A NEW GENERATION OF PLATFORM REGULATIONS

2023 Outcome of the UN IGF Coalition on Platform Responsibility

Luca Belli, Yasmin Curzi, Guy Berger, Rolf H. Weber, Shilpa Jaswant, Sofia Chang, Walter B. Gaspar and Clara Almeida

Abstract: This paper aims to discuss a selection of emerging platform regulations worldwide, scrutinizing their human rights implications. It stems from the analysis of different legislative initiatives, including the Brazilian Fake News Draft Bill (PL 2630/2020), the European Union's Digital Services Act (DSA), the Indian IT Act, and the Chinese Internet Information Service Algorithmic Recommendation Management Provisions. Our aim is to dissect their complexities, particularly their potential impacts on freedom of expression, privacy, due process, and other human rights. The primary objective of this paper is to discuss the extent to which these new (proposed) regulatory paradigms can reshape the digital ecosystem. In this sense, we aim to discern viable strategies and, considering the UNESCO's Guidelines for Regulating Digital Platforms (2023), provide a possible framework for fostering a universally safe, equitable digital environment that steadfastly respects and upholds fundamental human rights while avoiding internet fragmentation.

1. Introduction

The implications of social media platforms on our social, economic, and political lives have been increasingly significant, with the ubiquity of social media platforms in our daily lives. Their role in facilitating communication and their influence on how information is disseminated and consumed have drawn attention to the challenges they pose[1]. Perhaps most notably, online social media platforms have been raising complex concerns related to the protection and promotion of human rights, particularly with respect of disinformation, hate speech, and privacy violations. Simultaneously, the responsibility[2] of these platforms in mediating public discourse (Vickery & Everbach, 2018) as well as the adequacy of existing legal frameworks to govern their operation, remain topics of ongoing debate.

Within these debates, it is crucial to differentiate between approaches to regulation. On one hand, there is a notable push for more restrictive regulation targeting Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs), which often serve as centralized, large-scale entities in the digital landscape, such as in the Digital Services Act and the Digital Markets Act. It might encompass rules for mandated interoperability (Belli & Zingales, 2022), standards for co-regulation (Belli et al., 2021) and human rights impact assessments (UNESCO, 2023). This type of regulation aims at holding the large tech companies more accountable for their outsized impact on public discourse and human rights. On the other hand, there is an emerging discourse advocating for enabling regulation that fosters a more diversified, non-monopolistic social media environment – such as a user-centered approach (Hartmann, 2017), federated communities[3] (Belli et al., 2021), or self-regulation (Weber, 2021), and else.

In a broader perspective, these divergent regulatory paths illuminate the complexity of crafting laws that are both effective and equitable, shedding light on the blind spots in current regulatory frameworks. While much of the academic and policy discussion on platform responsibility centers on the unique challenges and opportunities presented by these companies, it is important to contextualize this within existing content regulation

[1] Cf. Berger et al. (2023). Platform problems and regulatory solutions: Findings from a comprehensive review of existing studies and investigations (CI-2023/WTR/1). UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000385813.

[2] See the Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie, (2011) Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, UN Human Rights Council Document A/HRC/17/31. Available at: https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

[3] Including not only smaller, niche platforms but also decentralized platforms that offer alternative models for data ownership and community governance.

frameworks that apply both online and offline.

Legal provisions for content regulation, which include aspects like hate speech, obscenity, and libel, exist in most jurisdictions and are not suspended in the online environment. Even in jurisdictions like the United States, where Section 230 of the Communications Decency Act provides platforms with a degree of immunity for user-generated content (UGC), there are exceptions that hold platforms accountable for illegal content such as child sexual abuse material (CSAM) and copyright infringement when they are aware and take no action.

A critical question, therefore, is not about the mere existence of regulations but the emerging regulatory frameworks designed for these platforms, given their expanding roles across societies, democracies, and economies. This distinction is especially relevant when considering diverse national contexts, such as China, where generic content regulations and platform-specific laws intersect in ways that have significant implications for digital rights and freedoms. Thus, any comprehensive analysis must account for both 'generic' and 'specific' regulatory frameworks, as the two are inextricably linked in shaping platforms' legal liabilities and, by extension, their impact on users' rights.

This paper delves into these issues by presenting and scrutinizing a selection of emerging regulations for platforms worldwide. Our discussion encompasses various legislative initiatives, including the Brazilian Fake News Draft Bill (PL 2630/2020), the European Union's Digital Services Act (DSA), the Indian IT Act, and the Chinese Internet Information Service Algorithmic Recommendation Management Provisions. We have selected these laws and draft bills as case studies, given that these jurisdictions account for the highest numbers of social media users. Each of these laws, in their unique contexts, reflects attempts to tackle the nuanced challenges posed by digital platforms, representing different perspectives and approaches to managing the digital world. In this sense, we delve into these complexities and dissect the intricate interplay between these regulations and human rights, as well as the risks for internet fragmentation[4] (De Gregorio & Radu, 2022)

A growing consensus recognizes the internet as a public space (UNESCO 2023), one that should be trustworthy, equitable, and respectful of all individuals' fundamental rights. Yet, the strategies to realize this vision remain elusive and contentious. The challenge lies not only in preserving the universal values inherent in our digital spaces but also in preventing internet fragmentation brought about by

[4] In addressing the concept of "internet fragmentation", it is important to clarify that our paper's primary focus is on legal and regulatory dimensions. However, we acknowledge that the term can be interpreted through multiple lenses—including technical aspects, political/legal layers, corporate "walled gardens" that limit interoperability, and data hoarding by companies. From a technical standpoint, different regulatory regimes do not inherently fragment the internet as a 'network of networks,' but rather introduce features of differentiation among national networks. Given the complexity of this term, we intend to further unpack its various facets in later sections of the paper.

varying legal and regulatory approaches across jurisdictions.

