065-02

Marcella de Pinho Pimenta Borges Ramos

Abstract
This essay analyzes the Brazilian Supreme Federal Court's (STF) decision declaring the partial and progressive unconstitutionality of Article 19 of *Marco Civil da Internet*[1], which had limited platform liability for third-party content to cases of disobedience to judicial orders. Based on a bibliographic review and analysis of constitutional provisions, case law, the study explores how freedom of expression, protected by the 1988 Federal Constitution, interacts with other fundamental rights in the context of digital communication. It examines the rise of social media as quasi-public spheres governed by private norms, the challenges posed by disinformation and algorithmic amplification, and the emergence of Digital Constitutionalism as a normative response. By analyzing the STF's recent rulings in two emblematic cases, the paper argues that while the Court's approach promotes accountability and the protection of dignity, it also reveals risks of legal uncertainty and judicial overreach. The essay concludes that a comprehensive, democratically debated legal framework remains essential.
Keywords: Freedom of expression. STF. *Marco Civil da Internet*. Digital Constitutionalism. Social media.

1. Introduction

In the past decade, the world has witnessed a profound transformation in the way public discourse is produced, disseminated, and consumed. No longer restricted to traditional media channels or physical public squares, the exercise of freedom of expression has migrated to digital platforms, particularly social media, which today

[1] The *Marco Civil da Internet* (Law No. 12,965/2014) is Brazil's foundational legal framework governing the use of the Internet, establishing principles, rights, and duties for users, service providers, and the State. It enshrines fundamental guarantees such as freedom of expression, protection of privacy and personal data, network neutrality, and due process in the removal of online content, while also regulating issues related to data retention, liability of application providers, and government access to digital information. Often described as a "Digital Bill of Rights," the statute seeks to balance innovation and economic development with the safeguarding of constitutional rights in the digital environment, providing legal certainty and normative guidance for Internet governance in Brazil.

concentrate unprecedented communicative and normative power in the hands of private actors.

This shift has challenged the very premises upon which constitutional democracies have historically regulated speech, raising urgent questions: How can freedom of expression be preserved in the face of algorithmic moderation, cross-border governance, and the viral spread of disinformation? Are existing legal frameworks equipped to mediate the tension between liberty and harm in the digital age?

This essay examines these questions from the perspective of Brazilian constitutional law, focusing on the country's landmark legislation on digital rights, *Marco Civil da Internet* (Law No. 12,965/2014), and the recent (re)interpretation of its Article 19 by the Federal Supreme Court (STF). It argues that Brazil is experiencing a process of judicial digital constitutionalism, whereby Courts are assuming an increasingly central role in defining the limits and guarantees of online expression, especially in the absence of updated legislation.

The objective of the essay is twofold. First, it seeks to analyze how the Brazilian Constitution conceptualizes freedom of expression, particularly in contexts of conflict with other fundamental rights. Second, it investigates how digital environments have reshaped this constitutional balance, culminating in the STF's ruling on the partial and progressively unconstitutionality of Article 19 of *Marco Civil da Internet*.

Methodologically, this study combines a review of scholarly literature on constitutional law, internet governance, and digital constitutionalism, with the analysis of primary legal sources, namely, the Brazilian Federal Constitution, Law No. 12.965/2014 (*Marco Civil da Internet*), and leading STF jurisprudence. By weaving together theoretical insight and doctrinal analysis, the essay aims to contribute to a deeper understanding of the legal dilemmas posed by the privatization of public discourse in the online age.

The structure of the essay unfolds in three main parts. The first section outlines the constitutional foundations of freedom of expression in Brazil, highlighting its scope, limitations, and the role of the Judiciary in mediating tensions between speech and other constitutional values such as dignity, equality, and

democratic order. The second section turns to the rise of social media and the emergence of digital constitutionalism, illustrating how new modes of communication have disrupted traditional regulatory paradigms and prompted normative responses both within and beyond the state. The third section focuses on the judicial (re)interpretation of Article 19 of *Marco Civil da Internet*, examining the STF's landmark ruling, its rationale, and its broader implications for platform responsibility and constitutional governance in cyberspace.

In traversing these three dimensions, the essay ultimately seeks to demonstrate that the digital public sphere, far from being legally neutral, demands renewed constitutional scrutiny, one that affirms rights while confronting the new architectures of harm and control that define our time.

2. Constitutional Right To Freedom Of Expression And Its Limitations

After a period marked by censorship and authoritarian control under Brazil's military regime (1964–1985), the 1988 Federal Constitution emerged as a constitutional milestone, reestablishing the right to express, receive, and disseminate ideas without fear of reprisal. As in most Western democracies, freedom of expression (commonly referred to as "freedom of speech") holds a central place in the Brazilian constitutional framework. It is not only an individual liberty, but also a structural pillar of democratic governance, particularly in a country whose democracy arose from the ashes of dictatorship. This is not a mere rhetorical claim, since, according to Filho and Sarlet (2016), "the role of freedom of speech in the democratic rule of law is widely acknowledged to be one of the most precious fundamental rights and corresponds to one of the oldest human needs".

Article 5 of the Brazilian Federal Constitution reflects this normative shift in the country's sociopolitical trajectory by safeguarding a broad spectrum of freedom to expression. These include freedom of conscience (Article 5, VI), freedom of thought (Article 5, IV), freedom of communication (Article 5, IX), and the confidentiality of private communications (Article 5, XII). Crucially, these rights are not framed in isolation, but rather within a constitutional architecture that also provides limits,

designed to ensure harmony with other fundamental values, such as dignity, privacy, and equality (Brazil, 1988).

This multifaceted right includes both positive dimensions (such as the ability to manifest one's beliefs through speech, writing, religion, or art) and negative dimensions (such as the right not to be compelled to express oneself or to remain silent in accordance with personal convictions) (Silva, 2003). Article 220 reinforces this framework by stating that "The manifestation of thought, the creation, the expression and the information, in any form, process or medium shall not be subject to any restriction" (Brazil, 1988). Nevertheless, no freedom is absolute, particularly when it conflicts with democratic principles or infringes upon the rights of others (Silva, 2003).

In this regard, the right to free expression is accompanied by a duty of responsibility. Article 5, IV and V, prohibits anonymity and ensures the right of reply and compensation for moral or material damage caused by unlawful speech (Brazil, 1988). Such provisions demonstrate that, in Brazil, freedom of expression is neither unconditional nor immune from accountability. As José Afonso da Silva (2003) asserts, "no one may be compelled to act contrary to their religious belief or political or philosophical conviction," but this right ends where another's begins.

According to Mendes and Gonet (2014), freedom of expression includes "any opinion, belief, commentary, assessment or judgment on any subject or person", regardless of whether the topic is of public relevance or private interest, as long as it does not violate other fundamental rights. This interpretation is consistent with jurisprudence from the Supreme Federal Court (STF), which has repeatedly emphasized that public liberties must be exercised in harmony, respecting constitutional boundaries and prevailing values such as dignity, equality, and honor (Brazil, 2003).

In this context, the judiciary plays a crucial role in mediating these constitutional tensions. Since the Constitution does not provide a closed list of expression restrictions, and arguably could not do so without becoming overreaching, it falls to Courts to evaluate, on a case-by-case basis, whether a particular act of expression crosses the line into abuse. This jurisprudential filter becomes particularly important in digital contexts, where speed, scale, and

anonymity intensify both the power and the risks of speech. As noted by the STF, "public freedoms are not unconditional" and must yield when they serve as cover for crimes such as hate speech or incitement to racism (Brazil, 2003).

This constitutional balance is especially relevant in contemporary Brazil, where freedom of expression is often invoked as a shield to legitimize unlawful or discriminatory discourse, particularly in online environments. The misleading assumption that speech on the Internet is free from consequences contradicts both constitutional doctrine and judicial precedents. As Justice Luiz Fux has stated, "the manifestation of thought cannot tolerate violence or threat, under the risk of turning against democracy itself – a fundamental precept for the guarantee of any constitutional freedom" (Brazil, 2021).

In sum, Brazil's constitutional framework protects expression as a core democratic right, but does so within a matrix of interconnected principles, including human dignity, legal equality, and institutional legitimacy. The STF has reinforced this interpretation in landmark decisions, including *Habeas Corpus* 82.424-RS, where it upheld the constitutionality of criminal penalties for racist publications and denied that hate speech could be protected under the umbrella of free expression (Brazil, 2021):

> *Freedom of speech is not absolute but limited by moral and law. Free speech cannot shelter manifestations that result in crime. No public liberty is unconditional, they all must be exercised in harmony with the limits prescribed by the Constitution. This means that freedom of expression does not include the right to incite racism, since an individual's right cannot be used in order to commit crimes. There is a clear prevalence of human dignity and judicial equality.*

In doing so, the Court drew clear boundaries around what it termed the "non-protectable core" of expression, particularly when speech violates the principles of dignity and equality. This does not happen only in Brazil. According to Filho and Sarlet (2016):

> *In turn, national, foreign and international practices have shown, in the case of restrictions imposed by court decisions, that they usually intend to solve conflicts in particular cases*

> *and seek to promote practical concordance (harmonization) between conflicting rights and principles, always applying the notion of restraints of fundamental rights and the criteria resulting therefrom with special observance of the criteria of proportionality and protection of the essential core of the rights concerned.*

As mentioned before, in recent years, the rise of disinformation and online radicalization has intensified debates over the limits of expression, especially as new forms of harm, such as targeted harassment, hate campaigns, and incitement to violence, emerge in digital spaces. These phenomena have tested the resilience of the constitutional framework and exposed gaps in the regulation of speech on global platforms. Importantly, while public discourse often embraces a simplistic and absolutist notion of freedom of expression, Brazilian constitutional law clearly mandates a more contextual and proportionate analysis.

Today, the challenge is not merely affirming freedom of expression, but defining how this right can coexist with new digital threats and systemic abuses. It is within this tension, between protection and responsibility, between liberty and harm, that Courts, lawmakers, and society must now operate.

3. Rise Of Social Media And Digital Constitutionalism

The digital revolution has fundamentally reshaped the landscape of communication, sociability, and public participation. What began as a decentralized and mostly passive network of static web pages has evolved into an interactive and immersive digital ecosystem, in which billions of users create and circulate content in real time (Santos, 2022). This transformation gave rise to the phenomenon of social media platforms, a shift that redefined the role of the individual in the digital sphere.

For the first time in the history of Internet, users were no longer mere consumers of online content; they became its primary creators. Social media has empowered individuals to simultaneously produce, share, and respond to content, thus establishing a dynamic feedback loop that has radically altered the conditions for public discourse. The intrinsic value of social media platforms lies not in the

infrastructure or curated content they offer, but in the collective output of users themselves, whose expressions animate and sustain these digital environments (Celeste, 2023).

Yet, this empowerment came at a cost. The digital spaces where individuals increasingly exercise fundamental rights, such as communicating, expressing political beliefs, protesting, or conducting business, are privately owned and governed by opaque contractual rules. In an era where online interactions are inseparable from one's physical and political existence, social media platforms have acquired almost constitutional character. As Celeste (2023) observes, social media has become "an integral component of the context where we live," so much so that "one could no longer think of exercising some of our core fundamental rights without resorting to social media".

However, the consolidation of social media as the main arena of public discourse has brought with it new normative and constitutional challenges. Companies such as Facebook and Google are not passive intermediaries; they actively shape the flow of information by using algorithms to amplify, suppress, or filter user-generated content (Mendes; Fernandes, 2022). These actions, often guided by business interests rather than democratic values, affect the conditions under which public freedoms are exercised. The technological infrastructure of platforms thus becomes a normative structure in itself, one that exerts influence on speech, behavior, and even political outcomes.

A striking example of this phenomenon was the Cambridge Analytica scandal, which exposed how personal data collected from social media could be transformed into a powerful tool of political manipulation. In 2016, the company illegally collected data from millions of Facebook users through personality quizzes. Unbeknownst to participants, the quizzes not only harvested their information but also accessed data from their entire contact network. This data was then used to create highly personalized political advertisements aimed at manipulating voter behavior in both the U.S. presidential election and the Brexit referendum. The case was widely publicized by the 2019 Netflix documentary *The Great Hack*, which revealed the scale and sophistication of these operations and raised serious concerns about the integrity of democratic processes and citizen autonomy (Amer; Nouajim, 2019).

The example is emblematic of how disinformation, when combined with the massive surveillance capacity of private platforms, can distort the public sphere in ways previously unimaginable. More than a violation of privacy, the case revealed a structural vulnerability in modern democracies, one in which digital tools originally designed for connection and communication were repurposed for targeted persuasion, ideological radicalization, and electoral interference (Martins; Tateoki, 2019). It also illustrated the dangers of placing the architecture of digital public discourse in the hands of actors motivated by profit rather than the public interest.

This concentration of communicative power in the hands of private corporations has led some scholars to describe platforms as emerging global actors rivaling nation-states. While these companies operate within territorial jurisdictions, their reach and governance mechanisms transcend borders. In response to overlapping legal regimes and regulatory ambiguity, platforms have increasingly adopted their own internal rules: terms of use and content policies that function, in practice, like quasi-legal instruments. These "boilerplate contracts" are unilaterally drafted, non-negotiable, and enforced without meaningful oversight (Celeste, 2023).

This regulatory self-governance has sparked criticism for its opacity and lack of accountability. By choosing which laws to comply with and when to intervene in content moderation, platforms sidestep democratic deliberation and blur the line between private policy and public normativity. As Celeste (2023) points out, this situation reveals a dangerous tension: on one hand, the risk of normative authoritarianism, where private interests dictate the boundaries of speech; on the other, the risk of anomie, where no clear legal standards apply. In both cases, the result is a fragmented, unpredictable legal landscape.

Against this backdrop, the concept of Digital Constitutionalism has emerged as a framework to restore balance between individual rights, state authority, and private power. Digital Constitutionalism is not merely a descriptive label; it constitutes a normative movement aimed at recognizing, affirming, and protecting fundamental rights in cyberspace, while reconfiguring the distribution of power among the actors that govern it (Mendes; Fernandes, 2022). Its objectives include guaranteeing freedom of expression online, ensuring transparency in content

moderation, and imposing limits on the discretionary authority of platforms.

Importantly, Digital Constitutionalism does not wait for legislation to catch up, since it provides normative guidance for judicial review, legislative drafting, and institutional accountability. As Mendes and Fernandes (2022) explain, "Digital Constitutionalism precedes normative reactions and can provide normative guidelines to judicial review". In Brazil, this is reflected in the *Marco Civil da Internet* (Law No. 12,965/2014), which incorporates several elements of this movement and was one of the early legislative responses to the challenges of regulating cyberspace (Mendes; Fernandes, 2022).

Still, the promise of Digital Constitutionalism is not without obstacles. The same platforms that enable speech also facilitate the spread of hate speech, cyberbullying, child exploitation, and disinformation, often at scale and with little oversight. These harms complicate the ideal of an open, participatory Internet, and have prompted increasing calls for national and regional regulation to address online abuse and algorithmic opacity (Celeste, 2023).

In this evolving scenario, states face the dual challenge of reclaiming their regulatory role without replicating authoritarian practices. The expansion of platform power, both communicative and normative, requires not only updated legislation but also a robust constitutional framework that ensures private governance remains accountable to democratic standards.

It is within this context that Courts, particularly Constitutional Courts, are being called to mediate the tension between technological autonomy and legal legitimacy. As Mendes and Fernandes (2022) argue, "the internet can alter the factual context of a given technology, raising questions about how the Constitution applies to it." In other words, the digital shift is not only technological, but also constitutional.

4. Judicial (Re)Interpretation Of Article 19 Of *Marco Civil Da Internet* (Law No. 12.965/2014): Judicial Digital Constitutionalism In Action

Approved in 2014, Brazil's *Marco Civil da Internet* (Law No. 12.965/2014) consolidated fundamental principles such as Internet neutrality, privacy protection, and freedom of expression in the digital environment (Brazil, 2014). Among its most

debated provisions was Article 19, which established that internet application providers could only be held civilly liable for third-party content upon a specific judicial order requiring its removal. This approach was intended to avoid the privatization of censorship by ensuring that platforms would not preemptively restrict content in fear of liability, and instead reserve the power of content legitimacy assessment to the Judiciary.

As Mendes and Fernandes (2022) point out, *Marco Civil da Internet* "incorporates several elements of the growing literature on digital constitutionalism" at the same time as it "erected general clauses of individual rights in cyberspace that serve as hermeneutic beacon for judicial review before the Federal Supreme Court (STF)". In its original design, Article 19 was a safeguard: a barrier against the privatization of censorship and a guarantee that only a Court could determine the unlawfulness of content.

However, as the online ecosystem became more complex and central to public life, this assumption of neutrality proved increasingly fragile. As Mendes and Fernandes (2022) argue, companies like Facebook, Google, and Amazon, far from acting as passive conduits, actively shape digital discourse through content curation algorithms, monetization policies, and opaque moderation practices. These actions often reflect corporate interests rather than democratic values, and thus influence the very conditions under which constitutional freedoms are exercised.

Over time, civil society, scholars, and judicial actors began to question whether Article 19 remained adequate to confront the new architecture of harm introduced by algorithmically amplified content and virality. In particular, the disinformation strategies deployed during Brazil's electoral cycles exposed the inability of the existing framework to contain the speed and scale of rights violations occurring online. Despite the law's intent to protect public discourse, its procedural safeguard, requiring a Court order, often acted as a structural delay that neutralized any meaningful intervention before damage was already inflicted.

It was in this context that the STF was called upon to reassess Article 19 in light of constitutional principles. The issue reached the Court through two extraordinary appeals with recognized repercussion: RE No. 1.037.396 (Theme 987) and RE No. 1.057.258 (Theme 533).

In the first case (Theme 987), a fake Facebook profile was created using the name of a person who did not even have an account on the platform. The account was used to attack and offend others. Although the platform was notified via its internal reporting system, it failed to act. The harmed individual then sought judicial intervention, requesting both the deletion of the profile and compensation for moral damages. The Court ordered the deletion, which was complied with, but denied the compensation. On appeal, however, the Court of Justice of São Paulo ordered Facebook to pay damages, stating that the company should have acted upon the extrajudicial notice. Facebook appealed to the STF, arguing that, under Article 19, it could only be held liable if it disobeyed a Court order, which it hadn't.

The second case (Theme 533) involved a community created on the now-defunct social network Orkut. The group contained offensive and derogatory comments about a school teacher. The teacher asked the platform to take the group down, citing damage to her honor and image. Orkut refused, saying the content didn't violate its policies or the law. She then filed a lawsuit. The court ordered the group's removal and awarded her moral damages. Google Brasil, which operated Orkut at the time, removed the group as required but appealed the indemnity, citing Article 19. It claimed it had acted as soon as it was ordered to and should not be held liable.

On June 26, 2025, by a majority of eight to three, the STF ruled that Article 19 of *Marco Civil da Internet* is partially and progressively unconstitutional. The decision recognized that while Article 19 aimed to protect freedom of expression, it failed to adequately safeguard other essential constitutional goods (such as the dignity of vulnerable groups, the protection of children and minorities, and the proper functioning of democratic institutions) when faced with the structural dynamics of digital harm.

Rather than invalidate Article 19 in full, the STF established a new interpretative framework in which Internet application providers may be held civilly liable without prior judicial order in specific scenarios. These include cases involving paid content promotion or artificial content amplification by bots, as well as the widespread circulation of gravely illicit materials, such as terrorist propaganda, hate speech targeting protected groups, child pornography, and incitement to violence or

suicide. In such contexts, platforms must act promptly to remove the offending content or face legal consequences for systemic failure. The Supreme Court further established a presumption of liability in these instances, which providers may rebut only by demonstrating that they took adequate, diligent, and timely steps in accordance with the current technological state of the art.

Additionally, STF refined the application of Article 19 to accommodate the diverse nature of online services. While content like personal emails and private messaging remains shielded due to constitutional confidentiality, content circulating on open social media feeds, especially when previously declared illicit by Court order, is subject to expedited removal, even without a fresh judicial intervention. In the case of repeated postings of the same unlawful content, platforms are now obliged to act following a single notification.

Beyond these interpretative changes, the STF also imposed a series of practical obligations on platforms operating in Brazil. These include implementing transparent internal regulations for moderation, publishing regular reports, ensuring user-friendly communication channels, and maintaining a legal representative in the country with full administrative and judicial capacity. While the Court stopped short of establishing strict liability, it emphasized that platforms' omission in the face of foreseeable harm could trigger civil responsibility, provided there is evidence of systemic failure.

To preserve legal certainty, the effects of the decision were modulated to apply only to future events, without retroactive implications. In parallel, the Court explicitly appealed to the National Congress to legislate on the matter, recognizing that the Judiciary alone cannot define the contours of digital constitutional governance in the long term.

The ruling marks a pivotal moment in Brazil's digital regulation landscape, reshaping the legal contours of platform responsibility. By declaring Article 19 of the *Marco Civil da Internet* partially and progressively unconstitutional, the Court signaled a departure from a model centered on judicial inertia toward a more proactive accountability regime. As Mendes and Fernandes (2022) assert:

> *From the incorporation of the values of Digital Constitutionalism, the constitutional judicial review of art. 19 of MCI must consider the degree of commitment of private actors to the constitutional precept of freedom of expression (art. 5, item IV of C.F./88). That perspective may eventually mean an opening of constitutional jurisdiction to the concrete evaluation of the practices of ruling personality rights by digital platforms. The experience accumulated by the Judiciary in dealing with these issues can undoubtedly contribute to an assessment of the risks and benefits of the regime of personal liability of internet providers.*

The Court's decision thus redefines the role of social media companies, no longer viewing them merely as passive intermediaries, but as co-responsible agents in safeguarding the constitutional order and the fundamental rights it protects. However, the ruling is not devoid of complexity. While it seeks to prevent harm arising from platform inaction, it simultaneously introduces ambiguities, particularly in delineating the boundary between legitimate content moderation and undue censorship. In the absence of democratically debated, comprehensive legislation, the decision leaves open the possibility of legal uncertainty and judicial overexposure to contentious speech disputes. The burden now falls on National Congress to craft a normative framework capable of reconciling the competing imperatives of freedom of expression and the effective protection of dignity, equality, and democratic integrity in the digital sphere.

In this light, the STF's intervention must be seen both as a constitutional necessity, in order to protect fundamental rights in a hyperconnected environment, and as a constitutional risk, due to the uncertain boundaries of judicial regulation. It is precisely this duality that renders the ruling so pivotal: the Supreme Court positioned itself not only as guardian of individual rights but also as architect of digital governance, assuming a central role in a space still largely undefined by formal legislation.

Ultimately, STF's partial invalidation of Article 19 represents an important step toward balancing freedom of expression with accountability in the digital age. It aligns Brazil with the global constitutional trend of assigning more responsibility to private platforms while reaffirming the need for transparent, rights-oriented governance. However, its success will depend not only on judicial vigilance, but on

the political will to enact comprehensive, legitimate, and technologically grounded legislation. In the meantime, digital constitutionalism remains a judicially driven project, necessary, perhaps, but far from ideal.

5. Conclusion

The transformation of the public sphere in the digital age has brought freedom of expression into a new and complex constitutional arena. In Brazil, this evolution has required not only a reaffirmation of fundamental rights, but also a reinterpretation of how these rights operate in a world where private platforms increasingly mediate the terms of public discourse.

The Federal Supreme Court's recent decision on the partial and progressively unconstitutionality of Article 19 of *Marco Civil da Internet* represents a turning point in this process: it confronts the normative inadequacies of a system that, while originally designed to safeguard expression, became structurally incapable of addressing the velocity, scale, and severity of digital harms.

The STF's ruling is significant not merely for what it declares, but for what it signals. By partially invalidating Article 19, the Court reframes platforms not as neutral conduits, but as actors bearing constitutional responsibilities. In doing so, it aligns with the values of digital constitutionalism and acknowledges that algorithmic amplification, monetization strategies, and omission in the face of grave rights violations are not private business matters alone, they are constitutional concerns.

At the same time, however, the decision raises critical issues about judicial reach, legal certainty, and the risks of substituting judicial improvisation for democratic deliberation. This is where the cautionary note must be emphasized. While the STF fulfilled an urgent role in responding to a legislative gap with immediate constitutional consequences, its decision also underscores the limitations of judicial governance in the long term.

The interpretation of fundamental rights, especially in contexts as mutable and technically complex as digital speech, cannot remain indefinitely in the hands of the Judiciary alone. Courts may offer protection, but only the National Congress can provide the normative stability, plural debate, and democratic legitimacy required to

regulate digital communication with constitutional depth and foresight.

Therefore, the ruling must be understood not as a final solution, but as a transitional step. It opens an essential legal path, but that path must now be paved by legislation that both reaffirms freedom of expression and imposes clear, proportionate, and democratically defined duties on digital platforms.

The future of digital rights in Brazil depends on the ability of constitutional institutions to act in coordination: courts to safeguard, legislators to regulate, and civil society to participate in the construction of a normative model that reflects not only technological realities, but also democratic aspirations.

References

AMER, Karim; NOUAJIM, Jehane (Dir.). **Privacidade hackeada**. Produção: Netflix. Estados Unidos, 2019. 1 vídeo (114 min). Documentário. Disponível em: https://www.netflix.com. Acesso em: 17 jun. 2025.

SANTOS, Gustavo Ferreira. Social media, disinformation, and regulation of the electoral process: a study based on 2018 Brazilian election experience. **Revista de Investigações Constitucionais**, Curitiba, v. 7, n. 2, p. 429-449, maio/ago. 2020.

BBC NEWS BRASIL. Cambridge Analytica: como funcionava o 'lado sombrio' do marketing político que ajudou Trump e Brexit. 10 abr. 2018. Disponível em: https://www.bbc.com/portuguese/geral-43705839. Acesso em: 20 jun. 2025.

BRASIL. [Constituição (1988)]. **Constitution of the Federative Republic of Brazil**. Brasília: Supremo Tribunal Federal, Secretaria de Altos Estudos, Pesquisas e Gestão da Informação, 2022.

BRASIL. **Lei n. 12.965**, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Diário Oficial da União: seção 1, Brasília, DF, 24 abr. 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 26 jun. 2025.

_____. Supremo Tribunal Federal (STF). **Case law compilation** [recurso eletrônico]: freedom of speech. Brasília: STF, Secretaria de Altos Estudos, Pesquisas e Gestão da Informação, 2021. eBook (v. 2, 170 p.).

_____. Supremo Tribunal Federal (Tribunal pleno). **Habeas Corpus n. 82.424-RS**. Habeas corpus. Publicação de livros: antissemitismo. Racismo. Crime

imprescritível. Conceituação. Abrangência constitucional. Liberdade de expressão. Limites. Ordem denegada. [...]. Paciente: Siegfried Ellwanger. Impetrante: Werner Cantalício João Becker. Coator: Superior Tribunal de Justiça. Relator: Ministro Moreira Alves. Relator para acórdão: Maurício Corrêa, 17 de setembro de 2003. Disponível em: https://jurisprudencia.stf.jus.br/pages/search/sjur96610/false. Acesso em: 21 jun. 2025.

_____. Supremo Tribunal Federal (STF). **Informação à sociedade**: julgamento sobre o artigo 19 do Marco Civil da Internet. Disponível em: https://www.stf.jus.br/arquivo/cms/noticiaNoticiaStf/anexo/Informac807a771oa768SociedadeArt19MCI_vRev.pdf. Acesso em: 26 jun. 2025.

CELESTE, Edoardo; PALLADINO, Nicola; REDEKER, Dennis; YILMA, Kinfe. **The content governance dilemma**: digital constitutionalism, social media and the search for a global standard. Cham: Palgrave Macmillan, 2023. (Information Technology and Global Governance). DOI: https://doi.org/10.1007/978-3-031-32924-1. Acesso em: 26 jun. 2025.

FILHO, Ilton Robl; SARLET, Ingo Wolfgang. Freedom of speech in the Federal Constitution of Brazil and the problem of its collision with other fundamental rights, particularly personality rights. **Przegląd Prawa Konstytucyjnego**, [s.l.], n. 6(34), p. 133–163, 31 dez. 2016. Disponível em: https://bibliotekanauki.pl/articles/940741. Acesso em: 26 jun. 2025.

MARTINS, Marcelo Guerra; TATEOKI, Victor Augusto. Proteção de dados pessoais e democracia: fake news, manipulação do eleitor e o caso da Cambridge Analytica. Revista Eletrônica Direito e Sociedade - REDES, v. 7, n. 3, 2019. Disponível em: https://revistas.unilasalle.edu.br/index.php/redes/article/view/5610. Acesso em: 29 jun. 2025.

MENDES, Gilmar Ferreira; OLIVEIRA FERNANDES, Victor. Digital Constitutionalism and Constitutional Jurisdiction: A Research Agenda for the Brazilian Case. **Rev. Just. Direito**, v. 34, p. 6, 2020.

MENDES, Gilmar Ferreira; GONET BRANCO, Paulo Gustavo. **Curso de direito constitucional**. *E-book*. 9. ed. rev. e atual. São Paulo: Saraiva, 2014.

MIGALHAS. STF e Big Techs: Barroso nega ativismo e diz que não agradará a todos. São Paulo, 26 jun. 2025. Disponível em: https://www.migalhas.com.br/quentes/433508/stf-e-big-techs-barroso-nega-ativismo-e-diz-que-nao-agradara-a-todos. Acesso em: 26 jun. 2025.

OLIVEIRA, André Soares; GOMES, Patrícia Oliveira. Os limites da liberdade de expressão: fake news como ameaça à democracia. **Revista de Direitos e Garantias Fundamentais**, Vitória, v. 20, n. 2, p. 93-118, maio/ago. 2019.

SANTOS, Lucas Francisco Gomes. **Web 3.0**: um ensaio sobre a evolução da

internet. 2022. Artigo apresentado ao curso de Sistemas de Informação do Centro Universitário Doctum de Teófilo Otoni – Unidoctum, na disciplina de Trabalho de Conclusão de Curso II, como requisito parcial para obtenção de título de Bacharel em Sistemas de Informação. Disponível em: https://dspace.doctum.edu.br/handle/123456789/4812. Acesso em: 20 jun. 2025.

SILVA, José Afonso da. **Curso de direito constitucional positivo**. 22. ed. rev. e atual. São Paulo: Malheiros, 2003.

SUPREMO TRIBUNAL FEDERAL (Brasil). STF define parâmetros para responsabilização de plataformas por conteúdos de terceiros. Notícias STF, Brasília, 26 jun. 2025. Disponível em: https://noticias.stf.jus.br/postsnoticias/stf-define-parametros-para-responsabilizacao-de-plataformas-por-conteudos-de-terceiros/. Acesso em: 26 jun. 2025.

4. FREEDOM OF SPEECH AND ITS LIMITS: IDEAL OR MYTH?

https://doi.org/10.36592/9786554603065-03

Cyntia Melo Rosa

Abstract
This essay explores the principle of freedom of expression and the challenges of establishing its boundaries within democratic societies. It investigates whether freedom of expression is an ideal to be universally upheld or a myth compromised by the practical need to balance it against other constitutional values such as human dignity, public security, and social order. Drawing from Brazilian jurisprudence, digital constitutionalism, and international case law, it examines the evolving nature of speech in the digital era and its regulation. The analysis shows that although freedom of expression is a cornerstone of democracy, it is not absolute and must coexist with competing rights and principles in both physical and digital realms.
Keywords : Freedom of Expression, Constitutional Law, Digital Rights, Limits, Democracy, Brazil, Jurisprudence.

1. Introduction

Freedom of expression is often celebrated as a fundamental pillar of democratic societies, enabling individuals to express their thoughts, opinions and beliefs without undue interference. It is embedded in international treaties, national constitutions, and the collective conscience of civil society. In Brazil, Article 5 of the 1988 Federal Constitution guarantees this right, prohibiting censorship and protecting various forms of expression. However, the ideal of absolute freedom is continuously tested by the realities of social coexistence and the necessity to balance this right with other values, such as the protection of honor, privacy, and public safety.

In the digital age, the issue becomes even more complex. The internet amplifies voices but also accelerates the spread of misinformation, hate speech, and coordinated manipulation. Legal systems around the world have struggled to reconcile the need for open dialogue with the imperative to prevent harm. This essay aims to interrogate whether freedom of expression can truly exist without limits or whether such an ideal is inherently flawed.

In Brazil, the Supreme Court (STF), through its extensive case law, has consistently emphasized the importance of balancing freedom of speech with its limitations, ensuring that the right does not become a tool for harm or the suppression of democracy itself. From landmark rulings on hate speech and anti-Semitism to decisions addressing homophobia, contempt and de dissemination of fake news, the Brazilian Supreme Court has demonstrated a nuanced approach to safeguarding free expression. While curbing its misuse.

This essay explores the tension between the ideal of unrestricted freedom of speech and the practical necessity of its limits drawing insights from the philosophical and legal foundations of freedom of expression (chapter 1); the understanding of freedom of expression in Brazilian jurisprudence (chapter 2); the ideal of Digital Constitutionalism and its challenges (chapter 3); the international perspectives and its comparatives (chapter 4); the limits to expression: necessity and proportionality (chapter 5); and ends with the conclusion.

2. The Philosophical and Legal Foundations of Freedom of Expression

Freedom of expression is a cornerstone of democratic societies, rooted in philosophical ideals and legal principles that have evolved over centuries. It embodies the right to express thoughts, opinions, and beliefs without fear of censorship or retaliation, serving as a vital mechanism for individual self-fulfillment, societal progress, and the preservation of democracy.

The philosophical underpinnings of freedom of expression trace back to Enlightenment thinkers such as John Locke, John Stuart Mill, and Voltaire. Locke (1689) emphasized the natural rights of individuals, including the right to freely express their ideas as an extension of their autonomy. Mill (1859), in his seminal work *On Liberty*, argued that the free exchange of ideas is essential for the pursuit of truth, as even unpopular or erroneous opinions contribute to the collective understanding of society. He emphasized its role in the "marketplace of ideas," suggesting that truth emerges from the contest of competing viewpoints. Legally, this principle is enshrined in the Universal Declaration of Human Rights (Article 19) and echoed in the American Convention on Human Rights, to which Brazil is a signatory. Voltaire

famously championed the principle of tolerating dissent, encapsulated in his assertion, "*I disapprove of what you say, but I will defend to the death your right to say it*" (Tallentyre, 1906)

These philosophical ideals underscore the intrinsic value of free expression as a means of fostering intellectual diversity, challenging authority, and promoting societal progress. They also highlight the importance of tolerating dissenting voices, even when they conflict with prevailing norms or beliefs.

Legally, freedom of expression is enshrined in constitutions, international treaties, and judicial precedents worldwide. In Brazil, the 1988 Federal Constitution represents a milestone in the protection of this right, explicitly guaranteeing the freedom of thought, speech, press, and artistic expression under Article 5, item IV. It prohibits anonymity and prior censorship, ensuring that individuals can freely express their ideas while remaining accountable for their actions.

