NON-FINAL PREPRINT VERSION of Luca Belli, Yasmin Curzi, Walter Gaspar. Online Content Regulation in the BRICS Countries: A Cybersecurity Approach to Responsible Social Media Platforms. In François Delerue, Arun Sukumar and Dennis Broeders (Eds.) Responsible Behaviour in Cyberspace: Global Narratives and Practice. Brussels: Publication Office of the European Union. (2023).

Online Content Regulation in the BRICS Countries: A Cybersecurity Approach to Responsible Social Media Platforms

Luca Belli; Yasmin Curzi; Walter Gaspar

Contact email: luca.belli [at] fgv.br

Abstract

This article aims to evaluate recent events regarding social media governance in the BRICS (Brazil, Russia, India, China, and South Africa) countries, focusing on recent developments and highlighting common trends. With documental and literature review, we aim to analyse the most recent policies and institutional arrangements directly affecting what could be deemed a responsible social media platform. Furthermore, by looking at the selected developments in these countries, we aim to identify convergence and divergence between the BRICS country approaches, providing clarity on the existing and proposed regimes and contributing to the identification of viable paths forward that strike a balance between all interests involved.

Keywords

Platform regulation; cybersecurity; platform responsibilities; intermediary liability; BRICS; Brazil; Russia; India; China; South Africa.

Table of contents

1. Introduction 3
2. Recent developments in the BRICS countries 8

2.1. Brazil 8

2.1.1 Brazilian Civil Rights Framework for the Internet 9

2.1.2. Draft Bill on "Freedom, Responsibility and Transparency" of application providers 10

2.1.3. Final considerations regarding Brazil's social media regulation 11

2.2. Russia 12

2.2.1. From a liberal to a sovereignty-led approach 12

2.2.2. Regulating terrorist content, fake news, and insults to public officials 13

2.2.3. The consequences of the Ukraine war 15

2.3. India 16

2.4. China 21

2.5. South Africa 25

2.5.1. Social Media Legislation and the “Internet Censorship Bill.” 27

2.5.2. Other relevant legislations 27

a) Cybercrimes Act 27

b) South African Disaster Management Regulations 28

3. Conclusion: Choosing between a sledgehammer and a scalpel to regulate content 28

1. Introduction

The increasing relevance of digital platforms for everyday societal activities has been generating concerns regarding the concentration of political and economic power in a few private enterprises. The substantial risk of electoral interferences, manipulation, and widespread circulation of harmful content have led several countries to draft and enact regulations targeting primarily social media platforms[1] to regain control over such sensitive matters.

Online content regulation is a core cybersecurity issue as it is instrumental in preserving the security of political infrastructures.[2] Particularly, when dealing with the phenomenon of disinformation, there are significant overlaps and even similarities and synergies between the tools and mechanisms through which information disorder is organised and other cyber threats[3].

This article analyses the regulatory *state of the art* in the BRICS grouping, composed of Brazil, Russia, India, China, and South Africa. We consider that, although keeping a low profile as a group, the BRICS countries have acquired an increasing relevance at both regional and global levels, crafting impactful policies and enhancing their cooperation on digital matters. Importantly, their relevance is not only due to their economic weight but also to their mounting influence as policy setters.

Furthermore, it is interesting to highlight that some BRICS countries, notably China and Russia, started defining their content regulation frameworks in the early 2000s and aligned them

[1] Platforms can be seen as the technical and governance structures that facilitate relationships and exchange of value between different categories of users. Digital platforms provide a governance structure, via their private ordering, and a technical architecture, via a wide range of standards, protocols, and algorithms. See Belli, L. "Platform" in Belli, L, Zingales N. and Curzi Y. (Eds.). Glossary of platform law and policy terms. Rio de Janeiro: FGV Direito Rio, 2021. https://platformglossary.info/

[2] Usually, literature identifies four macro-areas of cybersecurity: data protection, safeguards of financial interests, protection of public and political infrastructures, and control of information and communication flows. See Fichtner, L. (2018). What kind of cyber security? Theorising cyber security and mapping approaches. Internet Policy Review, 7(2), 1-19. https://doi.org/DOI:10.14763/2018.2.788

[3] Caramancion, K. M., Li, Y., Dubois, E., & Jung, E. S. (2022). The Missing Case of Disinformation from the Cybersecurity Risk Continuum: A Comparative Assessment of Disinformation with Other Cyber Threats. *Data*, *7*(4). https://doi.org/10.3390/data7040049

internationally through the Shanghai Cooperation Organisation (SCO).[4] Indeed, since 2011, the SCO has elaborated upon an International Code of Conduct for Information Security[5] – updated in 2015[6] – recognising that information security includes content control within digital media and reaffirming that "policy authority for Internet-related public policy issues is the sovereign right of States."

Since 2011, the SCO, which India joined as a full member in 2016, has emphasised that the International Human Rights Law (IHRL) allows restrictions to freedom of expression under specific circumstances stated in article 19.3 of the International Covenant on Civil and Political Rights. However, limitations to freedom of expression must be necessary and proportionate to a legitimate aim. As we will discuss, BRICS countries achieve mixed results as regards meeting the tests of necessity and proportionality. SCO states tend to have more pervasive information controls, content restrictions, and sanctions – even resulting in criminal punishment. Brazil and South Africa are struggling, so far with limited results, to design frameworks to regulate content effectively.

After having provided a brief introduction to the BRICS grouping and the growing importance of digital policies in BRICS fora, stressing the relevance of cybersecurity in the bloc's agenda, we discuss the countries' most recent policy development at the national level. In this sense, this work's research question is to identify the common trends among the BRICS countries regarding cybersecurity and online platforms regulation.

1. The BRICS and their Cybersecurity landscape

The BRICS acronym, first coined by Goldman Sachs economist Jim O'Neill, refers to four large emerging economies that would have experienced a similar and acute phase of economic

[4] The SCO is an intergovernmental organisation aimed at political, economic, and security cooperation. It covers three-fifths of the Eurasian continent and was established in 1996, in Shanghai, by China, Russia, Kazakhstan, Kyrgyzstan, and Tajikistan. See http://eng.sectsco.org
[5] See https://digitallibrary.un.org/record/710973
[6] For a comparison of the differences between the 2011 and 2015 versions of the Code, see https://openeffect.ca/code-conduct/

development: Brazil, Russia, India, and China. South Africa would only join the grouping later.[7] After getting acquainted with club governance as key emerging leaders invited to the G7/8 summits via the so-called "outreach process",[8] the BRICS countries started to increase their synergies.

Since the creation of the grouping, the number and type of BRICS governmental and multistakeholder gatherings, partnerships, and initiatives have grown considerably.[9] In 2014, the bloc established the BRICS-led New Development Bank (NBD)[10] and Contingent Reserve Arrangement – one of its most prominent institutional achievements. Moreover, BRICS heads of state have never missed any of the group summits, thus witnessing its importance for them.

Regarding cybersecurity, the 2013 revelations of NSA contractor Edward Snowden represented a particularly salient event for the BRICS. Most prominently, these illegal activities included wiretapping illegally the Brazilian President's personal phone[11] and the communications of a vast number of members of the Brazilian government. They triggered the elaboration and implementation of an ample range of cybersecurity policies in the countries and enhanced their cooperation. [12]

[7] See Jim O'Neill. (2001). Building Better Global Economic BRICs. (66) Goldman Sachs Global Economic Papers <https://www.goldmansachs.com/insights/archive/archive-pdfs/build-better-brics.pdf>

[8] The most relevant of such processes was the "G8 Outreach Five", which included Brazil, China, India, Mexico, and South Africa to the 2005 G8 summit (Russia was still part of the G group itself). However, while the outreach model recognised the relevance of emerging economies – notably the future BRICS members– it also perpetrated a shared sense of exclusion, as the countries kept on being merely invited as guest with marginal role, compared to the G members

[9] For detailed overviews of the evolution of BRICS, see Stuenkel O. The BRICS and the Future of Global Order. Lexington Books. (2016); and o the same author, quoted supra at n. 4.