While human rights principles are universally applicable, their interpretation can be subject to legitimate variations, as indicated by the International Covenant on Civil and Political Rights' general comment on Article 19. These variations, often influenced by differing cultural and societal conceptions of 'public morals' or 'safety,' are generally subjected to a three-part test for acceptable restrictions, namely that they are provided for in law, necessary and proportionate, and pursue a legitimate objective. Importantly, this nuanced interpretation allows for some diversity in applying human rights standards without necessarily causing fragmentation at a universal level.

Within this context, the challenge extends beyond merely preserving universal human rights in digital spaces to also preventing internet fragmentation due to disparate legal and regulatory frameworks across the selected jurisdictions[5]. To address this complex landscape, this paper evaluates legislative strategies from around the world. Its aim is to contribute to the development of a robust framework for platform regulation, designed to assist platforms in safeguarding fundamental human rights while navigating the intricacies of legal and jurisdictional differences, thereby minimizing the risk of internet fragmentation.

2. Large Online Platforms and its Challenges: A New Digital Landscape

The digital landscape has experienced a shift in power dynamics over the recent decades, largely due to the evolution of new technologies and the deployment of artificial intelligence systems by digital platforms (Gorwa et al., 2020; Llansó, 2020; Tufekci, 2017). This transformation has paved the way for what Zuboff terms as 'behavioral surplus' (2019) – the ever-expanding methods that digital platforms, especially commercial platforms (e.g., Google, Meta, Twitter/X, TikTok, etc.), employ to amass and monetize the residual data of users.

To grasp this emergent phenomenon, it's essential to delve into the concept of "structural power"[6], a term coined by Susan Strange (1988) and applied to digital platforms by Belli (2022).This idea encapsulates the profound capability of

[5] For further discussion on this topic, please refer to "The universal norm of freedom of expression – towards an unfragmented Internet. In: Berger, Guy. The Net and the Nation State. Multidisciplinary Perspectives on Internet Governance. (Ed: Kohl, U). Cambridge: Cambridge University Press, 2017.

[6] The concept was first elaborated by Susan Strange in her 1988 book "States and Markets." According to the British Political Scientist, power does not manifest itself in the sole form of command and control, but the structures that underpin the functioning of states and markets can be seen labyrinths shaped by the actors holding structural power. Such a metaphor is useful to understand the structural power as the power of those who are able to define where walls are in the labyrinth or when doors can be open or closed, thus ultimately controlling how the mice that are inside can move.

digital platforms not just to influence user behaviors, but also to sculpt the very contours of our global digital landscape. Belli states that digital platforms wield considerable structural power in a trifold manner: they possess quasi-legislative, quasi-judicial, and quasi-executive power. Respectively, through their self-titled community policies and terms of service, they enact rules – almost like a Legislative body – that regulate user behavior on their platforms. When disputes arise or actions are called into question, platforms also assume the role of a judicial body, making judgments that can have a significant impact on users' fundamental rights, thereby employing their quasi-judicial power[7]. Finally, they also have the power to enforce their rules and guidelines, and even levy sanctions on users based on their discretion – exhibiting a quasi-executive power.

This power lies in the hands of these digital platforms without effective checks and balances to prevent potential abuses, democratic legitimacy, or transparency (Haggart & Keller, 2021). As a result, these entities have the freedom to act as legislators, judges, and executors all at once.

Ultimately, this unchecked power within digital platforms can lead to an imbalanced online environment, where platforms exercise disproportionate authority and users' rights and freedoms remain inadequately protected. Furthermore, in many jurisdictions, States fail to fulfill the role of ensuring users' rights[8], creating a precarious situation where users are vulnerable to violations of their rights by both private and public powers. Consequently, the regulatory implications are severe and demand stringent criteria for evaluation.

Given this pressing situation, the alarm has been raised by various sections of society – journalists, whistleblowers, and civil society – about the multiple risks and harms to fundamental rights caused by these platforms. In response, national and international legislative initiatives have begun to emerge to curb potential abuses by digital platforms, such as UNESCO's Guidelines for Regulating Digital Platforms (UNESCO, 2023). These standards focus on

[7] Additionally, it is essential to spotlight that jurisdictions like the United States, Brazil, and Europe have historically established liability exemption regimes for digital platforms. These legal frameworks serve dual purposes. First, they enable platforms to execute internal content moderation policies without immediate legal repercussions, thereby encouraging a responsible approach to managing user-generated content. Second, these regimes indirectly lead to a de facto outsourcing of certain judicial functions to the platforms themselves. They set their own operational standards for content, effectively filling a role traditionally occupied by judicial institutions. This shifting of roles has significant implications, as it effectively leaves traditional judicial power on the sidelines, particularly when it comes to monitoring, investigating, and acting on various violations like harassment and hate speech. In this context, the judicial system's power is manifest not through intervention but rather through its conspicuous absence. This dynamic raise critical questions about the broader administration of justice and the shifting locus of regulatory authorities.

[8] Some jurisdictions prioritize public safety over freedom of expression rights, adding another layer of complexity to the assessment of regulatory efforts.

several important areas: they seek to establish and protect the rights of users, demand transparency from the platforms, impose a duty of care and establish obligations to diligently respect and uphold human rights. This wave of efforts represents a promising start to addressing the challenges of our new digital landscape, although much work remains to be done.

The current digital landscape presents us with an array of challenges stemming from this structural power wielded by commercial digital platforms, and only recently understood by regulators, pushing for normative measures that aim to put checks on these private actors and protect users. As we step into the next section, we will dive deeper into the specifics of these emerging legislative initiatives, and how they are changing the face of our digital world.