Internationally, the American Convention on Human Rights, ratified by Brazil in 1992, further reinforces this right. Article 13 of the Convention affirms that every person has the right to freedom of thought and expression, including the ability to seek, receive, and impart information without barriers. This legal framework aligns with global human rights standards, emphasizing the universality of free expression as a fundamental right.

While freedom of expression is a fundamental right, it is not absolute. Legal systems worldwide recognize the need to balance this right with other principles, such as human dignity, equality, and public safety (Sarlet et al., 2021). The Brazilian Federal Supreme Court (STF) has played a pivotal role in defining the boundaries of free expression, addressing cases where it intersects with hate speech, discrimination, and threats to democracy.

For instance, the STF has ruled that anti-Semitic publications constitute a crime of racism, emphasizing that freedom of speech cannot be used to incite hatred or discrimination. Similarly, the Court has extended the legal definition of racism to include homophobia and transphobia, underscoring the importance of protecting vulnerable groups from harmful speech. These decisions reflect the dynamic interplay between freedom and responsibility, ensuring that the exercise of free expression does not undermine the rights and dignity of others.

The philosophical and legal foundations of freedom of expression highlight its dual role as both a personal right and a societal necessity. It empowers individuals to voice their ideas, fosters the exchange of diverse perspectives, and safeguards democracy against authoritarianism. However, its limits are equally important, ensuring that this right is exercised responsibly and does not infringe upon the rights of others. As the Brazilian Federal Supreme Court has demonstrated, the balance between freedom and responsibility is essential for the preservation of democratic values and the protection of human dignity.

The Brazilian Supreme Court interprets freedom of expression as a fundamental right essential to democracy, individual self-fulfillment and the protection of human dignity. Its jurisprudence emphasizes that freedom of expression is not absolute and must be balanced against other constitutional principles, such as equality, privacy and public safety (Sarlet et al., 2021).

3. Freedom of Expression in Brazilian Jurisprudence

Brazil's Supreme Court has consistently upheld freedom of speech while articulating clear boundaries. As core principle for the Supreme Court's interpretation, the democratic foundation can be named, as the court views freedom of expression as a pillar of democracy, enabling the free exchange of ideas, public debate and criticism of government and institutions. The democratic foundation protects not only popular or conventional opinions but also dissenting, controversial or minority views.

In ADPF 130 (2009), for instance, the Court invalidated the Press Law from the military dictatorship era, ruling it incompatible with democratic values (STF, 2021).

Another key aspect of the court's interpretation is the prohibition of prior censorship, once the court consistently rejects it, affirming that judicial control over speech must occur only after its manifestation, ensuring accountability without suppressing expression beforehand. As an example, in the anti-Semitic speech case (HC 82424), the Court affirmed that incitement to hatred violates constitutional protections and qualifies as a crime of racism.

These rulings illustrate that while the STF robustly defends expression, it draws lines at content that undermines other protected rights or democratic order, demonstrating the accountability: while freedom of expression is protected, individuals remain accountable for abuses, such as defamation, hate speech or incitement to violence.

The Brazilian Federal Supreme Court (STF) has ruled on several landmark cases that define the scope and limits of freedom of speech. In the Marijuana March Case (ADPF 187), the court ruled that peaceful demonstrations calling for legislative changes, such as drug legalization, are protected under freedom of expression and assembly. It emphasized that the State cannot suppress ideas or debates, even if they challenge existing laws (STF, 2021).

Also, the court decided that biographies can be published without prior authorization from the subject of their family (ADI 4815). The Court declared that requiring prior authorization constitutes censorship and violates freedom of expression. It highlighted the importance of biographies in preserving historical knowledge and national memory (STF, 2021).

Equally defending the freedom of speech, the Supreme Court ruled that a journalism degree is not required to work as a journalist (RE 511961). The Court ruled that such a requirement unjustifiably restricts freedom of speech and information. It emphasized that journalism is an extension of free expression and should not be subject to unnecessary barriers (STF, 2021). In the same way was the rule about cartoons during the electoral period (ADI 4451), when the court emphasized that political satire is a vital form of expression in a democracy, fostering critical debate and public participation (STF, 2021).

Brazil's Supreme Court also decided that the "right to be forgotten" is incompatible with the Constitution (RE 1010606). The court ruled that truthful and lawfully obtained information cannot be suppressed due to the passage of time, as it would infringe on freedom of speech and the public's right to information (STF, 2021).

Protecting the freedom of speech, the court decided that religious proselytism on community radio stations is protected (ADI 2566), provided it does not incite hatred or discrimination. Identically, in ADPF 548, the court reinforced universities'

spaces for free debates and expression, prohibiting censorship of political activities (STF, 2021).

On the other hand, in the Anti-Semitic Case (HC 82424) the court showed the limits of the principle, establishing that publishing anti-Semitic content constitutes a crime of racism, which is not protected by freedom of speech, emphasising the importance of human dignity and equality over unrestricted speech. With a similar foundation, the court ruled the criminalization of homophobia (ADO 26), extending the legal definition of racism to include discrimination based on sexual orientation and gender identity, reinforcing protections for the LGBTQ+ community.

A great remark must be made in the fake news investigation (ADPF 572 and INQ 4781), when the court established that freedom of speech does not protect actions aimed at undermining democracy or spreading false information. The court ruled that speech inciting violence, hatred, or anti-democratic actions is unconstitutional, emphasizing the need to protect democratic institutions.

These cases collectively demonstrate the STF's commitment to safeguarding freedom of speech while balancing it against other constitutional principles, such as human dignity, equality, and public safety. They highlight the dynamic nature of free expression in a democratic society.

4. Digital Constitutionalism and Emerging Challenges

With the rise of digital platforms, freedom of expression has entered a new terrain. Social media companies act as both facilitators and regulators of speech, creating "private constitutions" (Celeste et al., 2023). This raises the issue of normative legitimacy: should private corporations define the limits of public discourse?

Social media platforms define freedom of expression through their content policies, community guidelines, and terms of service, which outline the types of speech and content allowed on their platforms. These definitions often reflect a balance between promoting open communication and addressing concerns such as safety, privacy, and public order. To justify what was stated above, here is an illustration of how the most important platforms approach freedom of expression.

Meta (Facebook and Instagram) emphasizes freedom of expression as a core value, stating that its platforms aim to create "a place for expression and giving people voice" (Meta 2022a). However, this principle is limited by other values, such as authenticity, safety, privacy, and dignity (Celeste et al., 2023). Meta's Community Standards prohibit content that violates these values, including hate speech, incitement to violence, and misinformation. While Meta claims to respect international human rights standards, its policies are tailored to its business interests and operational needs, leading to accusations of censorship or inconsistency.

Twitter defines freedom of expression as enabling "public conversation" while ensuring that users can participate "freely and safely" (Twitter 2020). Its Twitter Rules prohibit abusive behavior, harassment, and content that promotes violence or extremism. Twitter also claims that its moderation practices are informed by international human rights standards, such as the United Nations Guiding Principles on Business and Human Rights. However, the platform has faced criticism for inconsistent enforcement and for allowing harmful content to persist under the guise of free speech (Celeste et al., 2023).

TikTok's Community Guidelines focus on creating a "safe and inclusive environment" while allowing users to express themselves creatively. The platform prohibits content that promotes hate speech, harassment, or misinformation. TikTok's human rights statement acknowledges the importance of freedom of expression but emphasizes its responsibility to prioritize safety and respect for local laws. This utilitarian approach often leads to stricter moderation practices, particularly in regions with restrictive legal frameworks (Celeste et al., 2023).

YouTube defines freedom of expression as the ability to share ideas and information through video content (Celeste et al., 2023). Its Community Guidelines prohibit harmful or dangerous content, hate speech, and misinformation, particularly related to elections and public health. YouTube claims to respect international human rights standards but has faced criticism for allowing harmful content to spread and for inconsistent enforcement of its policies.

As can be seen, the common themes across platforms are balancing free speech and safety (as platforms often prioritize safety, privacy, and dignity over unrestricted freedom of expression, leading to limitations on harmful or offensive

content), Context-Specific Rules (content policies are tailored to address specific challenges, such as hate speech, misinformation, and harassment, while allowing room for creative and political expression), influence of International Standards (many platforms claim to be informed by international human rights standards, such as the Universal Declaration of Human Rights and the Ruggie Principles, but their implementation varies widely) and private governance (Platforms act as private regulators of speech, creating their own rules and enforcement mechanisms, which can lead to accusations of censorship or bias).

Therefore, social media platforms face criticism for inconsistent enforcement, lack of transparency and the tension between promoting free expression and addressing harmful content. Their definitions of freedom of expression are shaped by business interests, legal obligations and public pressure, making it difficult to establish a unified standard.

The Marco Civil da Internet (Law 12.965/2014) is Brazil's attempt to address this. It sets out principles like net neutrality and user privacy while affirming freedom of expression online. Yet the law is tested daily as courts confront issues like fake news, electoral manipulation, and algorithmic bias (Mendes & Fernandes, 2021).

Digital constitutionalism advocates for embedding human rights into the technical architecture of online platforms. This includes procedural fairness in content moderation and transparency in algorithmic decisions. However, the implementation of such norms remains inconsistent.

5. International Perspectives and Comparative Jurisprudence

Comparative insights from the U.S. and Europe offer contrasting models. The U.S. maintains an expansive view of free speech under the First Amendment, even tolerating hate speech. In contrast, European systems, guided by the European Court of Human Rights, permit more restrictions, especially concerning hate speech and defamation.

Brazil navigates a middle path, recognizing the importance of freedom but imposing limitations to safeguard dignity and social harmony. The STF's

jurisprudence frequently reflects this balance, positioning the Court as a moderator of constitutional values.

Freedom of expression is a cornerstone of democratic societies and a fundamental human right enshrined in international legal instruments such as the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR) (Celeste et al, 2023). However, its interpretation and application vary significantly across jurisdictions, influenced by cultural, political, and legal traditions (Mégret, 2013). This chapter explores international perspectives and comparative jurisprudence on freedom of expression, highlighting the tensions between universal principles and local adaptations.

The ICCPR, ratified by 173 states, provides the most comprehensive international framework for freedom of expression (Celeste et al, 2023). Article 19 guarantees the right to hold opinions and express ideas without interference, while Article 19(3) outlines permissible restrictions based on legality, necessity, and legitimacy. These restrictions aim to balance freedom of expression with other rights, such as privacy, non-discrimination, and public order.

Regional instruments further elaborate on these principles. The European Convention on Human Rights (ECHR) balances freedom of expression with competing rights under Article 10, emphasizing proportionality and necessity (Celeste et al, 2023). Similarly, the African Charter on Human and Peoples' Rights and the American Convention on Human Rights provide regional interpretations tailored to specific socio-political contexts.

The United States adopts a robust approach to freedom of expression, rooted in the First Amendment of its Constitution (Pollicino, 2019). The jurisprudence emphasizes minimal restrictions, allowing even controversial speech, such as hate speech, unless it incites imminent violence (Celeste et al, 2023). Landmark cases like *Brandenburg v. Ohio* (1969) established the "imminent lawless action" test, underscoring the high threshold for limiting speech.

In contrast, European jurisprudence under the ECHR adopts a more balanced approach. The European Court of Human Rights (ECtHR) has ruled in cases like *Handyside v. United Kingdom* (1976) that freedom of expression includes the right to offend, shock, or disturb. However, the ECtHR also permits restrictions to protect

public order, morality, and the rights of others, as seen in *Delfi AS v. Estonia* (2015), where liability for online hate speech was upheld (Celeste et al, 2023).

India's approach to freedom of expression under Article 19(1)(a) of its Constitution is nuanced, allowing reasonable restrictions for public order, decency, and sovereignty. Cases like *Shreya Singhal v. Union of India* (2015) struck down vague provisions of the Information Technology Act, emphasizing the need for clarity in restrictions (Celeste et al, 2023).

China represents a stark contrast, with freedom of expression heavily curtailed under its legal framework (Celeste et al, 2023). The government prioritizes social stability and state security, employing extensive censorship mechanisms. The absence of judicial independence further limits the scope for challenging restrictions.

The rise of social media platforms has complicated the application of freedom of expression. Platforms like Facebook and Twitter operate transnationally, creating a "global public sphere" but also raising questions about private governance (Celeste et al, 2023). As discussed in *The Content Governance Dilemma*, platforms often adopt their own rules, leading to accusations of censorship or laxity. The lack of a unified international standard exacerbates these challenges.

Civil society initiatives, such as Internet Bills of Rights, advocate for embedding human rights standards into platform governance. These efforts aim to reconcile the tensions between universal principles and local adaptations. The concept of "digital constitutionalism" seeks to instill human rights guarantees into the socio-technical architecture of online platforms, fostering a more inclusive and rights-respecting digital environment.

Freedom of expression remains a dynamic and contested right, shaped by diverse legal traditions and socio-political contexts. Comparative jurisprudence reveals the complexities of balancing this right with competing interests, particularly in the digital age. While international frameworks provide a foundation, the path toward a global standard requires multi-stakeholder collaboration, informed by both universal principles and local realities.

6. Limits to Expression: Necessity and Proportionality

Freedom of expression is a fundamental human right, but it is not absolute. International human rights law recognizes that this right can be limited under specific circumstances to protect other rights and societal interests. The principles of necessity and proportionality are central to determining when restrictions on expression are justified. This chapter explores these principles, their application in international and national contexts, and their relevance in the digital age.

The International Covenant on Civil and Political Rights (ICCPR) provides the most widely accepted framework for limiting freedom of expression. Article 19(3) states that restrictions must meet three cumulative criteria, just as legality (restrictions must be provided by law); necessity (Restrictions must be necessary to achieve a legitimate aim); and proportionality (Restrictions must be proportionate to the aim pursued).

Legitimate aims include protecting the rights or reputations of others, national security, public order, public health, or public morals. These principles are echoed in regional human rights instruments, such as the European Convention on Human Rights (ECHR) and the American Convention on Human Rights.

The principle of necessity requires that restrictions on expression address a pressing social need and are essential to achieving a legitimate aim. This means that the restriction must be narrowly tailored to address the specific harm it seeks to prevent. The necessity test requires governments and other actors to demonstrate that the harm caused by unrestricted expression outweighs the benefits of protecting free speech. Vague or overly broad restrictions fail this test, as they risk suppressing legitimate expression (Celeste et al, 2023).

The principle of proportionality ensures that restrictions on expression do not exceed what is required to achieve the legitimate aim. It involves balancing the severity of the restriction against the importance of the protected interest. Proportionality is particularly important in cases involving political speech, which is afforded higher protection under international law. Courts often emphasize that restrictions on political expression must be subject to strict scrutiny.

The ECtHR has developed a robust body of jurisprudence on necessity and proportionality. In *Handyside v. United Kingdom* (1976), the court emphasized that freedom of expression includes the right to offend, shock, or disturb. However, in *Delfi AS v. Estonia* (2015), the court upheld restrictions on online hate speech, finding them necessary and proportionate to protect the rights of others (Celeste et al, 2023).

The U.S. Supreme Court applies a high threshold for limiting expression, particularly political speech. In *Brandenburg v. Ohio* (1969), the court ruled that speech can only be restricted if it incites imminent lawless action. This approach reflects a strong commitment to the principle of proportionality, favoring minimal interference with free speech (Celeste et al, 2023).

India's Supreme Court has emphasized the importance of necessity and proportionality in cases involving restrictions on expression. In *Shreya Singhal v. Union of India* (2015), the court struck down vague provisions of the Information Technology Act, finding them disproportionate to the aim of preventing harm (Celeste et al, 2023).

The rise of social media platforms has complicated the application of necessity and proportionality. Platforms often act as private regulators of speech, creating their own rules for content moderation. While these rules aim to balance free expression with safety, they often lack transparency and accountability.

Automated systems used by platforms to detect harmful content raise concerns about necessity and proportionality. These systems often fail to consider context, leading to over-removal of legitimate speech. For example, posts discussing sensitive topics like breast cancer awareness have been mistakenly flagged as violating nudity policies (Oversight Board 2020).

Civil society advocates have called for greater transparency in content moderation to ensure that restrictions meet the principles of necessity and proportionality. Initiatives like the Santa Clara Principles (ACLU 2018) emphasize the need for clear rules, human oversight, and appeal mechanisms.

The principles of necessity and proportionality are essential safeguards against arbitrary restrictions on freedom of expression. They ensure that limitations are narrowly tailored and balanced against competing interests. However, their application in the digital age presents new challenges, particularly in the context of

private governance by social media platforms. Achieving a fair balance requires multi-stakeholder collaboration, informed by international human rights standards and comparative jurisprudence (Celeste et al, 2023).

Therefore, any limitation on freedom must satisfy tests of necessity and proportionality. These principles ensure that restrictions are not arbitrary and that they minimally impair the right in question. In Brazil, this is evident in rulings on electoral communication, public demonstrations, and artistic expression.

For instance, the STF struck down laws requiring prior authorization for biographies, arguing such demands constitute prior censorship (RE 593.343). However, it has upheld sanctions in cases of misinformation that harm democratic integrity, such as spreading lies during electoral periods.

7. Conclusion

Freedom of expression stands as a foundational right in democratic societies, essential to individual autonomy, civic engagement, and the health of public discourse. Yet, as this essay demonstrates, it is not,and cannot be,an unqualified liberty. The evolution of legal, social, and technological landscapes demands a more nuanced understanding, one that upholds the right to speak while recognizing the legitimate need to protect other core values such as dignity, equality, and public security.

Brazilian jurisprudence, especially through the work of the Federal Supreme Court, offers a compelling model of this balance. From striking down repressive press laws to upholding protections against hate speech and disinformation, the STF has continuously emphasized that expression must serve democracy,not undermine it. These decisions reflect a constitutional maturity that sees rights not as isolated absolutes, but as elements in constant dialogue and tension with one another.

In the digital sphere, the stakes are even higher. Social media platforms now function as global public squares but are governed by private rules and opaque algorithms. This shift has made digital constitutionalism not just a theoretical concern, but a necessary framework for ensuring that fundamental rights are upheld

in virtual spaces. Brazil's Marco Civil da Internet represents a pioneering step in this direction, yet its effective enforcement remains a work in progress.

International comparisons reveal diverse approaches to managing this right,from the maximalist protections in the United States to the more balanced frameworks in Europe and the cautious but increasingly rights-focused jurisprudence emerging in the Global South. These varied models remind us that freedom of expression, while universal in principle, must be interpreted through the lens of each society's legal, cultural, and historical context.

The principles of necessity and proportionality offer essential guardrails for restricting speech in democratic societies. They demand that any limitation be grounded in law, aimed at legitimate objectives, and minimally intrusive. These principles are particularly crucial in an age where automated content moderation and transnational platform governance often bypass democratic accountability.

Ultimately, freedom of expression is not a myth,but it is also not a license. It is a right that thrives when exercised responsibly, defended vigilantly, and limited judiciously. As new challenges arise,from deepfakes and algorithmic manipulation to global disinformation networks,societies must remain committed to a legal and moral framework that protects speech while preventing its abuse. The path forward lies not in absolutism, but in balance: a vision of liberty that is as mindful of its power as it is of its limits.

References

ACLU Foundation et al. 2018. The Santa Clara Principles on Transparency and Accountability in Content Moderation. https://santaclaraprinciples.org.

Celeste, E., Palladino, N., Redeker, D., & Yilma, K. (2023). **The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard.** Palgrave Macmillan. https://doi.org/10.1007/978-3-031-32924-1

Celeste, E., Palladino, N., Redeker, D., & Yilma, K. (2023). *The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard*. Palgrave Macmillan.

Mégret, Frédéric. 2013. **International Human Rights and Global Legal Pluralism: A Research Agenda.** In *Dialogues on Human Rights and Legal Pluralism*, Ius Gentium: Comparative Perspectives on Law and Justice, ed. René Provost e Colleen Sheppard, 69–95. Dordrecht: Springer Netherlands.

Mendes, G. F., & Fernandes, V. O. (2021). *Digital Constitutionalism and Constitutional Jurisdiction: A Research Agenda For The Brazilian Case*. SSRN. https://ssrn.com/abstract=3769947

Meta. 2022a. **Facebook Community Standards.** transparency.fb.com. https://transparency.fb.com/policies/community-standards/. Accessed December 21, 2022.

Oversight Board. 2020. **Breast Cancer Symptoms and Nudity. Case 2020-004-IG-UA.** https://www.oversightboard.com/decision/IG-7THR3SI1

Pollicino, O. (2019). **Judicial Protection of Fundamental Rights in the Transition from the World of Atoms to the World of Bits: The Case of Freedom of Speech.** *European Law Journal*, 25(2), 155–168.

Pollicino, O. (2021). **Judicial Protection of Fundamental Rights on the Internet: A Road Towards Digital Constitutionalism?** Oxford: Hart.
Sarlet, I. W., Marinoni, L. G., & Mitidiero, D. (2021). *Curso de direito constitucional* (10th ed.). Saraiva.

STF (Supremo Tribunal Federal). (2021). *Case Law Compilation: Freedom of Speech*. Brazilian Federal Supreme Court. http://www.stf.jus.br

Tallentyre, S. G. (1906). *The Friends of Voltaire*. A referência ao Twitter de 2020 no documento é a seguinte:
Twitter**. 2020. The Twitter Rules (Version of 28 October 2020). Platform Governance Archive.** Disponível em: https://github.com/PlatformGovernanceArchive/pga-corpus/tree/main/Versions/PDF/Twitter. Acessado em: 16 de junho de 2023.

5. THE MANDATORY DISCLOSURE OF GEOLOCATION DATA BY INTERNET SEARCH PROVIDERS: A CONSTITUTIONAL ANALYSIS OF GEOFENCE WARRANTS UNDER BRAZILIAN LAW

https://doi.org/10.36592/9786554603065-04

Yury Rufino Queiroz

Abstract
The social dynamics shaped by contemporary hyperconnectivity have fostered the emergence of digital investigative techniques that reshape the traditional contours of criminal procedure. Among these innovations, particular attention is drawn to geofence warrants,judicial orders directed at digital service providers compelling the disclosure of geolocation data from multiple users, even in the absence of prior individualization. This investigative modality challenges constitutional boundaries related to the protection of privacy, personal data, and the guarantees inherent to due process of law, and has become the subject of judicial scrutiny by digital platforms operating in Brazil. This article critically examines the feasibility of applying this technique within the Brazilian legal system, analyzing its normative and jurisprudential framework in light of the system of safeguards established by the 1988 Constitution. The study demonstrates that, although potentially useful for criminal investigations, geofence warrants must be subjected to strict criteria of legality, necessity, and proportionality, so as to prevent their transformation into tools of broad and indiscriminate surveillance. In conclusion, the article proposes interpretive guidelines aimed at balancing investigative efficiency and constitutional safeguards.
Keywords: geofence warrants; fundamental rights; digital criminal procedure; data protection; proportionality; geolocation; surveillance.

1. Introduction

The digital age has profoundly transformed social interactions and, consequently, methods of criminal prosecution. New investigative techniques have emerged, promising greater efficiency but also raising serious concerns about their compatibility with fundamental rights. Among these innovations, geofence warrants stand out. Initially applied in the United States, these warrants allow authorities to compel internet search providers to disclose geolocation data from all devices present in a specific area during a defined period, for criminal investigative purposes. Because they do not require individualized suspicion at the outset, such measures

raise significant constitutional concerns and have faced judicial scrutiny by digital platforms operating in Brazil.

This article examines the compatibility of this technique with the 1988 Brazilian Federal Constitution, especially regarding the rights to privacy (Article 5, X), intimacy, personal data protection (Article 5, LXXIX), and due process of law (Article 5, LIV). It questions whether, due to their breadth and abstraction, such warrants constitute mass surveillance, violating the reasonable expectation of privacy and the constitutional requirement of judicial specificity.

The main objective is to critically evaluate the constitutionality of geofence warrants under Brazilian law. Specific goals include: (i) defining and explaining the technique; (ii) identifying the applicable legal framework; (iii) analyzing relevant jurisprudence in the U.S. and Brazil; (iv) applying the principle of proportionality as a test of legitimacy; and (v) proposing interpretive criteria to align the technique with fundamental rights.

This research is warranted by the growing use of geofence warrants and the urgent need to define legal limits for tools that, while potentially useful to criminal investigations, may endanger essential liberties. Avoiding both uncritical acceptance and categorical rejection demands a mature and well-grounded constitutional reflection.

Methodologically, the study adopts a qualitative, deductive, and analytical-critical approach, drawing on doctrinal and jurisprudential review. It begins with the Brazilian constitutional framework, engages in comparative legal analysis, and seeks to define the risks and boundaries of non-individualized digital investigations within the context of algorithmic surveillance.

The article is structured in four parts: this introduction which is followed by a section on the theoretical and normative foundations of data protection and criminal procedure, including national and comparative case law; a third section offers interpretive parameters for the technique's use in Brazil; and a final section presents conclusions and legal policy proposals.

2. Theoretical and Normative Foundations of Data Protection and Criminal Investigation

A constitutional analysis of geofence warrants requires a careful examination of the theoretical and normative foundations that simultaneously inform the protection of personal data and the legitimacy of state investigative activity. This inquiry necessarily encompasses the right to privacy, the emergence of data protection as an autonomous fundamental right, the sensitivity of location information, the contours of due process of law, and the broader national and international legal framework.

The right to privacy, enshrined in Article 5, X of the 1988 Brazilian Constitution, has evolved from the classical notion of the "right to be let alone" to the principle of informational self-determination,the right to control the flow of one's own data. As Rodotà emphasizes, it serves as an essential tool for safeguarding individual autonomy in the face of surveillance and social control. Intimacy, in turn, refers to the most private sphere of personality, requiring even more intense protection. Location data directly impact both dimensions, as they reveal habits, movements, personal relationships, and intimate convictions (Rodotà, 2008).

The Constitution guarantees the inviolability of intimacy, private life, and data secrecy (Art. 5, X, XII, and LXXIX). While data secrecy is a relevant manifestation of a fundamental right linked to human dignity, both legal doctrine and jurisprudence recognize that this right is not absolute (Supremo Tribunal Federal [STF], 2000). It may be restricted, exceptionally, when doing so is essential to the protection of the public interest, particularly in the context of investigations into serious and complex crimes (Sarlet et al, 2016).

In such cases, due process of law (Art. 5, LIV of the Constitution) requires that any restriction of rights be supported by a duly reasoned judicial decision (Art. 93, IX). Regarding access to digital data, the Marco Civil Law of the Internet in Brazil (Law No. 12.965/2014) obliges providers to disclose connection and application access records, as well as personal data and the content of private communications (Art. 10, §1), upon a properly substantiated judicial order (Art. 22). Meanwhile, the LGPD - *Lei Geral de Proteção de Dados Pessoais*, General Data Protection Law (Law No.

13.709/2018) establishes key principles,such as purpose limitation, necessity, transparency, and accountability,that also apply to criminal investigations, albeit with certain reservations (Art. 4, III).

In this context, the STF - *Supremo Tribunal Federal*, Supreme Federal Court, in its judgment in ADI No. 6387, established important parameters for the legal treatment of personal data. The Court declared the unconstitutionality of Provisional Measure No. 954/2020, which had authorized telecommunications companies to share user data with the IBGE - *Instituto Brasileiro de Geografia e Estatística*, Brazilian Institute of Geography and Statistics, due to the absence of minimum legal safeguards. The Court affirmed that there is no constitutional distinction between data in transit and data at rest, recognizing an autonomous fundamental right to data protection. It further held that any data processing must comply cumulatively with the principles of adequacy, necessity, and proportionality, grounded in clear legal provisions and subjected to institutional oversight (Supremo Tribunal Federal [STF], 2020a).

This interpretive shift consolidated the understanding that the rights to privacy and data protection contain both a subjective dimension,as individual guarantees against interference,and an objective dimension, imposing duties upon the state to ensure institutional protection. As Minister Gilmar Mendes succinctly put it, these rights unfold into negative liberties and positive obligations, intrinsically linked to human dignity and informational self-determination (Supremo Tribunal Federal [STF], 2020b).

Accordingly, the constitutional assessment of geofence warrants must focus on their compatibility with the constitutional principles governing criminal procedure, especially with respect to the existence of probable cause, specific delimitation of the object of search, and robust judicial justification for any restriction on fundamental rights.

The United States' pioneering experience with geofence warrants provides valuable insights into the constitutional challenges associated with this technique. From a normative standpoint, these warrants are evaluated under the Fourth Amendment to the U.S. Constitution, which prohibits unreasonable searches and seizures and requires probable cause for warrant issuance. This safeguard seeks to

curb abusive investigative practices, and it is historically rooted in the rejection of general warrants, which were widely used under British colonial rule(Kerr 2024).

In this regard, the U.S. Supreme Court, in Stanford v. Texas, 379 U.S. 476 (1965), emphasized that the Fourth Amendment was in part a response to the issuance of such arbitrary warrants, which conferred broad powers upon authorities to arrest and search individuals without objective criteria concerning the nature of the offense or the identity of the suspect (U.S. Supreme Court, 1965).

In Carpenter v. United States (2018), the Court held that accessing historical location data constitutes a search subject to Fourth Amendment protections, due to the reasonable expectation of privacy involved. This decision underscores the gravity of privacy risks associated with continuous geographic tracking (U.S. Supreme Court, 2018).

Critics of geofence warrants argue that by compelling the disclosure of data from all devices present in a certain area and timeframe, without identifying specific suspects, these warrants are excessively broad and indistinct, resembling the very general searches that constitutional safeguards were designed to prevent. Moreover, they question whether mere physical presence at the scene of a crime constitutes sufficient probable cause to justify the collection and subsequent identification of users (Fussel, 2021).

Indeed, geofence warrants invert the traditional logic of criminal investigation: they begin with a mass of anonymized data from unknown individuals located in a particular area, in order to retrospectively identify potential suspects. However, when properly constrained to a specific and well-defined criminal investigation, and supported by rigorous justification, this does not necessarily amount to arbitrary state action.

Therefore, although geofence warrants do not begin with individualized suspicion, their constitutionality does not hinge on their conceptual structure. What matters is whether a sufficient degree of specificity allows for the clear delimitation of the object of the search, thereby distancing it from indiscriminate investigative practices. Nonetheless, abstract formulations and imprecise boundaries must be rejected, as they risk legitimizing speculative investigative efforts, known as fishing expeditions (Silva et al., 2022).

Even in the United States, the issue remains controversial. Cases such as United States v. Chatrie illustrate a growing judicial demand for greater rigor, including precise spatial and temporal limitations, sequential review stages for data analysis, and a heightened standard for probable cause (U.S. District Court for the Eastern District of Virginia, 2019).

Conversely, United States v. Rhine, which was decided in January 2023 by the U.S. District Court for the District of Columbia, the Court upheld the legality of geofence warrants in the context of the investigation into the January 6, 2021, attack on the U.S. Capitol (Schmitz, 2023).

In Brazil, the Superior Courts have consistently rejected investigative measures with vague or overly broad scopes, especially those lacking a concrete connection to the facts under investigation[1]. Nevertheless, by structuring geofence warrants within a constitutionally compatible investigative model**,** the Superior Court of Justice (STJ)**,** in RMS No. 61.302/RJ**,** has consolidated a position favorable to their constitutionality. The Court emphasized the relevance of the technique for solving serious crimes and recognized objective criteria for assessing the legitimacy of such warrants,notably the requirement for robust judicial reasoning**,** precise delimitation and proportionality between the measure and the investigative goal (Superior Tribunal de Justiça [STJ], 2022).

It is worth noting that the jurisprudence of STJ on this issue has largely emerged from mandamus actions filed not by individuals directly affected by the data orders, but by digital platforms reluctant to comply with judicial demands.

Although the Supreme Federal Court has not yet declared directly on the constitutionality of geofence warrants, its pending decision in General Repercussion Theme No. 1148, which concerns the legality of keyword warrants, may offer relevant insights into the constitutional parameters applicable to analogous reverse search techniques[2].

---

[1] Examples include decisions rendered by the Brazilian Federal Supreme Court (STF) in HC 144.159/PR**,** HC 163.461/PR, and ARE 1.535.677, as well as by the Superior Court of Justice (STJ) in AgRg in RHC 195.496/PR, reported by Justice Ribeiro Dantas, with the opinion delivered by Justice Joel Ilan Paciornik, 5th Panel, adjudicated on April 8, 2025, published in the *DJEN* on April 24, 2025.

[2] As of the date this article is written (May 17, 2025), the judgment of Extraordinary Appeal No. 1.301.250/RJ, listed under General Repercussion Theme No. 1148, remains suspended due to a request for review by Justice Gilmar Mendes. Thus far, two votes have been cast in favor of granting

This is the subject of the following section.

3. Interpretive Parameters for the Use of Geofence Warrants in Brazil: A Critical Analysis

The parameters that guide the assessment of the constitutional legitimacy of geofence warrants lie at the core of the debate surrounding their compatibility with the framework for the protection of fundamental rights. Although such orders pertain to the access of static location data from electronic devices, this information falls within the protective sphere of personal data, thereby requiring adequate judicial reasoning pursuant to Article 22 of the Marco Civil Law of the Internet in Brazil (Law No. 12.965/2014) and its respective provisions.

Although Law No. 12.965/2014 does not expressly require prior individualization of the data subject at the time of the request, the high degree of intrusion into the individual's private sphere necessitates adherence to strict criteria. The balancing of fundamental rights demands the existence of probable cause and the specific delimitation of the object of the measure**,** both of which must be supported by a duly substantiated judicial decision—a prerequisite for the legitimacy of state action in criminal investigations.

In this context, probable cause corresponds to the demonstration of objective, concrete, and preexisting elements that justify the necessity of the invasive measure in light of the ongoing criminal investigation. Judicially compelled measures cannot be based on vague suspicions or speculative reasoning. The requesting authority must demonstrate the existence of a criminal offense, articulate the relevance of the requested data to the offense under investigation, and provide a precise justification for the geographic and temporal scope of the warrant, including the expected level of specificity.

The specific delimitation of a geofence warrant is inseparable from its legality and must be directly linked to the criminal event under investigation. Such delimitation requires a clear, precise, and technically justified description of the

---

the appeal**,** by Justice Rosa Weber and Justice André Mendonça, and two votes against, by Justices Alexandre de Moraes and Cristiano Zanin.

geographic area covered by the warrant, along with the definition of a strict time interval, with the goal of preventing arbitrary or generic investigations and ensuring compliance with the principles of proportionality, legality, and necessity that conduct criminal procedure in a democratic society.

Finally, the constitutional assessment of geofence warrants requires the rigorous application of the principle of proportionality, especially in light of their potential impact on privacy, intimacy, and personal data protection. As a canonical limitation on state power, this principle unfolds into the dimensions of suitability**,** necessity**,** and proportionality in a strict sense, offering concrete standards for evaluating the legitimacy of investigative measures.

