IN THE ISLAMABAD HIGH COURT, ISLAMABAD

AMICUS BRIEF SUBMITTED BY:

NIGHAT DAD

IN

Writ Petition No. 3028/2020

Muhammad Ashfaq Jutt vs. Federation of Pakistan, etc.

For hearing on 03.02.2022

(Petition under Article 199 of the Constitution of the Islamic Republic of Pakistan)

Table of Contents

Table of References

16	Near v. Minnesota, 283 U.S. 697 (1931)	12
17	New York Times Company v. United States, 403 U.S. 713 (1971)	12
18	Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; and the Special Rapporteur on the situation of human rights defenders," OL PAK 3/2020 (2020)	15
19	Civil Aviation Authority Case, PLD 1997 SC 781	19
20	Mohtarma Benazir Bhutto vs. President of Pakistan, PLD 998 SC 388	21

Annex #	
1	Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; and the Special Rapporteur on the situation of human rights defenders," OL PAK 3/2020 (2020)
2	UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Online Content Regulation, A/HRC/38/35 (2018)
3	Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/74/486 (2019)
4	UN Guiding Principles on Business and Human Rights (2011)
6	The Johannesburg Principles on National Security, Freedom of Expression and Access to Information (1996)
7	The Manila Principles on Intermediary Liability (2014)

Summary of Arguments

1. This amicus curiae brief has been shared in response to the questions framed by the Honourable Islamabad High Court in the case of *Muhammad Ashfaq Jutt v. Federation of Pakistan, etc.*, W.P. No.3028/2020 through order dated 06-01-2022. This brief will address the two central questions framed by the Honourable High Court regarding the *Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules, 2021*, dated 13-10-2021 (henceforth referred to as "the 2021 Rules"):

    a. *Whether the Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules, 2021 are in consonance with the fundamental rights guaranteed under Article 19 and 19A of the Constitution of Islamic Republic of Pakistan, 1973?*
    b. *Whether they are in conflict with the parent statute, i.e. the Prevention of Electronic Crimes Act, 2016?*

2. It is contended that the 2021 Rules are in conflict with the right to freedom of expression and access to information, as enshrined in both the *Constitution of Islamic Republic of Pakistan, 1973* (henceforth referred to as "the Constitution ") and Pakistan's international human rights commitments as they are overbroad, contain disproportionate sanctions, lack regulatory independence, do not have meaningful appeal mechanisms, and take a one-size-fits-all approach to a wide array of content. The Rules violate other fundamental rights of the Constitution, particularly Article 14, 16, 17 in addition to Articles 19 and 19A.

3. It is also posited that the 2021 Rules go beyond the ambit of the *Prevention of Electronic Crimes Act, 2021* (henceforth referred to as "PECA"). Furtherstill, it is posited that Section 37 of PECA itself runs afoul international standards of free speech and access to information.

Case Background

4. In February 2020 when the cabinet notified the '*Citizens Protection (Against Online Harm) Rules 2020*' without any consultations, there was immense backlash and pressure from citizens and civil society actors over the lack of transparency from the Government.[1] This resulted in the first draft of the Rules being withdrawn, though there

[1] "No Consultation without Withdrawal of Cabinet Approval of Online Protection (Against Online Harm) Rules 2020", March 1, 2020, https://digitalrightsfoundation.pk/no-consultation-without-withdrawal-of-cabinet-approval-of-online-protection-against-online-harm-rules-2020/.

was no formal de-notification, and the government publicly committed to consultations. The consultations, conducted by the Ministry of Information Technology & Telecommunication (henceforth referred to as "MoITT"), were initiated in July 2020. These were limited and closed consultations boycotted by several civil society organisations, including all major digital rights organisations, over concerns regarding transparency and de-notification of the original Rules.[2]

5. Following the consultations in July 2020, no revised draft was shared with the public or consultation participants, and the '*Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules, 2020*' were published in the Extraordinary Gazette on 20-10-2020. Furthermore, the document was changed on 27-11-2020 again without any formal de-notification of the October 2020 version. After this publication, several courts were approached to challenge the Rules, including the august Islamabad High Court.

6. On 31-03-2021 the Prime Minister's office constituted an inter-ministerial committee to examine the Rules. Given that consultations had already taken place in July 2020, no assurances were provided to make sure that this round of consultation will be substantially different, more inclusive and less of an eyewash than the previous iteration. The amicus, Nighat Dad, along with other civil society members, took part in this round of consultations to ensure that their concerns regarding the Rules were raised on the record.[3]

7. On 18-06-2021, another draft was shared with the public accompanied by a call for feedback on the MOITT website.[4] This version of the draft did not contain any substantive changes vis a vis the November 2020 version of the Rules. The *amicus curiae*, Nighat Dad, submitted written recommendations and comments on the draft via email to the MOIT on July 5, 2021. In October 2021, the latest version of the Rules, i.e. '*Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards), Rules 2021'* were notified.[5]

8. The following are links to previous submissions and analysis of earlier versions of the Rules made by the *amicus curiae*:

[2] "Comments on the Consultation & Objections to the Rules," July 1, 2020, https://digitalrightsfoundation.pk/comments-on-the-consultation-objections-to-the-rules/.
[3] "Legal Analysis: Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules, 2020," Digital Rights Foundation, https://digitalrightsfoundation.pk/wp-content/uploads/2020/12/Removal-and-Blocking-of-Unlawful-Online-Content-Procedure-Oversight-and-Safeguards-Rules-2020_-Legal-Analysis.pdf.
[4] https://moitt.gov.pk/Detail/YjVmNzU0MWMtYzBkMC00Yjg5LTk1ODktOTJiODYzZTY5ZWRk.
[5] "Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards), Rules 2021", https://moitt.gov.pk/SiteImage/Misc/files/Removal%20Blocking%20of%20Unlawful%20Online%20Content%20Rules%202021.PDF.