3. Emerging platform regulations: case studies

3.1 The Brazilian Scenario

Regarding the challenges posed by the pervasive use of digital platforms in public communication, the Brazilian Legislature is still attempting to regulate this space, by introducing Draft Bill 2630 in 2020, a.k.a. "Fake News Bill".

Until now, the main law regulating platforms in Brazil is the "Marco Civil da Internet", a.k.a. MCI, which creates a liability exemption for platforms. This regime is somewhat similar to the US Section 230 and the EU ECommerce Directive. However, under MCI, the liability exemption can be revoked if application providers fail to comply with a specific judicial order to remove illegal content in a specific period and if this removal request is within their technical capacities, in the terms of its article 19.

While addressing only the specific case of intermediary liability for illegal content, the Brazilian framework lacks adequate tools to deal with disinformation or the amplification of content that harms human rights. In this sense, the emergence of digital platforms as significant players in the propagation of information, including misinformation and hate speech, led to an amendment in the Electoral Law (Law 9.504/1997) in 2017. The reform, regarding social media platforms and messaging apps, focused mainly on paid internet advertising, an attempt to stem the tide of digitally mediated disinformation.

However, as Lefèvre (2023) points out, this move unintentionally ended up cementing the position of digital platforms as crucial intermediaries in online political debates. This happened predominantly due to the insertion of Article 57-C, a result of lobbying by private sector interest groups, into Law 9.504/1997. This article effectively normalized the practice of 'content boosting', a tool leveraged by political candidates on social platforms to advertise their campaigns.

In institutionalizing platforms as mediators of political campaign advertising, the electoral reform, however, failed to establish clear rules for such procedure. As such, it overlooked mechanisms to mitigate the potential negative effects of boosted posts, paving the way for the wide dissemination of heavily funded and biased content disguised as journalism, favoring

certain candidates. It also ignored the responsibility of platforms to be transparent in the hiring and use of such services. The noticeable absence of clearly defined legal obligations, capable of operationalizing and structuring transparency practices for digital platforms, plays a critical role in perpetuating the lack of social control and the absence of democratic legitimacy of these entities (Haggart & Keller, 2021).

As of now, the Brazilian legal system heavily leans on the Consumer Protection Code (CDC) to tackle issues related to platform abuses. For instance, Article 14 has been invoked to hold companies liable for service provision failures. Although this is a valuable tool in instances where there is evident harm due to a platform's conduct, such as the unjustified cessation of a content creator's monetization service, it proves inadequate in more nebulous scenarios. One such scenario involves changes in content recommendation due to complex algorithmic systems that are difficult to inspect, control, or understand – even for their developers[9] (Seaver, 2017).

Additionally, the current legislation struggles to define and objectively evaluate what constitutes a 'risk' in the digital sphere. Brazilian law has yet to sufficiently enforce obligations of care and transparency upon digital platforms, especially regarding their operational and content moderation practices. These legislative gaps have allowed platforms to continue making arbitrary, opaque, and potentially abusive decisions, critically undermining the exercise of human rights.

3.2 The New EU Regulations

In recent years, the European Union has instated new regulations aimed at redefining the digital landscape. Two critical pillars of this regulatory transformation are the Digital Markets Act (DMA) and the Digital Services Act (DSA).

3.2.1 Digital Markets Act

The Digital Markets Act (DMA), which has been in force since May 2023, has introduced rules for platforms that act as “gatekeepers” in the digital sectors. Up to the enactment of the DMA, gatekeepers have largely been unregulated or have only been regulated based on scattered national rules, some of which pre-dated the digital economy. The DMA is establishing a set of narrowly defined objective criteria for qualifying a large online platform as a so-called gatekeeper. Platforms must have a significant impact on the internal market, serve as an important gateway for business users to reach their end users, and enjoy a durable position.

The DMA contains three main cumulative criteria that must be fulfilled to fall under its scope: (i) The size of the platform must

[9] At this juncture, it's worth distinguishing between paid and unpaid content, especially considering the focus of the Brazil case on paid content. Paid content operates under separate dynamics, often influenced by parallel ad-tech algorithms owned by major corporations. These algorithms not only determine ad placements within their platforms but also potentially direct ads to third-party sites based on data analytics. This added layer of complexity warrants a differentiated legal approach and complicates the task of ensuring accountability and fairness.

make it possible to impact the internal market (turnover of at least Euro 7,5 billion in three financial years). (ii) The platform must control an important gateway for business users towards financial consumers (more than 45 million monthly active end-users). (iii) The platform must hold an entrenched and durable position.

The DMA aims to prevent gatekeepers from imposing unfair conditions on businesses and end-users[10]. The openness of important digital services should be ensured. In September 2023, the European Commission listed six platforms as gatekeepers: Amazon, Apple, Alphabet, Meta, Microsoft and ByteDance.

Gatekeepers are obliged to conduct behavior that ensures an open online environment that is fair for businesses and consumers and open to innovation. Furthermore, companies identified as gatekeepers will have to comply with several “dos and don’ts”. Amongst others, gatekeepers must allow end users to easily uninstall pre-installed apps to change default settings on operating systems, install third-party apps, and unsubscribe from core platform services, etc. The negative side mainly concerns bans, such as with respect to data use, algorithmic ranking, etc.).

The ten core platform services within the scope of the DMA are online intermediation services, online search engines, online social networking services, video-sharing platform services, number-independent interpersonal communication services, operating systems, cloud computing services, the advertising services, web browsers, and the virtual assistance.