Suitability requires that the chosen measure should be appropriate to achieve a legitimate objective. In the case of reverse location searches, the identification of devices within areas of investigative interest may, effectively, aid in the investigation of serious crimes. However, the effectiveness of the technique depends on the reliability of geolocation data, which is subjected to technological, environmental, and operational interference. False positives and negatives can undermine the probative value of the data and increase the risk of judicial error.

Necessity, or requirement of using the least rights-restrictive means, mandates that any restriction of fundamental rights be minimal in light of the goal pursued. Because this technique affects the privacy of a wide range of non-suspect individuals, its use must be exceptional and subsidiary. Judicial authorization must be based on robust reasoning, including a clear demonstration of the unavailability or exhaustion of less intrusive alternatives.

Proportionality in a strict sense requires a substantive balancing between the impact of the measure on individual rights and the importance of the legal interest at stake. Accordingly, its use must be limited to the investigation of serious and complex offenses. The Superior Court of Justice (STJ), in RMS No. 73398/AM, rejected the issuance of a geofence warrant for the investigation of a less serious crime such as defamation. Moreover, although the seriousness of the offense is a relevant factor, it does not, by itself, justify invasive and generic investigative measures. The request for sensitive data must be directly connected to the alleged offense and take into account: (i) the estimated number of affected individuals, (ii) the geographic and

temporal scope of the data collection, (iii) the sensitivity of the information, and (iv) the procedural safeguards adopted—such as anonymization and the discarding of irrelevant data.

All of these parameters must be subjected to strict and effective judicial oversight to ensure that geofence warrants do not result in generalized surveillance or give rise to a chilling effect on constitutionally protected freedoms,such as freedom of association and expression,ultimately threatening the integrity of the democratic public sphere (U.S. Supreme Court, 1965a).

4. Conclusion

The analysis undertaken throughout this article has revealed the complex constitutional challenges posed by geofence warrants in Brazil. Theoretical, jurisprudential, and principled inquiry has shown that, although this technique may contribute to the elucidation of serious crimes, its adoption requires strict safeguards, lest it undermine core fundamental rights such as privacy, intimacy, personal data protection, and due process of law.

The main conclusion points to a model of conditional constitutionality: while the use of geofence warrants may find legal support in the 1988 Brazilian Constitution (Article 5, X, XII, and LXXIX), their legitimacy is contingent upon strict compliance with interpretive parameters that mitigate their invasive nature. In the absence of adequate safeguards, these instruments may become mechanisms of unchecked surveillance, incompatible with the values of Brazil's constitutional order.

Given the relative nature of fundamental rights, any breach of data confidentiality must be supported by a duly substantiated judicial order. Such judicial reasoning must clearly demonstrate probable cause and the specific delimitation of the object of the measure, directly tied to the investigation of a specific criminal offense.

Applying a proportionality test between investigative interest and fundamental rights, the use of such warrants must be restricted to serious criminal cases, with a clear demonstration of subsidiarity and precise geographic and temporal delimitation, so as to avoid indiscriminate data sweeps.

The procedure should follow progressive stages, beginning with the disclosure of aggregated and anonymized data. Individual identification should only be authorized by a new judicial order, based on concrete evidence of involvement. The confidentiality and disposal of data pertaining to innocent third parties must be guaranteed, along with notification of affected individuals.

Given the transnational nature of digital platforms, the international harmonization of safeguards is essential, including the definition of clear protocols for cooperation with law enforcement authorities, while preserving the rights of users.

Finally, the limitations of this study must be acknowledged, especially in light of the evolving nature of the topic and the scarcity of consolidated jurisprudence in Brazil. Future research may explore the European experience, conduct empirical studies on the effectiveness of such warrants, assess the application of anonymization and filtering technologies in investigative contexts, and further analyze the potential emergence of a chilling effect on constitutionally protected freedoms.

Ongoing critical oversight by academia, civil society, and legal institutions will be decisive in ensuring that technological advances do not erode the foundational guarantees of the Constitutional Rule of Law.

References

Brasil. Supremo Tribunal Federal. (2020a). **Ação Direta de Inconstitucionalidade n. 6.387 MC-Ref** [Direct Action of Unconstitutionality No. 6.387 MC-Ref] (Relatora: Ministra Rosa Weber). Diário da Justiça Eletrônico. https://portal.stf.jus.br/processos/

Brasil. Supremo Tribunal Federal. (2020b). **Voto do Ministro Gilmar Mendes na Ação Direta de Inconstitucionalidade n. 6.387 MC-Ref/DF** [Minister Gilmar Mendes' Vote in Direct Action of Unconstitutionality No. 6.387 MC-Ref/DF]. Diário da Justiça Eletrônico. Retrieved May 11, 2025, from https://www.stf.jus.br/portal/jurisprudencia/listarJurisprudencia.asp?s1=6387&base=baseAcordaos

Brasil. Supremo Tribunal Federal. (2000). **Mandado de Segurança n. 23.452/RJ** [Writ of Security No. 23.452/RJ] (Relator: Ministro Celso de Mello). Diário da Justiça. https://jurisprudencia.stf.jus.br/pages/search/sjur100956/false

Fussell, S. (2021, August 27). An explosion in geofence warrants threatens privacy across the US. Wired. Retrieved May 16, 2025, from https://www.wired.com/story/fbi-google-geofence-warrant-january-6/

Kerr, O. S. (2024, August 13). The Fifth Circuit shuts down geofence warrants—and maybe a lot more [Blog]. The Volokh Conspiracy. https://reason.com/volokh/2024/08/13/fifth-circuit-shuts-down-geofence-warrants-and-maybe-a-lot-more/

Rodotà, S. (2008). A vida na sociedade da vigilância: A privacidade hoje (L. Repa, Trans.). Editora Contexto. (Original work published 2008)

Sarlet, I. W., Marinoni, L. G., & Mitidiero, D. (2016). Curso de direito constitucional. Saraiva.

Schmitz, A. (2023, January 25). U.S. federal judge denies motion to suppress Jan. 6 location data. Lawfare. https://www.lawfaremedia.org/article/us-federal-judge-denies-motion-suppress-jan-6-location-data

Silva, V. G. da, Silva, P. B. M. e, & Rosa, M. da. (2022). Fishing expedition e encontro fortuito na busca e na apreensão: Um dilema oculto do processo penal (2nd ed.). Editora Emais.

U.S. District Court for the Eastern District of Virginia. (2019). United States v. Chatrie (No. 3:19-cr-00130). Retrieved May 15, 2025, from https://www.courtlistener.com/docket/16215471/united-states-v-chatrie/

U.S. Supreme Court. (2018). Carpenter v. United States, 585 U.S. __. https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf

U.S. Supreme Court. (1965a). Lamont v. Postmaster General, 381 U.S. 301. https://supreme.justia.com/cases/federal/us/381/301/

U.S. Supreme Court. (1965b). Stanford v. Texas, 379 U.S. 476. https://caselaw.findlaw.com/court/us-supreme-court/379/476.html

Part II - State Functions vs Private Governance

6. JUDICIAL RANSOMWARE AND DEMOCRATIC CONTINUITY WITHIN BRAZIL'S DIGITAL ELECTORAL PROCESSES

https://doi.org/10.36592/9786554603065-05

Celso Reic Urbieta

Abstract:

The increasing digitization of judicial systems has created unprecedented vulnerabilities that threaten the foundations of democratic governance. This essay examines how ransomware attacks, specifically targeting judicial institutions, can compromise democratic continuity, particularly during critical electoral periods when courts play essential roles in resolving disputes related to social media regulation and electoral integrity. Drawing on theoretical frameworks from surveillance capitalism, algorithmic governance, and cybersecurity practices, combined with empirical analysis of recent cyberattacks against Brazilian judicial institutions, this analysis demonstrates that judicial ransomware represents a new form of hybrid threat that transcends traditional cybersecurity concerns to become a fundamental challenge to democratic resilience. The essay argues that protecting judicial digital infrastructure is not merely a technical issue but a constitutional imperative for maintaining democratic legitimacy in the digital age.
Keywords: judicial cybersecurity, ransomware attacks, democratic continuity, electoral integrity, digital governance, social media regulation, Brazilian judiciary.

1. Introduction

The intersection of cybersecurity and democratic governance has emerged as one of the most critical challenges facing contemporary democratic states. As Lessig (2006) presciently observed, "code is law" in the digital age, meaning that the technical architectures governing our digital systems function as powerful forms of regulation that can either support or undermine democratic institutions. This observation has taken on new urgency as judicial systems worldwide have become increasingly dependent on digital infrastructure, creating novel vulnerabilities that malicious actors can exploit to disrupt the administration of justice itself.

The Brazilian experience provides a particularly compelling case study for understanding these challenges. Between November 2021 and May 2022, the Supreme Federal Court (STF) suffered 2,434,627 critical cyberattacks representing

93.8% of all attacks received during this seven-month period (JOTA, 2022). More recently, in September 2024, coordinated cyberattacks targeted the STF, Federal Police, and the National Telecommunications Agency (Anatel) following Justice Alexandre de Moraes's decision to suspend the X platform (formerly Twitter) in Brazil. These attacks rendered the STF's systems inoperative for ten minutes, demonstrating how cyberattacks can directly disrupt judicial functions during politically sensitive periods (O Globo, 2024).

The concept of "resurgent states" in the context of social media regulation reflects governments' attempts to reassert sovereignty over digital platforms that have traditionally operated beyond effective regulatory control. However, this reassertion of state power faces a fundamental paradox: the very digital infrastructure that enables states to regulate social media platforms also creates new attack surfaces that can be exploited to undermine state capacity. The Brazilian case demonstrates this clearly: during the 2022 electoral period, a group calling itself "BolsonaroCyberMafia" conducted systematic defacement attacks against websites of opposition political parties and candidates, replacing content with pro-Bolsonaro messages and electoral propaganda (Correio Braziliense, 2022).

This essay argues that judicial ransomware represents an emerging form of hybrid threat that requires a fundamental reconceptualization of cybersecurity as a democratic imperative rather than merely a technical challenge. The analysis proceeds through four main sections: first, an examination of the theoretical foundations linking cybersecurity to democratic governance; second, an analysis of the specific vulnerabilities created by the digitization of judicial systems, illustrated through Brazilian case studies; third, an exploration of how ransomware attacks can compromise democratic continuity; and fourth, a discussion of the implications for policy and institutional design.

Theoretical Foundations: Cybersecurity as Democratic Infrastructure

The theoretical foundation for understanding judicial ransomware as a threat to democratic governance begins with Lawrence Lessig's foundational insight that

"code is law." In "Code and Other Laws of Cyberspace," Lessig (2006) demonstrates that the technical architectures of digital systems function as regulatory mechanisms that can be as powerful as traditional legal frameworks. This insight proves particularly relevant to judicial systems, where digital infrastructure increasingly mediates fundamental democratic processes, from case management and evidence handling to the publication of judicial decisions and the conduct of virtual hearings. The Brazilian experience illustrates this principle with stark clarity: when cyberattacks rendered the STF's systems inoperative for ten minutes in September 2024, they effectively suspended the Court's ability to perform its constitutional functions during a period of intense political controversy over social media regulation. The precise timing of these attacks immediately following a judicial decision to suspend a major social media platform, demonstrates how technical disruptions can serve as sophisticated tools of political pressure and institutional intimidation.

Julie Cohen's (2019) analysis in "Between Truth and Power: The Legal Constructions of Informational Capitalism" provides crucial theoretical context for understanding how traditional legal processes have become embedded within digital architectures that are subject to the same market logics and vulnerabilities that characterize other forms of digital infrastructure. This embedding creates what Frank Pasquale (2015) identifies in "The Black Box Society" as opaque systems where the technical complexity of digital judicial infrastructure makes it difficult for legal professionals, policymakers, and citizens to understand how these systems operate and where their vulnerabilities lie. The Brazilian case demonstrates this opacity concretely: the STF's massive increase in cybersecurity spending from R$ 10.8 million in 2023, a six-fold increase, occurred with little public debate about its implications for judicial independence or democratic governance.

Shoshana Zuboff's (2019) comprehensive analysis of "surveillance capitalism" in "The Age of Surveillance Capitalism" provides an additional theoretical framework for understanding how the commodification and concentration of digital infrastructure creates systemic vulnerabilities that can be exploited by malicious actors. While Zuboff's primary focus is on the extraction of behavioral data for commercial purposes, her analysis reveals how the

concentration of digital infrastructure in the hands of a small number of technology companies creates single points of failure that can be exploited for various purposes, including the disruption of democratic processes. The Brazilian case demonstrates this principle through the massive scale of attacks against the STF, 2.4 million critical attacks in seven months, suggesting a sustained campaign designed to test and potentially overwhelm the Court's defensive capabilities rather than achieve immediate financial gain.

The concept of hybrid threats, which combine conventional and unconventional methods to achieve strategic objectives, provides a crucial framework for understanding how ransomware attacks on judicial systems can serve broader political purposes beyond immediate monetary rewards. Thilini Herath, Prashant Khanna, and Monjur Ahmed's (2022) systematic literature review of cybersecurity practices for social media users identifies platforms as particularly vulnerable to hybrid threats because of their role in mediating political communication and their technical complexity, which creates multiple attack vectors for malicious actors. The Brazilian electoral period of 2022 provides compelling empirical evidence of how cyberattacks can function as hybrid threats in practice.

The systematic targeting of opposition political websites by the "BolsonaroCyberMafia" group demonstrates how technical attacks can be coordinated with broader political campaigns to influence electoral outcomes. These attacks went beyond simple vandalism to constitute a sophisticated form of digital voter suppression. The group targeted high-profile politicians including Eduardo Suplicy, Paulo Teixeira, Natália Bonavides, and Benedita da Silva, as well as the official PT party website and Lula's campaign website. The content of these attacks reveals their political motivation clearly with messages. These messages demonstrate how cyberattacks can function as tools of political communication and voter manipulation, replacing legitimate political discourse with propaganda designed to influence electoral outcomes.

Safiya Noble's (2018) analysis in "Algorithms of Oppression" demonstrates how the strategic deployment of technical disruptions can have disproportionate impacts on democratic processes, particularly when they target vulnerable

populations or critical moments in democratic cycles. The September 2024 attacks on Brazilian judicial institutions, timed to coincide with a controversial decision about social media regulation, exemplify how cyberattacks can be weaponized to maximize their disruptive impact on democratic governance. The coordination of attacks against the STF, Federal Police, and Anatel following Justice Moraes's decision shows how judicial cybersecurity has become intertwined with broader questions of digital sovereignty and platform regulation.

The digital transformation of judicial systems has accelerated dramatically in recent years, driven by efficiency considerations, cost pressures, and the COVID-19 pandemic's requirement for remote proceedings. Brazil's experience illustrates both the benefits and risks of this transformation. The country has been a pioneer in electronic voting systems and digital judicial processes, but this early adoption has also created significant vulnerabilities that malicious actors have learned to exploit systematically. Carlos Blanco de Morais, Gilmar Ferreira Mendes, and Thomas Vesting's (2023) comprehensive analysis in "The Rule of Law in Cyberspace" reveals how this digital transformation has created new categories of legal and constitutional questions about the relationship between technical infrastructure and judicial independence. The Brazilian case demonstrates these challenges concretely: when the STF's systems were attacked in September 2024, the Court faced not only a technical problem but a constitutional crisis about its ability to maintain independence in the face of digital intimidation. The vulnerability of judicial systems to ransomware attacks becomes particularly acute during electoral periods, when courts play essential roles in maintaining democratic continuity. The Brazilian experience during the 2022 elections illustrates how cyberattacks can target the entire ecosystem of democratic institutions, from electoral courts to political party websites to individual candidate platforms. The attempted attacks on the TSE on election day itself, while unsuccessful, demonstrate the strategic thinking behind these campaigns. By targeting the electoral authority responsible for vote counting and electoral oversight, these attacks aimed to create doubt about the legitimacy of the electoral process itself, even when they did not achieve their immediate technical objectives.

The broader context of the 2022 electoral period included a 50% increase in

cyberattacks against Brazilian websites during the electoral period, suggesting that the heightened political tensions created opportunities for various actors to exploit democratic vulnerabilities for their own purposes. This pattern reveals how electoral periods create windows of vulnerability that extend beyond the immediate targets of attacks to affect the entire democratic ecosystem.

While traditional ransomware attacks are primarily motivated by financial gain, the Brazilian experience demonstrates how ransomware targeting judicial systems during electoral periods can serve broader strategic objectives that extend far beyond immediate monetary rewards. These attacks can be understood as a form of what Cathy O'Neil (2016) terms "weapons of math destruction" in her analysis of algorithmic systems that can cause widespread social harm while maintaining plausible deniability about their true purposes and effects.

The strategic value of judicial ransomware lies in its ability to create multiple forms of disruption simultaneously. At the most immediate level, such attacks can prevent courts from functioning normally, delaying or preventing the resolution of time-sensitive legal disputes. The September 2024 attacks on the STF, which occurred immediately after a controversial decision about social media regulation, demonstrate how even brief disruptions can send powerful messages about the consequences of judicial decisions that displease certain political actors. At a deeper level, these attacks can undermine public confidence in the reliability and security of democratic institutions, creating lasting damage to democratic legitimacy that extends far beyond the immediate technical disruption.

Timnit Gebru's (2023) analysis in "Data Conscience: Algorithmic Siege on Our Humanity" provides insights into how technical attacks on digital infrastructure can serve broader information warfare objectives. The Brazilian experience demonstrates how ransomware attacks on judicial systems can be understood as part of broader hybrid threat campaigns that combine technical disruption with information manipulation to achieve strategic political objectives. The information warfare dimension of judicial ransomware operates on multiple levels. At the most direct level, attacks that prevent courts from functioning normally can be used to support narratives about the incompetence or illegitimacy of democratic institutions. The timing of the September 2024 attacks on the STF, immediately

following a controversial decision about social media regulation, allowed critics to argue that the Court was unable to protect itself and therefore unfit to regulate digital platforms.

The constitutional implications of ransomware attacks on judicial systems are profound and largely unexplored in existing legal scholarship. The Brazilian experience provides concrete examples of how cyberattacks can raise fundamental questions about the continuity of democratic governance and the rule of law. Cohen's analysis of the legal constructions of informational capitalism provides a framework for understanding how technical disruptions of judicial infrastructure can constitute attacks on the constitutional order itself. When the September 2024 ransomware attacks prevented the STF from functioning normally, they effectively suspended the operation of constitutional rights and democratic processes, even if only briefly.

The temporal dimension of these attacks is particularly important from a constitutional perspective. Electoral calendars and statutory deadlines create time-sensitive requirements for judicial action that cannot be easily postponed or rescheduled. The attempted attacks on the TSE during the 2022 election day, while unsuccessful, demonstrate how cyberattacks timed to coincide with critical democratic moments can threaten the constitutional order even when they do not achieve their immediate technical objectives.

The Brazilian federal system creates particular challenges for coordinating cybersecurity responses across different levels of government and different branches of the judiciary. The experience of regional courts illustrates these challenges and their implications for democratic governance. The Regional Labor Court of Rio Grande do Sul (TRT-RS) reported blocking 135,000 malicious messages between November and December 2024, demonstrating that cyberattacks are not limited to high-profile federal institutions but affect the entire judicial system. The TRT-17 in Espírito Santo suffered a significant cyber incident in February 2022 that created “voluminous work demands" in the post-incident period, illustrating how cyberattacks can have lasting effects on judicial operations beyond the immediate disruption.

These regional experiences highlight the uneven distribution of cybersecurity

capabilities across the Brazilian judicial system. While the STF can afford to increase its cybersecurity budget six-fold, smaller regional courts may lack the resources to implement comprehensive defensive measures, creating vulnerabilities that can be exploited by attackers seeking to disrupt the broader judicial system. The National Council of Justice (CNJ) has attempted to address these coordination challenges through Resolution 396/2021, which establishes a national strategy against cyberattacks on the judiciary. However, implementation remains uneven, and the challenge is to balance the need for coordinated defense with the independence and autonomy of different judicial institutions.

The empirical analysis of Brazilian cyberattacks reveals several critical patterns that have broader implications for democratic societies worldwide. First, the scale and persistence of attacks, 2.4 million critical attacks against the STF in seven months, demonstrate that judicial institutions face sustained campaigns rather than isolated incidents. This pattern suggests that attackers view judicial systems as high-value targets worth sustained effort and investment, indicating that the threat will likely intensify rather than diminish over time. Second, the timing of attacks, coordinated with controversial judicial decisions and electoral periods, shows how cyberattacks can function as sophisticated tools of political pressure and institutional intimidation. Third, the coordination of attacks across multiple institutions, STF, Federal Police, Anatel ,demonstrates the strategic sophistication of these threats and their potential to disrupt entire governmental ecosystems rather than isolated institutions.

The policy implications of this analysis suggest that democratic societies need to fundamentally reconceptualize judicial cybersecurity as a constitutional imperative rather than a technical challenge. The Brazilian experience shows that this reconceptualization requires massive increases in resources, the STF's six-fold increase in cybersecurity spending represents just one institution's response to escalating threats. However, this also raises critical questions about the sustainability and democratic accountability of such investments. The scale of investment required raises questions about the sustainability of purely defensive approaches to judicial cybersecurity and the potential for such investments to compromise other judicial functions or create dependencies on private

cybersecurity companies that could themselves become sources of vulnerability.

The Brazilian context provides particularly relevant insights for understanding these challenges in federal democratic systems. The country's experience with electronic voting, its federal judicial structure, and its ongoing efforts to regulate social media platforms during electoral periods make it a critical case study for other democracies facing similar challenges. The uneven distribution of cybersecurity capabilities across different levels of government and different judicial institutions highlights the need for new forms of institutional coordination that can address the transnational and interconnected nature of cyber threats while respecting traditional principles of judicial independence and federalism.

Looking forward, the threat of judicial ransomware is likely to grow as democratic societies become increasingly dependent on digital infrastructure for core governmental functions. The Brazilian experience suggests that this threat will become more sophisticated and more strategically targeted, requiring new forms of institutional response that can balance security needs with democratic principles. The challenge for democratic societies is to develop cybersecurity capabilities that can protect judicial institutions without undermining the principles of transparency, accountability, and independence that are essential to democratic governance.

Conclusion

This essay has demonstrated that ransomware attacks targeting judicial systems during electoral periods represent a fundamental threat to democratic continuity that extends far beyond traditional cybersecurity concerns. The Brazilian experience provides compelling empirical evidence through documented cases,including 2.4 million critical attacks against the STF in seven months, the coordinated September 2024 attacks following the X platform suspension, and the systematic "BolsonaroCyberMafia" campaign during the 2022 elections, that judicial ransomware constitutes a new form of hybrid threat requiring fundamental reconceptualization of cybersecurity as democratic infrastructure.

The theoretical frameworks of Lessig, Zuboff, Pasquale, Cohen, and others reveal how cybersecurity threats to judicial systems should be understood as

attacks on the constitutional infrastructure of democratic governance rather than merely technical problems. The Brazilian case study demonstrates concretely how the security and reliability of digital judicial infrastructure has become a prerequisite for maintaining the rule of law in democratic societies, with technical disruptions capable of suspending constitutional governance even briefly.

The empirical patterns identified sustained campaigns rather than isolated incidents, strategic timing coordinated with controversial decisions and electoral periods, and sophisticated coordination across multiple institutions, indicate that this threat will intensify as democratic societies become increasingly dependent on digital infrastructure. The Brazilian experience suggests that addressing these challenges requires new forms of institutional response that can balance security needs with democratic principles while developing cybersecurity capabilities that protect judicial institutions without undermining transparency, accountability, and independence.

As the Brazilian case ultimately demonstrates, in the digital age, the security of democratic institutions and the security of digital infrastructure have become indistinguishable. Protecting democracy now requires protecting the code that increasingly governs democratic processes, making cybersecurity not just a technical necessity but a democratic imperative for the twenty-first century.

References

Bietti, E. (2023). A Genealogy of Digital Platform Regulation**. Georgetown Law Technology Review**, 7(1), 1-68.

Cohen, J. E. (2019). **Between Truth and Power: The Legal Constructions of Informational Capitalism.** Oxford University Press.

Correio Braziliense. (2022, October 13). **Hackers atacam sites petistas e pedem voto em Bolsonaro.** Retrieved from https://www.correiobraziliense.com.br/politica/2022/10/5043988-hackers-atacam-sites- petistas-e-pedem-voto-em-bolsonaro.html

Gebru, T. (2023). **Data Conscience: Algorithmic Siege on Our Humanity**. MIT Press.
Herath, T. B. G., Khanna, P., & Ahmed, M. (2022). Cybersecurity Practices for Social

Media Users: A Systematic Literature Review. **Journal of Cybersecurity and Privacy**, 2(1), 1-18.

Internet Society. (2024). Regulation Digital Platforms in Brazil: Potential Impacts on the Internet. Internet Society Brazil Chapter.

JOTA. (2022, August 22). **STF sofreu mais de 2,4 milhões de ataques cibernéticos críticos em 7 meses**. Retrieved from https://www.jota.info/stf/do-supremo/stf-sofreu- mais-de-24-milhoes-de-ataques-ciberneticos-criticos-em-7-meses Lessig, L. (2006). **Code and Other Laws of Cyberspace**, Version 2.0. Basic Books.

Morais, C. B., Mendes, G. F., & Vesting, T. (Eds.). (2023). The Rule of Law in Cyberspace. Springer. (**Law, Governance and Technology Series, Vol. 49)**.

Noble, S. U. (2018). **Algorithms of Oppression: How Search Engines Reinforce Racism**. NYU Press.

O Globo. (2024, September 3). **Sistemas do STF, Anatel e PF são alvo de ataque após decisão de Moraes sobre o X**. Retrieved from https://oglobo.globo.com/politica/noticia/2024/09/03/sistemas-do-stf-sao-alvos-de- ataque-apos-decisao-de-moraes-suspendendo-o-x.ghtml

O'Neil, C. (2016). **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. Crown Books.

Pasquale, F. (2015). **The Black Box Society: The Secret Algorithms That Control Money and Information**. Harvard University Press.

Zuboff, S. (2019). **The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power**. PublicAffairs.

7. RESURGENT STATES, DISINFORMATION AND THE COVID-19 INFODEMIC

https://doi.org/10.36592/9786554603065-06

Thiago Lopes Cardoso Campos[1]

Abstract
The COVID-19 pandemic revealed both the power and peril of social media in shaping public health responses and political discourse. While platforms enabled timely information sharing, they also became vehicles for widespread disinformation, including promotion of scientifically unsubstantiated medications such as hydroxychloroquine and ivermectin. This essay critically examines the emerging regulatory challenges posed by false messages on social media during the pandemic, highlighting the intensified role of "*resurgent states*" in content governance. It explores the tension between free expression and public safety, and investigates how state actors in Brazil, the United States, and India influenced the narrative around COVID-19 treatments. Using case studies and scholarly research, the paper shows how regulatory failures and political opportunism led to an erosion of trust in science, rise in medical populism, and widening of the digital divide. The essay argues for a rethinking of regulatory paradigms where social media companies, governments, and civil society co-develop norms that balance platform autonomy, truth standards, and democratic accountability. Ultimately, it calls for international regulatory frameworks that address transnational health misinformation while resisting state overreach and censorship.
Keywords: Social Media Regulation, Disinformation, COVID-19, Health Misinformation, Resurgent States, Medical Populism, Digital Governance, Free Speech

1. Introduction

In early 2020, as the COVID-19 pandemic escalated into a global public health emergency, a parallel crisis began to unfold online. The World Health Organization (WHO) characterized it as an "*infodemic*" – a deluge of false or misleading information that made it difficult for citizens to access reliable health guidance (WHO 2020). Unlike previous outbreaks, the COVID-19 crisis occurred during a digital age dominated by algorithm-driven social media platforms. These platforms, built on

[1] Public health lawyer, specialist in health law, and master's student in Public Health at the State University of Campinas – UNICAMP and in Constitutional Law at the Brazilian Institute of Public Law – IDP.

engagement-maximizing architectures, prioritized virality over veracity. As a result, disinformation regarding the virus's origin, transmission, and treatment spread faster and wider than authoritative medical content. A study by the Reuters Institute found that 59% of misinformation was actively reshared on Facebook and WhatsApp, even when flagged by experts (Brennen et al. 2020). These platforms, once heralded as tools for democratizing knowledge, inadvertently became accelerants of confusion and public health risks.

Amid this surge of digital disorder, particular attention must be paid to the widespread promotion of scientifically unsubstantiated treatments. Public figures in countries like Brazil, India, and the United States endorsed drugs such as hydroxychloroquine and ivermectin despite a lack of clinical evidence. In Brazil, President Jair Bolsonaro's government heavily promoted these medications through official channels and social media posts, leading to their mass consumption despite warnings from the national health agency, ANVISA. Scientific studies, including randomized control trials conducted by the WHO Solidarity Trial and Brazil's own COALITION COVID-19 Brazil I Study, found no therapeutic benefit from these drugs (Caetano et al. 2020). Yet, misinformation about these treatments proliferated, showing how political leadership and online ecosystems can converge to undermine evidence-based public health.

This convergence exposed a critical regulatory vacuum. While platforms like YouTube and Twitter took steps to moderate health misinformation, their efforts were fragmented, delayed, and inconsistent. In turn, governments around the world began to reassert themselves, challenging the prevailing model of platform self-regulation. These "*resurgent states*" proposed and implemented legislation to hold platforms accountable, introduce content moderation mandates, and, in some cases, directly influence what content could be shared. Brazil's proposed Bill 2630/2020,popularly known as the "*Fake News Bill*",illustrates this new regulatory posture. The bill sought to impose obligations on messaging platforms and social media companies to trace, flag, and remove disinformation. This legal and political shift reflects a broader trend: the reconfiguration of internet governance to include stronger state oversight amid crises of trust and truth.

This essay critically examines this phenomenon by tracing how the COVID-19 pandemic exposed structural weaknesses in platform governance and catalyzed a wave of regulatory interventions. It begins by exploring the nature of the COVID-19 infodemic, followed by an analysis of regulatory challenges, state interventions, and global legal implications. Drawing on case studies from Brazil, and guided by the principles of digital constitutionalism, this paper argues for a balanced regulatory framework that upholds public health and democratic freedoms without enabling censorship or authoritarian control.

2. The Infodemic and the Disinformation Ecosystem

The COVID-19 pandemic generated a profound crisis not only in the health systems of nation-states but also in the epistemic foundations of the digital public sphere. According to the World Health Organization (2020), the term *infodemic* refers to an excess of information – some accurate and some not – that makes it hard for people to find trustworthy sources and reliable guidance. This phenomenon was particularly exacerbated by the algorithmic logic of social media platforms, which prioritize engagement and visibility over informational integrity (Donovan, 2020). As a result, falsehoods surrounding the Covid-19 virus, its origin, and particularly its treatment,such as the advocacy of hydroxychloroquine and ivermectin,spread widely, often overshadowing scientifically verified content.

Digital platforms like Facebook, WhatsApp, YouTube, and Twitter were instrumental in this diffusion, due not only to their global reach but also their design. Their recommendation systems, based on maximizing user interaction, allowed emotionally charged and sensational content to circulate more rapidly than verified scientific information. Brennen et al. (2020) observed that over 59% of widely shared COVID-19 misinformation remained visible and uncorrected across multiple platforms. Islam et al. (2020) further linked the prevalence of false health information to increased anxiety, vaccine hesitancy, and even loss of life, emphasizing the tangible public health consequences of the infodemic.

In Brazil, the situation was uniquely aggravated by the political use of misinformation. President Jair Bolsonaro's public endorsement of unproven treatments such as hydroxychloroquine and ivermectin was repeatedly disseminated through official channels and social media. These endorsements occurred despite robust clinical evidence contradicting their efficacy. The COALITION COVID-19 Brazil I Study, coordinated by major Brazilian research institutions, concluded that hydroxychloroquine had no clinical benefit in treating COVID-19 (Caetano et al., 2020). Nonetheless, these treatments were not only recommended but also distributed through the public health system, supported by a digital campaign that included disinformation, selective data interpretation, and rhetorical attacks on institutions such as ANVISA.

This pattern fits within what Keller (2022) has identified as “state-infused disinformation": misinformation originating from or legitimized by political authorities. Unlike user-generated disinformation, this type is more difficult to regulate because it blurs the line between political discourse and public deception. In this context, social media platforms operated as amplifiers of state narratives rather than neutral intermediaries. In a comparative analysis, Celeste et al. (2023) argue that the failure of platforms to address health misinformation consistently across jurisdictions is symptomatic of a broader normative vacuum in global content governance.

Brazil is not an isolated case. Similar patterns were observed in India, where government-affiliated figures promoted unproven Ayurvedic solutions, and in the United States, where former President Donald Trump's claims regarding hydroxychloroquine led to its mass purchase and consumption by the public. The cross-national recurrence of misinformation tied to state rhetoric suggests a systemic vulnerability in the current digital ecosystem. According to Mendes and Fernandes (2020), the COVID-19 crisis exposed the limitations of the existing regulatory framework, particularly the reliance on platform self-regulation and the inadequacy of constitutional safeguards to mediate the balance between freedom of expression and public health.

In sum, the COVID-19 infodemic must be understood not merely as a spontaneous outgrowth of social media use but as a politically mediated phenomenon embedded in the architecture of digital communication. The instrumental use of these platforms by state actors, combined with the platforms' algorithmic logic, produced an informational environment in which falsehoods could dominate the public sphere. The next section will examine how regulatory responses,both by platforms and states, attempted to address this governance failure.

3. Resurgent States and the Push for Digital Sovereignty

The failure of platform self-regulation to curb misinformation during the COVID-19 pandemic catalyzed a shift in the regulatory landscape, marking the resurgence of the nation-state as a central actor in digital governance. This resurgence, often framed as an assertion of digital sovereignty, reflects states' growing efforts to reclaim normative and legal control over online spaces traditionally dominated by private corporations (Celeste et al., 2023). In this context, the pandemic functioned as both a catalyst and a justification for intensified state intervention in content moderation, particularly in countries where political elites instrumentalized the regulatory vacuum to implement new mechanisms of control.

In Brazil, this trend culminated in the proposal of Bill 2630/2020, also known as the "*Fake News Bill*." The legislation aimed to combat the dissemination of false content by requiring messaging applications and social networks to implement transparency mechanisms, user identification processes, and traceability of forwarded messages (Internet Society, 2024). Although framed as a tool to safeguard democratic integrity and public health, the bill was widely criticized for its potential to undermine privacy, chill legitimate speech, and enable disproportionate state surveillance. According to the Internet Society (2024), the bill risks creating "an architecture of control" that may be used selectively against political opposition and minority groups, especially in contexts of low institutional trust.