a. Citizens Protection (Against Online Harm) Rules 2020: https://digitalrightsfoundation.pk/wp-content/uploads/2020/02/Legal-Analysis-Harm-Rules-1.pdf.
b. Removal and Blocking of Unlawful Online Content (Procedure, Oversight and Safeguards) Rules, 2020: https://digitalrightsfoundation.pk/wp-content/uploads/2020/12/Removal-and-Blocking-of-Unlawful-Online-Content-Procedure-Oversight-and-Safeguards-Rules-2020_-Legal-Analysis.pdf.

About the Amicus

9. Nighat Dad is the founder and Executive Director of Digital Rights Foundation (DRF), a Lahore-based non-profit working on issues of online free speech, privacy and digital safety. Nighat is currently serving as a member of the Facebook Oversight Board, working on content moderation on the platform, a serving member of the Advisory Council on Human Rights and Technology at Microsoft, and part of the Trust and Safety Council in Twitter. She has been working for over a decade on issues of online content moderation, advocating with governments and tech companies to tackle online violence against women, minorities and other vulnerable groups. Nighat is a TED Fellow and has received numerous prestigious awards including the Dutch Human Rights Tulip Award and a TIME's Next Generation Leader.

10. In light of this experience in the field of digital rights and internet governance, this brief has been prepared to reflect the human rights perspectives and international best practices on content moderation.

International Human Rights Standards

11. The following international instruments and principles have been referred to in this document:
    a. International Covenant on Civil and Political Rights (ICCPR).[6]
    b. UN Guiding Principles on Business and Human Rights.[7]
    c. Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises,

[6] International Covenant on Civil and Political Rights, Adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966 entry into force 23 March 1976, in accordance with Article 49, https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx.
[7] "Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework", HR/PUB/11/04, https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf.

"Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework," A/HRC/17/31 (2011).[8]

d. Human Rights Committee, General Comment No. 34, 2011 (General Comment 34).[9]
e. UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Online Content Regulation, A/HRC/38/35 (2018).[10]
f. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/74/486 (2019).[11]
g. The Manila Principles on Intermediary Liability (2014).[12]
h. The Johannesburg Principles on National Security, Freedom of Expression and Access to Information (1996).[13]

Article 19 of ICCPR

12. Article 19 of the *International Covenant on Civil and Political Rights* (henceforth referred to as the "ICCPR") guarantees the freedom of expression and opinion. Pakistan became a signatory of the ICCPR in 2008 and ratified the convention in 2010. Article 19 states:

> "*1. Everyone shall have the right to hold opinions without interference.*
> *2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.*
> *3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may*

[8] "Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework", Report of the Special Representative of the SecretaryGeneral on the issue of human rights and transnational corporations and other business enterprises, Human Rights Council, John Ruggie, 21 March 2011, A/HRC/17/31, https://www.ohchr.org/documents/issues/business/a-hrc-17-31_aev.pdf.
[9] "General comment No. 34: Article 19: Freedoms of opinion and expression", *Human Rights Committee*, 12 September 2011, United Nations, CCPR/C/GC/34, https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.
[10] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, "Online Content Regulation", Human Rights Council, A/HRC/38/35 (2018), https://www.undocs.org/A/HRC/38/35.
[11] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/74/486 (2019), https://undocs.org/A/74/486.
[12] "Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation," Global Civil Society Initiative, March 24, 2015, https://www.eff.org/files/2015/10/31/manila_principles_1.0.pdf. .
[13] "The Johannesburg Principles on National Security, Freedom of Expression and Access to Information," *Article 19*, November 1996, https://www.article19.org/wp-content/uploads/2018/02/joburg-principles.pdf.

> *therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:*
> *(a) For respect of the rights or reputations of others;*
> *(b) For the protection of national security or of public order (ordre public), or of public health or morals.*"[14]

13. Interpreting Article 19 in the context of content moderation, the *UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* stated in their 2018 report that while companies have an obligation to conduct online content moderation in some cases, regulations by States on the pretext of content moderation should not result in "*censorship and criminalization to shape the online regulatory environment*" nor rely on *broadly worded restrictive laws*.[15]

14. The Special Rapporteur noted that overly broad rules on content moderation:

> "*[I]nvolve risks to freedom of expression, putting significant pressure on companies such that they may remove lawful content in a broad effort to avoid liability. They also involve the delegation of regulatory functions to private actors that lack basic tools of accountability. Demands for quick, automatic removals risk new forms of prior restraint that already threaten creative endeavours in the context of copyright. Complex questions of fact and law should generally be adjudicated by public institutions, not private actors whose current processes may be inconsistent with due process standards and whose motives are principally economic*."[16]

15. Applying these standards to the 2021 Rules, Rule 5 does not allow intermediaries sufficient time to analyse the take-down request or seek any further judicial remedy. The Rules have the potential of a chilling effect on the content removal process as social media companies will rush content regulation decisions to comply with the restrictive time limit (12-48 hours), leading to hasty decisions on particularly complicated cases of free speech that require deliberation and legal opinions. Given the massive volume of content shared online every day, platforms may feel obliged to take a 'better safe than sorry' approach–which in this case would mean 'take down first, ask questions later (or never).'