At present, it is premature to assess the effectiveness of the DMA. However, recent history shows that large platforms do not always comply with the EU regulatory obligations.

3.2.2 Digital Services Act

The recently passed "Digital Services Act" (DSA) is the regulatory framework responsible for the comprehensive regulation of digital services, serving as a legislative replacement for the two-decade-old eCommerce Directive. Intended to meet the contemporary challenges of the digital era, the DSA delineates obligations and responsibilities for digital service providers, ensuring they are congruent with the European Union's commitment to safeguarding fundamental rights and ensuring a transparent digital marketplace. The law innovates by introducing new obligations of due diligence, transparency, and responsibility in how companies perform their content moderation, alongside complaint and redress systems for users affected by content decisions.

The DSA architecture aims at regulating a comprehensive array of digital services, ranging from micro-enterprises to tech titans – the so-called Very Large Online

[10] It is necessary to highlight that as it seeks to restrict monopoly without providing for incentivization and support for a plurality of platforms, it is possible to argue that the fetters on the gatekeepers are not per se sufficient to empower competitors, nor to prevent the dominant practice of buying-up smaller companies.

Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) – that reach at least 45 million monthly active users. Notably, it employs a gradated approach, ensuring that the rigidity of obligations scales commensurately with a platform's reach and societal influence.

Regarding content moderation, platforms (especially VLOPs and VLOSEs) are now under a mandate to elucidate how their content moderation and recommendation systems work. Furthermore, an essential stipulation is the unambiguous differentiation between organic content and paid promotional materials, ensuring users retain discernment over their digital interactions.

Anchoring the DSA is its robust commitment to the expeditious handling of illegal content. While platforms are entrusted with the duty of swiftly addressing transgressions upon cognizance, the Act is simultaneously emphatic in its stance against indiscriminate content surveillance. Thereby, while platforms are charged with cultivating user-friendly mechanisms to report infringements, content creators and providers are also assured avenues to redress unwarranted content exclusions.

Confirming the DSA's concern with the protection of human rights on content moderation, the Act contains an explicit obligation for intermediaries to respect the fundamental rights outlined in the EU Charter of Fundamental Rights when enforcing their content restriction rules. This new provision is a shift in the regulator's comprehension of the protection of fundamental rights, especially in contrast to the eCommerce Directive's preamble, which only emphasized the importance of freedom of expression.

Ensuring compliance and adherence to its directives, the DSA underscores the importance of regular internal audits, especially for platforms of significant influence. Post-audit, platforms are expected to submit their evaluations to regulatory authorities, thereby establishing a continuum of accountability. In the realm of oversight, the Act envisages an intricate architecture of regulatory supervision. Member States will anoint Digital Services Coordinators, responsible for the Act's enforcement. In tandem, the European Board for Digital Services will operate, ensuring doctrinal consistency across the European expanse.

Another noteworthy aspect in the DSA, worth mentioning, is the requirement not only for audits (post-hoc), but also Human Rights Impact Assessments which can inform companies when taking advanced measures to mitigate assessed risk to specific human rights. How well this requirement is implemented remains to be seen, but that companies are obligated to undertake human rights impact assessments is important, including because it is a concrete mechanism that creates greater platform duty of care.

Recognizing the magnified societal impact of Very Large Online Platforms (VLOPs), the DSA ascribes additional mandates. These entities are compelled to undertake comprehensive evaluations of systemic risks emanating from their operations, further instituting mitigation measures as corrective strategies.

Finally, given the potential of digital platforms to contribute to rapid societal

shifts, the DSA conceives a crisis management protocol. This envisions a synergistic intervention mechanism, allowing platforms, Digital Services Coordinators, and the European Commission to act concertedly in addressing emergent digital perturbations.

In encapsulation, the Digital Services Act (DSA) is emblematic of the European Union's vision for a balanced digital horizon, where platform autonomy harmoniously coexists with user rights and societal welfare. However, this optimistic assessment merits some qualifications. For instance, the DSA's provision for researchers to access platform data for assessing systemic risk is a critical yet often overlooked aspect that could be a cornerstone in the pursuit of digital justice. On another note, the so-called 'Brussels effect,' the idea that EU regulations have a global standard-setting impact, may be overstated. While it is true that platforms will likely incur substantial costs to comply with DSA regulations within the EU, there is no guarantee that they will extend such compliance beyond EU jurisdiction. Recent deviations, like WhatsApp and Threads bypassing GDPR regulations outside the EU[11], serve as examples. Thus, while the DSA stands as a beacon in the quest for a transparent and equitable digital realm, its influence and effectiveness remains to be seen.

3.3 The Indian IT Act

The Indian legal system lacks a dedicated comprehensive framework to regulate intermediaries but provides certain obligations under the IT Act and its rules regarding the purposes for user content removal by intermediaries. Section 79 of the Information Technology Act 2000 (a.k.a. IT Act), when read along with Rule 3 of the Intermediary Guidelines and Digital Media Ethics Code Rules (2021), requires that the intermediary act with due diligence and publish rules and regulations on its platform. The published rules, which must be accessible by the users, cover the kinds of content it will ‘not host, display, upload, modify, publish, transmit, store, update or share’.

Rule 3, under sub-sections (b) and (d), mentions the list of information that falls within the purview of prohibited or unlawful content that an intermediary can act against. Sub-rule (8) of Rule 4 requires the intermediary to explain the action being taken and the grounds or reasons for such action. It also mandates the creator of the content is provided with ‘an adequate and reasonable opportunity to dispute the action being taken by such intermediary and request for the reinstatement of access to such information, data or communication link, which may be decided within a reasonable time'. For this purpose, the intermediary must appoint a Grievance officer, who is an employee of the intermediary responsible for hearing the

[11] Cf. https://www.theverge.com/23789754/threads-meta-twitter-eu-dma-digital-markets and https://www.euractiv.com/section/data-privacy/news/whatsapp-shifts-legal-basis-for-processing-personal-data-in-europe/.

complaints from users and other functions listed within the Rules.