Brazilian judicial responses during the pandemic further illustrate the complex interplay between state authority and fundamental rights in the digital sphere. The Federal Supreme Court (STF) issued a series of decisions balancing freedom of expression with the need to protect public health and democratic stability. One such decision upheld the suspension of social media accounts used to spread false information about the pandemic, citing constitutional values related to human dignity, life, and public security (STF, 2021). This case lawsuggests an evolving understanding within Brazilian constitutionalism, one that acknowledges the horizontal effect (*eficácia horizontal*) of fundamental rights in the regulation of private digital actors, especially when these actors affect the public sphere at scale (Mendes & Fernandes, 2020).

The Brazilian case, however, is not isolated. In Germany, the Network Enforcement Act (NetzDG), adopted in 2017 and reinforced during the pandemic, obliges platforms to remove illegal content within 24 hours or face substantial fines. The German model is often praised for balancing state regulation with judicial safeguards and procedural transparency (Celeste et al., 2023). In contrast, India's Information Technology Rules (2021) have been criticized for enabling executive overreach, compelling platforms to take down content deemed offensive by the government without clear legal or constitutional standards. This form of regulatory authoritarianism, according to Keller (2022), weaponizes misinformation regulation to silence dissent and consolidate political control.

The risk of regulatory overreach is particularly acute when states act unilaterally and without multistakeholder input. While the need for public oversight of digital platforms is undeniable, especially in crises involving public health, unilateral state action can lead to censorship, political manipulation, and the erosion of fundamental rights. As Celeste et al. (2023) argue, any legitimate framework for content governance must adhere to the principles of proportionality, legality, necessity, and democratic accountability. In this regard, the notion of digital constitutionalism offers a potential pathway: a model in which both private and public actors are bound by constitutional principles, ensuring that the regulation of digital spaces aligns with the protection of rights, procedural fairness, and public reason.

Thus, the pandemic revealed not only the limits of private platform governance but also the ambivalence of state-led regulation. While resurgent states have the capacity to fill governance gaps, they also pose new threats when their interventions lack transparency, legality, or respect for fundamental rights. The challenge lies in designing regulatory ecosystems that preserve the emancipatory potential of the internet while mitigating its risks.especially those related to health, democracy, and epistemic integrity.

4. Toward a Normative Framework for Balanced Regulation

The regulatory reactions observed during the COVID-19 pandemic,ranging from platform inertia to assertive state intervention,demonstrate that the global governance of digital content remains normatively fragmented and politically contested. Amid the proliferation of both health-related disinformation and authoritarian overreach, it has become clear that neither self-regulation by private corporations nor unilateral legalism by nation-states can alone ensure the integrity of the digital public sphere. The need for a balanced regulatory architecture,one that respects fundamental rights, guarantees transparency, and enables institutional pluralism,has become imperative (Celeste et al., 2023).

In this context, the concept of digital constitutionalism emerges as a promising normative paradigm. Grounded in constitutional law and human rights theory, digital constitutionalism proposes a multilevel governance framework in which both state and non-state actors are bound by constitutional principles such as legality, proportionality, accountability, and due process (Mendes & Fernandes, 2020). This framework emphasizes that content governance cannot be left to opaque algorithmic systems or populist legislative responses. Rather, it must be structured around democratic norms, judicial oversight, and institutional checks. According to Celeste et al. (2023), the core of digital constitutionalism lies in the idea that digital spaces should reflect the same normative guarantees that apply in the physical public sphere.

One key component of this approach is the enhancement of procedural transparency in content moderation practices. Platforms must be required to explain and justify their moderation decisions, especially in cases involving sensitive content such as public health information or political speech. Keller (2022) highlights that without transparency, users are unable to exercise their rights to contest, appeal, or understand the logic behind platform decisions. The implementation of notice-and-appeal mechanisms, independent oversight boards, and algorithmic audits are essential tools in aligning corporate practices with public law standards.

Equally important is the role of multistakeholder governance, which includes civil society, academia, international organizations, and user communities in the design, implementation, and monitoring of content governance norms. During the pandemic, civil society organizations played a pivotal role in fact-checking misinformation, supporting vulnerable populations, and advocating for rights respecting regulation. As the Internet Society (2024) notes, meaningful civil society participation is crucial for preventing regulatory capture and ensuring that digital regulation reflects the values of inclusiveness, fairness, and proportionality. Without such pluralism, regulation risks degenerating into censorship or technocratic elitism.

Finally, the global and cross-border nature of digital platforms necessitates international cooperation and normative harmonization. While national sovereignty is legitimate, isolated and conflicting national laws can create regulatory fragmentation and weaken enforcement. As proposed by Celeste et al. (2023), international instruments such as the EU's Digital Services Act (DSA) or UNESCO's Recommendation on the Ethics of Artificial Intelligence provide useful templates for building transnational principles that respect both local legal traditions and universal rights.

In sum, the path forward lies not in strengthening unilateral power,whether corporate or state,but in building a constitutional order for the digital age: one that balances security with liberty, truth with pluralism, and efficiency with accountability. The regulatory legacy of the pandemic must not be authoritarianism masked as emergency response, but rather a renewed commitment to the democratic governance of information ecosystems.

5. Conclusion

The COVID-19 pandemic exposed fundamental weaknesses in both digital infrastructure and regulatory capacity worldwide. While platforms facilitated the rapid exchange of public health information, they also became vectors for the dissemination of harmful disinformation,particularly regarding unproven treatments like hydroxychloroquine and ivermectin. This duality underscored the ambivalence of technological mediation in crisis contexts. As demonstrated in the Brazilian case, when state actors themselves become sources or amplifiers of falsehoods, the complexity of regulation deepens. The state's role oscillates between being a guarantor of rights and an agent of informational disruption.

The notion of resurgent states, examined throughout this work, reflects the reassertion of state power over digital spaces traditionally governed by private standards and opaque algorithms. The legislative and judicial responses in Brazil, such as the proposed Bill 2630/2020 and Supreme Court rulings on speech and misinformation, highlight both the possibilities and the perils of this resurgence. While legal intervention is necessary to prevent harm and promote democratic accountability, it must be guided by constitutional principles that prevent the misuse of power for partisan or authoritarian ends.

The failure of platform self-regulation,marked by inconsistency, lack of transparency, and minimal public oversight,further justifies the call for a comprehensive and balanced regulatory framework. However, as this study has argued, such a framework must transcend binary logics of public versus private control. Instead, it must be rooted in digital constitutionalism, a normative model that subjects both platforms and states to standards of legality, proportionality, transparency, and multistakeholder governance (Celeste et al., 2023; Mendes & Fernandes, 2020).

Looking ahead, the regulatory legacy of the pandemic offers a cautionary tale. It reveals not only the risks of inaction but also the dangers of overreach. The challenge is to design legal architectures that protect democratic deliberation without enabling surveillance, censorship, or information manipulation. This requires not merely reactive regulation, but proactive institutional design, guided by rights-

based principles and sustained through public engagement and international coordination.

Ultimately, the governance of digital spaces in times of crisis must reflect the same democratic values we seek to defend offline. The pandemic should not be remembered as a moment of informational collapse, but as the impetus for constructing a constitutional order in cyberspace that affirms dignity, truth, and justice for all.

REFERENCES

BRENNEN, J. Scott et al. *Types, sources, and claims of COVID-19 misinformation*. Oxford: Reuters Institute, 2020.

CAETANO, Rosângela et al. Use of chloroquine and hydroxychloroquine in

COVID-19 and the resurgence of medical populism. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 36, n. 8, 2020.

CELESTE, Edoardo et al. *The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard*. Cham: Palgrave Macmillan, 2023.

DONOVAN, Joan. Social-media companies must flatten the curve of misinformation. *Nature*, London, v. 580, p. 590, 2020.

EUROPEAN UNION. *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)*. Official Journal of the European Union, L 277, p. 1–102, 27 Oct. 2022.

INTERNET SOCIETY. *Regulação de Plataformas Digitais no Brasil: Análise do PL 2630/2020*. Brasília: ISOC Brasil, 2024.

ISLAM, Md. Saiful et al. COVID-19–related infodemic and its impact on public health: a global social media analysis. *The American Journal of Tropical Medicine and Hygiene*, v. 103, n. 4, p. 1621–1629, 2020.

KELLER, Clara I. Don't Shoot the Message: Regulating Disinformation Beyond Content. In: MORAIS, Carlos Blanco de; MENDES, Gilmar Ferreira; VESTING, Thomas (org.). *The Rule of Law in Cyberspace*. Cham: Springer, 2022. p. 309– 330.

MENDES, Gilmar F.; FERNANDES, Victor O. Digital Constitutionalism and Constitutional Jurisdiction: A Research Agenda for the Brazilian Case. *SSRN*, 2020. Available at: https://ssrn.com/abstract=3769947. Accessed: 1 June 2025.

SUPREMO TRIBUNAL FEDERAL. *Coletânea de Jurisprudência: Liberdade de Expressão e Desinformação durante a Pandemia*. Brasília: STF, 2021.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO, 2021.

WORLD HEALTH ORGANIZATION. *Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation*. Geneva: WHO, 2020.

8. RESURGENT STATES AND NEW CHALLENGES ON ALGORITHMIC DECISION-MAKING REGULATION

https://doi.org/10.36592/9786554603065-07

Pedro Nilson Moreira Viana

Abstract
This paper evaluates the use of Artificial Intelligence (AI) as a support tool for drafting judicial decisions, specifically in the realm of criminal law. In light of the increasing digitalization of the Judiciary, it is observed that the application of such models is influencing not only administrative tasks but also the structure of judicial reasoning. Although the adoption of such technologies promises gains in efficiency, precision, and systematization, concrete risks must be addressed, such as algorithmic opacity, data leakage and algorithmic bias. Based on data from Conselho Nacional de Justiça and also specialized literature, it is argued that the use of AI may be compatible with judicial decision-making, when it comes to supporting and drafting activities, provided that an institutional model is previously established. This model must be grounded in the standardization of legal reasoning, data mapping, prompt-based calibrated training, and continuous human validation. It is concluded that, if properly regulated, AI may serve as an auxiliary tool in decision drafting, without replacing or compromising a judge's prudential reasoning.
Keywords: Artificial Intelligence. Judicial Decision. Criminal Procedure.

1. Introduction

Applying Artificial Intelligence on adjudicatory functions represents one of the most significant phenomena of digital transformation in judicial institutions.

Gradually, as technical-instrumental innovations become increasingly integrated into daily routine, a paradigmatic shift in how adjudicative functions are carried out has been observed, affecting tasks ranging from bureaucratic procedures to more sensitive aspects of the decision-making process, particularly the drafting of judicial decisions.

These new tools, capable of producing coherent and logically structured natural language content, have sparked an important debate regarding their (in)ability, to properly support decision-making, especially because AI raises critical ethical, technical, and legal questions, notably concerning the protection of sensitive

data, bias and the need for transparency.

In such a scenario, this study is oriented towards evaluating the potential contributions of AI to enhancing both efficiency and quality of judicial decision-making.

At the same time, it proposes practical alternatives for a normative standard capable of ensuring an ethical and secure use of AI when it comes to decision-drafting.

2. Understating the Power of AI over users and institutions

There is an undisputable growing realization that Artificial Intelligence tools are no longer to be considered merely technical instruments. Rather, they are, perhaps, becoming some sort of *"gravitational center"* for human decisions, whether in government, market or judicial structure. Naturally, this transformation has (re)shaped how individuals behave when facing choices and, more importantly, when performing tasks.

Through the mediation of AI, it has become entirely feasible to *"maximize expectations"* both quantitatively and qualitatively, in many fields subjected to human reasoning.

This has facilitated how people, generally speaking, can organize, e.g., routines and tasks of low to medium complexity, which seem to be an undeniable (and, frankly irreversible) reality.

Such *"easiness"*, in fact, is one of the key attractions behind the rapid diffusion of language models, especially those freely accessible, as the human condition is intrinsically inclined to pursue means of achievement requiring the *"least effort"*.

From a purely phylogenetic standpoint of behaviorist theory, Hull (1943, p. 14) explains: *"If two or more sequences of behavior, each involving a different amount of energy expenditure or work, are equally well reinforced an equal number of times, the organism will gradually learn to choose the less laborious sequence of behavior"*.

There is no reason, a priori, to criticize this ontological view of how individuals typically behave. After all, such an inquiry would presuppose a deeper investigation into the nature of being itself, something far beyond the scope of this analysis.

However, the fact remains that this increasingly pervasive algorithmic automation leads to a *"reduction in human agency".*

In other words, the power and burden of deliberation, once exclusively based on the weighing of subjective criteria, is now being displaced to purely statistical models. This corresponds, by definition, to the concept of *"governance by statistics"* (Bucher, 2018, p. 49).

Curiously, these same statistical models are derived from patterns, probabilistic correlations, and inferences extracted from the human experience itself. This is exactly what Zuboff (2019, p. 249), brilliantly defined as the *"expropriation of human experience".*

By definition, this expropriation consists in the process through which human experience is appropriated as a free raw material, transfigured into behavioral data that are extracted, analyzed, and classified.

It is a systematic dispossession whereby everyday life, our interactions, habits, emotions, desires, and displacements, ceases to be merely lived and is converted into computational assets.

The consequence of this, when projected onto the judicial sphere, is the consolidation of an automated decision-making model wherein purely technical rationality overrides human discretion.

Ultimately, it is AI that provides users with conclusions based on data it alone has both collected and evaluated. Beer (2017, p. 12), for instance, believes on the matter that: *"Critical decisions are made not on the basis of the data per se, but on the basis of data analyzed algorithmically"*.

If this were not problematic enough, the consolidation of AI within judicial adjudication becomes even more complex due to the fact that the interpretation and selection of data necessary for generating known outputs are conducted via algorithmic processes governed by logics unknown to most of the public.

This state of *"unawareness"* does not stem solely from the complexity of the subject but rather from the very design of the tool, something alike to what Stanley (2015) described as *"ignorance by design".*

A study conducted by researchers from the Universities of York and Connecticut confirms this. Among 543 participants in an empirical survey, 74% of

online service users admitted to accepting terms of service without reading them, using the "*quick join*" option.

Among those who did read the terms, most still selected "*accept terms*". The documents, which would typically require 40 to 50 minutes of reading, were "*examined*" in less than 14 seconds.

This superficial reading resulted in a failure to notice a deliberately inserted clause in the test, which stated:

> *By agreeing to these Terms of Service, and in exchange for using the service, all users agree to immediately surrender their firstborn child to NameDrop, Inc. If the user has no children, the agreement shall remain valid and enforceable until the year 2050. All individuals surrendered to NameDrop become the property of NameDrop, Inc., without exception (…) (Obar & Oeldorf-Hirsch (2018, p. 19).*

At first glance, this evidence may suggest user negligence. But in fact, most participants were deterred by the "hard tech language" and overly detailed nature of such policies.

According to Obar & Oeldorf-Hirsch (2018, p. 28), the clickwrap mechanism was frequently praised for being "easy," "fast," "simple," and "convenient."

By contrast, the policies themselves were often criticized as "*too long*" and "*overly verbose*". Users expressed apathy and futility, often linked to the perception that these documents would remain unintelligible even if so read.

This perception is reinforced by Wagner's (2022, p. 217) analytical study, which, after 25 years of observation, concludes that privacy policies and terms of service have evolved to become increasingly longer and harder to read.

> *We find evidence of inflated length and decreased readability in privacy policies, especially after the introduction of new privacy regulations. We also identify worrying trends in data handling practices, such as increased collection and sharing of sensitive data, along with the absence of meaningful user choice. Most concerning is that such practices are hidden in lengthy policies requiring university-level training to understand, and over an hour per day to read (Wagner, 2022, p. 217).*

In other words, institutions that aim to benefit from free-access AI platforms must acknowledge that an asymmetry of power exists between platforms and users regarding the information gathered and processed, even if such asymmetry is masked by the legal fiction of conscious adhesion.

This concern becomes more acute when AI tools are institutionalized in government environments, where highly sensitive, strategic, or legally protected data circulate. The automated and potentially irreversible absorption of such data exposes the judiciary, for instance, to structural security vulnerabilities.

Indeed, the inadvertent disclosure of confidential or legally protected information (such as sensitive personal data) not only violates the fundamental right to data protection, but also "opens a gate" for external actors to *"hack"* institutional knowledge and capture patterns, preferences, and internal decision-making logic.

In addition to this asymmetric "disenchanted adherence", it is essential to observe that AI tools operate through completely opaque processes. So opaque, in fact, that they are often not disclosed to users, usually under the *"trade/business secret"* clause.

Based on the theories of Pasquale (2015), the opacity of AI tools, for example, can be systematized into two concentric circles: technical and epistemic.

Technically, these tools are based on billions of parameters within deep neural networks, resulting in systems whose inner workings are not entirely understood, even by their own developers.

The relationships between inputs and outputs are non-deterministic, and the inferential pathways are inaccessible to external observation.

This means that AI systems perform a multitude of interlinked inferences on massive datasets, in such a way that at a certain point, it becomes untraceable to determine which data led to which result (IBM, 2023).

Epistemically, opacity manifests through the concealment of sources, internal rationalizations, and logic used to produce automated outputs. This makes it impossible to validate results through truth, logic, or proportionality standards, causing cognitive harm akin to what Fricker (2023, p. 52) has called "epistemic injustice."

According to Fricker:

> The first harm involves the prejudicial exclusion of people from participation in the dissemination of knowledge. The second concerns the erosion of trust in what is said. The third is the threat to the very intelligibility of discourse" (Fricker, 2023, p. 52).

Therefore, for the institutional use of such tools to be justifiable, it is imperative to understand that algorithms operate on biased data and often lack accountability, that is, without clear obligations to inform or, most importantly, to explain (Schedler, 1999).

A Judiciary that intends to incorporate AI into its decision-making processes must be aware that the absence of oversight and contestability mechanisms only deepens power asymmetries between system developers and the subjects affected by such systems.

3. A practical approach on how to regulate algorithmic decision-making

It does not seem feasible, nor realistic, to prevent Artificial Intelligence (AI) and other disruptive technological apparatuses within courts; especially given the ever-present imperative of efficiency, in favor of which the State and its government branches, especially the Judiciary, must remain receptive to innovation and new strategies.

This scenario is particularly more interesting when it comes to Latin American courts, such as in Brazil, whose institutions suffer from chronic backlog and procedural delays, according to the World Bank's Diagnostic Report N. 319 (1999).

Observations like such may lead to the fact that instead of discussing *whether* such tools should be used, developing States would rather concentrate efforts on *how* they can be developed, regulated and employed ethically and safely.

Nonetheless, according to the most recent literature, this does not appear to be the predominant approach among scholars in Brazil.

In fact, due to the unfamiliarity of this new technological frontier, it is still quite uncertain how institutions could actually employ AI on a day-to-day basis. This is precisely why "implementing Artificial Intelligence in various sectors is somewhat

regarded as a negative development" (Simões & Morais, 2024, p. 26).

Regarding AI within processes of judicial adjudication, there is a much growing concern, and even some sort of criticism, with many authors pointing out great risks such as, e.g., "violations of fundamental rights," "deficient motivation of judicial decisions," "algorithmic bias," and "inequality".

For instance, some scholars even argue that, due to its "incapacity of processing and dealing with the interactive or communicative aspects of socio-legal relations in the concrete dimensions of life", AI would [never] be structurally compatible with adjudicative functions (Toledo & Pessoa, 2023, p. 14).

Although those perspectives arise from legitimate concerns on the (de)humanization of the judiciary as an institution, this should not lead to a radical approach such as excluding AI from a "decision-making environment".

Mainly because this perspective grossly disregards the auxiliary and non-autonomous role of AI in the decision-making process , one that could, surely, contribute positively to judicial efficiency.

Thus, this particular discussion should be established on a more ontological level: it is not, broadly speaking, a matter of surrendering "adjudicative authority" to machines, but rather on how emerging states can enhance the role of judges with tools for systematization, support, and drafting the so called "*judicial discourse*", which, far from constituting an abdication of responsibility, actually reaffirms the judge's hermeneutic role.

The use of AI as a support mechanism, specifically for drafting elements of judicial decisions, may help overcome operational externalities, ensure coherence in legal application, and (probably) improve judicial reasoning.

From this perspective, it is, at least, possible to reconcile AI with fundamental principles that govern judicial adjudication, above all, in developing countries, given that the "final word", as an expression of legal valuation and prudential judgment, will belong to the "human" judge vested with authority.

Within such an approach, "the judicial decision [is maintained as] an act of prudence, rather than a 'manufacture of things'" (Dip, 2021, p. 27).

To some degree, interestingly, this is already occurring in Brazil. According to recent reports, around 80% of judges and 70% of court staff consider AI tools to be

useful in their professional activities (Brasil, 2024).

The National Council of Justice, for instance, indicates as well that approximately one-third of judges and court's staff have already used generative AI directly in professional tasks, including, and most importantly, drafting of judicial documents.

However, transforming generative AI tools into reliable and efficient instruments capable of supporting the drafting of complex judicial decisions with reasonable precision depends on adopting a few essential measures, particularly aimed at mitigating bias and ensuring appropriate data protection.

It is precisely for this reason that this article will examine strategies for the harmonization of Artificial Intelligence within the judicial adjudication process in developing legal systems.

The first of these measures seems to be the logical "standardization" of judicial reasoning through structural uniformity. This stage is essential for ensuring clarity, precision, and predictability.

To that end, Courts ought to be encouraged, for instance, to develop well-defined templates and structural rules to judicial discourse, based on the most common procedural classes.

Organizing and systematizing the textual elements of decisions is key to AI support. It is crucial that the reference data (*i.e.,* "rawdata" used to "teach" the generative model) be composed of both fixed and variable segments (with the latter in lesser proportion). Without such customized parameters, the risk of bias increases significantly, as the model, lacking specific guidance, surely will resort to uncontrolled and unauthorized data sources.

In the absence of standardization, it is common for generative AI tools to rely on publicly available online data, generic repositories, and random linguistic associations. This is one of the main reasons why outputs may be outdated, decontextualized, or even inaccurate, which is a great risk when it comes to judicial adjudication.

So, once the macro-structure of the decision is somewhat established, the next equally challenging step is to "teach" the language model some "desired patterns". To do so, the expected results must be defined in advance.

In other words, the operator's reasoning must be reversed: first, users must determine the desired output, then create the corresponding prompts. At that point, a specific generative AI tool will likely have been chosen. Among the many available options, final selection depends on each user's personal experience and, of course, affinity.

Another important step is "process mapping". This serves to understand the organization's real operational flow, without creative alterations or changes in specifications. This stage aims to identify how processes actually develop in court, their decision points, operational risks and vulnerabilities.

The analysis must include a concise description of the main steps, actions, and outcomes, as well as their general and specific objectives. It is essential to identify the normative framework that guides the activities, departments responsible for execution, and the current operational context, including data such as historical and current caseloads, values, subjects, and timeframes of past and current cases.

Once the actual process flow is understood, it is necessary to conduct a sort of "data depuration". Balanced examples drawn from reliable sources are essential for building a model that reflects the diversity and complexity of legal issues dealt with day-to-day.

This practice is aimed at preventing AI from operating with outdated information, thus promoting greater accuracy in the drafting of reports, headnotes, or legal opinions.

In order to reach that, it is crucial to create a systematic repository of decisions issued by a specific judicial unit. A detailed survey of the past 12 months is recommended, from which the user will compile a table containing, among other elements: (1) procedural classes judged; (2) legal categories; (3) recurring subjects and key issues; (4) legal provisions applied; (5) case duration and timeline; and (6) main arguments and legal reasoning (thesis) applied.

Such systematization not only facilitates data depuration but also provides the operator with a comprehensive overview of the unit's functioning and demands. Based on this data, AI model training can be directed to more accurately reflect actual judicial practices, thereby enhancing response quality and relevance to the specific forensic environment of a given judge's office.

Validation and review of AI-generated results is another critical step. Using peer review to evaluate outputs provides an additional layer of security by identifying inconsistencies, errors, or biases that the system itself might overlook.

Creating detailed checklists to detect bias patterns is a practical measure for standardizing review procedures, while also promoting transparency in the validation process. This ensures that the model adheres to standards of impartiality and technical quality.

Furthermore, at the early stages of implementation, AI tools should be used by a limited number of trained users, who are equipped with specific prompts and shared *"memory logs".*

These trigger prompts ensure that users interact with the technology appropriately, unlocking its full potential while minimizing operational failures, which are common in collective and repetitive use.

During this initial stage, strict control allows the tool to mature, with ongoing adjustments based on practical feedback. Only after achieving a satisfactory level of maturity and reliability should the tool be extended to a larger group of users.

But, it's important to highlight that, even at such a point, all users must be equipped with shared usage memory to ensure that best practices adopted during the initial phase are preserved and replicated.

A gradual expansion, supported by a robust system of curation and validation, promotes an effective and responsible integration of technology.

4. Conclusions

Advances of new digital technologies, especially Artificial Intelligence, requires States and its Institutions to undergo a (re)shaping of their operational methods and decision-making rationality.

Far from constituting a threat in itself, AI seems to be a potentially valuable instrument for enhancing the efficiency of judicial services, provided that its use is guided by technical, normative, and ethical standards compatible with the principles of the democratic rule of law.

This essay has somewhat demonstrated that the application of AI within the Judiciary is not only inevitable but necessary. However, for such use to be both legitimate and functional, it is essential to critically (re)examine both the resistance to technology and the naïve enthusiasm for the unrestricted automation of adjudicative functions.

The central question, therefore, is not whether AI should be employed, but rather how and to what extent its integration can occur without compromising the impartiality, reasoning, and legitimacy of judicial decisions.

In this context, generative models may serve as an important auxiliary tool (never autonomously) in supporting the decision-making process, especially in the tasks of systematization and textual production.

To achieve this, the adoption of structuring measures is required, including: (i) the logical-rational standardization of judicial decisions; (ii) the instruction of interpretative and argumentative patterns to the model through calibrated prompts; (iii) the mapping and curation of data derived from actual judicial experience; and (iv) the gradual implementation of AI, subject to peer validation and ongoing human supervision.

Disregarding such standards poses significant risks, due to the sensitivity of the data involved, the opacity of algorithmic processes, and the potential for automated reproduction of historical and social biases, some of which are already well catalogued.

These risks, however, do not justify exclusion; rather, they demand developing an institutional model anchored in responsibility, traceability, and (for the time being) constant human oversight.

References

Beer, D. (2017). *The social power of algorithms*. *Information, Communication & Society*.

Brasil. Conselho Nacional de Justiça. (2024). *Pesquisa uso de inteligência artificial (IA) no Poder Judiciário: 2023*. CNJ.

Bucher, T. (2018). *If… then: Algorithmic power and politics*. Oxford University Press.

Dip, R. (2018). *Prudência (...)*. Dykinson.

Fricker, M. (2023). *Injustiça epistêmica: O poder e a ética do conhecimento*. Edusp.

Hull, C. L. (2024). *Principles of behavior: An introduction to behavior theory*. Appleton-Century-Crofts.

IBM. (n.d.). *What is explainable AI (XAI)?*

Obar, J. A., & Oeldorf-Hirsch, A. (2018). The clickwrap: A political economic mechanism for manufacturing consent on social media. *Social Media + Society*, 4(3), 1–14.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Schedler, A. (1999). Conceptualizing accountability. In A. Schedler, L. Diamond, & M. F. Plattner (Eds.), *The self-restraining state: Power and accountability in new democracies*. Lynne Rienner Publishers.

Simões, N. C., & Morais, L. F. (2024). As reflexões da inteligência artificial no Poder Judiciário e a sua efetividade. *Revista Jurídica da Faculdade Santa Rita de Cássia*.

Stanley, J. (2015). *How propaganda works*. Princeton University Press.

Toledo, C., & Pessoa, D. (2023). O uso de inteligência artificial na tomada de decisão judicial. *Revista de Investigações Constitucionais*, 10(1), e237.

Wagner, I. (2022). *Privacy policies across the ages: Content and readability of privacy policies 1996–2021*.

World Bank. (1999). *Judicial reform: A process of change through pilot courts* (Diagnostic Report No. 319). International Bank for Reconstruction and Development / World Bank.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

9. A DELEGATED GEOGRAPHICAL REGIME FOR SOCIAL MEDIA GOVERNANCE

https://doi.org/10.36592/9786554603065-08

João Pedro Barbosa Mota

Abstract:
Since the XXth century, the internet has influenced human lives by improving communication between individuals and access to information. As a consequence, more and more individuals are joining social media platforms, a movement that is also changing the legal order. Due to the social issues that are happening inside social media platforms, states are trying to figure out a way to best regulate these cyberspaces aiming to protect human rights, exposing a dilemma concerning social media governance. Trying to solve this issue, this article proposes a delegated regime where geographical rules should establish principles to guide social media governance while states should individually have the competence to edit procedural rules to provide for social media platforms operation. In addition to that, the internal terms of use should dictate the particularities of each platform in order to guarantee private liberty.
Keywords: Social media; Governance; Internet;

1. INTRODUCTION

Communication through language,has always been the mechanism that allows an individual to exercise its rationality and understand the concepts in the world[1]. Language has been used by humanity as an instrument to develop itself and improve its relation with and within the world. On account of that, as society develops people have continued finding ways to get more connected and share information with each other.

Since the Advanced Research Projects Agency (ARPA) developed the first computer network in the mid XXth century, precisely in 1969 (Araújo, 2014), the internet has been used by humankind to improve its ways to communicate and share

[1] This concept about language is brought by Ernildo Stein (1996, p. 16), a Brazilian philosopher who says that: *human beings only know through concepts, only know through language, [...] human beings are only rational because their access to the world is through meaning, through significance, through concepts, through words, through language* [...] "*o ser humano só conhece através dos conceitos, só conhece através da linguagem, [...] o ser humano somente é racional porque seu acesso ao mundo se dá via sentido, via significado, via conceitos, via palavras, via linguagem*".

information without being in the same place, an instrument that continues to influence people's lives .

As a result of Web 2.0, in the mid 90's and in the early 00's, social media platforms emerged as a development of internet use to improve communications therefore, information could flow easier and faster over longer distances, bringing an exponentially increasing number of people to social media platforms over time.

Social media platforms can be conceptualized as an organized cyberspace which allows people to gather, share common thoughts, interests (Zenha, 2018) or even discuss conflicting thoughts or just to share information; this creates an informational network phenomena (Zenha, 2018).

The information that flows in social media can be true or false or even dissipate bad preconceptions about a group of people. From this latter kind of information which is usually negatively valued, emerges a discussion about regulating social media to guarantee fundamental rights protection of its users.

Starting from the point of view about the necessity of social media regulation for human rights protection as the information published online reaches a large amount of people because of the number of its users and, as a consequence, can cause social conflicts, this essay briefly exposes the discussion about the dilemma of social media regulation and tries help solving it.

2. THE GOVERNANCE DILEMMA

It is known that nowadays about 5-6 billion[2] people use social media to connect with other individuals across the world, to share information, shop, find jobs among other things.

These places - social media platforms - used to be seen as mere spaces where users post the information they want to share or send to other users, often depicting a private communication (user x user and user x platform).

[2] In 2023 5,04 billion people got connected in social media, which corresponds to 62% of the worldwide population (Ministerio Publico do Estado do Mato Grosso [MPMT], 2024).

As mere private relations, in a first view, the state would not have interest in enacting laws to regulate social media interactions, as individuals must have freedom to establish rules to dictate private interactions. This freedom attributed to private governance is subsidised in the Human Rights Declaration and the Brazilian Federal Constitution, which protects the private freedom of individuals to have autonomy to regulate their private relations and interactions[3] as well as private property is also guaranteed. Due to this understanding, social media platforms also have the autonomy to establish the rules that will dictate the interactions within their own platforms in their terms of use.[4]

Establishing a connection with what Celeste *et al.* (2023) says in the book "The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard", internal regulations developed by digital platforms constitute a micro governance of social media and they "represent private constitutions as they regulate the exercise of users rights in these virtual spaces" (Celeste *et al.,* 2023, p. 13).

However, micro governance is not sufficient to regulate the use of social media, as these platforms are private actors developed by economic investments in their majority, their internal terms of use are designed to protect the private interests of these platforms, which can be at odds with the human rights of its users. Rereading Suzor (2019), Celeste *et al.* (2023) say that in "the social media environment, the decision of private platforms to adopt their own internal rules has been accused of arbitrariness and lack of accountability, being even associated with a 'no law' scenario" (p. 13).

This insufficiency of microgovernance can also be seen when we look at how platforms moderate the content published by users. Nowadays, the information flow on the internet is moderated by algorithms "used to monitor our behaviour and interests and to predict our necessities and future actions. These algorithms direct

[3] It is possible to infer this private interest protection of the Federal Constitution of Brazil from the regulations established by the article 5 around the guarantee of individual freedom.
[4] This conclusion is also subsidised on the major function of social media as just a space provided where users generate the content that flows within cyberspace (Celeste *et al*, 2023).

our actions and thus determine, among other things, the economic success of products and services" (Hoffmann-Riem, 2022, p. 2)[5].

An algorithm's content moderation can influence the number of users that can see the content published by other social media users. At first glance the platforms algorithm used to moderate the interactions and the information flow in social media contexts could violate human rights such as freedom of speech and the right to information[6] as some users could have disproportionate visibility on their published information in detriment of other users, without transparency from the platforms side on why the advantaged individual had the privilege previously mentioned.

In the same way, when an algorithm can control the access of the online information it can directly impact societies (G. Mendes, personal communication, march 31st of 2025)[7] an example is the 2015 scandal (Cadwalladr & Graham-Harison, 2018), around Facebook data collection of its users and the usage of this data to later influence the USA presidential election of 2016 (Possa, 2023).

Micro governance has not been shown to be an efficient means of protecting social media user's human rights, due to the high number of users in social media platforms The notion of a public interest in the interactions that happen within social media emerged created a necessity of state intervention to protect the social order[8]

This switch of social media from a mere private space to a public interest space, can also be seen when Public Administrations started to use social media platforms to share official information about government operations, natural disasters warnings, etc; demonstrating the high relevance that social media has within contemporary society. As an example of the importance of digital platforms nowadays, in 2023 the Brazilian Government released a Good Practice Guide for Federal Government behaviour in social media which established recommendations

[5] This understanding was translated from the portuguese version of the book written by Hoffmann-Riem Wolfgang (2022), who says that "Algoritmos são usados para monitorar nosso comportamento e interesses e para prever nossas necessidades e ações futuras. Eles orientam nossas ações e assim determinam, entre outras coisas, o sucesso econômico dos produtos e serviços" (p. 2).
[6] Article 19th of Human Rights Universal Declaration.
[7] Lecture delivered by Gilmar Ferreira Mendes to the class of Emerging Challenges in Social Media Regulation: Resurgent States on 03.31.25, IDP/PUCRS/DCU.
[8] This change of idea about a state intervention on social media throughout regulation is emphasized by the digital constitutionalism movement, which aims to protect fundamental rights in the cyberspace from normative prescriptions (Mendes, 2020)

to improve the communication capacity between the state and the society through social media (Good Practice Guide for Federal Government, 2023).[9]

Coming back to the right conceptualization around micro governance when the social media platforms impose their regulations inside terms of use, in another turn, Celeste *et al.* (2023) say that the regulations of social media imposed by external actors can be conceptualised as macro governance[10]. Macro governance is usually made by states individually[11] and set by international rules that aim to protect human rights within social media platforms.