[10] See https://www.ndb.int.

[11] See Bridi S. Glenn Greenwald G. "Documentos revelam esquema de agência dos EUA para espionar Dilma" (*Fantástico*, 1 September 2013) http://g1.globo.com/fantastico/noticia/2013/09/documentos-revelam-esquema-de-agencia-dos-eua-para-espionar-dilma-rousseff.html

[12] For an analysis of BRICS digital policies and most recent developments particularly in the field of cybersecurity, see Belli L. (Ed.), *CyberBRICS: Cybersecurity Regulations in the BRICS Countries*. Springer (2021); Belli, L. Cybersecurity Policymaking in the BRICS Countries: From Addressing National Priorities to Seeking International Cooperation. The African Journal of Information and Communication, v. 28. (2021); Belli L. and Doneda D. Data Protection in the BRICS Countries: Legal Interoperability through Innovative Practices and Convergence. In International Data Privacy Law. (2022)

Tellingly, the eThekwini Declaration issued as an outcome of the 2013 Durban Summit of the BRICS included, for the first time, an explicit reference to cybersecurity, stressing the "paramount importance" of the "security in the use of Information and Communication Technologies (ICTs)."[13] Furthermore, in 2014, the BRICS technology and communication ministers started a cooperation process establishing the BRICS Working Group on the Security of ICTs[14], and adopting the BRICS Memorandum of Understanding on Cooperation in Science, Technology, and Innovation.[15] Such yearning for cooperation seems to have recently acquired a renewed impetus, with the 2021 BRICS Declaration calling for establishing "legal frameworks of cooperation among the BRICS States [and] a BRICS intergovernmental agreement on cooperation." [16]

While the Ukrainian war has indubitably put under strain all diplomatic initiatives involving Russia, it is safe to state that BRICS members' commitment to the grouping remains unchanged. The entire calendar of events was confirmed under the 2022 Chinese rotating presidency. BRICS members continue to consider the group a diplomatic priority, despite the divergence of opinions regarding the Ukrainian war. A meeting of the BRICS Ministers of Foreign Affairs in May 2022 was remarkably cooperative, culminating with the release of a Joint Statement on "Strengthen BRICS Solidarity and Cooperation, Respond to New Features and Challenges in International Situation."[17]

---

[13] See BRICS (Fifth BRICS Summit) 'eThekwini Declaration' (Durban 2013) para 34. http://mea.gov.in/bilateral-documents.htm?dtl/21482

[14] For an analysis of such documents and their impact see Vladimir Kiselev and Elena Nechaeva, 'Priorities and Possible Risks of the BRICS Countries' Cooperation in Science, Technology and Innovation' [2018] 5(4) BRICS Law Journal <https://doi.org/10.21684/2412-2343-2018-5-4-33-60> accessed 8 October 2021.

[15] See BRICS (Second BRICS Science, Technology and Innovation Ministerial Meeting) 'BRICS Memorandum of Understanding on Cooperation in Science, Technology and Innovation' (Brasília, 18 March 2015) <https://www.gov.br/mre/pt-br/canais_atendimento/imprensa/notas-a-imprensa/ii-reuniao-de-ministros-de-ciencia-tecnologia-e-inovacao-do-brics-documentos-aprovados-brasilia-18-de-marco-de-2015> accessed 8 October 2021

[16] BRICS (XIII BRICS Summit) 'New Delhi Declaration' (9 September 2021) <https://brics2021.gov.in/brics/public/uploads/docpdf/getdocu-51.pdf> accessed 8 October 2021

[17] See BRICS Joint Statement on "Strengthen BRICS Solidarity and Cooperation, Respond to New Features and Challenges in International Situation". PRESS RELEASE N. 76. Published on May 19, 2022. https://www.gov.br/mre/en/contact-us/press-area/press-releases/brics-joint-statement-on-201cstrengthen-brics-solidarity-and-cooperation-respond-to-new-features-and-challenges-in-international-situation201d

After the above-mentioned BRICS meeting, the Brazilian Ministry for Foreign Affairs "reiterated its support for intra-BRICS cooperation"[18] and highlighted that the grouping has "shown concrete results"[19], emphasising that BRICS is a forum focused on international cooperation and sustainable development and on building a more robust multipolar order and inclusive global governance for the benefit of developing countries.

Recent developments in the BRICS provide substantial evidence that these countries' roles and interactions are starting to acquire global relevance for digital policymaking, besides their national and regional impact. Notably, the 13th BRICS Summit, hosted by India in September 2021, gave particular prominence to cybersecurity.[20]

While the five countries' national approaches diverge in many aspects, it is possible to identify several points of overlap and even tendencies towards convergence. Remarkably, their approaches to cybersecurity have started to converge and intensify ever since the creation of the "Working Group of Experts of the BRICS States on security in the use of ICTs" in 2014, with a mandate to, *inter alia*, "develop practical cooperation with each other in order to address common security challenges in the use of ICTs."[21]

While agreeing on shared principles and high-level objectives through the annual declarations, the countries have crafted a unique blend of normative and developmental approaches to shape how (cybersecurity) cooperation and regulation should unfold.[22] However, such an approach is not immediately intelligible for an observer to consider only the normative side of regulation, *i.e.,* regulation by prohibiting undesired behaviours and oversight by a specific authority. Indeed,

[18] See the official Twitter account of the Brazilian Foreign Affairs Ministry https://mobile.twitter.com/Itamaraty_EN/status/1527398486454460417
[19] Idem.
[20] See BRICS (2021). BRICS India 2021 - XIII BRICS Summit - New Delhi Declaration. https://brics2021.gov.in/brics/public/uploads/docpdf/getdocu-51.pdf
[21] BRICS (2015). VII BRICS Summit - Ufa Declaration. https://www.brics2021.gov.in/BRICSDocuments/2015/Ufa-Declaration-2015.pdf
[22] See Belli, L. (2020). Data protection in the BRICS countries: Enhanced cooperation and convergence towards legal interoperability. New Media Journal. Chinese Academy for Cyberspace Studies. https://cyberbrics.info/data-protection-in-the-brics-countries-enhanced-cooperation-and-convergence-towards-legal-interoperability/; Belli, L. (2021b). CyberBRICS: A multidimensional approach to cybersecurity for the BRICS. In L. Belli (Ed.), CyberBRICS: Cybersecurity regulations in the BRICS countries. Springer

cooperation and regulation, be they on cybersecurity or any other matters, can be achieved, arguably more effectively, through other means than mere norm-making, such as investments and standardisation.

Lastly, it is essential to emphasise that, despite the ambitions and intentions expressed in the BRICS annual declarations and official documents, the ease with which intra-BRICS cooperation on cybersecurity issues can occur remains unclear. On the one hand, most content regulation issues are highly sensitive, and national policymakers' decisions regarding content restrictions represent the quintessence of domestic cultural, political, and legal peculiarities, thus making them less than ideal candidates for international consensus.[23] Nevertheless, the likeliest rapprochement is in the form of information and good practice (or bad practices, depending on the observers' standpoint) for which a dedicated intra-BRICS body already exists.

Hence, it is crucial to evaluate the domestic approach of the various BRICS members to cybersecurity to understand in which areas and to what extent coordination, convergence, or divergence are most likely to occur. In addition, content regulation and online platform responsibility have become prominent in national debates, mainly due to disinformation. The following sections provide an overview of the latest national developments to shed light on what BRICS approaches converge, or even reproduce each other's, and on what elements the countries are taking different paths.