These Rules allow intermediaries to deploy technology-based measures, like automated tools to identify prohibited and unlawful content on its platform. These tools must be used appropriately to not violate the freedom of speech and expression and privacy of users. The intermediary must also ensure the accuracy and fairness of such tools, the propensity of bias and discrimination, and the impact on the privacy and security of users.

The IT Act, under Section 69A, also authorizes the government to order an intermediary to remove or block any content from its platform for "the interest of sovereignty and integrity of India, defense of India, security of the State, friendly relations with foreign States or public order or for preventing incitement to the commission of any cognizable offense". Before giving an order under Section 69A, the government must follow the procedure provided in the IT (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, (2009) for examining the content in question, providing the opportunity to hear the intermediary and to order blocking or removing the content. Additionally, the rules give leeway to the government from the general rules in case of emergency for removing certain content.

Critics have raised questions about the transparency of the intermediaries' content removal activities, and their balance with freedom of speech and expression (Annappa, 2021). The intermediaries are expected to be legally responsible when they host or publish content and not unnecessarily interfere with freedom of speech. However, the problem is not limited to the powers of the intermediaries since the state has also been authorized to limit this right, following Article 19.3 of the ICCPR.

Given these extended content removal capabilities, the Supreme Court of India, in the case of *Shreya Singhal v. Union of India [AIR 2015 SC 1523],* clarified the conditions under which the rules apply. The Court determined that orders from the government or courts, under Section 79(3) of the IT Act, must strictly conform to the restrictions laid down in Article 19(2) of the Indian Constitution when curtailing the fundamental freedom of speech and expression, in addition to the exceptions listed in the IT Act.

Despite this effort to check the power of the government when controlling and regulating content, in 2021, hundreds of Twitter accounts of farmers, activists, and news websites were suspended for using the hashtag #modiplanningfarmersgenocide concerning the farmers' protest of the controversial Farm Bills by order of the government. Twitter challenged these orders of the government to suspend or remove the content from its platform before the Karnataka High Court, but the petition was ultimately dismissed by the court.

The existing framework in India raises significant concerns about the impartiality and transparency of digital content oversight. Unlike systems that vest authority in independent regulatory bodies, the Indian legal structure currently places the power to arbitrate on content issues predominantly in the hands of government officials. This concentration of

power has not been complemented by effective enforcement of obligations of care and transparency, either upon the digital platforms or the government itself. The result is a legislative gap that allows both platforms and governmental authorities to make arbitrary, non-transparent, and potentially detrimental decisions. This lack of balanced oversight and absence of an independent regulatory mechanism significantly erode freedom of speech and democratic discourse.

Table 1. Relevant regulation in India

Year	Legal reference	Highlights
2000	Information Technology Act	**Section 67**: Publishing of information that is obscene in electronic form. **Section 79**: Network service providers are not liable for third-party content "if he proves that the offense or contravention was committed without his knowledge or that he had exercised all due diligence to prevent the commission of such offense or contravention".
2008	IT Act Amendment	**Section 66A**: Punishment for sending offensive messages through communication service (see Shreya Singhal v. Union of India). **Section 69A**: Power to issue directions for blocking public access of any information through any computer resource: for 'the interest of sovereignty and integrity of India, defense of India, security of the State, friendly relations with foreign States or public order or for preventing incitement to the commission of any cognizable offense'. **Section 72A**: Punishment for disclosure of information in breach of lawful contract.
2009	IT (Procedure and Safeguards for Blocking for Access of	Before giving an order under Section 69A of the IT Act, the government must follow the procedure provided in these Rules. There is leeway for the government in case of emergency to remove certain content.

Year	Legal reference	Highlights
	Information by Public) Rules	"Bhandari (2022) notes that the interplay between Section 69A (IT Act) and these rules, as interpreted by the government, creates an opaque system whereby content creators face an 'arduous legal process to first try and secure a copy of the blocking order and then challenge it'".
2011	IT (Intermediaries Guidelines) Rules	Sets due diligence rules to be observed by intermediaries.
2015	Shreya Singhal v. Union of India	"The court declared Section 66A of the IT Act unconstitutional under article 19(1)(a) of the Indian Constitution. The court found that the section's vagueness in terms such as 'annoyance' and 'inconvenience' could create a chilling effect over a 'large amount of protected and innocent speech' (para. 83)". The Court determined that orders from the government or courts, under Section 79(3) of the IT Act, must strictly conform to the restrictions laid down in Article 19(2) of the Indian Constitution when curtailing the fundamental freedom of speech and expression, in addition to the exceptions listed in the IT Act.
2021	IT (Intermediary Guidelines and Digital Media Ethics Code) Rules	"The Rules create due diligence duties for social media intermediaries and 'significant' social media intermediaries, thus specifying the conditions of liability immunity (according to section 79 of the IT Act) for these actors. Intermediaries are required to publish monthly grievance reports and to appoint a Chief Compliance Officer, a Grievance Officer, and a Nodal Contact Person, all residing in India."

Year	Legal reference	Highlights
		Rule 3: requires intermediaries to act with due diligence and publish rules and regulations on its platform. The published rules cover the kinds of content it will 'not host, display, upload, modify, publish, transmit, store, update or share'. **Rule 3, sub-sections (b) and (d)**: list of information that falls within the purview of prohibited or unlawful content that an intermediary can act against. **Rule 3(1)(d)**: content takedown within tight deadlines (up to 36 hours) upon court order or governmental notice. **Rule 4(4)**: automated content filtering. **Rule 4(7)**: voluntary identification of users on social media intermediaries. **Rule 4(8)**: requires the intermediary to explain the action being taken and the grounds or reasons for such action, as well as the possibility to dispute such action.