Macro governance by states seems to be a good solution to protect human rights in social media platforms because these regulations tend to be more adequate to fit in society.

Each society has its own cultural, historical and moral standards whose patterns evolve over the years, and this change influences their interpretation. As an example, the USA's legal interpretations[12] usually approach freedom of speech as a human right that has privilege in relation to others human rights via the first

[9] The guide spoken established five recommendations to improve the communication with society and help improving public politics: "(i) Digital Presence: amplification of the government presence in digital platforms; (ii) Transversality in nets: migration of government content through the multiple digital accounts to improve the communications; (iii) Strategy: each government organ has to create its own actuation objective within digital platforms; Iiv) Content Production: the content production must be done daily and planned in a serial way to spare political publics set out and public relevance informations; (v) Digital Listening: daily digital channels monitoring to supervise published content, interactions and results analysis. (vi) Accessibility: communication must be accessible and inclusive, aiming to guarantee the access of information and participation of people with any disability as well as recognising the diversity of regional, age group etc. language" (Good Practice Guide for Federal Government, 2023).

[10] Macro governance can be "represented by the mechanisms developed in conjunction with external actors, such as governments and advocacy groups" (Celeste, *et al.*, 2023, p. 10). It also can be seen as the governance "of platforms" (Gillespie, 2018, re-writed by Celeste, *et al.*, 2023, p. 10).

[11] As some examples: Network Enforcement Act - NetzDG (Germany, 2018,), Internet Civil Mark - MCI (Brasil, 2014), Law on the Protection of Personal Data - LPPD (Chile, 2024).

[12] Whitney v. California (1919); Twitter v. Taamneh (2023).

amendment[13]. However, in a restrictive approach[14], Brazil's legal interpretations[15] do not confer this privilege to freedom of speech in detriment of other human rights.

When a state individually has the competence to modify the rules that govern social media, these regulations should have more efficiency on their application since they fit better in each society's own cultural, historical and moral standards. Furthermore, state regulations also protect the sovereignty of jurisdiction from each country.

Despite these positive effects, state regulations seem not to respond properly to the problem of social media governance as they also have negative consequences. First, as each state could regulate social media, these platforms would have to create mechanisms to adapt their terms of use to each state's regulations and legal interpretations as they can be different, which can lead to an enhancement of production costs and create operational challenges to social media platforms. Second, this kind of macro governance could also cause the phenomenon of internet fragmentation, which can compromise the integrity of the user experience inside the internet (Mendes & Fernandes, 2020).

Regarding Internet Society's understanding, the internet was built as an open net that must provide integrity, accessibility and global reach to any virtual point within the planet (Internet Society, 2012, re-writed by Wachowicz & Lana, 2024). This important role aims to provide an interoperable access and distribution of information worldwide so the internet could be innovated anytime (Wachowicz & Lana, 2024), a standard that nowadays has a crucial importance to the development of globalized society.

---

[13] Constitution of United States, First Amendment (1789/2025), says that : "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances".

[14] Until 2003-2004, Brazil's Federal Constitution conferred freedom of speech as a preferred position in comparison to other fundamental rights. Although, for the last 4-5 years, Supreme Court jurisprudence related to using of freedom of speech to express fake news and democracy attacks shipped to interpret Brazil's Federal Constitution in a restrictive approach of freedom of speech (Lecture delivered by Ingo Wolfgang Sarlet to the class of Emerging Challenges in Social Media Regulation: Resurgent States on 03.31.25, IDP/PUCRS/DCU).

[15] ADPF n° 572 (2020).

Since the internet must provide open worldwide access to information to its users, the edition of regulations individually by the states could not guarantee its important roles as these norms and legal interpretations protect human rights in different dimensions. To clarify this idea and coming back to the examples exposed before to demonstrate the differences between the USA and Brazil's approach to freedom of speech, an information published online in the US that does not fit in the Brazilian regulation of freedom of speech may have to be adapted by social media platform to suit in the Brazilian legal order or even be prohibited of being accessed[16], macro governance that could cause the fragmentation of the internet, which does not follow its scope.

In order to protect human rights in social media and stop the fragmentation of the internet, establishing international rules to regulate it seems to give a solution to the problems pointed out until now.

Although, establishing international rules could fix the problems around protecting human rights in social media and stop the fragmentation of the internet, this solution could also create a problem with the effectiveness of norms application.

Generally, international laws aim to establish principles, rights and obligations with a more open substance on its statements - soft laws -, which requires a legal interpretation to provide its effectiveness. As said by Celeste *et al.* (2023, p. 18), these "norms, as they are, could not offer explicit guidance of behaviour to the actors involved in the social media environment and could not be directly applicable without a preliminary work of interpretation and recontextualisation".

As mentioned before, each society has its own cultural, historical and moral standards, which influences the interpretation of human rights. Thus, by the time that international rules could solve some of the problems pointed out around social media governance, they could also be inapplicable when the society patterns do not fit to the norms edited because of the human rights interpretation differences between the states.

[16] As another example, the territorialization of internet by states individual regulations could cause the phenomenon which the users of one kind of nation could get connected just with the users within the same state and get access only to the information published in the same country and even cause the consequence of the disconnection of the users to the rest of the world, what occur to the russian RuNET (Wachowicz & Lana, 2024).

Whereas international laws are also not sufficient to regulate social media in contemporary times as they could not be efficiently applicable to solve the cases involving human rights, a geographic bill of rights with a delegated regime might be a solution to the social media's governance dilemma as a union of legal theories around the regulation of social media.

This geographical movement has already been taken up by some entities such as the European Union who edited, for example, the Digital Services Act - DSA and the Digital Markets Act - DMA, which traced some directives aiming to protect human rights in the social media environment.

Compiling the effectiveness of nationals and international regulations and the juridic theories around social media governance, geographical norms established in blocks as Latin American, Northern America, European, Asian, Arabian, African and Oceanic blocks might seem as a kind of a solution to the dilemma of social media governance.

International regulations established in blocks by the countries situated in the same geographical area could minimize the differences around human rights interpretation as these states might share some similar cultural, historical and moral standards, which could lead to an effectiveness on the application of these norms. In addition, these regulations may decrease the problem involving the state's individually social media governance as the platforms would have to obey a lower quantity of laws with similar interpretations, which could ease the operational mechanisms and decrease the fragmentation of the internet. Moreover, geographical rules have to work collaboratively with micro governance established by the platforms aiming to secure fundamental rights in cyberspace.

Therefore, geographical rules should establish principles to guide social media governance while the states individually should have the competence just to edit procedural rules to provide the social media platforms operation. In addition to that, the internal terms of use should dictate only the particularities of each platform in order to guarantee their private liberty.[17]

[17] Besides the conclusion above-mentioned, it is important to expose the theory shared by some academics when they say that one of the solutions to the problem of social media governance is to "treat Internet as a different space and a proper system inside the nationals internationals juridic

3. **CONCLUSION**

Since the XXth century, the internet has been influencing human lives with the improvement of communication through the individuals and the access of information. As a consequence, more and more individuals are subscribing to social media platforms, a movement that is also changing the juridic order.

Due to the social issues that are happening inside social media platforms, the states are trying to figure out a way to best regulate these cyberspaces aiming to protect human rights. A lot of work has been done and there is still much more to be done trying to solve the dilemma of social media governance.

This essay is a brief exposition of this dilemma as well as an instrument to cooperate with this juridic issue resolution when it establishes a delegated regime of geographical social media regulation in collaboration with internal terms of use.

In this sense, geographical rules should establish principles to guide social media governance while the states individually should have the competence just to edit procedural rules to provide the social media platforms operation. In addition to that, the internal terms of use should dictate the particularities of each platform in order to guarantee private liberty.

REFERENCES

ARAÚJO, George Zeidan. Ler, pesquisar e escrever história em tempos de internet: desafios e possibilidades. **Revista Tempo e Argumento**, Florianópolis, v. 6, n. 12, p. 151-164. DOI: 10.5965/2175180306122014151. Disponível em: <https://revistas.udesc.br/index.php/tempo/article/view/2175180306122014151>. Acesso em: 02 jul. 2025.

BRASIL, Social Communication Office, **Good Practice Guide to Federal Gonvernment social media actuation**. Gov, Central of Content, Nets, set. 2023. Available on: <https://www.gov.br/secom/pt-br/central-de-conteudo/redes/guia>. Access on: jul. 4th. 2025.

systems, with particularities enough to cause a separation, in a big or small extent, from the traditional orders" (Wachowicz & Lana, 2024, p. 6). It is been discussed that the internet should be treated separately, with personal rules, courts and standards that should govern social media in order to establish applicable regulations and guarantee the integrity and the open access of the internet, which also might seem like a solution.

CADWALLADR, Carole; GRAHAM-HARISON, Emma. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Cambridge Analytica Files, The Guardian, sat. 17 mar. 2018. Available on: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. Access on: jul. 4th 2025.

CAMELO, Ana Paula *et al*. Internet Impact Brief. Proposals to Regulate Digital Platforms in Brazil: Potential Impacts for the Internet. São Paulo: FGV Direito SP, 2024. Available at: <https://repositorio.fgv.br/server/api/core/bitstreams/c56c7c31-0b5b-4ee2-951f-4efb41db972a/content>. Access in: july 2nd 2025.

CELESTE, Edoardo, PALLADINIO, Nicolla, REDEKER, Dennis, Yilma, Kinfe. The Content Governance Dilemma. In: **The Content Governance Dilemma.** Information Technology and Global Governance. Cham: Palgrave Macmillan, 2023. DOI: https://doi.org/10.1007/978-3-031-32924-1.

HOFFMANN-RIEM, WOLFGANG. Teoria Geral do Direito Digital: desafios para o direito. 2ª ed. Rio de Janeiro: Forense, 2022.

MENDES, GILMAR FERREIRA. Lecture delivered to the class of Emerging Challenges in Social Media Regulation: Resurgent States on 03.31.25, IDP/PUCRS/DCU, 2025.

MENDES, Gilmar Ferreira; FERNANDES, Victor Oliveira. Constitucionalismo digital e jurisdição constitucional: uma agenda de pesquisa para o caso brasileiro. **Revista Brasileira de Direito**, v. 16, n. 1, p. 1-33, 2020, DOI: 10.18256/2238-0604.2020.v16i1.4103. Available on: <https://seer.atitus.edu.br/index.php/revistadedireito/article/view/4103/2571>. Access on: jul. 3rd. 2025.

MPMT. 62% da população global está nas redes sociais, diz estudo. Conteúdo, notícias, 02.02.2024. Available on: <https://www.mpmt.mp.br/conteudo/1217/134871/62-da-populacao-global-esta-nas-redes-sociais-diz-estudo>. Access on: jul. 03rd 25.

POSSA, Alisson Alexandre. **A concretização da dignidade humana na era das nanotecnologias: o direito à liberdade cognitiva como neurodireito na ordem constitucional brasileira**. 2023. 100f. Dissertação (Mestrado Acadêmico em Direito) - Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, Brasília, 2023.

SARLET, Ingo Wolfgang. Lecture delivered to the class of Emerging Challenges in Social Media Regulation: Resurgent States on 03.31.25, IDP/PUCRS/DCU, 2025.

STEIN, Ernildo. **Aproximações Sobre Hermenêutica**. 1996. Porto Alegre: EDIPUCRS.

USA. Constitution of the United States, First Amendment, Congress, Constitution Annotated, 1789. Available on:

<https://constitution.congress.gov/constitution/amendment-1/>. Access on: jul. 4th. 2025.

WACHOWICZ, Marcos; LANA, Pedro de Perdigão. Entendendo a fragmentação da Internet a partir de aspectos fundamentais sobre regulação, soberania digital e a experiência da União Europeia. *In:* **Direito e Ciberespaço**: Coletânea de Artigos da Revista Digital *Cyberlaw by CIJIC.* Coord. Eduardo Vera-Cruz Pinto e Marco Antonio Marques da Silva. Paraná: Editora Quartier Latin. ISBN-10: 6555752068, 2024. Available on: <https://gedai.ufpr.br/wp-content/uploads/2024/03/Artigo_Internet-Regulacao-Fragmentacao.pdf>. Access on: jul. 06. 2025

ZENHA, LUCIANA. Redes sociais online: o que são as redes sociais e como se organizam?. **Revista UEMG**, Caderno de Educação, Belo Horizonte, ano 20, n. 49, v. 1, 2017/2018, p. 19-42.

10. CONTENT MODERATION IN BRAZIL: PLATFORM SELF-REGULATION AND THE 2025 JUDICIAL SHIFT ON LIABILITY

https://doi.org/10.36592/9786554603065-09

Larissa de Lima e Campos[1]

Abstract:
This work analyzes the content moderation regime of digital platforms in Brazil, addressing its challenges, the predominance of self-regulation, and the recent interpretations of the Brazilian Supreme Federal Court (STF) on the Brazilian Civil Rights Framework for the Internet (MCI). Initially, the text explores seven reflections on the future of content moderation on digital platforms made by the Institute of Technology and Society. Subsequently, it details the Brazilian self-regulatory model, highlighting its advantages compared with state regulation. Finally, the work discusses the liability of application providers under the MCI, contrasting the general rule of Article 19 with the exception of Article 21, and presents the significant changes introduced by the thesis established by the STF in June 2025.
Keywords: Content Moderation, Brazilian Civil Rights Framework for the Internet, Self-Regulation, Digital Platforms, Civil Liability, Brazilian Supreme Federal Court

1. Introduction

Internet and cyberspace regulation has been a persistent challenge since early attempts, such as the Telecommunications Act of 1996 in the United States. The approval of this law prompted Barlow (1996) to draft the *"Declaration of the Independence of Cyberspace"*, a classic text that, despite marking the beginning of internet history, is not immune to criticism (Morrison, 2009).

The Declaration conceived the internet as *"the new home of mind"*: a place with no state, *"no elected government"*, governed solely by the *"Golden Rule"*. In essence, it was a call for freedom and self-regulation, arguing that cyberspace should develop without government interference, as governments did not understand the internet and users had not consented to any form of regulation.

[1] Holds a European Master in Law, Data and AI (EMILDAI) from Dublin City University. Graduated from FGV Direito Rio. Lawyer specializing in Digital Law at BFBM Advogados.

Almost 30 years later, searching for the best way to regulate digital platforms remains a global challenge for governments. Regulating these platforms involves a set of norms that shape and control content moderation. But, limiting ourselves to the Brazilian scenario, how does content moderation currently operate on digital platforms?

2. Understanding content moderation

According to Silva et al. (2025), content moderation can be viewed in two ways. A more restricted view, predominant in public debate, refers to the activity platforms undertake to ensure compliance with their own behavioral rules, established in their Terms of Use and Community Policies. The broader view includes, in addition to applying sanctions, all tools that interfere with user interactions, such as content recommendation. This work adopts the more restricted view of content moderation.

Currently in Brazil, content moderation operates predominantly through self-regulation, with some legal obligations to be observed. Platforms have the freedom to define their internal rules but must adhere to limits imposed by Brazilian law. Before exploring self-regulation and legal obligations, it is essential to delve deeper into the concept of content moderation.

Recently, in January 2025, the Attorney General's Office (AGU) held a public hearing on the Digital Platforms Content Moderation Policy, and contributing to this debate, the Institute of Technology and Society (ITS) presented seven reflections on the future of content moderation on digital platforms (Souza, 2025a). As an independent, non-profit research institution with over ten years of expertise in the digital universe[2], analyzing an ITS document is of great value.

The document presents seven crucial reflections on the future of the content moderation debate. The first reflection addresses who should perform content moderation. The answer is clear: content moderation is not an exclusive activity of large technology companies. ITS reminds that online encyclopedias, public authorities, and companies developing AI applications, among others, also perform

[2] Available at: https://itsrio.org/en/institutional/. Accessed on: 28 June 2025.

content moderation. This distinction is vital because regulating moderation solely with large tech companies in mind can significantly burden smaller companies that also engage in this activity. Any law or public policy on the matter must be clear about the different figures involved.

The institute's second reflection addresses the legal nature of content moderation rules. Although the topic will be further explored later, the document states that the Terms of Use and Community Guidelines are a result of the private autonomy exercised by those responsible for these private spaces. By agreeing to these norms, the user enters into a contractual relationship. However, for the past decade, there has been a discussion about whether these private norms should promote fundamental rights and serve public interests. According to Mendes & Fernandes (2020), there is no clarity in Brazil as to whether the application of the direct effect doctrine in content moderation issues would be in line with the Supreme Federal Court's jurisprudence.

The third reflection suggests viewing moderation as a process, not limiting it to a single decision about content. Moderation involves prior and subsequent activities that must also be observed. ITS lists ten constituent elements of this moderation process, suggesting them as a checklist of good practices.

The fourth reflection warns against the risk of evaluating content moderation based on an isolated error or success, without considering its overall functioning. It is impossible to be 100% accurate, and when the volume of content to be analyzed is very large, the chances of error also increase. Illustratively, the institute presents some concrete data: from July to November 2024, YouTube removed over 9 million videos, Meta moderated over 1 billion accounts, and TikTok removed, in Brazil alone, 5,325,026 videos, with 89.7% of them within 24 hours.

The fifth reflection highlights that moderation is a product of time and space. Moderation decisions are directly influenced by the country and culture in question, as well as by the social perception of a given topic over time. For example, ITS mentions the USA which, due to its construction regarding freedom of expression, allows some content that would be considered illicit in other jurisdictions.

The penultimate reflection addresses how moderation is affected by technological trends. The emergence of new technologies, such as artificial intelligence applications, modifies how companies can perform content moderation. ITS also mentions the use, by some social networks, of a decentralized moderation model known as *"community notes"*.

Finally, the last reflection emphasizes the important role of the Oversight Boards. The purpose of these boards is to create a space for critical analysis of the moderation practices adopted by companies, avoiding a casuistic analysis defined by an error that gained media prominence. Having explored the topic, the next section will develop how digital platforms self-regulate in Brazil.

3. Self-regulation of digital platforms in Brazil

In the Brazilian context, content moderation by digital platforms operates predominantly through self-regulation, via a set of internal policies and content moderation mechanisms established by the companies themselves. Internal policies are expressed through Terms of Use, which, upon acceptance by users, establish what is permitted and what is prohibited.

Acting in the regular exercise of their rights and guided by their terms of use, digital platforms manage content in several ways: they remove user-reported content, reject complaints, take down algorithm-flagged material, and legally challenge content decisions. Harmful content can be identified and removed proactively using platform tech, by judicial order, or via user reports. The reported content is subjected to a review process, often aided by AI and human moderators, which varies by platform.

According to Silva et al. (2025), with the growth of platforms and large-scale content production, the only effective solution for content moderation is the use of automatic content detection tools as support for review teams or as an autonomous decision system. However, the authors emphasize that this is a very recent technology that still needs improvement for its decisions to be reliable and have a low error rate.

According to Zittrain (2020, *apud* Barroso, 2022), since 2010 there has been a growing critical movement challenging the approach taken by digital platforms and States concerning content in the online environment. The material displayed by platforms began to be controlled by algorithms that seek to maximize engagement, regardless of the author's good faith or caution. As the human tendency is to engage with sensationalist, defamatory, and misleading content, the digital environment has become conducive to user manipulation (Barroso, 2022).

However, despite the dissatisfaction, it was observed that state regulation on its own is also inefficient and dangerous. First, effective regulation requires collaboration with the companies that develop cyberspace codes; as their algorithms directly impact law enforcement (Keller, 2019, *apud* Barroso, 2022). Second, statelaws struggle to keep pace with rapid technological advancements, quickly becoming obsolete. Platforms, however, offer the flexibility and speed needed to innovate and address emerging issues like disinformation and hate speech (Barroso, 2022). Third, governments often lack a deep technical understanding of the virtual environment, leading to generic regulations that stifle innovation (Belli et al., 2017, *apud* Barroso, 2022). Tech companies possess this crucial expertise to develop tailored solutions. Fourth, empowering the State to define "illicit" content risks information control and suppressing dissent. Platforms, being non-political, are better positioned to foster environments that uphold freedom of expression and other fundamental rights (Zittrain, 2020, *apud* Barroso, 2022).

However, as mentioned earlier, although content moderation by digital platforms functions predominantly through self-regulation, some legal obligations need to be observed. The following section is dedicated to explaining some of these obligations.

4. Legal obligations and new interpretations

The main law governing internet use in Brazil is the Brazilian Civil Rights Framework for the Internet (MCI). The MCI emerged from a popular construction through online consultations promoted by the Ministry of Justice. Among its provisions, the MCI regulates intermediary companies that provide the structure for

third parties to post their content in the digital environment. These companies, classified as application providers, can be held responsible for content generated by third parties; that is, they can be held responsible if they do not correctly exercise content moderation on their digital platforms.

Recently, in June 2025, the Brazilian Supreme Federal Court (STF) judged the constitutionality of Article 19 of the MCI. The text of Article 19 of the MCI establishes the general rule of liability: application providers will only be civilly liable for damages caused by third-party content if, after a judicial order, they fail to take steps for its removal. Its first paragraph stipulates that a judicial order must clearly and specifically identify the infringing content for unequivocal removal.

It is clear that, based on the law, there is no prior obligation to monitor or remove content, and an extrajudicial notification is not sufficient to generate the liability of the application provider. This type of norm privileges freedom of expression by minimizing the risks of collateral censorship and excessive removal. This model does not require companies to make value judgments about the legality of certain content for fear of being held liable later (Barroso, 2022).

As mentioned, Article 19 is the general rule of liability. Article 21 of the MCI is the exception to this rule, applying to the content of nudity or private sexual acts published without the victim's consent (e.g., revenge porn). In this case, unlike Article 19, the application provider must remove the content after extrajudicial notification from the participant or their legal representative. Here, therefore, no prior judicial decision is required. The logic behind this exception is threefold: sensitive content, the need for rapid removal for the immediate protection of the victim, and the assumption that the illegality of this type of content is more objectively verifiable than in cases of violation of honor, privacy, disinformation, and others (Barroso, 2022).

The Brazilian Civil Rights Framework for the Internet does not prohibit companies from spontaneously removing content based on the violation of their Terms of Service. The Superior Court of Justice already recognized the legitimacy of this proactive conduct by platforms (Superior Tribunal de Justiça, 2024).

In a competitive environment like the virtual one, reputation speaks louder. Platforms that constantly fail to regulate freedom of expression become nests of disinformation and hate speech, leading to the exodus of users and advertisers.

Therefore, even without extrajudicial notification or judicial decision, platforms can identify and remove content that violates their private norms. If the user disagrees with the content moderation performed by the platform, they may resort to the Judiciary to question the removal. An example is the case of Canal Terça Livre on YouTube, where the Judiciary overturned the platform's decision to suspend the channel (Sestrem, 2021). According to Souza (2025b), the Judiciary will always have the final word. The above provisions prove that, although digital platforms self-regulate in Brazil, there is no scenario of impunity.

However, not everyone was satisfied with the regulatory framework defined by the MCI. Authors like Schreiber (2015) argue for the unconstitutionality of Article 19 for violating the constitutional guarantee of full and integral reparation for damages to honor, privacy, and image, by limiting the reparation of damage to the non-compliance with a judicial order. According to Schreiber (2015), damages are caused at the moment of publication and not when the non-compliance with the judicial order occurs. It is evident, however, that Article 19 does not exempt the user who published the removed content from accountability for damages incurred since the time of publication. Platforms can even assist by providing IP addresses that allow the identification of those responsible for the publications.

The debate is not simple, and there are strong arguments on both sides. Attorney Eduardo Mendonça, in his oral argument before the STF[3], pointed out that disinformation, hate speech, anti-democratic manifestations, and intolerance are not problems created by the Internet; they exist outside it and would not disappear if Article 19 was declared unconstitutional. For Mendonça, by requiring a court order, Article 19 safeguards controversial content from removal, thus ensuring it is subject to judicial challenge. In his words: "*It would make no sense to hold a platform responsible for not having removed content whose examination is controversial and subject to subjective valuations, which are often a matter of division within the Judiciary itself*" (free translation).

[3] Available at: https://www.youtube.com/watch?v=NH3hfqkmoKE. Accessed on: 28 June 2025.

As mentioned, the STF judged the constitutionality of Article 19. According to the Court's president, Luís Roberto Barroso, the tribunal waited for years for a regulation on digital platforms made by the National Congress, but as it did not occur, "*we have to decide the cases that arrive here*" (Boechat, 2025). This statement by the minister is an attempt to counter the main criticism at the STF: that the Judiciary is legislating and, therefore, interfering with the other Powers of the Republic (Cavalcante, 2025).

This work was completed the same week the STF established its thesis on Article 19 (Rocha, 2025). Therefore, the author opts to only present the thesis introduced by the Court. The STF understood, by majority vote, that Article 19 is partially unconstitutional because it does not provide sufficient protection for fundamental rights and democracy.

Until a new legislation arrives, Article 19 of the MCI holds internet application providers civilly liable, with specific exceptions for electoral law. This work also covers crimes against honor, allowing content removal via extrajudicial notification. The STF clarifies that under Article 21, platforms are civilly liable for damages from third-party content in cases of crime or illicit acts and still must remove it. This includes inauthentic accounts and repeated uploads of offensive acts already ruled on by courts.

The STF also announced a hypothesis of presumed liability in cases of illicit content in (i) paid advertisements and boosts; or (ii) artificial distribution networks (chatbots or robots). In these cases, removal must occur independently of judicial decision and notification. Platforms can be exempt from liability if they prove that they acted diligently and within a reasonable time to remove the content.

The ruling also established a duty of care for platforms regarding the widespread circulation of serious illicit content. Providers will be liable if they don't immediately remove content related to an exhaustive list of severe crimes: anti-democratic acts, terrorism, inducement to self-harm, discrimination (based on race, color, ethnicity, religion, national origin, sexuality, or gender identity), crimes against women, sexual crimes against vulnerable persons/child pornography/crimes against children and adolescents, and human trafficking.

Liability in this case would arise from a systemic failure. If the content exists in isolation, the liability regime of Article 21 will apply. The party responsible for the publication of the removed content may judicially request its restoration by demonstrating the absence of illegality, but even if the content is restored, no indemnity will be imposed on the provider.

The STF also created additional duties for application providers, who must establish self-regulation encompassing notification systems, due process, and annual transparency reports. Specific service channels must also be permanently available, and providers operating in Brazil must establish and maintain a headquarters and representative in the country. Finally, the established thesis explicitly states that there will be no objective liability and that the effects were modulated so they do not affect past cases.

5. Conclusion

This work aimed to present the content moderation regime on digital platforms in Brazil. Based on seven reflections drafted by ITS, the analysis first addressed important points on content moderation. Subsequently, it detailed how self-regulation of digital platforms is carried out in Brazil and its main advantages over state regulation. Finally, the work presented the main forms of liability for application providers regarding content moderation, according to the text of the law and the new thesis established by the STF.

Although the text focused on content regulation and moderation within the Brazilian legal system, the digital environment is singular. Consequently, decisions made in Brazil influence (and are influenced by) other jurisdictions. Two practical examples of this globalization within the Brazilian scenario are the blocking of WhatsApp by Brazil in 2019, which technically impacted the application in Argentina and Chile (Caputo, 2015), and Minister Alexandre de Moraes's determination for content moderation on the "X" platform, even when posts were made by individuals located in the United States (Martins, 2025). The truth is that a message written in Brazil can be read in Japan, routed through a European server, and on a US-based

platform. This intricate global web means digital content regulation and moderation always have worldwide repercussions.

6. References

Barlow, P. J. (1996). *A Declaration of the Independence of Cyberspace*. https://www.eff.org/cyberspace-independence

Barroso, L. V. B. (2022). *Liberdade de expressão e democracia na era digital* [Freedom of expression and democracy in the digital age]. Fórum.

Belli, L., et al. (2017). *Platform regulations: how platforms are regulated and how they regulate us*. Leeds. https://hdl.handle.net/10438/19402

Boechat, G. (2025). *Barroso diz que STF julga redes por falta de lei aprovada pelo Congresso* [Barroso says Supreme Court judges social networks due to lack of law passed by Congress]. CNN Brasil. https://www.cnnbrasil.com.br/politica/barroso-diz-que-stf-julga-redes-por-falta-de-leiaprovada-pelo-congresso/

Caputo, V. (2015). *Bloqueio no Brasil tira WhatsApp do ar na Argentina e Chile* [Block in Brazil takes WhatsApp off air in Argentina and Chile]. Exame. https://exame.com/tecnologia/bloqueio-no-brasil-tira-whatsapp-do-ar-na-argentina-echile/

Cavalcante, I. (2025). *'Judiciário não está legislando', diz Barroso sobre julgamento do Marco Civil* ['Judiciary is not legislating,' says Barroso about the Civil Rights Framework judgment]. Consultor Jurídico. https://www.conjur.com.br/2025-jun-04/judiciario-nao-esta-legislando-diz-barroso-sob re-julgamento-do-marco-civil/

Keller, C. I. (2019). *Regulação nacional de serviços na Internet: exceção, legitimidade e o papel do Estado.* [National regulation of Internet services: Exception, legitimacy and the role of the State] Universidade do Estado do Rio de Janeiro. http://www.bdtd.uerj.br/handle/1/9210

Martins, L. (2025). *Moraes quer censurar contas dentro dos EUA, diz advogado do Rumble* [Moraes wants to censor accounts within the US, says Rumble's lawyer]. CNN Brasil. https://www.cnnbrasil.com.br/politica/moraes-quer-censurar-contas-dentro-dos-eua-d iz-advogado-do-rumble/

Mendes, G. F., & Fernandes, V. O. (2022). Eficácia dos direitos fundamentais nas relações privadas da internet: o dilema da moderação de conteúdo em redes sociais na perspectiva comparada Brasil-Alemanha [Efficacy of fundamental rights in

private internet relations: The dilemma of content moderation on social media in a comparative Brazil-Germany perspective]. *Revista de Direito Civil Contemporâneo*, (31), 33–68. https://ojs.direitocivilcontemporaneo.com/index.php/rdcc/article/view/1107

Morrison, A. H. (2009). An impossible future: John Perry Barlow's 'Declaration of the Independence of Cyberspace'. *New Media & Society*, *11*(1-2), 53–71. https://doi.org/10.1177/146144480810016

Rocha, P. (2025). *STF define parâmetros para responsabilização de plataformas por conteúdos de terceiros* [Supreme Court defines parameters for platform liability for third-party content]. Supremo Tribunal Federal. https://noticias.stf.jus.br/postsnoticias/stf-define-parametros-para-responsabilizacaode-plataformas-por-conteudos-de-terceiros/

Schreiber, A. (2015). *Marco Civil da Internet: avanço ou retrocesso? A responsabilidade civil por dano derivado do conteúdo gerado por terceiro.* [Brazilian Civil Rights Framework for the Internet: Advance or setback? Civil liability for damages derived from third-party generated content].

Sestrem, G. (2021). *Justiça determina que Google reative canal do Terça Livre no YouTube* [Court orders Google to reactivate Terça Livre's channel on YouTube]. Gazeta do Povo. https://www.gazetadopovo.com.br/vida-e-cidadania/justica-determina-que-google-re ative-canal-terca-livre-youtube/

Silva, A. P., Guimarães, T., & Salvador, J. P. F. (2025). *Moderação de conteúdo: desafios jurídicos e sociais* [Content moderation: Legal and social challenges]. Almedina Brasil.

Souza, C. A. P. de. (2025a). *7 reflexões para o futuro do debate sobre moderação de conteúdo em plataformas digitais* [7 reflections for the future of the debate on content moderation on digital platforms]. ITS Rio. https://itsrio.org/wp-content/uploads/2017/01/Relatorio-Reflexoes-Sobre-Moderacaode-Conteudo.pdf

Souza, C. A. P. de. (2025b). *Tribunais devem prestigiar o discurso humorístico, inclusive o crítico* [Courts should privilege humorous discourse, including critical]. Consultor Jurídico. https://www.conjur.com.br/2015-jan-22/carlos-souza-tribunais-prestigiar-discurso-hu moristico/

Superior Tribunal de Justiça. (2024). *Provedor não precisa de ordem judicial para remover conteúdo contrário aos seus termos de uso* [Provider does not need a court order to remove content contrary to its terms of use]. Superior Tribunal de Justiça. https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2024/111120

24-Pr ovedor-nao-precisa-de-ordem-judicial-para-remover-conteudo-contrario-aos-seus-ter mos-de-uso-.aspx

11. THE PRIVATIZATION OF THE PUBLIC AND EFFICIENCY AS A SMOKESCREEN: THE ELON MUSK EFFECT IN CENTER-RIGHT MUNICIPAL ADMINISTRATIONS IN BAHIA

https://doi.org/10.36592/9786554603065-10

Vitória Andréa De Almeida Nicolau

Abstract: This paper proposes a reflection on the contours that must be observed in the search for administrative efficiency, which should, yes, aim at greater productivity, but cannot replace other values of equal legal relevance, such as the principles of legality and impersonality. At this point, the first topic of the work presents a brief report on the context of the positivization of the principle of efficiency in article 37 of the Federal Constitution of 1988, promoted through Constitutional Amendment No. 19. Next, the vision of efficiency that has been disseminated by the center-right municipal administrations in Bahia, notably by the União Brasil party, is exposed. Finally, considerations are made about the case of the Municipality of Ilhéus, in the State of Bahia, which, in serious violation of the legal and constitutional order in force, withdrew legal powers of action from effective prosecutors through an infra-legal act, under the allegation that the measure sought to increase the productivity of the legal advisory body.
Keywords: Government efficiency; center-right public administration; Municipality of Ilhéus.

1. Introduction

According to the Michaelis Dictionary, one of the meanings of the word efficiency is *"attribute or condition of what is productive; performance, productivity, yield: Nothing compares to the efficiency of these imported machines"* (MICHAELIS, 2025).

It can be said that, currently, efficiency is one of the greatest objectives to be achieved by individuals, who aim to do more and more in less time, achieve their personal and professional goals earlier and earlier, overcome the goals established by their superiors in record time, and be recognized for surpassing themselves.

Producing more in less time is seen as being competent, and an example to

be followed. Those who fail to achieve the productivity standards that are successively expanded by an increasingly competitive society are left behind.

This has always been the logic of the private sector, of the business sector. However, what was once related to the corporate sphere has now extended to the public service and even to the personal lives of individuals. At work, in social circles, and in life, to exist is to produce!