2. Recent developments in the BRICS countries

2.1. Brazil

The Brazilian social media regulation relies on the Brazilian Civil Rights Framework for the Internet, Law n. 12,965/2014, a.k.a. "Marco Civil da Internet", or "MCI", which is in the process of being supplemented by Draft Bill n. 2,630/2020, a.k.a. the "Fake News Bill."

[23] See Belli, L. Cybersecurity Policymaking in the BRICS Countries: From Addressing National Priorities to Seeking International Cooperation. The African Journal of Information and Communication, v. 28. (2021).

2.1.1 Brazilian Civil Rights Framework for the Internet

The MCI is Brazil's primary law regarding internet regulation and the first and only general Law for internet governance adopted in Latin America. It establishes rules and principles for a democratic, plural, and neutral internet and defines general provisions for application providers. Article 19 establishes a general regime[24] of a judicial notice-and-takedown[25] system where application providers can only be liable for user-generated content (UGC) if failing to comply with court orders for the removal of specified content within 24 hours, granted they have the technical capacity to do so. The rationale was that, by imposing such legal procedure, abusive requests would not follow through, and only valid demands would come to the Judiciary[26], ensuring legal certainty for the companies.

The Brazilian Supreme Court will soon assess article 19's constitutionality in the Extraordinary Appeals ("RE") n. 1,037,396/SP and n. 1,057,258/MG[27] – both questioning intermediaries' role

[24] . The exceptions are articles 19.2 and 21, which respectively refer to copyright infringement and intimate imagery and provide a notice-and-takedown regime.

[25] Before MCI, the Brazilian Superior Court of Justice (STJ) was in the process of "unifying" its jurisprudence to establish the notice-and-takedown regime as the general regime in the country, influenced by the North American Digital Millennium Copyright Act (DMCA). In its session 512, DMCA enacts a "safe harbour" for service providers, which are exempt from liability if they have set notice-and-takedown procedures enabling users (copyrights holders) to request a quick removal of infringing content. STJ justice Nancy Andrighi even mentioned such legislation in a case against Google to condemn the search engine for not complying with a takedown request by an offended user. However, this majority opinion neglected the massive number of requests for content removal – not all valid and lawful.

[26] It is also relevant to mention that Brazil has a relatively functional public judicial system. "Access to justice" is, in fact, a constitutional right (article 5º, XXXV), and in article 19.3, MCI assures that users can refer their cases to Special Courts, where they can count on free legal assistance and an expedited judicial procedure.

[27] In the first case, a Facebook user had a fake account created in her name and issued a lawsuit for Facebook to delete it, requesting compensation. The regional appeals court not only sentenced Facebook to delete the fake profiles and to pay for damages but also declared the "incidental unconstitutionality" of article 19, considering that Facebook did not act expeditiously – before the lawsuit. The regional judges argue that article 19 is incompatible with the Brazilian Federal Constitution regarding consumer protection (art. 5°, XXXII) and general civil rights provisions, such as intimacy, privacy, honour and reputation (art. 5°, X). Facebook appealed to the Supreme Court, remarking that Article 19 determines that intermediary liability should only stem from failing to comply with a judicial request when proven that the company could do so. In the second case, students from a school in Minas Gerais created a forum on the social network Orkut (controlled by Google) to criticise a teacher. She demanded the page's removal to Orkut. This case followed Facebook's in much the same way, with Google losing and appealing to the Supreme Court – which joined both the appeals, due for judgment in June 2022. The whole lawsuit can be accessed at thte Brazilian Federal

in amplifying users' rights violations. Furthermore, following other countries' initiatives to curb disinformation and other harms, the Brazilian legislators started drafting bills towards this goal, establishing platforms' responsibilities and transparency. The main result is the – now-under-discussion in the Federal Congress – Draft Bill on Freedom, Responsibility and Transparency on the Internet, PL n. 2,630, presented in 2020, a.k.a., "PL das *Fake News*".

2.1.2. Draft Bill on "Freedom, Responsibility and Transparency" of application providers

In 2020, Senator Alessandro Coronel presented to the Brazilian Federal Senate the Draft Bill n. 2,630/2020, submitting it to the National Chamber on July 3rd for appreciation. Experts and civil society organisations criticised the Draft's first version due to problematic provisions such as traceability of communications for tackling disinformation, criminalisation of disinformation spread, and the absence of more sophisticated transparency and users' rights provisions or a proper governance model.

The Draft Bill is currently under debate at the National Chamber, having as its rapporteur Deputy Orlando Silva. The Chamber held multiple public hearings in 2021, counting on the participation of civil society organisations and experts to improve the Draft, culminating in entirely new versions presented by its rapporteur on November 4th, 2021, and on March 31st, 2022.

Some improvements of the current version merit highlight: the provision on the criminalisation of disinformation dissemination now targets only coordinated actions by enterprises/companies, not individuals. In addition, it altered the traceability provision in compliance with due process in criminal law – it must be based on (1) previous intelligence work, (2) presumption of innocence and user privacy, and (3) security of communications.

Nevertheless, the Draft Bill left broad room for platforms' self-regulation. According to the current version, it is up to them to create their own codes of conduct to assure transparency and

Supreme Court website here: https://portal.stf.jus.br/jurisprudenciaRepercussao/verAndamentoProcesso.asp?incidente=5160549&numeroProcesso=1037396&classeProcesso=RE&numeroTema=987.

accountability – which the CGI.br (the Brazilian Internet Steering Committee), a multisectoral entity, must certify. However, the Steering Committee does not have enforcement tools or power under the law to enforce regulations. Therefore, there is a high risk that the codes of conduct will deviate entirely from what the law intended.

Regarding transparency reports duties, the Draft only requires numerical information on the total amount of moderation measures, which fail to provide meaningful transparency and do not allow the identification of biases and failures in moderation or recommending systems, according to several experts[28]. In addition, the Draft does not present a methodology or model for presenting reports, making it challenging to monitor failures and biases.

2.1.3. Final considerations regarding Brazil's social media regulation

Arbitrary removal, shadowbans[29], and lack of transparency are oft-pointed-out issues impacting free speech and democracy. With the growth of platforms' powers, governments must move toward platform observability[30] to assure non-discrimination and democratic legitimacy of their actions before civil society.

In this sense, within the constitutionality of MCI's article 19 debate, the Supreme Court[31] can propose the differentiation of duties between very large platforms and other actors, fostering fundamental rights and innovation. In addition, the Legislator could enact duties, such as the Digital Services Act[32], for those with power and technical capacities to implement efficient

[28] Suzor, Nicolas P. 2019. *Lawless*. Cambridge University Press. https://doi.org/10.1017/9781108666428.
[29] "Shadowban" refers to a relatively common moderation practice of lowering a user's visibility, content or ability to interact without them knowing it so that they can continue to use the platform normally. See See Radsch "Shadowban/Shadow Banning" in Belli, Zingales and Curzi, n(1).
[30] Rieder, Bernhard, and Jeanette Hofmann. 2020. "Towards Platform Observability." *Internet Policy Review* 9 (4): 1–28. https://doi.org/10.14763/2020.4.1535.
[31] The lawsuit at the Brazilian Federal Supreme Court can be accessed here: https://portal.stf.jus.br/jurisprudenciaRepercussao/verAndamentoProcesso.asp?incidente=5160549&numeroProcesso=1037396&classeProcesso=RE&numeroTema=987.
[32] Cf.: European Commission. Questions and Answers: Digital Services Act. Available at: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348.

monitoring of inappropriate content and risk assessment obligations, especially considering that content moderation technologies have improved with AI advances[33].

It could also define a new civil liability scheme, which should ensure both an innovative ecosystem and legal certainty for enterprises, as well as duties and increased responsibilities for very large platforms, in harmony with new regulations that attempt to tackle issues derived from the unprecedented economic and political power by such actors. Despite this, as it is possible to conclude from the analysis of the "Fake News" Draft Bill's most recent versions, Brazil did not make much progress in creating a governance model that would affect platforms' activities. As a result, the current version of the Draft Bill is not moderate but a conservative piece of legislation that enables platforms to regulate themselves at will.