Bhandari, V. 'Twitter case underlines web moderation issues', Hindustan Times (8 July 2022), available at: https://www.hindustantimes.com/opinion/twitter-case-underlines-web-moderation-issues-101657209298117.html.

Source: the authors, with excerpts from Belli L, Curzi Y and Gaspar WB, 'Online Content Regulation in the BRICS Countries. A Cybersecurity Approach to Responsible Social Media Platforms', *Responsible behavior in cyberspace. Global narratives and practice* (Bietlot 2023) <doi.org/10.2815/728569> accessed 28 June 2023.

Online interactive version available at: https://is.gd/exipop.

3.4 The Chinese Internet Information Service Algorithmic Recommendation Management Provisions

As of March 1, 2022, China's Internet Information Service Algorithmic Recommendation Management Provisions (Belli et al., 2023; Creemers et al., 2021) referred to as the “Provisions”, came into force. These new Chinese regulations, comprised of 35 articles, represent an inclusive endeavor to oversee the utilization of “algorithmic recommendation services” in various aspects of society. They encompass domains such as news, social media, e-commerce, fraud prevention, and platform operations, essentially governing almost all types of recommendation and decision-making algorithms. They also aim to regulate how technology companies employ algorithms to safeguard individual rights and the public interest. These algorithmic regulations are part of a broader regulatory push in the Chinese digital economy that began in October 2020, which included actions like the suspension of Ant Financial’s US IPO, restrictions on Didi Global[12], controls on the tech sector and video gaming, and the enactment of laws such as the Personal Information Protection Law (PIPL), in effect since November 2021 and the Data Security Law (DSL) in September 2021 (Rolf, 2023).

The Provisions encompass various articles aimed at safeguarding users' rights and interests. For instance, the Provisions offer Internet users the option to either opt into or opt out of algorithmic recommendations. This empowers users to make choices such as selecting or removing keywords used for targeting, as stated in Article 17. Simultaneously, the regulator emphasizes transparency in the provision of algorithmic recommendation services, ensuring users are informed about the services’ key operational mechanisms (Article 16). Additionally, Article 18 focuses on the protection of minors, obliging algorithmic recommendation service providers to create tailored models and services for minors while ensuring they do not promote unsafe conduct, actions contrary to social norms, or any behaviors detrimental to minors’ physical and mental well-being.

Moreover, the article prohibits the use of algorithmic recommendation services to encourage online addiction among minors. In a similar vein, Article 19 emphasizes the rights of the elderly, requiring providers to offer services aligned with elderly individuals’ specific needs and security measures to combat fraud. Article 20 pertains to workers, obliging providers to protect labor rights, including fair remuneration, rest, and work allocation, through the development of appropriate algorithms. Lastly, Article 21 is dedicated to consumers, ensuring their fair-trading rights by prohibiting the use of algorithms

12 Didi Global is a Chinese ride-hailing company that operates on a digital platform, much like its Western counterparts Uber and Lyft. The company was founded in 2012 and has since expanded its operations to include food delivery, financial services, and other mobility solutions. It went public in the United States in June 2021. However, Didi has faced significant regulatory scrutiny from the Chinese government, particularly in the context of data privacy and national security concerns. Shortly after its U.S. IPO, China's cybersecurity regulators launched an investigation into the company, resulting in the removal of Didi's app from various app stores in China. The actions against Didi Global were part of a larger regulatory clampdown on technology companies in China that began in October 2020, aimed at addressing issues such as antitrust behavior, data protection, and the societal impact of technology platforms.

for unfair practices like price discrimination based on consumer tendencies or habits. These articles collectively demonstrate the Provisions' commitment to safeguarding user rights and interests across various demographics and scenarios.

On top of the explicit requirements of user protection, however, the Provisions also stipulate a series of norms applicable to the service providers when it comes to the management and supervision of their algorithmic recommendation systems: algorithmic recommendation service providers must enhance their management of user profiles and tagging, ensuring that interests logged in user profiles are free from unlawful or harmful keywords (Article 10). Furthermore, service providers must allow for manual intervention and user choices, as well as prioritize mainstream values in displaying content on front pages, main screens, hot search terms, selected topics, topic lists, and pop-up windows. Lastly, in this context, Article 12 encourages providers to use various tactics like reducing the importance of certain content and implementing interventions to enhance transparency in search, ranking, selection, push notifications, and display norms. The aim is to prevent negative effects on users and reduce controversies or disputes. Nonetheless, the potential ambiguity surrounding terms such as “controversies or disputes”, as well as “mainstream values” may lead to differing interpretations, possibly resulting in restrictive measures that could hinder freedom of expression.

The Provisions also seek to combat the proliferation of disinformation, by requiring algorithmic recommendation service providers to obtain an Internet news information service permit and regulate their operations in collecting, editing, sharing, and broadcasting news information (Article 13). While the article appears to be a proactive step in addressing the issue of disinformation, the extent to which it will be applied democratically remains unclear. There may be concerns regarding potential restrictions on the granting of permits, affecting the flow of information. Striking a balance between curbing disinformation and safeguarding freedom of expression is a challenge that requires careful implementation and oversight. Questions linger about how the Provisions will impact the free exchange of ideas and news dissemination.