It so happens that the Public Administration has particularities not shared with private bodies, such as the need to act in accordance with the principle of legality and legality. This is not saying that the private sector can act illegally. The point is that acting in the private sphere does not suffer the constitutional and legal implications that are typical of an entity that acts to achieve and protect the public interest, and not primarily to obtain profit.

In this step, this paper proposes a reflection on the contours that must be observed in the search for administrative efficiency, which should indeed aim at greater productivity, but cannot replace other values of equal legal relevance.

To this end, the first topic of the work presents a brief report on the context of the positivity of the principle of efficiency in article 37 of the Federal Constitution of 1988, through Constitutional Amendment nº. 19.

The second topic, in turn, exposes the vision of efficiency that has been disseminated by center-right municipal administrations in Bahia, and which is influenced by the perspective of government efficiency disseminated by Elon Musk during his brief stint in Donald Trump's second administration.

Finally, considerations are made about the case of the Municipality of Ilhéus, in the State of Bahia, which, in the name of productivity and in serious violation of the legal and constitutional order in force, published a decree attributing competence to a commissioned servant to issue legal opinions in bidding processes, removing such attribution from the effective attorneys who worked in the municipality's legal consultancy and advice until then.

2. The (In)Efficiency Of The State And The Public Servant In Check

The original wording of article 37 of the Federal Constitution provided that the direct, indirect or foundational public administration of any of the Powers of the Union, the States, the Federal District and the Municipalities shall obey the principles of legality, impersonality, morality and publicity.

By means of Constitutional Amendment No. 19, of June 4, 1998, the wording of the aforementioned provision was changed to include efficiency among the principles to be observed by the public administration, and modifications were promoted in relation to the legal regime of public servants, with changes, among others, to arts. 37, 38, 39 and 41 of the Federal Constitution.

In an explanatory memorandum addressed to the then President of the Republic, Fernando Henrique Cardoso, the Minister of Federal Administration and State Reform at the time, Luiz Carlos Bresser Pereira (1995), asserted that *"increasing the efficiency of the State apparatus is essential for the definitive overcoming of the fiscal crisis",* and that Constitutional Amendment No. 19 would be *"an innovation of bureaucratic administration with the purpose of combating patrimonialist administration and the use of dismissal as a political instrument".* Also according to Bresser Pereira, *"The intended objective is to value the civil service, increase its productivity, and reward the most competent more appropriately. Today, the civil service lives in the vicious circle of stability, inefficiency and low remuneration". It* is important to clarify that Justice Bresser was not against the stability of the civil servant; he only understood that it should vary according to the functional category. Thus, there should be: ***(i)*** a rigid stability for categories that perform exclusive functions of the State, whose members could only be dismissed for serious misconduct, through administrative or judicial proceedings, and; ***(ii)*** a more flexible stability for the other employees, who could be dismissed not only due to the practice of serious misconduct, but also due to insufficient performance or due to the need of the administration, resulting from an excess of staff or organizational or technological restructuring processes.

It should also be noted that the Administrative Reform promoted by Constitutional Amendment No. 19 was part of a broader project to redefine the role

of the State in the economic and social spheres. In fact, a year before the promulgation of EC No. 19, Luiz Carlos Bresser Pereira (1997) presented a paper to the second meeting of the Montevideo Circle and explained that:

> *The reform of the State involves four problems that, although interdependent, can be distinguished: (a) an economic-political problem - the delimitation of the size of the State; (b) another also economic-political, but which deserves special treatment - the redefinition of the regulatory role of the State; (c) an economic- administrative one - the recovery of governance or financial and administrative capacity to implement the political decisions made by the government; and (d) a political one - the increase in governability or the government's political capacity to mediate interests, ensure legitimacy, and govern. In the delimitation of the size of the State, the ideas of privatization, "publicization" and outsourcing are involved. The issue of deregulation concerns the greater or lesser degree of State intervention in the functioning of the market. In the increase of governance we have a financial aspect: overcoming the fiscal crisis; a strategic one: the redefinition of the forms of intervention in the economic and social sphere; and an administrative one: the overcoming of the bureaucratic way of administering the State. The increase in governability includes two aspects: the legitimacy of the government in the eyes of society, and the adequacy of political institutions for the intermediation of interests.*

Thus, the modification of the rules related to the civil service would be inserted in the State's strategy of increasing governance from an administrative perspective, with the overcoming of the bureaucratic way of administering the State, based on the idea of process control and not on the search for results or achievement of goals.

Despite the importance of the analysis of the other structural factors pointed out for the complete understanding of the administrative reform, the present work will focus on the aspects related to the civil service, an option that, despite consisting of the methodological reduction of the subject analyzed here, does not remove the relevance of the more specific approach undertaken. This is because, if it is true that the performance of civil servants is not the only element to be considered in the study of administrative efficiency, it is no less true that it is through these agents that the will of the State is manifested.

But how is the performance of public servants related to administrative efficiency? More than that, what is efficiency?

At the international level, traditionally, the principle of efficiency has come to be associated with good public administration, as can be seen from article 97 of the Italian Constitution of 1947[1] and article 103 of the Spanish Constitution of 1978[2]. However, the link between efficiency and good administration, far from bringing concreteness to the principle, amplified its imprecision. After all, the question "what is efficiency?" has been replaced by "what is good management?".

In this step, the efficiency of the Public Administration began to be correlated with its productivity. *"According to this conception, it would be up to the Administration to produce the maximum results at the lowest possible cost, just like an industry or private company, with a view to profit. However, the idea of productivity is associated with quantitative aspects, without concern for qualitative aspects"* (DIAS, 2009, p. 75). Seeking to explain what the principle of efficiency would represent, based on constitutional positivity, Odete Medauar argues that *"the word [efficiency] is linked to the idea of action, to produce results quickly and accurately"*, and that *"efficiency is opposed to slowness, negligence, negligence, omission – usual characteristics of the Brazilian Public Administration, with rare exceptions"* (2018, p. 127).

---

[1] **Articolo 97**
Le pubbliche amministrazioni, in coerenza con l'ordinamento dell'Unione europea, assicurano l'equilibrio dei bilanci e la sostenibilità del debito pubblico.
I pubblici uffici sono organizzati secondo disposizioni di legge, in modo che siano assicurati il buon tempo e l'imparzialità dell'amministrazione.
Nell'ordinamento degli uffici sono determinate le sfere di competenza, le attribuzioni e le responsabilità proprie dei funzionari.
Agli impieghi nelle pubbliche amministrazioni si accede mediante concorso, salvo i casi stabiliti dalla legge.

[2] **Artículo 103**
1. La Administración Pública sirve con objetividad los intereses generales y actúa de acuerdo con los principios de eficacia, jerarquía, descentralización, desconcentración y coordinación, con sometimiento pleno a la ley y al Derecho.
2. Los órganos de la Administración del Estado son creados, regidos y coordenados de acuerdo con la ley.
3. La ley regulará el estatuto de los funcionarios públicos, el acceso a la función pública de acuerdo con los principios de mérito y capacidad, las peculiaridades del ejercicio de su derecho a sindicación, el sistema de [3]incompatibilidades y las garantías para la imparcialidad en el ejercicio de sus funciones.

It should also be noted that, while the author recognizes the need for efficient action by the State, it does not fail to warn that the other administrative principles must be observed. In the words of the author, *"the principle of efficiency has been giving rise to an erroneous understanding in the sense that, in the name of efficiency, legality will be sacrificed*". However, *"the two constitutional principles of Administration must be reconciled, seeking to act efficiently, within the law".*

And it is in the wake of this alert that the present work will analyze the vision of efficiency that has been disseminated by the center-right municipal administrations in Bahia, in the terms outlined below.

3. The Elon Musk Effect On Public Service: Productivity, Regardless Of Illegality

Technology entrepreneur Elon Musk became even better known internationally after acquiring the X platform (formerly Twitter), which is currently owned by xAI, another company owned by the South African billionaire.

An advocate of broad freedom of expression on social networks, Musk is strictly opposed to the content moderation practices of digital platforms and state regulation on the subject. An avowed supporter of the President of the United States, Musk spent more than $250 million to help elect Donald Trump, who chose him to lead his Office of Government Efficiency, created in early 2025.

Having recently left office, Elon Musk's stint in the Trump II administration was controversial from the beginning. In February of this year, the businessman determined that an e-mail be sent to all federal government employees, with approximately 5 topics, asking them to explain the work they did in the last week. Failure to respond would be considered a resignation from office.

In response to the move, the national president of the American Federation of Public Employees, Everett Kelley, declared that *"it is cruel and disrespectful that hundreds of thousands of veterans who wear their second uniforms in the public service are forced to justify their professional duties to this incommunicado, privileged, and unelected billionaire who has never performed a single hour of honest public service in his life"* (SHELTON; MCKEND, 2025).

But the promise of productivity is today's siren song, especially with the popularization of generative AI. And if the song is echoed by a famous and successful businessman who is committed to transforming the public service by making it as efficient as private companies, it is heard in the four corners of the world.

In Bahia, Musk's speech was carefully observed by public agents affiliated with the União Brasil party.

Resulting from the merger between the Democrats Party (DEM) and the Social Liberal Party (PSL), União Brasil *"declares itself to be a social liberalist, considered a strong defender of human rights and civil liberties, believing that the State can play the role of regulator in the economy, in order to guarantee the population quality access to essential and fundamental public services. such as health, education, security, freedom, housing and sanitation"* (2021).

In view of the political orientations classified as right and left, União Brasil is framed as a party whose ideological option seeks a balance between individual freedom and social well-being, which is why it is considered a moderate right-wing or center-right party.

An association that opposes the Workers' Party - PT in Bahia, União Brasil won 21 of the 32 direct confrontations against the PT in the 2024 municipal elections.

The PT, in turn, won in eight municipalities, and three other cities had mayors elected from other parties who participated in the electoral dispute (MONTEIRO, 2024).

The municipalities in Bahia in which União Brasil won were: Araçás, Barreiras, Cairu, Conceição do Coité, Conde, Dom Macedo Costa, Feira de Santana, Ibicaraí, Ilhéus, Lauro de Freitas, Mansidão, Mata de São João, Ourolândia, Quixabeira, Rafael Jambeiro, Santo Amaro, Santo Estêvão, São Gonçalo dos Campos, Senhor do Bonfim, Uruçuca and Vitória da Conquista.

Considering the municipalities conquered by União, Feira de Santana, Vitória da Conquista, Lauro de Freitas and Ilhéus are those with the largest number of inhabitants.

The transition team of the mayor-elect of Feira de Santana declared that José Ronaldo should prioritize transparency and efficiency in the execution of public policies in his administration (SILVA, 2024). The mayor of Vitória da Conquista, Sheila Lemos, promoted an administrative reform in the municipal Executive and created, among other bodies, the Special Secretariat for Public Transformation, which *"will act in the promotion and coordination of studies and discussions on the transformation of the State, through measures on administrative organization, civil servants, employees, technology and provision of public services".* This Special Secretariat will also be responsible for *"proposing and coordinating projects and initiatives aimed at administrative simplification, efficiency, effectiveness in the provision of public services and the expansion of state capacity"* (PMVC, 2025).

In an article published on its official website on 04/17/2025, União Brasil declared that *"in the first months of the administration of Mayor Débora Regis, Lauro de Freitas has demonstrated a commitment to transparency and efficiency in public management, reorganizing services and adopting measures that should generate savings estimated at R$ 9 million this year"* (2025).

In the same sense, the mayor of Ilhéus began his administration with the publication of twenty-seven decrees that, according to Valderico Reis Júnior, *"aim to reorganize the municipal administration and promote improvements for the population. The measures, which are already in force with the publication in the Official Gazette of the municipality, seek to contain expenses, optimize the public budget and implement more efficient management"* (MANDATE BAHIA, 2025).

However, it was another decree published by the current Mayor of Ilhéus that gained prominence at the beginning of his term: Decree No. 268, of January 14, 2025, appointing a commissioned servant to occupy the position of Manager of the Preparatory Bidding Center and attributing to her the competence to issue legal opinions in bidding processes, replacing the effective attorneys who worked in the municipality's legal consultancy and advice until then.

It is about this peculiar situation, which involves productivity and illegality, that we will discuss below.

4. The Case Of The Municipality Of Ilhéus.

Modern Times, United States, 1936. Starring Charles Chaplin, the black-and-white film gains a new color with the demands of the present day. Right at the beginning of the film, a warning: *"Modern Times is a story about industry, private enterprise and humanity in search of happiness".*

Touched like cattle, workers enter a factory en masse. The president of the company observes the employees through a large screen, demanding that the controller of the gears increase the speed of the machines, and to check the increase in production. Chaplin's character, Little Tramp, has the job of tightening screws, which must be done almost uninterruptedly. Almost, because employees could still stop to eat.

But one day a sales representative presents the factory president with the Bellows Feeding Machine, a technological innovation that would allow employees to work while eating. In the words of the contraption's salesman, *"the feeder machine eliminates lunchtime, increases production and reduces idle time."*

Hoping to stay ahead of the competition by eliminating the lunch stop, Little Tramp's boss decides to test the machine's efficiency by choosing him to participate in the experiment. After fainting due to several problems presented by the device during the tests, the boss decides not to purchase the feeder machine because he considers it not practical.

But as the factory still wanted to reach the peak of productivity, in the late afternoon its president determined that the speed of the machines be increased to maximum capacity. Little Tramp had a nervous breakdown from tightening screws so much, and left the factory straight to the hospital after trying to tighten all the round objects he saw in front of him.

In a mix of drama and comedy, Modern Times is more than a critique of capitalism, Nazi-fascism and imperialism; It is a warning about the abuses and illegalities that the most varied production and management systems can inflict on workers in general, regardless of the nature of the relationship (public or private).

As already mentioned, Decree No. 268/2025, published by the Municipality of Ilhéus, conferred on the Manager of the Preparatory Bidding Center the competence

to issue a legal opinion in the processes of bidding, direct contracting, agreements, terms of cooperation, covenants, adjustments, adhesions to price registration minutes, other similar instruments and their addendums and other procedures linked to the bidding sector.

By means of an infra-legal act, the Head of the Executive Branch appointed a commissioned servant to act in the legal advice of any and all issues involving public bids and contracts, replacing career municipal prosecutors. And what was the justification for this? Control and productivity, as made clear by the "recitals" presented at the beginning of the decree:

> ***Considering*** *the* ***restricted staff of the Attorney General's Office and the need*** *to implement and execute better conducts to* ***meet the large scale of demands of the municipality****;* ***Considering*** *the need to value the staff of Attorneys, including the realization of a technical study for knowledge and presentation of solutions,* ***so that the management can achieve greater productivity*** *and better return for the municipality;*
> *Considering also the* ***need to speed up the bidding procedures*** *for the acquisition of goods and services for the Municipality, with a sufficient technical staff of employees before the Bidding Center;*
> ***Considering*** *the* ***need for greater monitoring and control over the contracting processes****, as well as other acts linked to the bidding sector.*

Thus, Municipal Decree No. 268/2025 not only removed from the Attorney General's Office of the Municipality of Ilhéus and career municipal prosecutors the attributions of examining and issuing an opinion on the bidding documents and legal instruments of contracts, agreements and other adjustments in which the Direct and Indirect Administration is a party or interested , which is illegal, for frontally violating Municipal Law No. 4,025/2019[3] – but created new attributions for the position of Manager of the Preparatory Bidding Center , contrary to article 37, item II, of the

[3] Law nº. 4,025/2019, which establishes the Organic Law of the Attorney General's Office of the Municipality: Article 5 It is incumbent upon the PGM: (...) X – To examine and give an opinion on the legal instruments of contracts, agreements and other adjustments in which the Direct, Autarchic and Foundational Administration is a party or interested; XI – To previously examine bidding notices of interest to the Direct, Autarchic and Foundational Administration.

Federal Constitution[4], and article 78 of Municipal Law No. 4,236/2023[5][6], being, therefore, illegal and unconstitutional , all so that, in the mayor's view, the management could achieve greater productivity.

In view of the distortion of the idea of public management and the numerous violations of the constitutional and legal order, the municipal prosecutors made a representation to the Public Prosecutor's Office of the State of Bahia, which in the person of the prosecutor Alícia Violeta Botelho Sgadari Passeggi, issued a recommendation in the following sense:

> *[...]*
> *CONSIDERING, the Federal Supreme Court, in the judgment of*
> *ADI No. 6331, consolidated the understanding that, once a proper prosecutorial body is established at the municipal level, the composition of its technical staff must comply with constitutional norms, especially the inescapable duty to fill positions by public competition, as provided for in article 37, item II, of the Federal Constitution;*
> *CONSIDERING, the designation of a public servant without a link to the Attorney General's Office of the Municipality to issue legal opinions in bidding processes, direct contracts, adjustments and similar instruments seriously violates the legality of administrative acts, compromises impartiality and exposes bidding procedures to legal risks and nullities;*

[4] Federal Constitution of 1988:
Article 37. The direct and indirect public administration of any of the Powers of the Union, of the States, of the Federal District and of the Municipalities shall obey the principles of legality, impersonality, morality, publicity and efficiency, and also the following:
[...]
II - investiture in public office or employment depends on prior approval in a public examination of tests or tests and titles, according to the nature and complexity of the position or employment, in the manner provided for by law, except for appointments to a position in commission declared in a law of free appointment and dismissal.

[5] Law nº. 4,236/2023, which provides for the Administrative Structure within the scope of the Municipal Executive Branch:
Article 78. It is incumbent upon the Manager of the Preparatory Bidding Nucleus: I - to direct the acts that are part of the bidding processes, in the various modalities for the acquisition of goods and contracting of services, supervising all stages; II – to determine the correct organization and archiving of the processes corresponding to bids or direct contracts; III - to coordinate the maintenance services of the suppliers' registration records, as well as the issuance of the respective certificates; IV - to make decisions in favor of the proper conduct of the bidding, boosting the procedure, including demanding from the internal areas of the purchasing sector, the reorganization of the preparatory phase, if necessary; V - to monitor the bidding procedures, promoting diligences, if applicable, so that the contracting calendar is fulfilled on the scheduled date; VI - to perform other activities related to the position.

CONSIDERING, the appointment of a person outside the staff of the Attorney General's Office to perform functions inherent to the body constitutes a flagrant violation of the principle of legality, compromising the regularity of administrative acts and affronting the constitutional duty that such functions be performed exclusively by professionals who have passed public examinations and are members of the legal career of the Municipality;

[...]

IT IS RECOMMENDED:

To the Honorable Mayor of the Municipality of Ilhéus, Mr.

Valderico Luiz dos Reis Júnior, who:

A. Repeal Municipal Decree No. 268/2025, published on 01/15/2025, due to the irremediable illegality existing in articles 1, 2 and 3, rendering ineffective the appointment of Ms. ANA CAROLINA MENEZES DANTAS to the Position of Manager of the Preparatory Bidding Center, extinguishing any attributions related to legal advice in bidding processes, direct hiring, agreements, terms of cooperation, agreements, adjustments, adhesions to price registration minutes and other procedures linked to the bidding sector, especially the issuance of legal opinions improperly instituted by decree;

B. Adopt the necessary measures to ensure the faithful compliance with Law No. 4,025/2019, refraining from appointing or maintaining the appointment of any public servant for the exercise of functions related to the Attorney General's Office of the Municipality, ensuring that the activities of consulting and issuing legal opinions are carried out exclusively by the members of the competent body, under the terms of the current legal system.

C. To determine the referral of ALL procedures related to public contracts in which an opinion was issued by the aforementioned nominee for regular analysis by the members of the Attorney General's Office of the Municipality of Ilhéus duly invested in public office, suspending their processing until the defect of absence is remedied.

A period of five (5) days, counted from the notification of this Recommendation, is established for compliance and sending to this Public Prosecutor's Office information regarding the adoption of effective measures for the faithful fulfillment of this Recommendation.

It should be noted that failure to comply with this Recommendation may lead to the adoption of judicial and extrajudicial measures applicable to the species.

After the repercussion of the ministerial recommendation in the press, and in view of the fear of filing a public civil action against him, the mayor of the Municipality of Ilhéus issued a new decree determining *"the return to the usual*

activities of the civil servant occupying the position of Manager of the Preparatory Bidding Center",[6] without recognizing the unconstitutionalities and illegalities that were perpetrated under his command.

Even with the constitutional and legal order reestablished, at least temporarily, it is necessary to remain alert. After all, in these "modern times" one never knows when the discourse of efficiency will emerge again, with the same aspect or in a new guise, snatching rights and subverting public administration.

5. Conclusion

As seductive as the corner of efficiency may be, one cannot lose sight of the need to observe the other guiding principles of public administration, such as morality and legality.

The public service also needs to be exercised with excellence, but not at the expense of the rights of its employees, and there must be a reconciliation between efficiency and morality, productivity and legality, so that the organizational advancement of the public administration is based on solid bases, capable of contributing to the economic and social development of the community in which it is inserted.

References

BRAZIL. **Constitutional Amendment No. 19, of June 4, 1998**. Modifies the regime and provides for the principles and rules of the Public Administration, civil servants and political agents, control of expenses and public finances and costing of activities under the responsibility of the Federal District, and provides for other provisions.
Available at: <https://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc19.htm>. Accessed on: 30 May. 2025.

BRAZIL. **Constitution of the Federative Republic of Brazil of 1988**. Available at: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Accessed

[6] Municipal Decree nº. 432, of March 10, 2025, which *"decrees the return to performance in the usual activities of the civil servant occupying the position of Manager of the Preparatory Bidding Center, pursuant to article 78 of Law No. 4.236, of August 01, 2023"*

on: 30 May. 2025.

CABRAL, Flávio Garcia. **The legal content of administrative efficiency**. 2nd ed. Belo Horizonte: Forum, 2024.

DIAS, Jefferson Aparecido. **Principle of Efficiency & Administrative Morality**: the submission of the principle of efficiency to administrative morality in the Federal Constitution of 1988. 2nd ed. Curitiba: Juruá, 2009.

ESPAÑA. **Constitución Española**. Available at: <https://www.tribunalconstitucional.es/es/tribunal/normativa/Normativa/CEportugu%C 3%A9s.pdf>. Accessed on: 30 May. 2025.

ISLANDERS. **Decree No. 432, of March 10, 2025**. Decrees the return to work in the usual activities of the civil servant occupying the position of Manager of the Preparatory Center
Bids, under the terms of article 78 of Law No. 4,236, of August 01, 2023. Available at: <https://www.ilheus.ba.gov.br/abrir_arquivo.aspx?cdLocal=12&arquivo={D6DBCCAC-CEB2-D4CA-5D1C-EDEAB5EA4C05}.pdf>. Accessed on: 20 May. 2025.

_____. **Decree No. 268, of January 14, 2025**. Designates a civil servant to act in legal advice to the contracting processes and other acts. Available at: <https://www.ilheus.ba.gov.br/abrir_arquivo.aspx?cdLocal=12&arquivo={6ABE76BA- 0BA2-03C2-D6A0-6E6E1BAEEABC}.pdf>. Accessed on: 20 May. 2025.

_____. **Law No. 4,236, of August 01, 2023**. Provides for the Administrative Structure within the scope of the Municipal Executive Branch, and provides for other provisions. Available at: <https://transparencia.ilheus.ba.gov.br/arquivo/legislacao/lei-ordinaria_4236_2023>. Accessed on: 25 May. 2025.

_____. **Law No. 4025, of August 29, 2019**. Establishes the Organic Law of the Attorney General's Office of the Municipality and provides for other provisions. Available at: https://www.ilheus.ba.gov.br/abrir_arquivo.aspx?cdLocal=12&arquivo={40BEADEA- EDBE-BA1E-5D60-BBCAB2CCA05C}.pdf>. Accessed on: 25 May. 2025.

ITALIA. **La Costituzione**. Available at: <https://www.senato.it/istituzione/la-costituzione/parte-ii/titolo-iii/sezione-ii/articolo-97>. Accessed on: 31 May. 2025.

MANDATE BAHIA. **Ilhéus under new management: decrees mark the beginning of the government Valderico Reis**: Actions aimed at austerity and improvements in public management. 02 Jan. 2025. Available at: <https://www.mandatobahia.com.br/noticia/valderico-reis-inicia-mandato-com-27- decretos-para-reorganizar-ilheus>. Accessed on: 27 May. 2025.

MEDAUAR, Odete. **Modern Administrative Law**. 21st ed. Belo Horizonte: Forum, 2018, p. 115-132.

MICHAELIS. **Brazilian Dictionary of the Portuguese Language**. Available at: <https://michaelis.uol.com.br/>. Accessed on: 01 Jun. 2025.

MONTEIRO, Raul. **In the direct confrontation, União Brasil defeated the PT in most of the municipalities of Bahia in this year 's elections**. Política Livre, 15 out. 2024. Available at: <https://politicalivre.com.br/2024/10/no-confronto-direto-uniao-brasil-derrotou-o-pt-na-maioria-dos-municipios-da-bahia-nas-eleicoes-deste-ano/#gsc.tab=0>. Accessed on: 28 May. 2025.

PEREIRA, Luiz Carlos Bresser. **The State Reform of the 90s**: logic and mechanisms of
Control. Brasília: Ministry of Federal Administration and State Reform, 1997. 58 p. (MARE Notebooks of the reform of the state; v. 1).

_____. **Explanatory Memorandum of Constitutional Amendment No. 19**. Available at: <https://www.bresserpereira.org.br/index.php/mare-ministerio-da- reforma-do-estado/documents-of-the-1995-managerial-reform/7869-1392>.
Accessed on: 31 May. 2025.

CITY HALL OF VITÓRIA DA CONQUISTA (PMVC).
City Hall now has new secretariats. Laws sanctioned today also restructure existing portfolios. 12 May. 2025. Available at: <https://www.pmvc.ba.gov.br/prefeitura-passa-a-contar-com-novas-secretarias-leis- sancionadas-hoje-tambem-reestruturam-pastas-ja-existentes/>. Accessed on: 27 May. 2025.

SHELTON, Shania; MCKEND, Eva. **Musk demands that federal employees justify jobs and threatens dismissal**: workers received an email demanding explanations about last week's activities. CNN Brasil, 22 Feb. 2025. Available at: <https://www.cnnbrasil.com.br/internacional/musk-exige-que-funcionarios-federais- justifiquem-empregos-e-ameaca-demissao/>. Accessed on: 28 May. 2025.

SILVA, Carlos Augusto Oliveira da. **Mayor-elect José Ronaldo schedules announcement of the secretariat and presents goals for the management of Feira de Santana from 2025 to 2028**. Jornal Grande Bahia, 20 dez. 2024. Available at: <https://jornalgrandebahia.com.br/2024/12/prefeito-eleito-jose-ronaldo-agenda- anuncio-do-secretariado-e-apresenta-metas-para-gestao-de-feira-de-santana-de- 2025-a-2028/>. Accessed on: 27 May. 2025.

UNIÃO BRASIL. **Under the leadership of Débora Regis, Lauro de Freitas advances in transparency and saving resources in the first 100 days of**

management. April 17, 2025. Available at: <https://uniaobrasil.org.br/2025/04/17/sob-a-lideranca-de-debora-regis-lauro-de- freitas-avanca-em-transparencia-e-economia-de-recursos-nos-primeiros-100-dias/>. Accessed on: 27 May. 2025.

_____. **Statute of the União Brasil Party**. 2021. Available at: <https://uniaobrasil.org.br/wp-content/uploads/2024/08/estatuto_registrado_cartorio.pdf>. Accessed on: 28 May. 2025.

Part III - National and Transnational Regulatory Frameworks

12. CIVIL LIABILITY OF SOCIAL NETWORKS FOR USER-GENERATED CONTENT IN THE BRAZILIAN LEGAL SYSTEM

https://doi.org/10.36592/9786554603065-11

Vitória Monego Sommer Santos

Abstract
Social networks belong to the category of digital platforms because they not only provide hosting services, but information dissemination services as well. Issues emerge when the social networks end up disseminating illicit material generated by its users which, on the internet, can gain exponential visibility, multiplying the damages, and generating the need to revisit the issue of civil liability in the light of new Technologies. This essay approaches the civil liability of social networks for third-party content in the Brazilian legal system.
Keywords: civil liability; content moderation; social networks; digital platforms; hosting providers; internet providers; Digital Services Act; Marco Civil da Internet; Brazilian law; user-generated content.

1. Introduction

The social networks, in view of their peculiar characteristics, bring new challenges to the law, requiring a revisiting of the issue of civil liability in the light of new media technologies. The main characteristic of social networks, which differentiates them from other internet providers, is that they not only memorize the information of their users, but also disseminate it, in such a way that the data can potentially reach an indeterminate number of people, without limitation of borders. The challenge arises when social networks end up disseminating illicit content from their users, which can lead to a multiplication of the damage suffered by victims, given the inner characteristics of the World Wide Web.

This essay aims to discuss the following questions: 1) What is most effective method to balance user rights on social networks?; 2) should social networks have a general obligation to moderate user content?; 3) can social networks be held civilly liable for the dissemination of illicit content published by their users? and 4) if so, should its liability be subjective or objective? The matter acquires even more

relevance in the context of the nationwide debates on the constitutionality of Article 19 of the "*Marco Civil da Internet* (MCI)'' (law nº 12.965/2014), the first law that regulated the civil liability of internet providers in Brazil. According to this article, hosting providers can only be held liable for third-party illicit content if they don't remove it after a judicial order. The constitutionality of the Marco Civil da Internet law was recently judged in the Appeal nº 1.037.396/SP[1], before the Brazilian Federal Supreme Court.

In order to analyze these questions, it is essential to first understand the nature of social networks, which belong to the broad category of "internet providers", known in the European Union law as information society "intermediary service providers". Marcel Leonardi clarifies that "internet provider" is "the gender of which the other categories (backbone provider, access provider, e-mail provider, hosting provider and content provider) are species" (Leonardi, 2015). In Marco Civil da Internet law, the broad category of "internet providers" was organized through the bipartition "connection provider" (*provedor de conexão*) and "application provider" (*provedor de aplicações*) (art. 5º, MCI).

Meanwhile the "connection provider" provides service of "enabling a terminal for sending and receiving data packets''[2], the "application provider" provides service of "a set of functionalities that can be accessed through a terminal connected to the Internet''[3], which includes digital platforms, such as social networks. In this way, the category of "hosting provider" of the Digital Services Act partially overlaps with the category of "application providers" of the Brazilian Marco Civil da Internet law, both of which include social networks. More specifically, social networks belong to the subcategory of "digital platforms" because, unlike other virtual hosting service providers, they not only store user information, but also disseminate it[4].

---

[1] Brazil. Federal Supreme Court (Supremo Tribunal Federal – STF). Appeal nº 1.037.396/SP (RE nº 1.037.396/SP).
[2] MCI, art. 5º.
[3] MCI, art. 5º.
[4] Recital 13, Digital Services Act: [...] it is necessary to distinguish, within the broader category of providers of hosting services as defined in this Regulation, the subcategory of online platforms. Online platforms, such as social networks or online platforms allowing consumers to conclude distance contracts with traders, should be defined as providers of hosting services that not only store information provided by the recipients of the service at their request, but that also disseminate that information to the public at the request of the recipients of the service. However, in order to avoid imposing overly broad obligations, providers of hosting services should not be considered as online

2. Civil liability of social networks for user-generated content in the Brazilian legal system

The theme of social network's civil liability for the content of its users has always been controversial in the Brazilian law, being the subject of major doctrinal and jurisprudential divergences. The Marco Civil da Internet law, published in 2014, was the first law to regulate the liability of internet providers in the national legal system. However, even after its publication, the disagreements regarding the civil liability regime to be adopted, as well as the requirements for its application, have not ceased. In this context, the Brazilian Federal Supreme Court recently judged the Appeal nº 1.037.396/SP (STF, RE nº 1.037.396/SP)[5], in which the constitutionality of Article 19 of the Marco Civil da Internet law, that regulates the liability regime of "application providers'', including social networks, was analyzed.

As Paulo Roberto Binicheski reports, for many years, the judiciary was unable to "adopt a jurisprudential consensus on the liability of Internet service intermediaries; its decisions vary widely, causing legal uncertainty" (Binicheski, 2011), which persists to this day. As mentioned by Justice Nancy Andrighi on the Appeal nº 1.642.997/RJ (STJ, REsp. nº 1.642.997/RJ, 2017), in Brazil there are three main doctrinal lines on the civil liability of "application providers'': 1) the unliability for illicit user-generated contents; 2) the objective civil liability for illicit user-generated contents (direct liability); 3) the subjective civil liability for illicit user-generated contents[6]. The subjective liability line, in its turn, is divided in two: 3.1) subjective liability if the application provider doesn't remove the illicit content after a notification by the user (extrajudicial); 3.2) subjective liability if the application provider doesn't remove the illicit content after a court order (judicial).

In other words, the strands vary substantially, from the thesis of the unliability of social networks for illicit user-generated content, to the diametrically opposed thesis of objective liability (direct liability). According to the first line, social networks are unliable for third-party content, since it was nor authored or edited by the platform

platforms where the dissemination to the public is merely a minor and purely ancillary feature that is intrinsically linked to another service, or a minor functionality of the principal service [...]".

[5] RESP nº 1.642.997/RJ.

(Getschko, 2015). This is how the "*Santa Catarina*'' State Tribunal decided, in 2011, the Civil Appeal nº 2011.018828-1, dismissing the civil liability of *Orkut* (social network managed by Google) for the creation of a fake profile by a user, which contained offensive material. In the words of Justice Fernando Carioni, although the relation between the offended party and Orkut is a consumer one, the social network cannot be held liable, since it is a mere intermediary, and because of "the absence of causal link between the conduct and the damage suffered by the appellant"[6].

According to the opposite doctrinal line, social networks would be objectively liable (direct liability) for the content of their users, based on the risk of the activity (art. 927 of the Brazilian Civil Code) or based on the defect in the service provided (art. 14 of the Brazilian Consumer Code). This implies that social networks would have a general obligation to monitor all content published on their platform. This is how the Tribunal of the "*Minas Gerais*'' State judged, in 2009, the Appeal nº 10701082216857001, where the social network *Orkut* was held objectively liable for a page that was offensive to the plaintiff's image and honor, based on the theory of the risk[7]. In support of the theory of objective liability, according to Guilherme Guimarães Martins, the lack of a general obligation to monitor social networks would represent a "step backwards towards guilt, in the midst of the age of risk" (Martins, 2014).

In mid-2011, still prior to the publication of the Marco Civil da Internet law, the jurisprudential line of the "subjective liability'' of social networks for third-party content was consolidated by the Brazilian "Superior Court of Justice'' (*Superior Tribunal de Justiça* - STJ), along very similar lines to that one of Directive 2000/31/EC, which at the time regulated the matter in the European Union, and to the Digital Services Act. According to this jurisprudence, social networks could only be held liable for illicit third-party content if, after a notification of the user, the platform remained inert, not removing the content. This means that the social network would not have a general obligation to monitor the content of its users, but to analyze the

[6] TJ-SC - AC: 20110188281 Itajaí 2011.018828-1. Presiding judge: Fernando Carioni. Trial date: 03/05/2011, Terceira Câmara de Direito Civil.
[7] TJ-MG - AC: 10701082216857001 Uberaba. Presiding judge: Saldanha da Fonseca. Trial date: 05/08/2009. 12ª Câmara Cível. Publication date: 24/08/2009).

content and, if applicable, to remove it, after having received a user notification (extrajudicial notice).