2.2. Russia

Over the past years, Russia has adopted multiple restrictive normative provisions crafting a vision of "Russian Internet Sovereignty"[34], consisting of provisions on personal data localisation, content regulation and a new type of "infrastructure-embedded control,"[35] inspiring governments and legislators globally.[36]

2.2.1. From a liberal to a sovereignty-led approach

The main goal of recent digital policies adopted at the Russian level has been the establishment of an autonomous Russian segment of the Internet, dubbed the "Runet", allowing increased

[33] Cf. Gorwa, Robert, Reuben Binns, and Christian Katzenbach. "Algorithmic content moderation: Technical and political challenges in the automation of platform governance." *Big Data & Society* 7.1 (2020): 2053951719897945.

[34] See Shcherbovich, A. Data protection and cybersecurity legislation of the Russian Federation in the context of the "sovereignisation" of the internet in Russia. In Belli, L. (Ed.), CyberBRICS: Cybersecurity regulations in the BRICS countries. Springer. (2021) p. 67-131. https://link.springer.com/chapter/10.1007/978-3-030-56405-6_3. Daucé, F. and Musiani F. (Eds.) Infrastructure-embedded control, circumvention and sovereignty in the Russian Internet. Vol. 26. N. 5 (May 2021). https://firstmonday.org/ojs/index.php/fm/issue/view/693

[35] See Daucé and Musiani (2021) *cit. supra*.

[36] See e.g. Cory N., and Dascoli L. How Barriers to Cross-Border Data Flows Are Spreading Globally, What They Cost, and How to Address Them. Information Technology & Innovation Foundation. (2021)

control on national digital infrastructures, largely reproducing the strategies deployed by China since the early 2000s, with the so-called "Great Firewall of China"[37].

Unlike China, however, the Internet in Russia remained relatively free from regulation for more than a decade, with the introduction of light regulation in the mid-2000s. Only in recent times has Russia tightened its control on online media. While a certain degree of censorship has always existed, until the early 2010s, Russia maintained a somewhat liberal[38] approach.

In 2006, Russia adopted Federal Law n. 149-FZ "On Information, Information Technologies and Protection of Information", based mainly on the EU approach to intermediary liability, exempting intermediaries from civil liability related to UGC. Since the early 2010s, however, the initial liberal approach was substituted by an increasingly heavy-handed approach.

2.2.2. Regulating terrorist content, fake news, and insults to public officials

Since March 2019, Russia has moved towards a new content regulation regime aimed at regulating disinformation and restricting opinions on public authorities. Federal Law n. 31-FZ introduced the first set of provisions on Amending Article 15.3 of the Federal Law "On Information, Information Technologies, and Protection of Information", March 18th, 2019, a.k.a. "Fake News Law." It prohibits publishing "socially important information" and defines disinformation[39].

[37] The Chinese approach led to the creation of an Internet with Chinese characteristics that observers compare to an extensive national intranet connected to the rest of the global Internet through limited channels.

[38] Especially considering the media regulation during the Soviet era.

[39] Disinformation is defined under the "Fake News Law" as "information of public interest, which is known to be unreliable, is disguised as accurate information and poses risks of harm to the life and/or health of citizens or property, mass disruption of public order and/or public safety, or impeding or halting the functioning of critical, transport or social infrastructures, lending institutions, or power generation, industrial or communications facilities".

As pointed out by Shcherbovich[40], the Explanatory Note to the Draft Bill states that the optimal way to implement it is vesting the Prosecutor General of the Russian Federation or his deputies with the power to request Roskomnadzor[41] – the Russian Media, Telecommunications and Information Regulator – to restrict access to information resources that disseminate disinformation. Hence, to implement the provisions, the Prosecutor General or his deputies request that Roskomnadzor order providers to remove information within a specific deadline. Failing to comply with this request allows the authority to add the corresponding IP address to one of the state registers, obliging providers to block the IP address and prevent users from accessing the content.

Hovyadinov highlights the hybrid nature of Russian social media governance, as internet businesses with close ties to the government play a key role in conducting "censorship and surveillance activities."[42] These partnerships, enabled through state bodies' purchase of tech companies' shares, allow the federal government to count on the cooperation of intermediaries to control information flows and user activities. For example, a leading state bank, Sberbank, is a majority shareholder controlling Russian search engine and e-commerce giant Yandex, while email portal Mail.ru and the social media platform VKontakte are controlled by entrepreneurs closely affiliated with the Kremlin.[43]

[40] See Shcherbovich, A.A. (5 April 2019) Exploring the New Russian Measures against "Fake News" and Online Insults. CyberBRICS.info. https://cyberbrics.info/exploring-the-new-russian-measures-against-fake-news-and-online-insults/

[41] Roskomnadzor is the Russian Federal Service for Supervision of Communications, Information Technology and Mass Media. This executive agency is responsible for controlling and regulating all Russian mass media, including online media and Internet networks, supervising the compliance with the data protection legislation, the implementation of content regulation and telecoms law, and the operation of the Russian Autonomous Internet Subnetwork better known as "RuNet" in compliance with the Russian Sovereign Internet Law.

[42] See Hovyadinov, S. Intermediary Liability in Russia and the Role of Private Business in the Enforcement of State Controls over the Internet. In Frosio, G. (2020). Oxford Handbook of Online Intermediary Liability 10.1093/oxfordhb/9780198837138.013.33.

[43] *Idem.*

Since 2019, Russia has limited the right to express 'disrespectful' opinions on public officials, society and symbols of the Russian Federation[44], passing Federal Law n. 30-FZ[45]. Under it, certain types of online content can be deemed illegal and taken down or blocked. Some cases do not even require a court order, thus allowing the government to directly instruct Roskomnadzor to request ISPs to block access to the webpage or websites. After the Roskomnadzor notification to the ISPs, they must inform the content removal. Finally, Roskomnadzor verifies if the illegal content is inaccessible and informs the access providers to restore the access resource.

In June 2020, the European Court of Human Rights (ECtHR) – to which Russia was subject, as a Member of the Council of Europe, until September 2022 – delivered a series of judgements assessing the implementation of Russia's Law on Information, Information Technologies, and Protection of Information. The ECtHR held that blocking entire websites was an extreme measure, which can be only justified in exceptional circumstances, as it is equivalent to banning a newspaper or a television station, having collateral effects on lawful content.[46] After the Court rulings, the Duma introduced new amendments[47] to regulate platforms. They entered into force in February 2021, requiring social media platforms to monitor content and "immediate restrict access" to users that post information about state secrets, justification of terrorism or calls to terrorism, pornography, violence and cruelty, obscene language, drugs manufacturing, information on methods to commit suicide, as well as calls for mass riots.

2.2.3. The consequences of the Ukraine war

Recent developments related to the Ukrainian war have had repercussions regarding online content regulation. First, the State Duma has adopted amendments to the Criminal Code of the

[44] The Amendment prohibits "the spreading of information which shows blatant disrespect for society, the government, official state symbols of the Russian Federation, the Constitution of the Russian Federation or authorities exercising governmental authority in the Russian Federation."
[45] "On Amendments to the Federal Law On Information, Information Technologies and Protection of Information".
[46] See Gurshabad Grover and Anna Liz Thomas. (22 February 2021). Notes From a Foreign Field: The European Court of Human Rights on Russia's Website Blocking. CyberBRICS.info. https://cyberbrics.info/notes-from-a-foreign-field-the-european-court-of-human-rights-on-russias-website-blocking/
[47] Law 149-FZ "On Information, IT and Protection of Information".