Article 24 stipulates that companies offering recommendation algorithms with characteristics related to public opinion or the ability to mobilize social engagement must submit essential information to the Internet information service algorithm filing system within 10 working days of commencing their services. In August 2022, the Cyberspace Administration of China (CAC) publicly released the first set of algorithm registrations, which lacked substantial information. However, a closer examination of the user manual unveiled more comprehensive disclosure requirements, such as listing the datasets used for training, conducting algorithm security assessments, and potentially undisclosed sections like “Algorithm Strategy” and “Algorithm Risk and Prevention Mechanism” (Sheehan & Du, 2022).

Sheehan's & Du’s assessment led to the conclusion that contrary to certain expectations, the Chinese government does not gain direct access to the algorithms or their underlying code

through this registry, which means that such transparency initiative in China can be likened to the European Union's Digital Services Act, which also underscores transparency - but with distinct specifics. According to the experts, China's approach diverges from the AI ethics community's model cards, which primarily emphasize performance evaluation and transparency over security. Furthermore, the registry would also reflect the Chinese Communist Party's role as the ultimate authority in determining matters of security and risk and establish a basis for the CAC to gradually intensify its demands for disclosure (Sheehan & Du, 2022). In this context, such an effort would follow a pattern reminiscent of prior efforts to oversee the Internet and online platforms.

According to Paul Triolo's comments on the Provisions, by formulating comprehensive regulations for a specific but crucial aspect of the commercial use of AI and big data-driven algorithms, the Cyberspace Administration of China and other Chinese authorities are conveying a commitment to scrutinizing all corners of the digital economy, unearthing any business practices that may be unpopular, exploit user data, or potentially result in adverse societal consequences. Concurrently, alongside the implementation of what appears to be one of the world's most robust data protection frameworks on paper, Chinese regulators are reasserting their proactive stance, which Triolo et al. (2021) see as a clear message that in, the digital economy, much like in the sphere of digital assets and cryptocurrencies, China is determined to avoid regulatory lag.

In this context, it is vital to recognize the broader scenario within which the regulations on algorithmic governance and data protection are implemented in China. This scenario revolves around a governance approach that emphasizes monitoring and intervention by governmental entities, impacting various facets of platform operations and prompting inquiries into their alignment with international principles related to freedom of expression. In this light, considerations regarding the broader scope of online freedoms maintains its significance as a subject of critical examination and analysis. However, it is imperative to approach these discussions with a truly global perspective, acknowledging the manifold cultural and regulatory frameworks that shape the interpretation and exercise of online freedoms.

4. Concluding remarks: a possible framework for a democratic platform regulation

The lexicon of global regulatory initiatives, as we discerned from our comparative analysis of the Brazilian Fake News Draft Bill, the EU's DSA, the Indian IT Act, and the Chinese algorithmic regulations, illuminates a common dilemma: the reconciliation of safeguarding human rights with the exigencies of effective regulation and robust governance, infused with tensions between public and private interests.

These initiatives, while shaped by their distinct socio-political contexts, offer valuable insights into the nuances of crafting digital governance in an era of increasing platform dominance. The

emphasis is not merely on enumerating these nuances but on deriving a strategic framework that can serve as a reference for future regulatory endeavors.

It is important to stress that the enactment of new regulatory frameworks geared explicitly towards digital platforms aims to reassert state sovereignty in the online environment, preserving the stability and security of the national political infrastructures. Unavoidable national specificities have a remarkable impact on how such frameworks are elaborated and implemented and, while most legislator would love to regulate platform in an effective and chirurgical fashion, the risk that such approaches translate into a regulatory sledgehammer is particularly concrete (Belli et al., 2023).

In light of these debates, UNESCO's Guidelines for Regulating Digital Platforms (2023), diverge from conventional governmental regulation, laying emphasis on a multistakeholder approach to platform governance. Importantly, its provisions do not merely refer to advisory multistakeholderism, but extend to rulemaking, monitoring, and review as well.

The table below elucidates key elements – some of them derived from the UNESCO Guidelines – envisaged to serve as foundational components for a robust and inclusive regulatory structure.

Table 2. Framework proposal

Objectives	Mechanisms	Recommendations for Domestic Legislation	Recommendations for International Legislation
Human Rights Equilibrium	Universal Declaration of Human Rights. United Nations Guiding Principles on Business and Human Rights. UNESCO Guidelines for Regulating Digital Platforms. International Human Rights Law.	Adopt UDHR in national digital constitutions. Mandate biennial HRESIA (Mantelero 2018) for major platforms.	Create international standards for HRESIA. Encourage regional alliances to uphold IHRL for platforms.
Advanced Transparency and Accountability	Meaningful and interoperable transparency on content moderation decisions (DCPR, 2022). Platforms' systemic risks to democracy. (Algorithmic Watch, 2023).	Legislate mandatory transparency reports and systemic risks assessments. Define standards for accountability structures.	Establish a global transparency framework. Institute international oversight bodies.
User-Centric Content Moderation	User-customizable content filters. Transparent appeal and grievance channels.	Mandate platforms to provide users with customizable content filters.	Advocate for global standards on user-centric content moderation.

Objectives	Mechanisms	Recommendations for Domestic Legislation	Recommendations for International Legislation
	User councils for content policy formulation.	Establish national guidelines for transparency in content moderation decisions. Encourage platforms to set up user councils to influence content policies.	Promote the adoption of user councils at an international scale.
Duty of Care and Liability	Code of Conduct on Countering Illegal Hate Speech Online (EU)	Legislate explicit 'duty of care' obligations for platforms operating within the country. Design liability regimes that balance platform responsibilities with user rights.	Develop international guidelines on 'duty of care' obligations. Create global consensus on balanced liability regimes considering platform size, reach, and potential impact.
Multi-Stakeholder Governance	Internet Governance Forum (IGF) DCs. Tunis Agenda.	Establish national multi-stakeholder advisory panels for tech regulations. Regularly update domestic regulations based on feedback from these panels.	Encourage the development of international multi-stakeholder advisory boards for platforms. Develop mechanisms for cross-border collaboration in tech regulations.