The Superior Court of Justice decision in the Appeal nº 1.308.830/RS (STJ, Resp. nº 1.308.830/RS, 2011), exemplifies the adoption of this jurisprudential line, based on the theory of the subjective liability of social networks for the omission of the platform on removing the illicit content after the user's notice. In the words of Justice Nancy Andrighi, "upon being informed that a certain text or image has illegal content, the provider must act energetically, removing the material from the air immediately, under penalty of being jointly liable with the direct author of the damage, due to the omission practiced"[8]. Notwithstanding, this strand is criticized by a significant part of the doctrine, according to which social networks would not be able to evaluate the rights of its users. In this sense, as stated by Fabiana Siviero and Guilherme Sanchez, "the complex task of balancing these rights, which are of equal constitutional importance, is the job of the judge and can only be carried out after a detailed analysis of the facts involved" (Siviero & Sanchez, 2015).

However, with the publication of the Marco Civil da Internet law in 2014, this consolidated jurisprudential understanding of the STJ had to be modified to be adapted to the text of the new law. According to Article 19 of the Marco Civil da Internet, as a means to ensure freedom of expression and prevent censorship, the application provider "can only be held civilly liable for damages arising from content generated by third parties if, after a specific court order, it fails to take steps to (...) make the content identified as infringing unavailable". In other words, social networks would only be obliged to remove content after specific judicial order. Consequently, they could only be held liable for inaction after a court order. Thus, despite maintaining the absence of a general obligation of vigilance and the theory of subjective liability, there has been a change in the criteria for establishing civil liability, with its bureaucratization.

The decision of the Superior Court of Justice in the Appeal nº 1.642.997/RJ (STJ, REsp. nº 1.642.997/RJ, 2017), brought by *Facebook Serviços Online do Brasil Ltda*. exemplifies the change in case law following the publication of the Marco Civil

[8] STJ - REsp: 1308830 RS 2011/0257434-5. Presiding judge: Justice Nancy Andrighi. Trial date: 08/05/2012. Terceira Turma. Publication date: 19/06/2012.

da Internet. According to Justice Nancy Andrighi's opinion, Article 19 of the Marco Civil da Internet aims to avoid the danger of application providers deciding "what remains online and what is removed", as well as the great subjectivity "of the criteria that can be used to remove content"[9]. As mentioned by Paulo Roberto Binicheski, who supports the MCI system, "it is not the role of Internet providers to act as network police, monitoring the steps of their users to prevent the commission of illegal acts", because, from his point of view, they would then act as judges, deciding what would be published (Binicheski, 2011).

The problem with Article 19 of the MCI is that judicial decisions are slow, during which time the victim's damages can be multiplied on the internet. The Marco Civil da Internet law has only one hypothesis in which a simple extrajudicial notification is sufficient to generate the obligation of the application providers to remove the illicit content, regardless of a court order: the occurrence of unauthorized disclosure of "images, videos or other materials containing scenes of nudity or private sexual acts", foreseen in its Article 21. In the words of Cristiano Colombo and Eugênio Facchini Neto, "this exception seeks to speed up the protection of victims of these types of posts, especially in the face of the wave of leaked intimate photos" (Colombo & Facchini, 2017).

Following the publication of the Marco Civil da Internet law, the social networks had no obligation to remove illicit user content prior to a court order. Therefore, they could only be held liable for disobedience to a judicial order to remove the illicit content. However, as mentioned, its Article 19 is so controversial that its constitutionality started being analyzed in the Appeal nº 1.037.396/SP (STF, RE nº 1.037.396/SP)[10], before the Brazilian Constitutional Supreme Court. The trial of the Appeal was concluded on June 26th, 2025, and Article 19 of the Marco Civil da Internet law was considered partially and progressively unconstitutional.

The main arguments used by the Federal Supreme Court to substantiate its decision are that Article 19 "is not fully capable of offering effective protection of fundamental rights and safeguarding fundamental constitutional principles and

[9] STJ - REsp: 1642997 RJ 2016/0272263-4. Presiding judge: Nancy Andrighi. Trial date: 12/09/2017. Terceira Turma. Publication date: 15/09/2017.
[10] Brazil. Federal Supreme Court (Supremo Tribunal Federal – STF). Appeal nº 1.037.396/SP.

values in virtual environments", and it is "insufficient to address the systemic risks that have arisen in these environments". Therefore, until a new law is approved, with the change of the MCI's dispositions, its Article 19 should be interpreted according to the decision of the Federal Supreme Court on this trial.

The Federal Supreme Court established different content moderation and civil liability regimes to the application providers, according to the nature of the user's illicit acts. However, the STF judges emphasize that Article 19 "is unconstitutional not because of what it provides for - that is, removal by court order - but because it does not provide for broader exceptions beyond those in Article 21", which sets the obligation to remove content after the notification of the user in the hypothesis of unauthorized disclosure of "images, videos or other materials containing scenes of nudity or private sexual acts".

In this context, Article 19 of the MCI is still applicable, but with an extensive list of exceptions, in a way that its content was largely emptied out. In judgement, three different regimes were established by the Court: *a*) hypothesis where Article 19 will be applied as it is (need of a court order to oblige the application provider to remove the illicit content); *b*) hypothesis where the user's notification is enough to oblige the application provider to remove the illicit content (notice and take down); and *c*) hypothesis where the application provider has a general obligation to monitor illicit content and take it down, regardless of notification.

According to the Federal Supreme Court, in the cases of "crime or illicit acts" and "fake accounts" the rule of Article 21, MCI, should be applied. This means in these hypotheses the notice and take down regime is to be used, and thus the application provider is subjectively liable if it does not remove the illicit content after the notice[11]. The exception to this rule regards the "crimes against honor", where Article 19 is still applicable, with the need of a court order to oblige the application provider to remove the illicit content, in order to avoid undue censorship. However, in the hypothesis where this "offensive content" was replicated multiple times (after the initial court

[11] Court decision translated from Portuguese to English by the author: "3. The internet application provider shall be held civilly liable, pursuant to art. 21 of the MCI, for damages arising from content generated by third parties in cases of crime or illicit acts, without prejudice to the duty to remove the content".

order of its removal), the application provider has the obligation of its removal according to the notice and take down system[12].

Successively, the Federal Supreme Court establishes a "presumption of liability" on the cases of "paid advertisements" and "robot" illicit contents, where there is an obligation of removal independently of notice, and the providers liability unless they prove having acted diligently and within a reasonable time[13]. Finally, the "presumption of liability" is also applicable to an extensive list of severe illicitudes, such as "anti-democratic acts", "terrorism", "incitement or assistance to suicide", "discrimination" of ethnicity, religion, national origin, sexuality, or gender identity, "crimes committed against women because of their gender", "sexual crimes against vulnerable persons", "child pornography" and "human trafficking"[14].

In broad terms, after the trial of the Appeal nº 1.037.396/SP, the Federal Supreme Court decided that Article 21 of the MCI - that establishes the notice and take down system, and thus liability in case the provider doesn't take down the illicit content after the user's notice - would be the general rule of "application provider" liability. As an exception, Article 19 will continue to be applicable to online offenses hypothesis. This Court decision raises questions about the Federal Supreme Court competence to turn Article 21 of the MCI into a general rule. However, this decision by the STF must be followed until a new law on the matter is published, replacing the MCI system, in a way that debates in Brazil on the subject are far from over.

---

[12] Court decision translated from Portuguese to English by the author: "3.1. In cases of crimes against honor, art. 19 of the MCI applies, without prejudice to the possibility of removal by extrajudicial notification. 3.2. In the case of successive replications of the offensive act already recognized by a court decision, all social network providers must remove posts with identical content, regardless of new court decisions".

[13] Court decision translated from Portuguese to English by the author: "4. The presumption of liability of providers in the case of illicit content is established when it involves (a) paid advertisements and boosts; or (b) artificial distribution networks (chatbots or robots). In these cases, liability may arise regardless of notification. Providers shall be excluded from liability if they prove that they acted diligently and within a reasonable time to make the content unavailable".

[14] Court decision translated from Portuguese to English by the author: "5. The internet application provider is liable when it fails to promote the immediate removal of content that constitutes the serious crimes listed in the following exhaustive list: (a) anti-democratic conduct and acts that fall under the types provided for in articles (...) of the Penal Code; (b) crimes of terrorism or preparatory acts of terrorism (...); (c) crimes of inducement, instigation, or assistance to suicide or self-mutilation (...); (d) incitement to discrimination based on race, color, ethnicity, religion, national origin, sexuality, or gender identity (homophobic and transphobic conduct) (...); (e) crimes committed against women on the basis of their female gender, including content that propagates hatred towards women (...); (f) sexual crimes against vulnerable persons, child pornography, and serious crimes against children and adolescents, pursuant to arts. (...); g) human trafficking (...)".

3. Conclusion

In European Union, according to the Digital Services Act, hosting providers, a category to which social networks belong, do not have a general obligation to monitor the content posted by their service recipients (art. 6º, DSA). Similarly, Marco Civil da Internet law does not provide for a general obligation for "application providers" to monitor the content of third parties (art. 19, MCI). As there is no general obligation to monitor, both systems adopt a regime of subjective liability of social networks for damages caused by illicit content of users on their platform. The regime of subjective liability, in fact, is not the most appropriate to be applied as a general rule to social networks, since monitoring all content posted (objective liability) is not an intrinsic part of their activity, which would de-characterize the very nature of these platforms.

However, the criteria for establishing subjective civil liability differ substantially between these two systems. Under the Digital Services Act, the "hosting provider" can only be held liable for third-party illicit content if, after an extrajudicial notification of the user, it fails to act diligently by disabling access to that content. In this way, the social network is obliged to remove the content after a simple user's notice of its unlawfulness, using a tool on the platform itself. In turn, in the light of the Marco Civil da Internet, the social networks can only be held liable for inaction after a court order of content removal (art. 19, MCI). Thus, the simple notification of the user on the platform would not be enough to generate the obligation to remove the content, requiring judicial notification, a much more bureaucratic and time-consuming system.

In fact, the system of Article 19 of the MCI is incompatible with the speed at which information spreads on the internet, and could cause an exponential multiplication of illicit content and, consequently, damage to victims. The only exception to the general rule of Article 19 of the MCI is the case of unauthorized disclosure of "images, videos or other materials containing scenes of nudity or sexual acts of a private nature", foreseen in Article 21 of Marco Civil da Internet law, which generates an obligation for the hosting provider to remove the content after a simple extrajudicial notification from the user.

In this context, the most appropriate approach seems to be the conversion of Article 21 of the MCI into a general rule, so that the simple notification of the user on the platform itself would generate their obligation to diligently evaluate and, if applicable, delete the content. If this system were adopted, Brazilian law would come closer to that of the European Union, and to the consolidated case law of the Superior Court of Justice prior to the publication of the Marco Civil da Internet law. On the other hand, this system involves risks related to censorship and content overblocking.

The recent decision of the Federal Supreme Court on the Appeal nº 1.037.396/SP established the partial unconstitutionality of Article 19 of the MCI, under the argument that its system was unable to protect fundamental rights in virtual environments. The sentence turned the Article 21 of the MCI, that established the notice and take down system as an exception, into the general rule for illicit user content. Consequently, Article 19 became an exception, applicable only to "crimes against honor". Therefore, it is important to note that for online offenses, the rule remains the same. Nevertheless, if this offensive content is replicated after the court order of its removal, the application provider has the obligation of its removal according to the notice and take down system.

Besides that, the Federal Supreme Court established a general obligation to monitor and take down, regardless of notice, the severe illicit contents listed in the sentence. The "presumption" of liability was established as well in the case of "paid advertisements" and "robot" illicit contents, unless the provider proves having acted diligently on its removal. The Federal Supreme Court sentence is highly controversial, raising questions on its competence to create an entire "application provider" obligation and liability system. However, this judicial decision must be followed until a new law is published by the National Congress, replacing Article 19 of the MCI, context in which the Brazilian debates on content moderation and digital platform's liability are far from over.

References

Binicheski, P. R. 2011. *Responsabilidade civil dos provedores de internet. Direito comparado e perspectivas de Regulamentação no Direito Brasileiro*. Juruá Editora.

Brazil. *Law nº 12.965, 23 April 2014* (Marco Civil da Internet). Available on: *https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm.* English version on https://www.cgi.br/pagina/marco-civil-law-of-the-internet-inbrazil/180

Brazil. *Federal Constitution*. Available on: *https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm*

Brazil. Superior Court of Justice (Superior Tribunal de Justiça – STJ). Appeal nº 1308830 RS 2011/02574345. Presiding Judge: Nancy Andrighi. Trial date: 08/05/2012. Terceira Turma. Publication date: 19/06/2012. Available on: https://scon.stj.jus.br/SCON/GetInteiroTeorDoAcordao?num_registro=201102574345&dt_publicacao=19/06/2012

Brazil. *Superior Court of Justice (Superior Tribunal de Justiça – STJ). Appeal nº 1642997 RJ 2016/02722634*. Presiding Judge: Nancy Andrighi. Trial date: 12/09/2017. Terceira Turma. Publication date: 15/09/2017. Available on: https://www.stj.jus.br/websecstj/cgi/revista/REJ.cgi/ATC?seq=76349712&tipo=0&nreg=

Brazil. *Federal Supreme Court (Supremo Tribunal Federal – STF). Appeal nº 1.037.396/SP*. Available on: https://portal.stf.jus.br/processos/detalhe.asp?incidente=5160549

Brazil. *Tribunal de Justiça de Santa Catarina (TJSC). AC: 20110188281 Itajaí 2011.018828-1*. Presiding Judge: Fernando Carioni. Trial date: 03/05/2011. Terceira Câmara de Direito Civil. Available on: https://www.jusbrasil.com.br/jurisprudencia/tj-sc/1101081638

Brazil. *Tribunal de Justiça de Minas Gerais (TJMG). AC: 10701082216857001 Uberaba*. Presiding Judge: Saldanha da Fonseca. Trial date: 05/08/2009. 12ª Câmara Cível. Publication date: 24/08/2009). Available on: https://www.jusbrasil.com.br/jurisprudencia/tj-mg/1123928935

Colombo, C.; Facchini Neto, E. 2017. Ciberespaço e conteúdo ofensivo gerado por terceiros: a proteção dos direitos de personalidade e a responsabilização civil dos provedores de aplicação, à luz da jurisprudência do Superior Tribunal de Justiça. In *Revista Brasileira de Políticas Públicas*, vol. 7, n. 3.

Conti, G. 2020. *Lineamenti di diritto delle piattaforme digitali. Volume 1. Le tutele del consumatore e dell'utente commerciale nei confronti dei cybermediary*. Maggioli Editore.

European Union. *Directive 2000/31/EC*. Available on: https://eur-lex.europa.eu/legal-content/PT/ALL/?uri=celex%3A32000L0031

European Union. *Regulation 2022/2065/EU* (*Digital Services Act*). Available on: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

Getschko, D. 2015. Marco Civil da Internet e os fundamentos de seus princípios. In Artese, G. (coord.). *Marco Civil da Internet. Análise Jurídica sob uma Perspectiva Empresarial*. Quartier Latin.

Leonardi, M. 2005. *Responsabilidade civil dos provedores de serviço da Internet*. Juarez de Oliveira.

Martins, G. M. 2014. *Responsabilidade civil por acidente de consumo na internet*. Revista dos Tribunais.

Santos, V. M. S. 2025. A responsabilidade civil das redes sociais por conteúdos gerados por terceiros. In: Angelone, M.; Colombo, C.; Martone, I.; Perlingieri, C.; Valongo, A. (Org.). *Intelligenza artificiale e piattaforme digitali: dalla co-regolamentazione alla co-vigilanza*. 1ed. Edizioni scientifiche italiane, p. 283-318.

Siviero, F.; Sanchez, G. C. 2015. O Novo Regime de Responsabilidade Civil dos Provedores de Aplicações de Internet. In Artese, G. (coord.). *Marco Civil da Internet. Análise Jurídica sob uma Perspectiva Empresarial*. Quartier Latin.

13. THE URGENCY OF SOCIAL MEDIA REGULATION: META'S DIVERGENT APPROACHES TO THE EU AND BRAZIL REGARDING AI TRAINING ON PUBLICLY AVAILABLE USER CONTENT

https://doi.org/10.36592/9786554603065-12

Luisa Maciel Perez

Abstract
In June 2024, Meta updated its privacy policy. Amongst the updates: a plan to train the company's AI on publicly available user content. In the EU, the company was already in contact with public authorities to ensure compliance with the legal framework, whereas in Brazil, neither the authorities nor users were informed. Meta's divergent approaches across the jurisdictions shall be the subject of analysis in this article. The research presents public authorities' convergent regulatory goals but divergent enforcement pathways in the EU and Brazil towards Meta's AI training and highlights how the different legal frameworks shape the company's actions. This article aims to stress the urgency of social media regulation in Brazil by demonstrating that Meta was allowed to proceed under lighter constraints, and disregard, due to the regulatory gap that exists, unlike in the EU where the DSA is in effect.
Keywords: Meta, AI training on user content, divergent approaches, EU, Brazil, social media regulation, GDPR, LGPD, DSA.

1. Introduction

Social media is undeniably part of today's world. Not to engage with it can be the equivalent of being a 21st-century hermit.

In Brazil, where social media is almost a religion, it is no different. The country stands out as the fifth-largest social media market worldwide, according to Statista[1] (BIANCHI, 2025). As of 2023, over 170 million people access social media in the country, which is over eighty per cent of its population (*idem*).

Zap, Insta, and Face might not sound familiar to a foreigner, but they are nicknames used daily by Brazilians to reference the country's most widely used platforms, all of which are Meta-owned, demonstrating the hegemony of this private company over the national social media market (*idem*).

[1] Statista describes itself as a "global data and business intelligence platform with an extensive collection of statistics, reports, and insights" (STATISTA, 2025).

Despite this intense presence, Brazil does not yet have a regulation to address social media platforms, as the European Union (EU) does with the Digital Services Act (DSA).

This directly impacts how public authorities interact, respond and issue orders to online platforms in Brazil and in the EU. Their divergent approach can be evidenced in the recent case of Meta's AI training.

In June 2024, Meta notified its users in the EU of its latest privacy policy changes. Within the new text, the platform informed its intention to use all content of its users publicly shared since 2007 in all its products, such as posts and comments, to train its AI models under the legitimate interest basis. The new processing activity was set to come into force on June 26 (noyb, 2024).

This somewhat silent move, just signalled in the privacy policy, sparked controversy worldwide. In the EU, just a few days later, on June 6, noyb (*idem*), an NGO aimed at enforcing data protection laws, announced it had filed complaints in 11 European countries' data protection authorities (DPAs) requesting the change to be stopped immediately.

In light of social pressure, on June 10, Meta officially announced on its website its plans to expand AI by training on their users' content (FRATTA, 2024). It is important to stress, however, that the announcement was only directed at European users, and in no way was it replicated to Brazilian audiences in Portuguese.

Meta then adopted different approaches towards the EU and Brazil, and its authorities reacted differently, the former preventive, the latter reactive.

This dichotomy shall be the subject of analysis in this article, structured in two parts. First, a brief overview of the measures taken by the EU and Brazil. Secondly, an analysis of how the legal framework of each impacts Meta's approach and authorities' response, especially Brazil's regulatory gap on online platforms.

2. EU *versus* Brazil: divergent response approaches

Meta's decision to start processing user content for training its AI was not a surprise to EU authorities. On the contrary, the company had informed European DPAs of its intention in March, and according to a press statement, had already

incorporated some regulatory feedback (FRATTA, 2024). At the time, no restriction was imposed by the Data Protection Commission (DPC), Meta's lead supervisory authority per the General Data Protection Regulation (GDPR)[2].

However, upon the public announcement of Meta AI, the DPC reverted its position. According to the authority, in cooperation with other EU DPAs, it engaged in intensive discussions with Meta on the issue that eventually led to the request for Meta to pause its plans to engage in AI training in the EU (DPC, 2024). The company accepted it on June 14th(FRATTA, 2024).

As highlighted by noyb (2024), no further context on these discussions was given. But it was enough to make the big tech halt its plans twelve days before enforcement.

Furthermore, on December 17, the European Data Protection Board (EDPB)[3] (2024) issued an opinion on the use of personal data for the development and deployment of AI models, per request of the DCP for Europe-wide regulatory harmonisation.

Opinion 28/2024 (EDPB, 2024) highlights that a three-step test shall be conducted by companies when assessing the use of legitimate interest as a legal basis, as Meta did. It additionally stresses that simply informing of AI model training in the privacy policy, although transparent, does not necessarily fall within the data subjects' reasonable expectations, which shall be considered.

Following this publication, Meta reassessed its plans and provided updated documentation to the DPC, which was followed by a number of recommendations by the authority (DPC, 2024).

In response, Meta implemented relevant measures and improvements. Most significantly, the company updated its transparency notices and objection form to allow easier exercise of data subjects' rights. It also updated measures to protect data subjects, such as filtering and de-identification, and especially its risk

[2] The Irish DPA acts as the lead supervisory authority to Meta because the company's headquarters are located in Dublin, per Article 56 of the General Data Protection Regulation (GDPR) (EU, 2016).
[3] The EDPB is the body of the EU responsible for ensuring the consistent application of the GDPR among its member states.

assessments under GDPR, including for legitimate interests as a processing basis, Data Protection Impact Assessment (DPIA) and Compatibility Assessment (*idem*).

Only upon fulfilling the requests and complying with GDPR, Meta was allowed by the DPC to proceed with AI training, starting on May 27, 2025 – almost a year later than the original date (*idem*).

Nevertheless, the launch of the new feature also depended on compliance with the Digital Services Act (DSA), the EU's regulation on online platforms and search engines.

The European Commission (EC), responsible for enforcement, confirmed in June 2025 that it was closely monitoring the deployment of Meta AI, engaging in continuous dialogue and inquiries with the company, which proactively cooperated (European Parliament, 2025). The authority clarified that Meta was expected to submit risk assessment documentation under the DSA shortly and will follow up as appropriate (*idem*).

In essence, the EU adopted a preventive approach. Brazil, on the contrary, approached reactively.

Despite Meta's block in the EU, the company proceeded with its plans to start AI training on user content in Brazil on June 26, 2024.

On the same day, lacking social media regulation or an intervention by public authorities, Idec[4], a civil society organization dedicated to consumers' rights , notified the National Secretary of Consumers (Senacon), the national data protection authority (ANPD) and the authority on competition law (CADE), requesting an investigation on Meta (Idec, 2024) under the Lei Geral de Proteção de Dados (LGDP)[5] and Código de Defesa do Consumidor (CDC)[6].

On July 2, the ANPD issued an injunctive relief suspending Meta's privacy policy in relation to AI training on user content (DOU, 2024).

In her vote, Director Miriam Wimmer highlighted that the processing of data could have already started, and its continuance could lead to difficulty in excluding

[4] Idec – Institute for the Defence of Consumers is a civil society.
[5] Brazil's national regulation on data protection.
[6] Brazil's code for the protection of consumers.

certain personal data from AI training, fulfilling the requirement for relief due to serious and irreparable damage or damage that is difficult to repair (ANPD, 2024).

Furthermore, Meta attempted to use the legitimate interest basis for processing, which was rejected by the ANPD due to the processing of sensitive data, the violation of legitimate expectations of data subjects, and the failure to comply with the principles of purpose and necessity (*idem*).

Not only that, ANPD highlighted the different treatment of Meta in the EU and Brazil, especially the lack of any notification to users or authorities in the latter with clear, precise and easily accessible information on the privacy policy and AI training. It also verified a great difficulty in exercising the right of objection[7] (*idem*).

And, differently from the EU, Meta did not announce it would not train its AI on children or adolescents, nor guarantee sufficient safety measures for such processing in Brazil (*idem*).

Still on the same day, Senacon and Cade took measures to inquire Meta for information (Idec, 2024).

On July 10, the ANPD decided to postpone Meta's request for reconsideration until the company presented its plans for compliance with LGDP (DOU, 2024).

Finally, on August 30, the ANPD issued a decision allowing Meta to proceed with AI training following the terms of its plans for compliance and continued monitoring by the authority (2024). Amongst the documents presented by Meta was the Impact Report on Data Protection (RIPD) – similar to a DPIA in the EU – which would address the potential risks of the AI training in Brazil's context, especially considering the processing was already in place when the injunctive relief was issued.

Director Rael (2024) explains in his vote that Meta not only committed to informing its users of AI training through notification before processing, but also enhanced its transparency measures, its objection form, and eliminated risks.

The company also provided the required balancing test for legitimate interests to be used as a legal basis under the LGPD, accepted by the Authority, and included a provision confirming that no data of users under the age of eighteen would be processed (*idem*).

[7] Not a clear path to users to find the objection form in its platform or information of its existence (ANPD, 2024).

In light of the above, it is evident that whereas the EU acted preventively, through direct negotiation with Meta and mere requests, Brazil acted reactively, requiring legal orders.

Nevertheless, their responses touched mainly on the same issues. A lack of transparency in Meta's plans to users and difficulty in exercising their right to object, along with the validity of employing legitimate interests as a processing basis and especially the presentation of risk assessments on data protection.

In essence, public authorities aimed for the same: compliance with their legal frameworks and respect for users' rights.

However, Meta adopted different approaches across jurisdictions. In the EU, the company informed authorities before its privacy policy update, whereas in Brazil, even after delaying its plans abroad, no communication was made, nor were users notified.

One of the reasons for these divergent approaches could be the DSA.

3. Social Media Regulation: a framework and a gap

Legal frameworks not only shape the power of public authorities but also a company's actions. Meta's divergent approaches reveal how the EU is one step forward to Brazil regarding social media and why their different frameworks might be to blame.

The EU, as it stands, is at the forefront of regulation globally. The GDPR (2016) is one of the first major trend-setters, establishing data protection standards later followed worldwide, throughthe so-called Brussels effect (BRADFORD, 2015).

Brazil is not an exception. The LGPD (2018), national regulation on data protection, was heavily inspired by the GDPR, to the point that companies compliant with the latter would already have implemented sufficient measures to be compliant with the former (SILVA, CABRAL, 2025).

As Wimmer highlighted (ANPD, 2024), it is due to these many similarities that Meta employed the same legitimate basis for processing user data in the EU and Brazil: legitimate interests – Article 6(f) GDPR (2016) and Article 7(IX) LGPD (2018).

This almost regulatory mirror can also be appointed as the reason for DPC's and ANPD's similar responses to Meta's AI training, which led to many of the same measures and improvements regarding transparency, data subjects' rights, and most significantly, the presentation of impact assessments.

Under LGPD (2016), Article 10§3º empowers ANPD to request from Meta an Impact Report on Data Protection (RIPD), to assess potential risks to civil liberties and fundamental rights, when processing is based on legitimate interests. Similarly, GDPR (2018) on Article 35 requires controllers to conduct a Data Protection Impact Assessment (DPIA) when processing is likely to result in high risks and freedoms of natural persons, obliging Meta to present such documentation under the principle of accountability – Article 5.

In light of all that similarity, there is one key difference that shall be analysed, the sanctioning powers of data protection authorities under GDPR and LGPD. In a capitalist world where companies aim at maximising profit, fines amongst all sanctions act as motivational factors for companies to be compliant.

LGPD (2016) establishes that ANPD is responsible for monitoring compliance with the regulation and is the sole authority allowed to apply sanctions in the case of its violation, through an administrative process that ensures the right to a fair hearing, full defence and the right of appeal , Article 55-J(IV) and 55-K.

Similarly, under GDPR (2018), supervisory authorities are to carry out investigations in the form of data protection audits and apply corrective powers as necessary – Article 58(1)(2).

However, their fines differ in weight. Whereas in Brazil, a violation of LGPD (2016) can result in a simple fine of up to 2% of a company's revenue in the country, limited to 50 million reais per infraction[8], in the EU infringements can result in fines up to 20 million euro or 4% of the total worldwide annual turnover of the preceding financial year, whichever is higher[9].

[8] Article 52(II) of LGPD.
[9] Article 83(4) of GDPR.

So, for example, a violation of the right of data subjects to object to Meta's AI processing would result in a fine of at most 7.8 million euros[10] in Brazil and 5.66 billion euros[11] in the EU, approximately 725 times more, considering Meta's 164.50 billion dollars turnover in 2024 (Meta, 2025).

This financial difference alone could potentially justify Meta's commitment to inform DPC to ensure its compliance with its AI training before enforcement, differently to taking chances in Brazil.

Nevertheless, the DSA (2022) also plays an important role in potentially shaping Meta's actions. This is evident because the Irish DPA is not the only authority Meta is in a continuous dialogue and being monitored by, but also the European Commission, the DSA's enforcer (European Parliament, 2025).

As a regulation, the DSA establishes rules for online intermediaries and platforms, such as social media, aiming to ensure a safe and accountable online environment in the EU, by both protecting consumers and their fundamental rights and rebalancing the roles of users, platforms and public authorities (European Commission, 2022).

Its relevance is notable in the present case.

Meta's products, Facebook and Instagram, which shall be the base for their new AI training, have both been appointed as Very Large Online Platforms (VLOPs) under the DSA (*idem*), once they possess more than 45 million users per month in the EU, per Article 33(1) of the DSA (2022).

Consequently, they are subject to Chapter III, Section 5 of the DSA (2022), which provides for additional obligations to manage systemic risks. Regarding Meta's AI training, a few articles stand out.

First, Article 34 establishes additional assessments to be conducted by Meta's VLOPs to identify, analyse and assess systemic risks in the EU stemming from the functioning of their services and their related systems, including any effects present or future to the exercise of fundamental rights.

[10] Approximate conversion rate of 50 million reais to euros on June 24, 2025, at: https://economia.uol.com.br/cotacoes/.
[11] Approximate conversion rate of 4% of Meta's turnover in dollars to euro on June 24, 2025, at: https://economia.uol.com.br/cotacoes/.

This obligation includes algorithmic systems such as Meta's AI (*idem*). These are the documents the European Commission are currently awaiting from Meta for review of compliance.

Once the systemic risks are identified, the VLOPs are obligated per Article 35 to put in place tailored, reasonable, proportionate and effective mitigation measures, particularly considering impacts on fundamental rights (*idem*).

Furthermore, Article 37 establishes that Meta's products shall, at least once a year, be subject to independent audits to assess their compliance with the obligations stated above and all others included in Chapter III, at their own expense (*idem*).

Internally, Facebook and Instagram shall establish a compliance function, with sufficient authority, stature, resources and access to the VLOPs to monitor compliance with the DSA – Article 41 (*idem*).

Ergo, the DSA is the one responsible for establishing clear guidelines on how Meta should implement its AI training on user content and creates a system of accountability, in which the company is not only responsible for identifying systemic risks, but also mitigating them, along with regular accountability checks by both the compliance function and independent audits and many other obligations.

Even more striking, the DSA raises the bar regarding its sanctions. In the event of non-compliance, the EC shall adopt a decision confirming the intentional or negligent infringement of relevant provisions and impose a fine on VLOPs not exceeding 6% of its total worldwide annual turnover in the preceding financial year – Articles 73 and 74 (*idem*).

Compared to the GDPR, the fine can be 2% higher; in the case of Meta, this would mean 2.8 billion euros more[12]. However, GDPR and DSA are not mutually exclusive, which means Meta, in regards to its AI training, in the worst-case scenario, could be fined for infringement of each regulation to the highest amount, resulting in a financial penalty of 10% of its turnover for a full year.

In conjunction with the GDPR, the DSA creates a cohesive legal framework to address innovative cases that not only challenge data protection but also pose risks

[12] Simple math solution of dividing 5.66 billion by 4, multiplying by 6. Then subtract the original number from the final one. This is based on the previous example regarding Meta's annual turnover in 2024.

in the online environment with algorithmic processes, and provides authorities with enforceable rules and penalties that are effective, proportionate and dissuasive.

It is this dissuasive nature that is so evident in Meta's actions in the EU. The company's ongoing discussion with the DPC and EC, along with the voluntary . not ordered , delay of implementation plans, demonstrates a respect for regulations and a desire to achieve compliance.

In Brazil, its actions demonstrate otherwise.

As of June 2025, the country does not have a regulation addressing social media or any other online platform, like the DSA.

Yet, attempts have been made. Bill 2630 was proposed in 2020, mid-covid crisis and issues with disinformation online, aimed to address freedom, liability and transparency on the Internet by establishing rules for online platforms. It gained force once more after the attacks on Brazilian congress on January 8, 2022, politically motivated and organised on social media (VALLE, 2023).

Originally scheduled to be voted on in May 2023 in the House of Representatives[13], the bill got postponed, and two years later, a new date has not been set.

With intense media coverage, the proposal was nicknamed “Fake News Bill”, although such a term was never included, and sparked an ideological dichotomy, also politically inspired, reducing the discussion to one side claiming for social media regulation, whereas the other argued potential impacts to freedom of speech and censorship (PINHEIRO; COSTA, 2025).

Big techs also applied intense lobbying. Google released a campaign during the week of voting, exhibiting on its front page the message: “The Fake News Bill can increase confusion about what is true and what is a lie in Brazil” (*idem*, p. 30434). Meta also released a statement claiming that the Bill would make it difficult for their services to be offered for free in Brazil (Meta, 2023).

[13] The Bill passed in the Senate and then is directed to the House of Representatives for a second voting, before becoming law.

In light of that, research conducted revealed only a third of representatives saw the Bill as a priority (CRUVINEL, 2025). This scenario contributed to legislative standstill and citizens' rejection (PINHEIRO; COSTA, 2025).

Whilst the Legislative branch does not legislate, the Judiciary suffers from this regulatory gap to address the many issues that arise, especially platforms' liability.

This is the reason the Federal Supreme Court (STF) recognised the discussion on the constitutionality of Article 19 of Law no. 12.965/2014 as a matter of general interest. That implies the judgment of its leading case, RE 1037396,will constitute a thesis with effect *erga omnes*,Theme 987 (STF, 2025).

Marco Civil da Internet (12.965/2014) is the national framework for internet users' civil rights, which establishes in its Article 19 that internet providers are only civilly liable for damages arising from content generated by third parties if, after a specific court order, within its capabilities and specified time frame, does not make the content unavailable. (BRASIL, 2014).

At the time this article is written,the judgment is currently ongoing and is expected to be finished by July 2025. Nevertheless, the majority of the Court's Ministers have already voted in the sense of partial or total unconstitutionality of Article 19, reinforcing the expansion of the liability framework to internet providers in Brazil (STF, 2025).