---

[14] Article 19, ICCPR.
[15] UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Online Content Regulation, A/HRC/38/35 (2018), para. 13.
[16] Online Content Regulation, A/HRC/38/35 (2018), para. 17.

16. There is no opportunity for the 48-hour time limit to be put on hold in case a Social Media Company or Service Provider decides to contest PTA's decision for removal of content, they would be required to comply with the decision first and raise contestations later. This threatens to incentivize the removal of legitimate content. Furthermore, smaller Social Media Companies, which do not have the resources and automated regulation capacities that big tech companies such as Meta or Google possess (defined as Significant Social Media Companies in the 2021 Rules), will be disproportionately burdened with urgent content removal orders.

17. Further, in situations of an emergency, such as sexually explicit content causing harm on the basis of a protected category or hate speech inciting violence against an individual or community, it may be tenable to impose certain median timelines, but for content that relates to private disputes/wrongs and has a free speech element, such as defamation, it would be unreasonable to impose such a strict timeline for intermediaries to act. In all instances, the provision should also contain "Stop the Clock" provisions by listing out a set of criteria (such as seeking clarifications, technical infeasibility, etc.) under which the time limit would cease to apply to allow for due process and fair play in enforcing such requests.

18. The *UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* has concluded any regulation on content moderation needs to be "smart", i.e. focusing on a. Transparency; b. Adequate Remedies; c. based on orders by Independent Judicial Authority; d. follow the tripartite rule of Legality, Necessity and Legitimacy under Article 19 of the ICCPR; and e. shall not contain Disproportionate Sanctions. The relevant text from the 2018 report has been reproduced below:

> "*[S]mart regulation, not heavy-handed viewpoint-based regulation, should be the norm, focused on ensuring company transparency and remediation ... [s]tates should only seek to restrict content pursuant to an order by an independent and impartial judicial authority, and in accordance with due process and standards of legality, necessity and legitimacy. States should refrain from imposing disproportionate sanctions, whether heavy fines or imprisonment, on Internet intermediaries, given their significant chilling effect on freedom of expression*."[17]

Principles for Private Companies

[17] Ibid, para. 66.

19. The 2021 Rules cannot be considered "smart regulation" based on any of the criteria laid down under international human rights law. Transparency and remediation have been emphasised by the *UN Guiding Principles on Business and Human Rights* (henceforth referred to as the "UN Guiding Principles"), endorsed by the UN Human Rights Council in 2011. The UN Guiding Principles establish a voluntary framework for private businesses to adhere to human rights standards, chief among them is freedom of expression. Principle 13 of the UN Guiding Principles state that companies "*avoid causing or contributing to adverse human rights impacts through their own activities, and address such impacts when they occur*".[18] They rest on a "Protect, Respect and Remedy" Framework where states have an obligation to protect against human rights abuses by third-parties, secondly, corporations have a responsibility to respect human rights, and lastly, individuals have access to an effective remedy, both judicial and non-judicial.[19] When compared with these Guiding Principles, the 2021 Rules fail to place respect for human rights at the center of content moderation or provide adequate remedy to end-users in Pakistan.

20. Applying the UN Guiding Principles, the *Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprise*s lays down substantive standards for content moderation which include obligations to ensure human rights by default; legality; necessity and proportionality; and non-discrimination.[20] Rule 7 of the UN 2021 Rules, on the other hand, lays down obligations to develop Community Guidelines in line with local law, with no guardrails for human rights protections.

<u>Automated Systems and Human Rights</u>

21. Rule 7(6)(f) requires Significant Social Media Companies to "deploy suitable content moderation methods including Artificial Intelligence (AI) based content moderation system(s) and content moderators well versed with the local laws." Furthermore, they place obligations on Service Providers and Social Media Companies to deploy mechanisms to ensure "immediate blocking" of live streams related to "terrorism, hate speech, pornographic, incitement to violence and detrimental to national security" (Rule 7(5)). This process lacks transparency, effective remedy and provides no human rights standards to companies. It bears repeating that most social media companies already have

[18] UN Guiding Principles on Business and Human Rights, Principle 13.
[19] Report of the Special Representative of the Secretary General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, A/HRC/17/31 (2011), para. 6.
[20] Online Content Regulation, A/HRC/38/35 (2018), para. 45-48.

systems in place to remove live streams containing violent and sexually explicit content that can result in breaches of privacy or harms, such as child pornography. For instance, in wake of the Christchurch attacks in New Zealand in 2019, Facebook (now Meta) updated its policies and AI systems to ensure that violent and suicide related content on livestreams was prioritised for content moderation.[21]

22. These obligations under the 2021 Rules regarding automated filtering are out of step with the recommendations by the *UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* which find that "*States and intergovernmental organizations should refrain from establishing laws or arrangements that would require the "proactive" monitoring or filtering of content, which is both inconsistent with the right to privacy and likely to amount to pre-publication censorship.*"[22]

23. Pre-publication censorship or "prior restraint" models for regulating speech are "*regulations which prevent the publication of speech prior to its distribution, including orders to remove an expression that has already been published*."[23] Prior restraint models are subject to strict scrutiny in in US free speech jurisdiction, as established by the US Supreme Court in *Near v. Minnesota.*[24] In the landmark case of *New York Times Company v. United States*,[25] the government's plea seeking to block the publication of the "Pentagon Papers" on the basis of national security was rejected as the court noted that: "*[a]ny system of prior restraints of expression comes to this Court bearing a heavy presumption against its constitutional validity" ... The Government "thus carries a heavy burden of showing justification for the imposition of such a restraint*."[26] Prior restraint models in the digital age are most common when it comes to the automatic blocking of content through automated systems and AI. These models have the potential of enacting widespread censorship online.