The proposed framework, anchored in the International Human Rights Law (IHRL), aims at giving a possible pathway for international and domestic legislations.

It is structured around five core objectives:

- Human Rights Equilibrium strives for a harmonious balance between platform autonomy and the protection of human rights, guided by international norms such as the Universal Declaration of Human Rights.
- Advanced Transparency and Accountability call for a mandatory framework for platforms to disclose their content moderation decisions and assess systemic risks to democracy.
- User-Centric Content Moderation focuses on empowering users through customizable content filters and transparent appeal mechanisms, ensuring a more equitable digital environment.
- Duty of Care and Liability underscores platforms' obligations to safeguard users from harm, advocating for explicit legislative mandates in this regard.
- Multi-Stakeholder Governance echoes UNESCO's emphasis on collective decision-making, recommending the formation of advisory panels comprising diverse stakeholders.

To operationalize these objectives, the framework provides concrete recommendations for both domestic and international lawmaking. For instance, it suggests that national laws should incorporate elements from the Universal Declaration of Human Rights, and also mandates biennial Human Rights and Environmental Social Impact Assessments (HRESIA) for major platforms (Mantelero, 2018). On an international scale, the framework advocates for the establishment of universally applicable standards for HRESIA, alongside fostering regional alliances to uphold human rights on platforms.

Additionally, it also calls for a domestic focus on mandatory transparency reports and systemic risk assessments, and internationally, for the institution of global oversight bodies. This multi-level approach ensures that platforms are held accountable within a coherent and standardized set of guidelines that also allows room for regional specificities.

By integrating these components into a cohesive whole, the framework aims to offer a comprehensive, balanced, and actionable guide for regulators, scholars, and practitioners in navigating the challenges and opportunities that digital platforms present in today's interconnected world.

References

Annappa, N. (2021). ENSURING RIGHT TO FREEDOM OF SPEECH AND EXPRESSION ON CYBER SPACE AS AGAINST STATE INTERVENTION - INDIAN EXPERIENCE. *Revista Direitos Fundamentais & Democracia*, *26*(1), 119–134. https://doi.org/10.25192/issn.1982-0496.rdfd.v26i12123

Belli, L. (2022). *Structural Power as a Critical Element of Social Media Platforms' Private Sovereignty*. 22.

Belli, L., Curzi, Y., & Britto Gaspar, W. (2023). *Online Content Regulation in the BRICS Countries: A Cybersecurity Approach to Responsible Social Media Platforms* (SSRN Scholarly Paper 4424913). https://doi.org/10.2139/ssrn.4424913

Belli, L., & Zingales, N. (2022). *Interoperability to foster open digital ecosystems in the BRICS.* Chinese Academy of Cyberspace Studies.

Belli, L., Zingales, N., & Curzi, Y. (2021). *Glossary of Platform Law and Policy Terms*. FGV Direito Rio.

Creemers, R., Webster, G., & Toner, H. (2021, August). Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022. *DigiChina*. https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/

De Gregorio, G., & Radu, R. (2022). Digital constitutionalism in the new era of Internet governance. *International Journal of Law and Information Technology*, *30*(1), 68–87. https://doi.org/10.1093/ijlit/eaac004

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1), 205395171989794. https://doi.org/10.1177/2053951719897945

Haggart, B., & Keller, C. I. (2021). Democratic legitimacy in global platform governance. *Telecommunications Policy*, *45*(6), 102152. https://doi.org/10.1016/j.telpol.2021.102152

Hartmann, I. A. (2017). Let the Users be the Filter? Crowdsourced Filtering to Avoid Online Intermediary Liability. *Legal Studies*, *1*, 27.

Lefevre, F. (2023, July 25). Desinformação nas eleições e o plano B do governo Lula. *Mobile Time*. https://www.mobiletime.com.br/colunistas/25/07/2023/desinformacao-nas-eleicoes-e-o-plano-b-do-governo-lula/

Llansó, E. J. (2020). No amount of "AI" in content moderation will solve filtering's prior-restraint problem. *Big Data & Society*, *7*(1), 205395172092068. https://doi.org/10.1177/2053951720920686

Mantelero, A. (2018). AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, *34*(4), 754–772. https://doi.org/10.1016/j.clsr.2018.05.017

Rolf, S. (2023). *China's regulations on algorithms—Context, impact, and comparisons with the EU*.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 2053951717738104. https://doi.org/10.1177/2053951717738104

Sheehan, M., & Du, S. (2022). *What China's Algorithm Registry Reveals about AI Governance*. https://policycommons.net/artifacts/3336740/what-chinas-algorithm-registry-reveals-about-ai-governance/4135571/

Strange, S. (1988). *States and Markets*. Pinter.

Toner, H., Creemers, R., & Triolo, P. (2021, August). Experts Examine China's Pioneering Draft Algorithm Regulations. *DigiChina*. https://digichina.stanford.edu/work/experts-examine-chinas-pioneering-draft-algorithm-regulations/

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

UNESCO. (2023). *Safeguarding freedom of expression and access to information: Guidelines for a multistakeholder approach in the context of regulating digital platforms*. https://unesdoc.unesco.org/ark:/48223/pf0000384031.locale=en

Vickery, J. R., & Everbach, T. (Eds.). (2018). *Mediating Misogyny*. Springer International Publishing. https://doi.org/10.1007/978-3-319-72917-6

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