Nevertheless, this is not enough, and the country urges for social media regulation. Hence, the Comitê Gestor da Internet no Brasil[14] opened until June 30, 2025, a public consultation on the regulation of digital platforms in Brazil (CGI, 2025).

Given the above, it is not surprising that "Meta treats Brazilians as citizens of second-class" (Idec, 2025), and even its authorities, disregarding any communication towards its plans, thus following through with the original enforcement date and requiring an order to suspend its AI training.

[14] "The Brazilian Internet Steering Committee is responsible for establishing strategic guidelines related to the use and development of the Internet in Brazil and guidelines for the registration of domain names, allocation of IP (Internet Protocol) addresses, and administration of the '.br' top-level domain. It also promotes studies and recommends procedures for Internet security and proposes research and development programmes that enable the maintenance of technical quality and innovation in the use of the Internet." (CGI, ?)

4. Conclusion

Meta's different approaches towards its plans to train AI on user content in the EU and Brazil are reflected by each jurisdiction's legal framework.

The combination of GDPR and DSA created a cohesive framework in the EU that encompasses users' rights, online platforms' obligations, and empowers public authorities to act and dissuade big techs like Meta, which, without any order, delayed its AI plans for almost a year.

Whereas in Brazil, without a DSA-like regulation and light penalties for non-compliance with the LGPD, compared to the financial power of Meta, the Brazilian legal framework does not encompass the complexities of the intersection of data protection, AI and social media, and leaves authorities with fewer, and less dissuasive, tools to address and hold liable online platforms.

This regulatory asymmetry allowed Meta to proceed under lighter constraints despite lacking regulatory consultation, without regard to rights or risks.

Lastly, Meta's AI training case reveals an undeniable, yet not new, truth: the urgency of social media regulation in Brazil.

References

'ANPD e Senacon reforçam argumentação do Idec e Meta está proibida de usar dados de brasileiros para treinar IA, Cade também investigará o caso' <https://idec.org.br/release/anpd-reforca-argumentacao-do-idec-e-proibe-meta-de-usar-dados-de-brasileiros-para-treinar-ia> accessed 20 June 2025.

Bianchi T, 'Social Media Usage in Brazil - Statistics & Facts' (*Statista*, 11 February 2025) <https://www.statista.com/topics/6949/social-media-usage-in-brazil/> accessed 19 June 2025.

Bradford A, 'Exporting Standards: The Externalization of the EU's Regulatory Power via Markets' (2015) 42 International Reviewof Law and Economics 158.

BRASIL, Lei nº 12.965 - Marco Civil da Internet 2014 [12.965].

BRASIL, Lei nº 13.709 - Lei Geral de Proteção de Dados Pessoais (LGDP) 2018 [13.709].

CGI.br, 'CGI.br - Comitê Gestor da Internet no Brasil' (*CGI.br - Comitê Gestor da Internet no Brasil*) <https://cgi.br> accessed 25 June 2025.

CGI.br, 'Consulta Sobre Princípios Para a Regulação de Redes Sociais' (*Diálogos CGI.br*, 2025) <https://dialogos.cgi.br/> accessed 19 June 2025.

Cruvinel S, 'Regulação de redes sociais encontra resistência política' (*Migalhas*, 23 May 2025) <https://www.migalhas.com.br/depeso/430847/regulacao-de-redes-sociais-encontra-resistencia-politica> accessed 19 June 2025.

'DPC Statement on Meta AI' (*Data Protection Commission*, 21 May 2025) <https://www.dataprotection.ie/news-media/latest-news/dpc-statement-meta-ai> accessed 19 June 2025.

'EDPB Opinion on AI Models: GDPR Principles Support Responsible AI | European Data Protection Board' (18 December 2024) <https://www.edpb.europa.eu/news/news/2024/edpb-opinion-ai-models-gdpr-principles-support-responsible-ai_en> accessed 19 June 2025.

'The EU's Digital Services Act' (27 October 2022) <https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en> accessed 19 June 2025.

European Data Protection Board, 'Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models | European Data Protection Board' <https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en> accessed 19 June 2025.

European Parliament, 'Parliamentary Question | Answer for Question E-001460/25 | E-001460/2025(ASW)' (2 June 2025) <https://www.europarl.europa.eu/doceo/document/E-10-2025-001460-ASW_EN.html> accessed 22 June 2025.

Fratta S, 'Building AI Technology for Europeans in a Transparent and Responsible Way' (*Meta*, 10 June 2024) <https://about.fb.com/news/2024/06/building-ai-technology-for-europeans-in-a-transparent-and-responsible-way/> accessed 20 June 2025.

Imprensa Nacional, 'Despacho Decisório nº 20/2024/PR/ANPD - DOU' <https://www.in.gov.br/web/dou> accessed 19 June 2025.

Imprensa Nacional, 'Despacho Decisório nº 24/2024/PR/ANPD - DOU' <https://www.in.gov.br/web/dou> accessed 19 June 2025.

Imprensa Nacional, 'Despacho Decisório nº 33/2024/PR/ANPD - DOU' <https://www.in.gov.br/web/dou> accessed 19 June 2025.

noyb, 'Noyb Urges 11 DPAs to Immediately Stop Meta's Abuse of Personal Data for AI' (6 June 2024) <https://noyb.eu/en/noyb-urges-11-dpas-immediately-stop-metas-abuse-personal-data-ai> accessed 19 June 2025.

European Parliament, 'Parliamentary Question | Answer for Question E-001460/25 | E-001460/2025(ASW)' (2 June 2025) <https://www.europarl.europa.eu/doceo/document/E-10-2025-001460-ASW_EN.html> accessed 22 June 2025.

'Idec notifica órgãos governamentais para que suspendam o uso de dados dos usuários para treinamento de IA pela Meta' <https://idec.org.br/release/idec-notifica-orgaos-governamentais-para-que-suspendam-o-uso-de-dados-dos-usuarios-para> accessed 20 June 2025.

Meta, 'PL 2630/2020 precisa de mudanças' (*Sobre a Meta*, 29 April 2023) <https://about.fb.com/br/news/2023/04/pl-2630-2020-precisa-de-mudancas/> accessed 25 June 2025.

Meta, 'Meta Reports Fourth Quarter and Full Year 2024 Results' (29 January 2025) <https://investor.atmeta.com/investor-news/press-release-details/2025/Meta-Reports-Fourth-Quarter-and-Full-Year-2024-Results/> accessed 24 June 2025.

Pinheiro C and Costa E, 'As Fake News sobre o PL das Fake News: manipulação algorítmica no debate sobre regulação das plataformas digitais no Brasil' (2025) 7 ARACÊ 30432.

'(Preliminary) Noyb WIN: Meta Stops AI Plans in the EU' (14 June 2024) <https://noyb.eu/en/preliminary-noyb-win-meta-stops-ai-plans-eu> accessed 19 June 2025.

'PL 2630/2020' (*Portal da Câmara dos Deputados*) <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735> accessed 25 June 2025.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) 2016 [2016/679].

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) 2022 [2022/2065].

Silva AP and Cabral RM, 'Beyond Copy-Paste: The Brussels Effect and the Evolution of Digital Law in Brazil' (2025) 16(1) Beijing Law Review 610.

Supremo Tribunal Federal, 'STF Avança Em Análise de Recursos Sobre Normas Do Marco Civil Da Internet' <https://noticias.stf.jus.br/postsnoticias/stf-avanca-em-analise-de-recursos-sobe-normas-do-marco-civil-da-internet/> accessed 19 June 2025.

Tema 987 - Discussão sobre a constitucionalidade do art 19 da Lei n 12965/2014 (Marco Civil da Internet) que determina a necessidade de prévia e específica ordem judicial de exclusão de conteúdo para a responsabilização civil de provedor de internet, websites e gestores de aplicativos de redes sociais por danos decorrentes de atos ilícitos praticados por terceiros (RE 1037396) Ministro Dias Toffoli (Supremo Tribunal Federal).

'The DPC's Engagement with Meta on AI' (*Data Protection Commission*, 14 June 2024) <https://www.dataprotection.ie/news-media/latest-news/dpcs-engagement-meta-ai> accessed 19 June 2025.

'UOL Economia' <https://economia.uol.com.br/cotacoes/> accessed 24 June 2025.

Valle Va, 'PL Das Fake News e Regulação Retórica Das Redes Sociais' (*Consultor Jurídico*, 4 May 2023) <https://www.conjur.com.br/2023-mai-04/interesse-publico-pl-fake-news-regulacao-retorica-redes-sociais/> accessed 25 June 2025.

VOTO No 11/2024/DIR-MW/CD [2024] ANPD PROCESSO No 00261.004509/2024-36, Miriam Wimmer.

VOTO No 23/2024/DIR-JR/CD [2024] ANPD PROCESSO No 00216.004529/2024-36, Joacil Basílio Rael.

'WHO WE ARE - The Company behind the Success' (*Statista*) <https://www.statista.com/aboutus/> accessed 19 June 2025.

14. REGULATING THE ATTENTION ECONOMY: META'S NON-COMPLIANCE WITH THE EU

https://doi.org/10.36592/9786554603065-13

Mahon McCann

Abstract:
In Nov 2023, Meta introduced a "pay or consent" option for users, creating a standard subscription price of €9.99 to prevent personal data combination for advertising. However, in March 2024, the EU opened a non-compliance investigation under the Digital Markets Act (2022) to investigate if this "pay or consent" model offered a genuine alternative to the data-driven advertising business model as required by the DMA (European Union, 2022). In April 2025, the EU concluded this investigation and fined Meta €200 million for breaching the Digital Markets Act because this subscription model did not provide a genuine alternative to the advertising business model. So far Meta has failed to provide a genuine alternative to their normal service that uses less personal data of users. Furthermore, in June 2025, Meta announced that after eleven years, its messaging platform WhatsApp will begin serving ads to users, raising further questions about Meta's future plans to cooperate with the EU's DSA and DMA. This paper argues that Meta's reliance on the extractive digital advertising model and its platform design prevents compliance with the DSA and DMA, leading to an adversarial relationship between the company and the EU.
Keywords: Meta, Social Networks, Digital Services Act, Digital Markets Act, Social Media, Social Media Regulation, Commodification of Attention, Attention Economy

1. Introduction

In November 2023, Meta introduced a "pay or consent" option for users, creating a standard subscription price of €9.99 to prevent personal data combinations for advertising purposes. In March 2024, the EU opened a non-compliance investigation under the Digital Markets Act (2022) to investigate if this "pay or consent" model offered a genuine alternative to the data-driven advertising business model, as required by the DMA (European Union, 2022). In April 2025, the EU concluded its investigation and fined Meta €200 million for breaching the Digital Markets Act because this subscription model did not provide a genuine alternative to the advertising business model.

So far, Meta has failed to provide a genuine alternative to its standard service that the EU requested, which uses less personal data from users and in June 2025 Meta announced that after eleven years, its messaging platform WhatsApp will begin serving ads to users, raising further questions about Meta's future plans to cooperate with the EU's DSA and DMA. As we will argue in this paper, the relationship between regulators and the social media giant will be adversarial because Meta has relied on the extractive digital advertising business model, and this business model has shaped the design and evolution of the platform. In this paper, we will (1) examine the advertising-driven business model of Meta, (2) how this business model has shaped the evolution of the platform design and (3) what the future holds for the relationship between Meta and the EU.

2. The Advertising Business Model

Ethan Zuckerman, inventor of the pop-up ad, described personalised advertising for revenue as the 'original sin' of the internet (Zuckerman, 2014). The advertising business model solved a problem for the internet - how was it going to make money? The advertising business model allowed tech companies to give away their services and products for free to users and instead generate profits from advertisers and third parties. The advertisers pay for user data and an opportunity to influence the user's behaviour to their preferences, which results in the advertising or attention economy business model becoming the de facto business model of the internet: "the business model relies on the data of users, extracted from attention, which is sold to third parties and advertisers for profit" (Myllylahti, 2018).

Traditionally, before the internet, this advertising business model was seen in television, newspapers and radio. However, the critical difference between the traditional attention economy and the internet model is the real-time collection of user data. User data has been described as the 'life-blood' of the advertising business model and comes from tracking the user's attentional habits to build a predictive model of their behaviour (Ghosh & Scott, 2018, 1). The more behavioural data that the company can collect, the better-targeted ads can be to their unique interests and the higher the chances of success (Ghosh & Scott, 2018). However, what this means for

social media platforms with the advertising business model is that the users are not their customers; rather, the advertisers are the customers, and hence, the users are the product. As former Facebook and Google Chief Engineer Justin Rosenstein said, "If you are not paying for the product, you are the product" (McDavid, 2020).

This advertising business model, which makes the customer the advertiser rather than the user, also prevents Meta from changing significantly to meet the EU's demands. In traditional media with the advertising business model (for example, newspapers or television), the revenue from advertising is re-invested into content production to attract more readers or viewers. However, digital platforms with the advertising business model do not produce their own content (the content is user generated). Hence, the platforms re-invest their revenue into the transformation of the medium itself: the persuasive design of the platform to capture more and more attention and action (Carah & Brodmerkel, 2022, 112). Meta's platform has evolved to become maximally efficient at capturing attention and action for advertisers while not necessarily pleasing their users, who get the apps for free.

We can better understand this difference between subscription-based business models and advertising models by comparing Meta to a subscription-based business like Netflix. The competition between Netflix and other close substitutes, such as Disney Plus or Amazon Prime, is still primarily on content. To attract more viewers to watch Netflix, the company needs a hit show, bigger actors and actresses or writing, etc. In other words, Netflix must innovate on the level of content, creating more engaging content for users to view for their monthly subscription payment, in order to be successful.

Meta, on the other hand, does not generate the content themselves, so the platform must instead provide real careers and opportunities for "influencers" to encourage these former users to create their content. On Social Media platforms, the "followers" or "friends" count generates social competition - creators are ranked and compared based on their success on the platform. Those who are successful in attaining higher social status become real celebrities and attain genuine economic and social opportunities in the offline world. The typical subscribers on Netflix are not also the creator, and there is a much greater gap between creators and consumers on Netflix, which is more akin to traditional film or television.

Social media platforms entice creators and influencers with the promise of overnight success through virality. This promise encourages frequent uploads which drive engagement and advertising. The use of virality and social success is integral to the business model to keep content creators uploading content on the sites for the users to watch and hence to generate revenue for the advertisers. An influencer who spends all their time on the platform creating content to attract others on the platform, essentially for free, is the ideal customer for the advertising business model!

The user on social media is also in a totally different position than the subscriber of Netflix. The user on social media has a free and public-facing profile (at least in some aspects of the profile), which displays information about the user which others can interact with. For the advertising business model, the profile functions to track the user's behaviour, likes and dislikes, views, engagement, etc and provides sentiment metrics for the advertisers (Hwang, 2020). The users on social media are both (1) the generators of data for the advertisers and (2) the prime targets of the advertising to purchase goods, services or messages. This dual role for the users as providers of data and consumers of goods, services and messages means having a large body of active users is essential to the success of platforms with the advertising business model; therefore, 'network effects' play a much bigger role in the success of the platform - the network gains more value the more people use it.

For the company with the advertising business model to be effective, the platform needs the largest user base possible, as active as possible, to generate data for the advertisers and the most opportunities for sales. The scale is crucial for providing sufficient data for advertisers to obtain deterministic results across large enough populations. In other words, in television advertising, the advertiser gets probabilistic results, as one television advertiser noted, "Half the money I spend on advertising is wasted; the trouble is I do not know which half" (Linford, 2020, p.129). However, through digital advertising across a large enough population with enough in-depth information on that population, advertisers can achieve predictable and deterministic returns through programmatic advertising. It is this perceived predictability and reliability that has earned programmatic advertising the lion's share of the advertising market; Digital channels now account for 72.7 percent of worldwide ad investment, with online spending exceeding US$790 billion in 2024 (Kemp, 2025).

Netflix itself has nearly 300 million paying subscribers, but with these numbers, a social media platform would not be viable. For example, when Twitter became X and lost many subscribers, falling to around 600 million, the platform also lost most of its advertisers (for both political and economic reasons). Consequently, it has tried to implement a subscription model, gaining around 1.4 million active subscribers. Subscription companies like Netflix succeed by pleasing their subscribers and investing in new and engaging content. Large social media platforms with the advertising business model succeed by generating data and sales for advertisers, while content and platform design aim at bringing in and keeping users and influencers for the advertisers. Therefore, the social media platform specialises in the curation, alteration and ultimately manipulation of content to gather as many users as possible and to keep them engaging for as long as possible. The social media platform invests in the medium to more efficiently capture the maximum amount of attention and action.

In summary, the success of Meta as a company has been interlinked with the advertising business model since the late 2000s. The user is the product, and the true customer is the advertiser or third party. This model incentivised social media to innovate on the level of the medium rather than content, driving developments like influencers, virality, and leveraging network effects. These features, such as limitless pools of user-generated content, viral videos, and economies of scale, drove the success of large social media platforms like Meta but now prevent Meta from changing significantly to meet the EU's demands for a safer platform.

3. The Evolution of Platform Design

If we zoom in on the platform design of social media platforms like Meta, we can see how the business model has shaped the evolution of their design and how this prevents compliance with the EU's demands. In this section, we will walk through some of these design features and show how they evolved to satisfy the short-term aims of advertisers over the long-term good of users. This section illustrates how the platform design itself presents an obstacle to Meta's ability to meet the EU's demands.

The incentives created by the advertising business model are such that once a company has leveraged network effects and attained the lion's share of the market, it must keep users regularly engaged for as long as possible on the platform. Because the advertisers pay per ad and therefore the more people who see the ads, the more possibility for engagement and clicks. Therefore, the social media platform employs the platform design to capture and hold user's attention for as long as possible, which leads to optimisation for *addiction*.

In Nir Eyal's behavioural design manual Hooked (2014), the goal of social media platforms is to dynamically optimise content to create "internal triggers" in users - urges for action that arise without perceptual stimuli to form habitual engagement (Eyal, 2014). Eyal defines habits as "automatic behaviours triggered by situational cues" and "behaviours we do with little to no conscious thought" (Eyal, 2014, 12). He argues that creating these internal triggers is the "brass ring" for social media companies and is the source of their "economic value", which "links their services to the users' daily routines and emotions" (Eyal, 2014, 1).

The internal triggers function by forming a link between particular emotions and the user's choice of actions. In other words, the goal of these internal triggers is that when users feel bored, they instantly open Twitter, or when loneliness arises, they scroll Facebook, etc (Eyal, 2014). If companies can create these internal triggers, the users will police themselves and be compelled to return to the platform to relieve uncomfortable emotions without the need for expensive and inefficient marketing campaigns.

We can understand how reinforcement learning creates these internal triggers from another example in gambling technologies. In gambling technologies, the persuasive design aims to create a closed loop known as the 'ludic loop', and within the ludic loop, addicts' emotional regulation functions through the repetition of certain behaviours (McKelvey & Hunt, 2019). The ludic loop is what we see in the 'Hooked model' (Eyal, 2014), where the goal is to create an association between the user's emotion and the product as a source of relief (Bruineberg, 2023). Social media platforms generate emotions through their design and the corresponding repetition of behaviours that relieve the user. The platform keeps users coming back by causing

the problem, the personalised pain point of emotional discomfort, and also providing the solution, readily available endless supplies of feel-good novel rewards.

The internal triggers are created through reinforcement learning: which refers to learning where agents learn to navigate an environment a specific type of information, reinforcers, which index how good or bad something is (Collins, 2024). This employment of reinforcement learning is why frequently in the public sphere, social media is considered a digital skinner box for modern humans and this claim is supported by research to show that behaviour on social media follows the principles of reward maximisation (Lindström et al, 2021, p. 7).

Social media platforms set up a controlled environment which provides a dynamic feedback system of positive social rewards (likes, shares, comments, novel content) that condition users to associate the product with high levels of reward, but then also negative feedback through social punishments (ostracisation, negative comments, FOMO, no social success, or just finding nothing entertaining), which feel bad and hence want to be avoided. Therefore, over time, the user is conditioned to associate the platform with easily available feel-good rewards, which are available 24/7 for free. This reinforcement learning is aided by carrying smartphones with us all of the time that carry the applications. Many examples of addictive design techniques are available on social media platforms with the same aims:

1. The infinite scroll: removes friction by endlessly populating new content rather than requiring the user to click on the next page, which can foster a state of "dissociation" or "doom scrolling". The infinite scroll has been compared to taking the bottom out of a glass of wine, creating a never-ending supply of content optimised and personalised for addiction.

2. Intermittent Variable rewards: Notifications, messages, and novel stimuli trigger the brain's reward system, similar to the anticipation of gambling or eating food, and reinforce "checking habits".

3. Manipulating Social Validation: Features such as"likes" and "shares" offer social rewards or punishments (social ostracism and loss of status) according to the platform's game rules, taking advantage of powerful human cognitive biases such as social comparison, the fear of missing out, and the desire for status and recognition.

Conveniently, these features also act as metrics on social media platforms to track sentiment for advertisers (Hwang, 2020)

In summary, the need to optimise for as much attention as possible, keeping users on the platform to generate data and sales for advertisers, acted as a selection pressure that drove the platform design evolution of the large social media platforms from relatively harmless in the early 2000s to as addictive as possible today. The current platform is the end-state of a process of iteratively refining digital addictions. Faced with the option of paying monthly for their fix, many users might choose a healthier option and leave the toxic design behind, or a platform without the toxic design would lose the frequent users who are genuinely addicted, and thus the best customers. Therefore, it is a lose-lose situation for Meta, and their only strategy is to continue with the advertising business model. For Meta to stop using this platform design that facilitates its business model would be suicide; consequently, the platform design itself is another reason why Meta cannot comply with the demands of the European Union.

4. The future of Meta and The EU

In conclusion, for the advertising business model, the user is not the customer but rather the product. The actual customer is the advertiser or third party that wants to pay for the user's data and the chance to influence their behaviour. The very large social media platforms require network effects to have a sufficiently large population to generate data and opportunities for advertisers to influence their behaviour. They must keep as many people on the platforms for as long as possible to remain financially viable. The platform's optimisation for generating data and tools for advertisers leads to massive investment in innovating at the medium level (the platform design) to be as addictive as possible.

If the company switches to a subscription model, most users will likely opt not to pay for the service and instead switch to a free alternative, a fact Meta is aware of. The implementation of a subscription model or an advertising-free model could potentially cause a situation for the social media giant that is not unlike a run on a bank: a loss of confidence in the bank due to rumoured insolvency causes customers

to withdraw their money, and this, in turn, causes the bank to fall into the predicted bankruptcy. The same thing could happen on social media: users do not pay for subscriptions and therefore cannot use the service, influencers receive less engagement on their content and switch to other platforms, and advertisers generate less revenue and spend their money elsewhere. The value of the company lies not in the software but in the critical mass of users and their frequent engagement on the platforms; therefore, if people start leaving due to a subscription or otherwise, Meta is in big trouble. We have seen social media giants such as Myspace, rise and fall before,, and these companies often die as quickly as they emerged in the first place. Therefore, Meta and the EU are likely to maintain a continuing adversarial relationship due to Meta's inability to shift away from its advertising business model and persuasive platform design.

This adversarial relationship between the EU and Meta highlights the broader challenge of regulating digital technologies with the advertising or attention economy business model. The successful path forward for regulators lies in linking the platform's success to its users' success. The business model and platform design certify an extractive relationship between Meta and its users. Therefore, the EU's attempts to decouple behavioural data collection from the programmatic advertising industry are necessary first steps to getting users to a better social media platform.

Despite Meta's opposition, in the long run, moving away from the unstable advertising business model is in their best interests as well. Like all companies in this area, Meta faces a collective action problem - if users must pay for their platform and not the others, their business will fail. However, If all social media platforms, now and in the future, were subject to the same regulations preventing behavioural data collection from being used for programmatic advertising, then these platforms and the whole industry would need to compete to please their users instead of the advertisers to succeed. In this sense, regulating the attention economy is only about one thing: stopping the platforms profiting from advertisers at the expense of their users.

Bibliography

Bruineberg, J. (2023). Adversarial inference: predictive minds in the attention economy. *Neuroscience of consciousness, 2023*. https://doi.org/10.1093/nc/niad019

Carah, N. and S. Brodmerkel (2022). Regulating Platforms' Algorithmic Brand Culture: The Instructive Case of Alcohol Marketers on Social Media. *Digital Platform Regulation, Springer*: 111-130.

Clark, L. (2024, May 16). EU probes Meta over its provisions for protecting children. *The Register*. https://www.theregister.com/2024/05/16/eu_investigates_meta_over_its/

Collins, A. G. (2024). Reinforcement Learning. In M. C. Frank & A. Majid (Eds.), Open *Encyclopedia of Cognitive Science*. MIT Press. https://doi.org/10.21428/e2759450.36d1ca92

European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (*Digital Services Act*). https://eur-lex.europa.eu/eli/reg/2022/2065/oj

European Union. (2022). Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (*Digital Markets Act*). https://eur-lex.europa.eu/eli/reg/2022/1925/oj

European Commission. (2024, April 30). Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act [Press release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373

Eyal, N., & Hoover, R. (2014). *Hooked : how to build habit-forming products*. Portfolio/Penguin.

Ghosh, D. & B. Scott (2018). "The technologies behind precision propaganda on the Internet." *New America*.

Hwang, T (2020). *Subprime Attention Crisis: Advertising and the Time Bomb At the Heart of the Internet*. FSG Originals x Logic, Farrar, Straus and Giroux.

Kemp, S. (2025, February 5). Digital 2025: Global advertising trends. *DataReportal*. https://datareportal.com/reports/digital-2025-sub-section-global-advertising-trends

Linford, J. (2020). "Copyright and Attention Scarcity." *Cardozo L. Rev. 42*: 143.

Lindström, B., Bellander, M., Schultner, D.T. et al (2021). A computational reward learning account of social media engagement. *Nat Commun 12*, 1311. https://doi.org/10.1038/s41467-020-19607-x

McKelvey, F. & Hunt, R (2019). "Discoverability: Toward a definition of content discovery through platforms." *Social Media+ Society 5*(1): 2056305118819188.

McDavid, J (2020) "The Social Dilemma," *Journal of Religion & Film: Vol. 24: Iss. 1, Article 22.*

Myllylahti, M. (2018). "An attention economy trap? An empirical investigation into four news companies' Facebook traffic and social media revenue." *Journal of Media Business Studies 15*(4): 237-253.

Narayanan, A. (2023). Understanding social media recommendation algorithms. *Knight First Amend. Inst.*

Zuckerman, E. (2014, August 14). The internet's original sin. *The Atlantic*. https://www.theatlantic.com/technology/archive/2014/08/advertising-is-the-internets-original-sin/376041/

EDITORS

Edoardo Celeste is an Associate Professor of Law, Technology and Innovation at the School of Law and Government of Dublin City University, Ireland. Edoardo is the Programme Chair of the Erasmus Mundus Master in Law, Data and Artificial Intelligence (EMILDAI), the Deputy Director of the Dublin European Law Institute, the coordinator of the DCU Law and Tech Research Cluster, and the Deputy Editor of the European Journal of Law and Technology (EJLT). He is the editor of the Routledge Series in Digital Law and Governance as well as the author of the monographs 'Digital Constitutionalism: The Role of Internet Bills of Rights' (Routledge 2022) and 'The Content Governance Dilemma' (Palgrave 2023).

Ilton Norberto Robl Filho is an Associate Professor at the School of Law of Federal University of Paraná (UFPR) and of the Brazilian Institute of Teaching, Development and Research (IDP). He has doctorate in Law (2012), having obtained post-doctorate certificates in Constitutional Law (2015) at the School of Law of Pontifical Catholic University of Rio Grande do Sul (PUCRS). Visiting Researcher at University of Toronto - Canada (2012), at Max Planck in Heidelberg - Germany (2013) and at Nova University Lisbon – Portugal (2022). Participates in the academic collaboration between IDP and Dublin City University LAW AND TECH RESEARCH CLUSTER since 2024. He is research group leader Peter Häberle in Brasília – IDP and participates in an international research network with several institutions such as the University of Granada, PUCRS and IDP.

Ingo Wolfgang Sarlet. PHD in Law at the Munich University Germany. Full professor for constitutional law at the Pontifical Catholic University Rio Grande do Sul - PUC-RS - where he coordinates the Master and PHD-Program in Law. Post-Doctoral Studies at the Munich University. Former visiting researcher at the Georgetown Law Center, Harvard Law School, Stellenbosh Institute for Advanced Studies, Max-Planck-Institute for Social Law, Max-Planck-Institute for Private Law. Former Judge at the State Appeal Court and State Electoral Court Rio Grande do Sul. Member of the Center for Constitutional Studies of the Brazilian Federal Supreme Court. Lawyer and legal advisor.

ABOUT THE AUTHORS

Marcella de Pinho Pimenta Borges Ramos
Master's candidate in Constitutional Law at the Instituto Brasiliense de Direito Público (IDP). Postgraduate degree in Contracts and Civil Liability from the Instituto Brasiliense de Direito Público (IDP). Postgraduate degree in Legal Order and Public Prosecution from the Fundação Escola Superior do Ministério Público do Distrito Federal e Territórios (FESMPDFT). Bachelor of Laws from Centro Universitário de Brasília (UniCEUB). Attorney since 2010, with specialized practice in strategic civil litigation. Currently a partner at Roque Khouri e Pinheiro Advogados.

Cyntia Melo Rosa
Master and PhD student at the Brazilian Institute of Education, Development and Research (IDP-DF)

Yury Rufino Queiroz
Yury Rufino Queiroz is a Brazilian lawyer and State Attorney for the State of Piauí. He is also a member of the Artificial Intelligence Commission of the Federal Council of the Brazilian Bar Association. He is currently pursuing a Doctorate in Constitutional Law at the Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP), Brasília, Brazil, and holds a Master's degree in Law and Conflict Resolution from the University of Fortaleza (UNIFOR). His professional and academic work focuses primarily on Constitutional Law, Criminal Law, Electoral Law, and Digital Rights. His research concentrates on state surveillance, mass data collection, digital privacy, keyword warrants, and the constitutional limits of investigative technologies in democratic societies. He has developed academic research and legal studies concerning fundamental rights, judicial review, artificial intelligence, and constitutional safeguards in the digital era, with emphasis on comparative constitutional law and the jurisprudence of the Brazilian Supreme Federal Court and international human rights courts.

Celso Reic Urbieta
Celso Reic Urbieta is a lawyer registered with the Brazilian Bar Association (OAB/MS) under number 15,958. Graduated from the State University of Maringá (UEM), he began his work with entrepreneurs due to his experience in a large export company. He then pursued a specialization in International and Economic Law at the State University of Londrina (UEL). He worked as a Legal Advisor at the Municipal Attorney's Office of Campo Grande/MS, in tax litigation. Currently, he is an Attorney for the City Council of Jateí/MS, holds a Master's degree in Public Management from the Federal University of Grande Dourados (UFGD), and is a Ph.D. candidate in Constitutional Law at IDP, academically developing his practical work in the public sector. He has published articles and books on the implementation of Compliance programs in city councils. He is also a Visiting Scholar at the University of Oklahoma.

Pedro Nilson Moreira Viana
Ph.D. Student in Constitutional Law at the Brazilian Institute of Education, Development, and Research (IDP - Brasília, 2025). Master of Laws (LL.M.) and Justice

Institutions from the Federal University of Maranhão (UFMA). Visiting international student at the Universidade Autónoma de Lisboa (UAL). Master of Laws at the Pontifical Catholic University of Rio Grande do Sul (PUCRS). Master in English from Cultural Norte Americano (CNA, 2017). German language student at the Goethe-Institut für Sprache und Kultur der Bundesrepublik Deutschland, São Paulo, Brazil (2025). Holds Postgraduate Diplomas in Notary and Public Registry Law from UFMA (2023) and in Constitutional, Administrative, and Tax Law from PUCRS (2025). Bachelor of Laws from UFMA (2019). Attorney at Law (currently on leave). Serves as a Law Clerk at Maranhão State Court of Appeal. Served as Public Notary at the 1st Notary and Registry Office of the Judicial District of Água Branca, Alagoas, Brazil (2025-2025). Member of the Land Regularization Committee of the Brazilian Bar Association (OAB/MA). Peer Reviewer for the Journal of Legal Studies of the Superior Court of Justice (REJuriSTJ), Reviewer for the Alagoas Court of Appeal's Journal and Peer Reviewer for the Pará Public Prosecutor's Office Journal.

João Pedro Barbosa Mota
Lawyer. Master's student in Constitutional Law at the Institute of Education, Development and Research (IDP). Postgraduate degree (lato sensu) in Procedural Practice in Superior Courts from the University Center of Brasilia - Ceub. Member of the Research Center on Constitutional Freedoms and the Research Group on Civil Procedure in light of the 1988 Federal Constitution, both at the Brazilian Institute for Development and Research - IDP.

Vitória Andréa De Almeida Nicolau
She holds a Bachelor's degree in Law from the Jorge Amado University Center (2007), a Postgraduate degree in State Law from the Federal University of Bahia - UFBA (2011), a Postgraduate degree in Public Bidding and Contracting from the Renato Saraiva Teaching Complex - CERS (2022), and a Postgraduate degree in Constitutional Law from the Brazilian Institute of Education, Development and Research - IDP (2022). She is a Government attorney for the Municipality of Ilhéus/BA, holds a Master's degree in Law from IDP (2023), and is a PhD candidate in Law at the same institution (2024). She has experience in the area of Constitutional and Administrative Law, with an emphasis on public bidding and contracting.

Vitória Monego Sommer Santos
Ph.D. in Law from the Università degli Studi di Perugia and Universidad de Salamanca. Currently pursuing a master's degree in Law at the Pontifícia Universidade Católica do Rio Grande do Sul (Capes scholarship), and conducting postdoctoral research in Law at the Universidade do Vale do Rio dos Sinos.

Luisa Maciel Perez
Luisa M. Perez is a Brazilian legal professional currently pursuing the European Master in Law, Data and AI (EMILDAI), an ERASMUS MUNDUS program coordinated by Dublin City University in partnership with Avignon Université and Università di Pisa, with a specialisation in data and AI governance. She has worked in major law firms in Brazil, including as a lawyer, and has more recently worked as a researcher for international organisations, including the ADAPT Research Centre at Trinity College

Dublin and the Privacy & Access Council of Canada. She also took part in the Center for AI and Digital Policy (CAIDP).

Mahon McCann
Mahon McCann is an Irish philosophy lecturer and PhD researcher based in Dublin. He holds a BA in Philosophy & Economics from University College Dublin and an MA in English from Queen's University Belfast. In 2022, he was awarded the DCU School of Theology, Philosophy, and Music Scholarship for his research on the attention economy, particularly how social media companies' commodification of user attention shapes users' moral character. His academic research is at the intersection of transformative technologies, cognitive science and ancient virtue ethic.