24. The *UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* in their report on online hate speech noted that automated tools for speech regulation "*serve as a form of pre-publication censorship… because such filters are notoriously unable to address the kind of natural language that typically constitutes hateful content, they can cause significant disproportionate outcomes.*

---

[21] Guy Rosen, A Further Update on New Zealand Terrorist Attack, Meta, March 20, 2019, https://about.fb.com/news/2019/03/technical-update-on-new-zealand/.
[22] Online Content Regulation, A/HRC/38/35 (2018), para. 67.
[23] Ariel L. Bendor and Michal Tamir, Prior Restraint in the Digital Age, 27 Wm. & Mary Bill Rts. J. 1155 (2019), https://scholarship.law.wm.edu/wmborj/vol27/iss4/7.
[24] 283 U.S. 697 (1931).
[25] 403 U.S. 713 (1971).
[26] Ibid.

Furthermore, there is research suggesting that such filters disproportionately harm historically underrepresented communities."[27]

Reasonable Restrictions on Content Moderation

25. Content moderation in digital spaces is both complicated and inevitable. The internet is host to an array of content, ranging from informative to harmful. The central challenge for content moderation is the subjectivity inherent in deciding which content is permissible and which should be removed or restricted. Article 19 of the ICCPR lays down a three-part test: (1) legality, (2) legitimate purpose and (3) necessity and proportionality. When determining whether the 2021 Rules are consistent with the right to freedom of expression, the honourable court would have to determine whether the criteria and processes in place to place content restrictions under the 2021 Rules, i.e. the powers to remove and block content, satisfies this tripartite test. It is contended that the Rules fail on all three aspects of this test.

26. *General Comment No. 34* on Article 19 of the ICCPR elaborates that the first arm of the test requires that "*a norm, to be characterized as a "law", must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly and it must be made accessible to the public*."[28] Additionally, the law must not "confer unfettered discretion for the restriction of freedom of expression on those charged with its execution."[29]

27. The requirement for any restriction on freedom of expression to be precisely drafted so that any member of the public can understand what constitutes a violation cannot be seen in the 2021 Rules under challenge. Rule 4 of the 2021 Rules defines unlawful content under five heads: glory of Islam; security of Pakistan; public order; decency and morality; and the integrity or defence of Pakistan. Many of these phrases are not sufficiently defined. For instance, the definition of "public order" includes the "dissemination of fake or false information that threatens public order, public health and public safety". The determination of what constitutes false information is extremely subjective especially when determined by a government body that practices little independance from the government. The Pakistan Telecommunications Authority (henceforth referred to as the "PTA"), tasked with determining whether the content is unlawful, has members appointed by the Federal Government.[30] Furthermore, Section 8

[27] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/74/486 (2019), para. 34.
[28] Human Rights Committee, General Comment No. 34, 2011 (General Comment 34), Para. 25.
[29] Ibid.
[30] S. 3(2), Pakistan Telecommunication (Re-Organization) Act, 1996.

of the *Pakistan Telecommunication (Re-Organization) Act, 1996* gives the Federal Government the powers to issue policy directives to the PTA.[31]

28. Rule 4 contains references to several criminal offences many of which are not speech acts. This means that offences meant to be interpreted by judicial officers will be interpreted by a statutory body. It is unclear how these will be interpreted by the PTA, which has sole discretion to decide which content constitutes unlawful content. Rule 5(2) states that it will give those adversely affected by a complaint the opportunity to be heard, however there is no procedure laid down especially in cases that are deemed "emergencies" and where the identity of the person behind a post is unclear.

29. Lastly, there are insufficient avenues for appeals against the PTA. Review against orders of the PTA also lies with the PTA (Rule 8(1)). While there is an appeal against a review order of the PTA with the High Court, the entire process of appealing an arbitrary order of the PTA would take several months at best, which would mean crucial content can remain removed for lengthy periods of time. In the context of digital spaces, where the reach of content is often time-bound given the algorithmically determined nature of content dissemination, such a delay is tantamount to censorship itself. In cases where those aggrieved by the PTA's orders do not have resources to re-appeal its orders, there will effectively be no remedy.

30. The second part of the test which requires that any restriction of content must pursue a legitimate aim is not clearly spelt out in the Rules. *General Comment No. 34* states that grounds such as public order (*ordre public*), or of public health or morals, should be interpreted with "extreme care":

> "*It is not compatible with paragraph 3, for instance, to invoke such laws to suppress or withhold from the public information of legitimate public interest that does not harm national security or to prosecute journalists, researchers, environmental activists, human rights defenders, or others, for having disseminated such information*."[32]

31. Specifically when regulating content on the basis of public order, international human rights law states that it may "*be permissible in certain circumstances to regulate speech-making in a particular public place*."[33] The 2021 Rules under challenge fail to provide guarantees of this specificity. Additionally, these Rules cannot be looked at in a vacuum; the fact that vague terms such as "morality" and "decency" have been used by

[31] S. 8, Pakistan Telecommunication (Re-Organization) Act, 1996.
[32] Human Rights Committee, General Comment No. 34, 2011 (General Comment 34), Para. 30.
[33] Ibid, Para. 31.

the PTA in the past to impose blanket bans on platforms and applications such as TikTok. The Rules fail to provide specific guidelines or safeguards to prevent such arbitrary decision-making.