Russian Federation, increasing responsibility for spreading "fake news" about Russian Armed Forces actions or callings for sanctions against Russia on social media.[48] They establish punishments with fines of 700 thousand to 1.5 million rubles or imprisonment for up to three years. Moreover, if the illegal behaviour derived from "abusing one's official position, based on political, ideological, racial, national or religious hatred or enmity, or based on hatred or enmity against any social group", then the term of imprisonment can be up to 10 years.

In addition, administrative sanctions and criminal liability might apply in case of "public actions aimed at discrediting the exercise by state bodies of the Russian Federation of their powers outside the territory of the Russian Federation."[49] Special additional sanctions apply in cases of threat to "public order"[50], where the Code of Administrative Offenses foresees administrative fines of 50 to 100 thousand rubles for individuals, from 200 to 300 thousand rubles for officials, and 500 thousand to 1 million rubles for legal entities.

Since the early 2010s, the liberal approach has been substituted by an increasingly heavy-handed approach. Russia has amended its national framework on content regulation, introducing "normative packages" to combat terrorism and preserve national sovereignty and, more recently, to regulate "fake news", online insults to public authorities, and war-related disinformation. The most recent amendments have confirmed a trend towards a stringent regime."[51]

2.3. India

A long line of rules and judicial decisions affect platform regulation in India, starting with the Information Technology Act of 2000 (IT Act[52]) and its subsequent Rules. In 2021, a new set of rules concerning media intermediaries was enacted, the IT (Intermediary Guidelines and Digital

[48] See http://duma.gov.ru/news/53620/

[49] The fine will range from 100-300 thousand rubles or imprisonment for up to three years. If these conducts generate concrete consequences beyond the circulation of the disinformation, then the maximum term of imprisonment is up to five years. See http://duma.gov.ru/news/53773/

[50] "Calls for holding unauthorised public events, as well as pose a threat of harm to the life and (or) health of citizens, property, a threat of mass disruption of public order and (or) public safety, or a threat to interfere with the functioning or termination of the functioning of life support facilities, transport or social infrastructure, credit institutions, energy, industry or communications facilities."

[51] See http://duma.gov.ru/news/53773/

[52] Amended in 2008.

Media Ethics Code) Rules. This scenario may soon change with the "Digital India Act", currently being drafted by the Minister of State for IT and expected to be publicly debated in 2023[53].

In terms of security and data protection concerns, the IT Act originally contained civil sanctions for "cyber contraventions" (Section 43(a)-(h)) and criminal sanctions for "cyber offences" (Sections 63-74). The Act was amended in 2008 to include Sections 43A ("Compensation for failure to protect data"[54]), 66A ("Punishment for sending offensive messages through communication service"[55]), and 72A ("Punishment for disclosure of information in breach of lawful contract"[56]).

Article 79 of the IT Act provides immunity to network service providers (meaning intermediaries) for UGC. This immunity is conditioned on their due diligence (according to applicable rules) and solely participation as an intermediary. It is lost if the intermediary "fails to expeditiously remove or disable access to that material" after having actual knowledge[57] or receiving a notification from a government agency. This provision was criticised at the time for casting too wide a net, potentially bringing liability to intermediaries conducting simple content moderation operations[58].

---

[53] PTI. (2022, November 6). *Significant work done, draft Digital India Act framework by early 2023: MoS IT*. The Hindu. https://www.thehindu.com/business/Economy/significant-work-done-draft-digital-india-act-framework-by-early-2023-mos-it/article66103357.ece

[54] India. (2009, February 5). *IT Amendment Act 2008*.

[55] India. (2009, February 5). *IT Amendment Act 2008*.

[56] India. (2009, February 5). *IT Amendment Act 2008*.

[57] The Indian Supreme Court clarified the meaning of "actual knowledge" in Shreya Singhal v. Union of India, addressing "the issue of intermediaries complying with takedown requests from non-government entities and has made government notifications and court orders to be consistent with reasonable restrictions in Article 19(2)". Panday, J. (2015, April 11). The Supreme Court Judgment in Shreya Singhal and What It Does for Intermediary Liability in India? The Centre for Internet and Society. https://cis-india.org/internet-governance/blog/sc-judgment-in-shreya-singhal-what-it-means-for-intermediary-liability.

[58] This immunity can be guaranteed according to their due diligence (following applicable rules) and their participation solely as an intermediary (i.e., without "select[ing] or modify[ing]" the information). It is lost if the intermediary "fails to expeditiously remove or disable access to that material" after having actual knowledge or receiving a notification from a government agency. This provision was criticised at the time for casting too wide a net, potentially bringing liability to intermediaries conducting simple content moderation operations. Prakash, P. (2009, February). *Short note on IT Amendment Act, 2008*. The Centre for Internet and Society. https://cis-india.org/internet-governance/publications/it-act/short-note-on-amendment-act-2008.

The relevant provision on content-blocking is Section 69A[59], which led to a judicial controversy between Twitter and the Ministry of Electronics & Information Technology (MeitY), where the company questioned the government's block notices of thousands of accounts[60]. It considers these orders procedurally and substantially flawed for not providing prior judicial review and hearings to content creators, besides failing to demonstrate the public interest necessity on a case-by-case basis[61][62]. Moreover, as commented by Bhandari (2022)[63], the interplay between section 69A and the IT (Procedure and Safeguards for Blocking for Access of Information by Public) Rules of 2009, 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".

From the free speech perspective, one important highlight is the decision in *Shreya Singhal v. Union* of India[64], whereby the Court declared Section 66A unconstitutional under article 19(1)(a) of the Indian Constitution[65]. The Court found that the Section's vagueness of terms such as "annoyance" and "inconvenience" could create a chilling effect over a "large amount of protected and innocent speech" (para. 83). More recently, the 2021 intermediary Rules[66] have raised attention in platform regulations. The Rules create due diligence duties for social media intermediaries and "significant"[67] social media intermediaries, thus specifying the conditions of liability immunity for these actors.

---

[59] India. (2000). *Section 69A in The Information Technology Act, 2000*. Indiankanoon.Org. https://indiankanoon.org/doc/10190353/.

[60] ETech. (2022, July 26). *Twitter-ministry hearing in Karnataka HC adjourned till August 25* . The Economic Times. https://economictimes.indiatimes.com/tech/technology/twitter-ministry-hearing-in-karnataka-hc-adjourned-till-august-25/articleshow/93129940.cms.

[61] Bhandari, V. (2022, July 8). Twitter case underlines web moderation issues. *The Hindustan Times*, 12.

[62] Ghosh, S. (2022, July 12). *Twitter's petition on Section 69A of the IT Act*. The Hindu. https://www.thehindu.com/sci-tech/technology/twitters-petition-on-section-69a-of-the-it-act/article65623202.ece.

[63] Bhandari, V. (2022), n. 50.

[64] *Shreya Singhal v. Union of India*. (2015, March). Columbia Global Freedom of Expression. https://globalfreedomofexpression.columbia.edu/cases/shreya-singhal-v-union-of-india/

[65] Subramaniam, A., & Das, S. (2020, October 22). *In a nutshell: data protection, privacy and cybersecurity in India*. Lexology. https://www.lexology.com/library/detail.aspx?g=04c38a97-f6cb-4d23-ae95-00df33df8a68.

[66] India. (2021). *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules*. Central Government. https://egazette.nic.in/WriteReadData/2021/225464.pdf.

[67] Distinguished by the number of users in India, according to a threshold determined by the Central Government (currently, it is set at 5 million or more registered users in India (Notification S. O. 942(E), Pub. L. No. S. O. 942(E), Gazette of India (2021). https://egazette.nic.in/WriteReadData/2021/225497.pdf).