32. Thirdly, it is required that the restrictions must be necessary for the legitimate purpose. It is stated that the restrictions must not be "overbroad" or disproportionate, meaning that they must be the least restrictive or least intrusive method to achieve the legitimate purpose. Any such restriction "*must demonstrate in specific and individualized fashion the precise nature of the threat*".[34] In the case of the 2021 Rules, the Government has not entertained other options such as safeguards for speech by journalists and other protected speech. Rule 5, which deals with the disposal of a complaint, contains no provisions which require the restrictions to be limited. In fact, the Rule contains the power to block the entire Online Information System, which is both a disproportionate power and goes beyond the scope of power accorded by Section 37 of PECA (discussed below). *General Comment No. 34* states that "*permissible restrictions generally should be content-specific; generic bans on the operation of certain sites and systems are not compatible with paragraph 3 [of Article 19 in ICCPR]*".[35] The *UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* has specifically cited the heavy-handed regulation of the Pakistani government and blanket bans in their report on content moderation:

> "*States [...] place pressure on companies to accelerate content removals through nonbinding efforts, most of which have limited transparency. A three-year ban on YouTube in Pakistan compelled Google to establish a local version susceptible to government demands for removals of "offensive" content*."[36]

33. In March 2020, the *UN Special Rapporteurs on freedom of opinion* and on *the situation of human rights defenders* wrote to the Pakistan Government regarding the '*Citizens Protection (Against Online Harm) Rules 2020*' to raise concerns regarding the Rules compliance with human rights obligations towards free expression. The letter noted that despite severe restrictions on speech and broad powers, "the Rules do not contain any procedureal safeguards against abuse."[37]

---

[34] Ibid, Para. 35.
[35] Ibid, Para. 43.
[36] Online Content Regulation, A/HRC/38/35 (2018), para. 20.
[37] "Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; and the Special Rapporteur on the situation of human rights defenders," OL PAK 3/2020, March 19, 2020, https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL_PAK.pdf.

34. In conclusion, the 2021 Rules do not satisfy standards set under international human rights law for restrictions on speech. The Rules are too broadly drafted and lack targeted approach to dealing with illegal online content. Furthermore, the Rules operate from a limited understanding of what constitutes content moderation, which includes both removal of content and protections against arbitrary and unreasonable removals. The 2021 Rules fail to come up with a mechanism that holds companies accountable for "false positives", where content is incorrectly flagged or removed. Content moderation is not restricted to simply removal or restriction of content, it is concerned with how content is disseminated to the audience through targeted ads, algorithmically curated timelines that determine the content consumption of individuals, and lack of timely and independent fact-checking for misinformation and disinformation. These Rules do not focus on these larger issues of equitable content moderation, rather focus on concentrating power in the hands of the PTA.

35. Crucially, the organization "Article 19" found in its report on online content regulation that there is a glaring lack of evidence about the effectiveness of content regulation measures. In order to curtail a fundamental right, the Government needs to demonstrate a strong nexus between the mechanism proposed by the 2021 Rules and the purported aims of Rules. The report found "*[a]ll too often, lawmakers seek to adopt laws to send a political message to the public that 'something is being done' to address an issue, rather than investing resources in less visible but more effective long-term solutions*."[38]

Rules are Ultra Vires the PECA

36. Notwithstanding the critiques of PECA and Section 37 that have been raised previously,[39] the 2021 Rules are *ultra vires* their parent statute, PECA as they exceed the scope of Section 37.

37. Section 37 states, "*The Authority shall have the power to remove or block or issue directions for removal or blocking of access to an information through any information system*". It is evident from a bare reading of Section 37 that it only grants limited powers to remove or block access to *only a particular information.* It nowhere states that PTA can remove or block access to Online Systems including virtual platforms or applications.

---

[38] "Watching the watchmen Content moderation, governance, and freedom of expression," Article 19, 2021, https://www.article19.org/wp-content/uploads/2021/12/Watching-the-watchmen_FINAL_8-Dec.pdf, p. 18.

[39] Concerns along lines of human rights have been raised regarding PECA since its drafting. In 2015, the UN Special Rapporteur on freedom of opinion and expression raised concerns regarding the then-draft PECA, which he warned "could result in censorship of, and self-censorship by, the media." Source: "UN expert urges Pakistan to ensure protection of freedom of expression in draft Cybercrime Bill," December 14, 2015, https://www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=16879&LangID=E.

38. The manner in which Section 37 of PECA is being interpreted and used by the PTA to block entire information systems is the result of gross and erroneous misreading of the said Section. It is a well-settled proposition that if the parent act defines the scope of the power to make delegated legislation relatively tightly, then the courts must intervene. Therefore, the authority granted to PTA under Rule 5 to block entire Online Systems goes beyond the scope of Section 37 of PECA and is *ultra vires* the enabling Act.

39. Secondly, other powers accorded to the PTA under the 2021 Rules, particularly the power to "*degrade the services of such Service Provider or Social Media Company or Significant Social Media Company for such period of time as deemed appropriate by the Authority*" (Rule 5(7)(ii)(a)) has not been granted under Section 37. This is an excessive power granted to PTA. Iit violates the established principle of "net neutrality" which requires that all data and content on the internet shall be treated equally and without discrimination, as explained below:

> "*Network neutrality (or net neutrality) is a design paradigm according to which the network must not prioritise some information over other, for example by charging different rates or providing different bandwidths. Network neutrality is closely related to the demand for openness of the Internet and can be violated by blocking, monopolistic pricing, preferential treatment to certain providers (or certain content) and failures in transparency. […] But checking the content that runs through the network is only possible through Deep Packet Inspection, a practice that meets serious human rights challenges, especially in light of the closed process of standard-setting in which the World Telecommunication Standardisation Assembly of the International Telecommunication Union passed its new Requirements for Deep Packet Inspection in Next Generation Networks"*[40]

40. Any efforts to "degrade" content on any platform (Rule 5(7)(ii)(a)), i.e. slowing down internet speeds to access it, are violative of the principle of net neutrality.[41] Various jurisdictions have adopted strong network neutrality principles. For instance, India is considered to have adopted some of the strongest rules on net neutrality in 2018.[42]

---

[40] Freedom of expression and the Internet by Wolfgang Benedek and Matthias C. Kettemann https://rm.coe.int/prems-167417-gbr-1201-freedom-of-expression-on-internet-web-16x24/1680984eae

[41] "Net Neutrality," The EDRi papers, issue 08, https://edri.org/files/EDRi_NetNeutrality.pdf.

[42] Rishi Iyengar, "India now has the 'world's strongest' net neutrality rules", July 12, 2018, CNN, https://money.cnn.com/2018/07/12/technology/india-net-neutrality-rules-telecom/index.html.

Intermediary Liability

41. Rule 7(3), which requires that Service Providers, Social Media Companies and Significant Social Media Companies "*shall not knowingly host, display, upload, publish, transmit, update or share any Online Content in violation of local laws*", is violative of Section 38 and *ultra vires* PECA. This provision violates the fundamental principle of intermediary liability which protects platforms and service providers from being held liable for content hosted on their platform. Section 38 of PECA states (underscore added):

> "No service provider shall be subject to any civil or criminal liability, unless it is established that the service provider had specific actual knowledge and willful intent to proactively and positively participate, and not merely through omission or failure to act."[43]

42. The specific inclusion of Section 38 in PECA is a clear indication that the legislature intended to protect against intermediary liability. Protections for intermediary liability against third-party content is often considered to be one of the foundational principles of the internet, with similar provisions found in other countries, such as Section 230 of the *Communications Decency Act* in the United States.

43. Even otherwise and without prejudice to the foregoing, under the 2021 Rules, the 48 and 12 hour time-frames are not sufficient for platforms to review content and failure to comply within such unreasonable and restrictive conditions is not lawful justification to establish "willful intent" on their part to "proactively and positively participate" in offences under PECA.[44] The Rules have the effect of rendering the intermediary liability protections under Section 38 of PECA meaningless. Furthermore, many standards draw distinctions between "actual" and "constructive" knowledge, with actual knowledge of illegality obtained through a court order rather than notice by a body such as the PTA.[45]

Constitutionality

44. Without conceding that the mandate given under Section 37 of PECA allows blocking of entire Online Systems, it is still submitted that the power to 'block' an Online System is a

[43] Section 38, PECA.
[44] Actual knowledge is deemed to be fulfilled through notice requirements in other jurisdictions in a limited number of circumstances such as copyright claims rather than larger speech content. The 'Digital Millennium Copyright Act, or 'DMCA', in the US one such example.
[45] "Watch the Watchmen," pg. 27.

violation of Article 19 of the Constitution as well. The said Article only allows for "reasonable restrictions" to be imposed on free expression in accordance with law. It was held in *Civil Aviation Authority Case*[46] that "*the predominant meanings of the said words (restrict and restriction) do not admit total prohibition. They connote the imposition of limitations of the bounds within which one can act*". Therefore, the power to 'block' cannot be read under, inferred from or assumed to be a part of the power to restrict free speech. While Article 19 of the Constitution allows imposition of "restrictions" on free speech, the power to "block" an information system entirely exceeds the boundaries of permissible limitations under it and renders Rule 5 inconsistent with the Constitution.

45. It is submitted that in today's digital age, Online Systems allow individuals to obtain information, form, express and exchange ideas and are mediums through which people express their speech. Hence, entirely blocking an Online System is synonymous with blocking speech itself. The blocking of Online Systems, as a blunt instrument will cause unintended consequences, including preventing Pakistani citizens and companies from benefiting from access to resources from the rest of the world, thus inhibiting the country and reinforcing a digital divide.

Privacy and Data Localisation

46. It is submitted that the requirement for "Significant Social Media Companies" to register with PTA, establish a permanent registered office in Pakistan, and "comply with the user privacy and data localisation in accordance with applicable laws" is a move towards "data localisation" and challenges the borderless nature of the internet – a feature that is intrinsic to the internet itself. Even otherwise, forcing businesses to create a local presence is outside normal global business practice and compels an investment without a business need.