Among the due diligence obligations in the 2021 Rules, it requires intermediaries to publish monthly grievance reports and to appoint a Chief Compliance Officer, a Grievance Officer and a Nodal Contact Person, all residing in India.[68] Furthermore, the 2021 Rules demanded intermediaries to implement "content takedown within tight deadlines [Rule 3(1)(d)], automated content filtering [Rule 4(4)] and voluntary identification of users on social media intermediaries [Rule 4(7)[69]. They also enacted a traceability obligation[70], which was criticised for its potential to break end-to-end encryption in messaging applications.

Although the Indian government has proposed two models that allegedly allow this traceability obligation without disclosing the content of messages, these might require breaking end-to-end encryption, nonetheless abandoning forward secrecy or simply being based on faulty assessments of how encrypted messaging applications work.[71] In addition, the rule has been criticised[72] for raising other operationalising costs – particularly data storage to trace every message on a messaging thread – thus, increasing barriers for smaller competitors. Another provision pointed out as problematic for similar reasons is Rule 4(4), which requires client-side scanning for

[68] Twitter, for example, had problems when the Rules were enacted. ET Bureau. (2021, August 10). *Twitter now in compliance with IT rules, govt tells court*. The Economic Times. http://www.ecoti.in/KAdCwb47; Peermohamed, A. (2021, July 28). *Delhi High Court gives Twitter "last opportunity" to show compliance with IT rules*. The Economic Times. http://www.ecoti.in/c0hCoZ; Peermohamed, A. (2021, July 6). *Twitter lost immunity under IT Act: Centre to HC*. The Economic Times. http://www.ecoti.in/3OGSqY.

[69] Biyani, N., & Choudhury, A. (2021, November 8). *Internet Impact Brief: 2021 Indian Intermediary Guidelines and the Internet Experience in India*. Internet Society. https://www.internetsociety.org/resources/2021/internet-impact-brief-2021-indian-intermediary-guidelines-and-the-internet-experience-in-india/.

[70] To identify the "first originator" (in India) of certain information shared through an intermediary's messaging application (such as WhatsApp or Signal) [Rule 4(2)]

[71] Biyani, N., & Choudhury, A. (2021), n. 53. Maheshwari, N., & Nojeim, G. (2021, June 4). *Part 2: New Intermediary Rules in India Imperil Free Expression, Privacy and Security - Center for Democracy and Technology*. Center for Democracy & Technology. https://cdt.org/insights/part-2-new-intermediary-rules-in-india-imperil-free-expression-privacy-and-security/. Pfefferkorn, R. (2021). *New intermediary rules jeopardize the security of Indian internet users*. Brooking's TechStream. https://www.brookings.edu/techstream/new-intermediary-rules-jeopardize-the-security-of-indian-internet-users/.

[72] Biyani, N., & Choudhury, A. (2021), n. 53.

matches against certain types of material (e.g., rape or child sexual abuse material) – an intrusive manner of content control and is not necessarily practical[73].

Finally, rule 3(1)(d) of the 2021 Rules requires content removal upon court order or governmental notice[74] in up to 36 hours. This provision aims to expedite content removal related to various subjects listed in the rule. However, in doing so, it creates a wide net of hypotheses for content removal based on open-ended juridical terms such as "public order" and "incitement to an offence" and subjective terms such as "decency" and "morality".

In summary, the 2021 Rules have been criticised for conflicting with the IT Act from whence it comes[75] and, through vague wording, creating space for arbitrariness. They also came under scrutiny for establishing obligations to implement technical procedures which have been widely regarded as incompatible with end-to-end encryption and data privacy, potentially creating a harmful chilling effect over legitimate forms of speech and exposing minority and sensitive political groups to risks online. Criticism over the traceability rule went beyond simple discourse: WhatsApp and the Foundation for Independent Journalism have filed suits questioning the IT Rules 2021's constitutionality and legality, respectively[76]. All this comes on top of an already contentious system of content-blocking and liability for content posted, with open concepts that

[73] This case brings to mind Apple's plan, revealed in 2021, to implement a similar mechanism on their iOS devices, which was promptly dropped after they were heavily criticised for infringing upon user's privacy and putting sensitive information at risk. See McKinney, I., & Portnoy, E. (2021, August 5). *Apple's Plan to "Think Different" About Encryption Opens a Backdoor to Your Private Life*. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2021/08/apples-plan-think-different-about-encryption-opens-backdoor-your-private-life.

[74] "[U]pon receiving actual knowledge in the form of an order by a court of competent jurisdiction or on being notified by the Appropriate Government or its agency".

[75] Behera, N. (2020). *Legal protection of right to privacy in cyberspace*. Biyani, N., & Choudhury, A. (2021), n. 53.

[76] Agarwal, S. (2021, May 29). *WhatsApp sues Government of India over new IT rules*. The Economic Times. https://economictimes.indiatimes.com/tech/technology/whatsapp-sues-india-govt-says-new-it-rules-mean-end-to-privacy/articleshow/82963637.cms. Menn, J. (2021, May 26). *WhatsApp sues Indian government over new privacy rules*. Reuters. https://www.reuters.com/world/india/exclusive-whatsapp-sues-india-govt-says-new-media-rules-mean-end-privacy-sources-2021-05-26/. The Wire Staff. (2021, March 9). *Why The Wire Wants the New IT Rules Struck Down*. The Wire. https://thewire.in/media/why-the-wire-wants-the-new-it-rules-struck-down.

give rise to curtailments on speech based on deficient procedural check-and-balances – all of which have been or are currently under litigation.

2.4. China

Over the past two years, China has considerably updated its cyberspace regulations. For example, it adopted the Provisions on the Governance of the Online Information Content Ecosystem in 2020, the Data Security Law (DSL, effective in September 2021), and the new Personal Information Protection Law (PIPL, effective in November 2021). In terms of cybersecurity, these build upon the foundations already established by the 2017 Cybersecurity Law (CSL). Taken together, they create a comprehensive cybersecurity framework[77] [78].

In January 2022, a regulation on algorithmic recommendation systems, published for comments in 2021[79], was adopted[80]. Press announced the algorithmic recommendation regulation as "pioneering" and "groundbreaking",[81] and it seems so: the closest existing norm at the time of its

[77] Zhang, D. (2022, April). *China: The interplay between the PIPL, DSL, and CSL*. DataGuidance. https://www.dataguidance.com/opinion/china-interplay-between-pipl-dsl-and-csl.
[78] Belli, L. (2021). Cybersecurity Policymaking in the BRICS Countries: From Addressing National Priorities to Seeking International Cooperation. *The African Journal of Information and Communication (AJIC)*, *28*(28), 1–14. https://doi.org/10.23962/10539/32208.
[79] China. (2021, August 27). *Notice of the state Internet Information Office on the provisions on the administration of internet algorithmic recommendation (draft for solicitation of comments)*. Cyberspace Administration of China. http://www.cac.gov.cn/2021-08/27/c_1631652502874117.htm.
[80] China. (2022, March 1). *Translation: Internet Information Service Algorithmic Recommendation Management Provisions* (R. Creemers, G. Webster, & H. Toner, Eds.). Digichina. https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/. Chambers, H., & Sun, J. (2022, March). *China: The Internet Information Service Algorithm Recommendation Management Regulations*. DataGuidance. https://www.dataguidance.com/opinion/china-internet-information-service-algorithm.
[81] Lu, S. (2022, March 1). *Chinese tech companies now have to tell users about their algorithms*. Protocol. https://www.protocol.com/bulletins/china-algorithm-rules-effective. Luo, Y., Liu, V., & Danescu, I. (2022, February 8). *China Takes the Lead on Regulating Novel Technologies: New Regulations on Algorithmic Recommendations and Deep Synthesis Technologies*. Covington Inside Privacy. https://www.insideprivacy.com/artificial-intelligence/china-takes-the-lead-on-regulating-novel-technologies-new-regulations-on-algorithmic-recommendations-and-deep-synthesis-technologies/. Toner, H., Creemers, R., & Triolo, P. (2021, August 27). *Experts Examine China's Pioneering Draft Algorithm Regulations*. Digichina. https://digichina.stanford.edu/work/experts-examine-chinas-pioneering-draft-algorithm-regulations/.

discussion would be the United Kingdom's algorithmic transparency standard[82]. This rule provides a useful example of the Chinese strategy toward platform regulation – containing strong bureaucratic, content, and technical controls, in a fashion similar to other specific regulation and to the more general "Internet Information Service Management Rules" and "Provisions on the Governance of the Online Information Content Ecosystem"[83]. Due to its novelty and specificity, as well as the growing importance of algorithmic recommendation systems underlying the operations of digital platforms of various kinds, it merits a detailed description.

The regulation defines algorithmic recommendation systems as "the use of generative or synthetic–type, personalised recommendation–type, ranking and selection–type, search filter–type, dispatching and decision-making–type, and other such algorithmic technologies to provide information to users" (art. 2). As such, it covers a wide array of standard practices in digital platforms' activities – content recommendation, ranking, selection, search filters and others.

Some highlights, divided by the authors into broader thematic categories below, include:

1. Platforms' duties:
    1. Duty to mark algorithmically generated or synthetic information before dissemination (art. 9);
    2. Duty to remove unlawful or harmful information, preserve records and alert cybersecurity and other competent authorities (art. 9);
    3. Control of algorithmic processes to the level of the tagging of user profiles/models, which shall avoid unlawful or harmful keywords (art. 10);
    4. Duty to establish systems of manual intervention by users in algorithmic recommendation processes directed at them and to promote "autonomous user choice" (art. 11);

[82] CDDO. (2021, November 29). *Algorithmic Transparency Standard*. GOV.UK. https://www.gov.uk/government/collections/algorithmic-transparency-standard.
[83] China. (2000, September 25). *Internet Information Service Management Rules*. Available at: https://chinacopyrightandmedia.wordpress.com/2000/09/25/internet-information-service-management-rules/; and China (2019, December 15). *Provisions on the Governance of the Online Information Content Ecosystem*. Cyberspace Administration of China. Available at: https://wilmap.stanford.edu/entries/provisions-governance-online-information-content-ecosystem.

5. Duty of transparency and understandability concerning algorithmic recommendation processes (art. 12);
6. Duty to provide users with complaint and reporting mechanisms (art. 22);
7. A general prohibition of various behaviours enabled by algorithms, such as account and likes/comments/shares manipulation (seemingly aimed at bot activity) and manipulative administration of listings and topics to influence public opinion (art 14);
8. A general prohibition on anti-competitive behaviours enabled by algorithms (art. 15);

2. User rights:
    1. Notification and information about algorithmic recommendation systems in use (art. 16, with special protection of minors and the elderly in art. 18 and 19, respectively);
    2. Granular control over algorithmic recommendation services, including the capacity to choose and delete user tags (art. 17);
    3. Special protection to workers in labour relations intermediated by algorithmic services, upholding interests "such as obtaining labour remuneration, rest and vacation, and others." (art. 20);
    4. Special protection to consumers in consumer relations, hinting at predatory marketing practices ("they may not use algorithms to commit acts of extending unreasonably differentiated treatment in trading conditions such as trading prices, and others.", art. 21);
    5. Duty to provide users with complaint and reporting mechanisms (art. 22);
3. Content control:
    1. A general duty to prevent harmful content (various articles), including through active technical measures such as "content de-weighting, scattering interventions, and others" (art. 12);
    2. Requirement of a permit for news information services, accompanied by a fake news prohibition: "They may not generate or synthesise fake news information, and may not disseminate news information not published by work units in the State-determined scope" (art. 13);

4. Security measures:
    1. Duty to establish security plans, incident response processes, and regular revisions of algorithms (art. 8);
    2. Graded and categorised algorithm security management system (art. 23);
    3. Exceptional cybersecurity and reporting/registering duties for "algorithmic recommendation services with public opinion properties or social mobilisation capabilities" (art. 24 to 27);
    4. Cybersecurity assessments by authorities and a duty to preserve network records "according to the law" (art. 28);

These are all provisions that stick out either for their qualities or their defects. Some provisions are too general (e.g., art. 6, 10, 14, 17, 21), becoming possibly over-inclusive and, thus, potentially harmful to innovative efforts, day-to-day operations of the regulated firms, or users' rights, such as free speech. Others are strongly pro-user and go into *minutiae* of the realisation of user rights (art. 16-22), revealing a clear view of how these algorithmic systems work and how their adverse effects might be halted or mitigated. Finally, other provisions seem aspirational or closer to public policy aims, such as observing "science and reason, and sincerity and trustworthiness" (art. 4) and advancing the use of algorithms "in the direction of good" (art. 6).

Overall, the regulation touches upon many subjects involved in using algorithmic systems. Its preoccupation with the generation of addiction and excessive consumption (art. 8 and 18) resonates with studies of the addictive effects of social media recommendation systems. Its inclusion of particular mention of the rights of workers mediated by algorithms (art. 20) seems to recognise potential vulnerabilities in the algorithmic labour organisation. Its inclusion of fake news (art. 12) and manipulative practices, including false likes, comments and shares (art. 14), echoes some pervasive practices that threaten political systems worldwide. Finally, the inclusion of granular user control of algorithmic systems, including the possibility to outright deactivate those systems (art. 17), marks a strong position in empowering users in the face of data-intensive digital platforms.

On the other hand, the regulation contains several references to State control of news media (art. 13) and overly broad provisions and terms insufficiently defined (e.g., art. 6, "mainstream value

orientations", "positive energy"); and does not go into detail on the administrative structure that will be needed to operate the level of control the regulation aims to implement. Moreover, the use of broad language in the definition of controlled content – including encouraged internet content – follows the previous tendency set by the Provisions on the Governance of the Online Information Content Ecosystem[84] enacted in 2020.

Overall, the Chinese framework provides interesting study cases for Western legislators[85] – in its dos and don'ts. The incisiveness in dealing with the technical details of algorithmic recommendation systems and their societal and economic consequences demonstrates possible strategies for dealing with the harms of surveillance capitalism and the attention economy. However, the use of overly broad legal terms, especially concerning content control, and the lack of an independent regulator implementing the provisions may lead to frameworks considerably unaligned with the West's paradigm on due process, necessity and proportionality in case of restrictions to speech.

2.5. South Africa

Regarding the legal and regulatory environment for intermediary liability in South Africa, Zingales[86] points out that the Republic of South Africa's Constitution enacts equality, dignity, freedom, and advancement of human rights as its central values. In its democratisation process, the country prioritised promoting equality, stating it on several legal provisions, such as the Promotion of Equality and Prevention of Unfair & Discrimination Act (PEPUDA), from 2000. This Act also defines hate speech, which binds application providers to combat explicitly hateful

[84] China. (2020, March 1). *Provisions on the Governance of the Online Information Content Ecosystem*. Wilmap. https://wilmap.stanford.edu/entries/provisions-governance-online-information-content-ecosystem.
[85] Wheeler, T. (2021, September 14). *China's new regulation of platforms: a message for American policymakers*. Brookings. https://www.brookings.edu/blog/techtank/2021/09/14/chinas-new-regulation-of-platforms-a-message-for-american-policymakers/.
[86] See Zingales, N. 2013. "Internet Intermediary Liability: Identifying Best Practices for Africa." Intermediary Liability in Africa Research Series, Association for Progressive Communications. November 26, 2013. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2359696.

content and content that might "be reasonably construed to have a clear intention to be hurtful"[87]. In this context, the DoC called, in 1999, for laws regarding intermediaries' liability, pointing out concerns about their roles in disseminating or allowing unlawful content. According to Zingales[88], this led to a public consultation resulting in the Electronic Communications and Transactions Act (ECTA), passed in 2001.