47. The requirement to establish database servers in Pakistan is alarming inasmuch as it threatens the state of privacy of citizens in Pakistan because there are no data protection laws within the country at the moment – leaving the data/information so collected or gathered to open abuse and misuse. The *United Nations High Commissioner for Human Right* recommends that:

> "*Strict data localization requirements that oblige all data processing entities to store all personal data within the country at issue should be avoided. Instead, States should focus on ways to*

[46] PLD 1997 SC 781.

ensure that personal data transferred to another State is protected at least at the level required by international human rights law."[47]

48. It is important to note that data localisation *per se* does not protect the safety of personal data. If other jurisdictions offer an adequate level of protection, there is no justification based on the safety of personal data for preventing their transfer or imposing the storage of personal data in a particular country. Research in other jurisdictions has shown that confining data to a few physical locations can often reduce the level of security rather than enhance it, making it vulnerable to hacking and cyber attacks. To effectively defend against cybercrimes and threats, companies protect user data and other critical information via a very small network of highly secure regional and global data centers staffed with uniquely skilled experts who are in scarce supply globally. These centers are equipped with advanced IT infrastructure that provides reliable and secure round-the-clock service. The clustering of highly-qualified staff and advanced equipment is a critical factor in the ability of institutions to safeguard data from increasingly sophisticated cyber attacks. Further, it has been noted that in other jurisdictions the imposition of data localisation has been introduced as a way to facilitate unlawful surveillance and limit the capacity of individuals to protect the confidentiality of their communications.[48]

49. Rule 7(4) requires social media companies to provide "decrypted, readable and comprehensible information" to the Federal Investigation Agency (hereinafter as the "FIA"). No requirement for judicial oversight of such requests is included. Encryption is considered to be an important component of the right to privacy and allows for the exercise of other rights. The *UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression* has noted that "*Encryption and anonymity are especially useful for the development and sharing of opinions, which often occur through online correspondence such as e-mail, text messaging, and other online interactions*."[49]

50. Rule 7(4) essentially amounts to a key disclosure law or mandatory key disclosure law, in that it requires social media companies, platforms and service providers to hand over encrypted data to law enforcement. While the FIA can send data requests to companies in order to investigate crimes under PECA, the requirement for data to be decrypted threatens the privacy expectation of users on encrypted platforms. The right to privacy

[47] Report of the United Nations High Commissioner for Human Right, "The right to privacy in the digital age," Human Rights Council, A/HRC/39/29, August 3, 2018, https://undocs.org/A/HRC/39/29, para. 32.
[48] "The Localisation Gambit: Unpacking Policy Measures for Sovereign Control of Data in India," 2019, https://cis-india.org/internet-governance/resources/the-localisation-gambit.pdf.
[49] "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression", UN Human Rights Council, May 22, 2015, A/HRC/29/32, https://www.refworld.org/docid/5576dcfc4.html.

has been guaranteed under Article 14 of the Constitution, which includes the privacy of correspondence:[50]

> "The dignity of man and, subject to law, the privacy of home, shall be inviolable."[51]

51. This provision also fails to account for the fact that platforms often do not have access to encryption keys themselves to effectively decrypt data and information. On the other hand, requiring companies to develop backdoors to encryption will expose entire platforms and services to attacks, thus undermining the overall privacy of all users. Instead of focusing on a rights-compliant data sharing mechanism that respects the privacy of users and focuses on timely retrieval of data for the most heinous of crimes, these Rules replicate the worst international practices focused on control of data rather than genuine issues faced by citizens.

Economic Harm

52. Such a regulation runs the risk of forcing international social media companies to exit the country rather than invest further in Pakistan. It is unreasonable to expect companies to set up infrastructure in the country when the nature of the internet allows for it to be easily administered remotely. Major tech companies have already expressed their reservations with the Rules and the possibility of exiting the country.[52] With an increase in compliance costs that comes with incorporation of a company in Pakistan, companies across the globe including start-ups may have to reconsider serving users in Pakistan. Consequently, users in Pakistan including the local private sector may not be able to avail a variety of services required for carrying out day-to-day communication, online transactions, and trade/business-related tasks. The 2021 Rules requiring local incorporation and physical offices will also have huge repercussions on taxation, foreign direct investment and other legal perspectives along with negatively impacting economic growth. The GNI notes that:

> "*Laws and regulations governing the ICT sector should also be targeted and narrowly framed. Lawmakers should pay careful attention to the ways laws and regulations will impact companies*

[50] Mohtarma Benazir Bhutto vs. President of Pakistan, PLD 998 SC 388.
[51] Article 14, Constitution of Islamic Republic of Pakistan, 1973.
[52] "[Pakistan] AIC Issues Media Statement on Removal and Blocking of Unlawful Content (Procedure, Oversight and Safeguards) Rules, 19 Nov 2020," Asia Internet Coalition, Novemver 19, 2020, https://aicasia.org/2020/11/20/pakistan-aic-issues-media-statement-on-removal-and-blocking-of-unlawful-content-procedure-oversight-and-safeguards-rules-20-nov-2020/.

> *with different business models, seeking to foster a diversity of digital services and avoid raising barriers to entry*."[53]

53. Requiring local incorporation and presence unnecessarily discriminates against foreign businesses, poses a non-tariff barrier to trade, and unfairly tilts the playing field in favour of domestic players. This is particularly stark in view of the nature of the services provided through the internet, which can be provided on a cross-border basis without the need for physical presence. By instituting local presence requirements, Pakistan is deviating from established international trade norms and practices, and erecting unnecessary barriers to cross-border services trade.

54. The global nature of the internet has democratized information that is available to anyone, anywhere around the world in an infinite variety of forms. The economies of scale achieved through globally located infrastructure have contributed to the affordability of services on the internet, where several prominent services are available for free. Companies are able to provide these services to users even in markets that may not be financially sustainable as they don't have to incur the additional cost of setting up and running local offices and legal entities in each country where they offer services. Therefore, the 2021 Rules will harm consumer experience on the open internet, increasing costs to an extent that offering services/technologies to consumers in Pakistan becomes financially unviable.