ECTA is considered "to date, the most articulate framework for dealing with intermediary[89] liability in Africa"[90]. Its development aimed explicitly to deal with the growth of e-commerce in the country, promoting legal certainty for enterprises with safe harbours similar to those present in the US's DMCA and the EU's E-Commerce Directive. It establishes that the law cannot require a service to actively monitor data, facts, or circumstances indicating unlawful activity. Nevertheless, it limits liability to two additional requirements: "(1) the intermediary's membership of an industry representative body (IRB); and (2) adoption and implementation of the corresponding code of conduct"[91]. Furthermore, the Minister of Communications issued a document titled "Guidelines for recognition of industry representative bodies of Information System Service Providers" in 2006, which integrates the code of conduct requirement. It states that "the only monitoring or control done by the State [...] is to ensure that the IRB and its ISPs meet certain minimum requirements"[92].

Despite such advanced provisions, intermediaries have been in relative juridical uncertainty, frequently subject to injunctions and lawsuits under criminal law for their users' behaviours. Moreover, in addition to the failures in the safe harbours' application, the hopes for building a more democratic social media governance are now on hold with the approval of several laws that constitute the so-called Internet Censorship Bill, as explored below.

[87] Zingales, Nicolo. 2020. "Intermediary Liability in Africa: Looking Back, Moving Forward?" In *Oxford Handbook of Online Intermediary Liability*, edited by Giancarlo Frosio, 213–35. Oxford: Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780198837138.013.11, p. 4.
[88] *Idem*, p. 5.
[89] Which is defined as "any person providing information system services" by its Chapter IX.
[90] *Ibidem*.
[91] *Idem*, p. 8.
[92] Zingales, Nicolo (2013), n. 71, p. 11.

2.5.1. Social Media Legislation and the "Internet Censorship Bill."

The South African primary legislation for regulating online content is the Film and Publications Act (FPA), 1996. The enactment of such a law aimed to repeal acts of prior legislation that aimed at censoring cultural productions under the apartheid context. It also established the Film and Publications Board (FPB) to receive complaints or applications to evaluate the classification of cultural production regarding its suitability for an audience.

By the end of 2019, the South African President, Cyril Ramaphosa, signed an Amendment to the FPA (a.k.a. the FPAA), dubbed the "Internet Censorship Bill" by opponents. The new version of the bill shifts the intermediary liability completely by imposing new duties and obligations to ISPs, which become obliged to monitor illicit, abusive, and harmful content, such as child exploitation and abuse imagery, war propaganda, incitement to violence, hate speech, and more. If the ISP fails to remove such content promptly, it could suffer sanctions such as fines of up to ZAR 50,000 (approximately 3,200 USD) and even imprisonment for six months. The bill also establishes criminal provisions for individuals that distribute prohibited content.

In addition, FPAA changes the role of the FPB, transforming it from a classification authority into a full regulator, with powers to renew or not the certificates of its applicants and request them to submit their content for evaluation. The FPB is also allowed, under the FPAA – that has started to take effect on March 2022 –, to issue takedown notices for ISPs regarding potentially prohibited content. But, experts at the ISPA[93] have been pointing out the possible censorship nature of FPB. Furthermore, they highlight the possible impacts of such measures, given that the body does not have the same capacities for weighing rights compared to the courts.

2.5.2. Other relevant legislations

a) Cybercrimes Act

The Cybercrimes Act, passed in May 2021, aims at tackling harmful speech in the online environment, including incitement of violence and other harms. It designates several specific

[93] Freedom House (2020), n. 70.

offences as cybercrimes and criminalises "malicious communication", such as sending data messages with violence, threats, harm, or non-consensual intimate imagery.

b) South African Disaster Management Regulations

The Covid-19 pandemic led the South African government to declare a "state of disaster" in March 2020, enacting the "Disaster Management Act", which criminalises disinformation. However, according to a report by Mawarire to USAID[94], "the Act had been amended at least three times within a month, making it difficult for ordinary citizens to interpret it". In addition, article 19 has pointed out some concerns with such measures, highlighting that it could be "a dangerous trend of countries using the Covid-19 pandemic to enforce disinformation laws in the region"[95].

3. Conclusion: Choosing between a sledgehammer and a scalpel to regulate content

With the increase in digital platforms' impact on political and economic systems, the BRICS countries are establishing regulations to tackle malicious activities and unlawful content, establishing intermediary obligations for transparency and accountability. The enactment of laws geared explicitly towards digital platforms aims to reassert state sovereignty in the online environment, preserving the stability and security of the national political infrastructures. Nevertheless, historical institutional complexities and disputes affect how such regulations and approaches are shaped and chosen for such sensitive matters.

Not surprisingly, we can remark on an inevitable overlap between the cultural specificities of the country at stake and its approach to content regulation. In Brazil, Draft Bill 2,630/2020 is extremely moderate because the Brazilian democratic model is historically sceptical towards

[94] Mawarire, Teldah. 2020. "'Things Will Never Be the Same Again' Covid-19 Effects on Freedom of Expression in Southern Africa, 2020 Research Report."
[95] Article 19. 2021. "South Africa: Prohibitions of False COVID-19 Information Must Be Amended." ARTICLE 19. April 23, 2021. https://www.article19.org/resources/prohibitions-of-false-covid-information-must-be-amended/.

media regulation and relatively permeable to lobbying from private companies. While the Brazilian Legislator might want to avoid using a sledgehammer to tackle disinformation with strict legislation, the proposed framework so far has failed to propose an effective scalpel to fight disinformation in a chirurgical fashion.

Interestingly, the Russian online content-blocking regime is remarkably similar to the Indian regime defined in Section 69A of the Information Technology (IT) Act. While there is no official document explicitly acknowledging the Russian influence, it is safe to assume knowledge of the Russian system from India (and other BRICS countries), given the existence of a specific intra-BRICS body for information exchange on cybersecurity for almost eight years. However, both regulatory frameworks have been criticised for their tendency towards a sledgehammer approach to platform regulation, which may easily be abused.

The Chinese approach seems to be the most coherent and structured, as well as the most innovative. While it adopts a rigorous approach to content regulation, it offers valuable food for thought regarding what practical measures can be considered and the tough time that legislators might have to regulate disinformation effectively without engaging in draconian norms. The South African approach is an example of how even countries that are internationally renowned for their commitment to democracy and human rights and strive to elaborate a well-articulated framework to regulate content properly will inevitably end up being criticised for censorship.

Despite the divergences in the BRICS online content regulations, some common trends can be highlighted. First, almost all countries are drafting or have already passed legislation outlawing specific types of online content and frequently defining transparency obligations, from moderate ones, such as the Brazilian, to stricter ones, such as the Chinese. Duties of care are present in most legal frameworks. The oversight mechanism allowing the implementation of the content regulation provisions is usually an administrative procedure. As such, this governance model may lead to concerns regarding the independence of the process and the proportionality and full respect of rule-of-law criteria, especially when the administrative body competent for the oversight is not an independent body.

To conclude, we provide the reader with a visual representation[96] of the primary norms regulating online content, the type of content deemed illegal, and the bodies competent for implementing the regulatory framework in each BRICS country. The regulatory choices of the BRICS members will naturally exert influence on the countries' regional neighbours, but these frameworks should also be carefully analysed by non-BRICS nations struggling with similar issues. While the BRICS have long been transplanting Western policy elements in their national frameworks, some of the BRICS countries are amongst the most "experienced" regarding content regulation. Their experiences offer valuable insights into what could, should or should not be reproduced by others.

[96] A detailed visual representation of the online content normative frameworks of the BRICS countries can be found at https://cyberbrics.info/map-online-content-normative-frameworks-in-the-brics/