Other Fundamental Rights

55. Restrictive content moderation runs the risk of restricting other rights such as freedom of assembly and association guaranteed under Articles 16 and 17 of the Constitution:

> "16. Freedom of assembly.
> Every citizen shall have the right to assemble peacefully and without arms, subject to any reasonable restrictions imposed by law in the interest of public order.
>
> 17. Freedom of association:
> (1) Every citizen shall have the right to form associations or unions, subject to any reasonable restrictions imposed by law in the interest of sovereignty or integrity of Pakistan, public order or morality.

[53] "Content Regulation and Human Rights: Analysis and Recommendations," Global Network Initiative, 2020, https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf, pg. 6.

> (2) Every citizen, not being in the service of Pakistan, shall have the right to form or be a member of a political party, subject to any reasonable restrictions imposed by law in the interest of the sovereignty or integrity of Pakistan and such law shall provide that where the Federal Government declares that any political party has been formed or is operating in a manner prejudicial to the sovereignty or integrity of Pakistan, the Federal Government shall, within fifteen days of such declaration, refer the matter to the Supreme Court whose decision on such reference shall be final.
> (3) Every political party shall account for the source of its funds in accordance with law."[54]

56. The *UN Special Rapporteur on the rights to freedom of peaceful assembly and of association* stated in their "Report on the rights to freedom of peaceful assembly and of association: The Digital Age" has found that "digital technology is integral to the exercise of the rights of peaceful assembly and association. *Technology serves both as a means to facilitate the exercise of the rights of assembly and association offline, and as virtual spaces where the rights themselves can be actively exercised*."[55] In Pakistan, digital spaces and platforms have been used as a means to raise awareness regarding issues, particularly those marginalized in the mainstream. For instance, women and survivors of sexual violence and harassment have been using the #MeToo hashtag to raise awareness regarding their experiences and the need for meaningful reform of the law. The potential impact of broad-based regulation on these freedoms should need to be considered before the 2021 Rules are deemed to pass Constitutional muster.

Alternate Models of Content Moderation

57. This is not to underplay the need for platform accountability, particularly the unchecked concentration of power in the hands of big tech companies, however, to suggest that alternate models, which place human rights at the center, exist and should be adopted. The '*Manila Principles on Intermediary Liability*' offer a good model that balances accountability and state power.[56] Principle 1 states that "*Intermediaries must never be made strictly liable for hosting unlawful third-party content, nor should they ever be required to monitor content proactively as part of an intermediary liability regime*."[57] The

[54] Article 16, 17, Constitution of Pakistan.
[55] "Report on the rights to freedom of peaceful assembly and of association: The Digital Age", Human Rights Council, 17 May 2019, A/HRC/41/41, https://undocs.org/A/HRC/41/41, para. 11.
[56] "Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation," Global Civil Society Initiative, March 24, 2015, https://www.eff.org/files/2015/10/31/manila_principles_1.0.pdf.
[57] Ibid.

2021 Rules place undue burdens on Service Providers and Social Media Companies to proactively monitor and regulate content. As discussed above, the Rules also violate Section 38 of PECA which deals with limitation of liability of service providers.

58. Furthermore, the Global Network Initiative (GNI), a multi-stakeholder platform consisting of information and communications technology (ICT) companies, human rights and press freedom organizations, academics, and investors, provides principles and guidelines on content moderation and intermediary liability based on human rights principles. The 'GNI Principles', along with constant advocacy with companies by the nework, seeks to ensure that free speech "*restrictions should be consistent with international human rights laws or standards, the rule of law and be necessary and proportionate for the relevant purpose*".[58] Many of the Significant Social Media Companies that the 2021 Rules seek to regulate are part of the GNI and regularly engage with recommendations made by the forum.

59. In its analysis of laws and legislative proposals from across the world, including Pakistan's "*The Citizens Protections (against Online Harms) Rules, 2020*", GNI has found that "there are no off-the-shelf solutions to complex regulatory challenges".[59] This is particularly true when clubbing various forms of content into the same regulatory mechanisms. For instance, there is international consensus and cooperation on the model of prior restraint when dealing with child pornography, however, the same model can not be applied to content such as defamation and hate speech which requires context and content specific decisions while balancing free speech concerns.

60. The "*Santa Clara Principles on Content Moderation*", which have been endorsed by a variety of civil society actors and companies, balance the need for content moderation with principles of due process.[60] For instance, Principle 1 emphasises the importance of clear, transparent and effective processes in content moderation. Automated process *"to identify or remove content or suspend accounts, whether supplemented by human review or not, [should only be used] <u>when there is sufficiently high confidence in the quality and accuracy of those processes</u>.*"[61] Laws that push social media companies and service providers to be more transparent about their processes should be welcome. The *Santa Carla Principles* require that companies take measures to ensure that all users of their platform understand their policies and community guidelines, thus, obligations requiring companies to invest more in user awareness would be permissible under these Principles.

[58] "The GNI Principles", Global Network Initiative, https://globalnetworkinitiative.org/gni-principles/.
[59] "Content Regulation and Human Rights", pg. 6.
[60] "Santa Clara Principles on Content Moderation", https://santaclaraprinciples.org/.
[61] "Santa Clara Principles on Content Moderation", Principle 1.

61. When dealing with subjects such as national security in Pakistan there is little definitional clarity or jurisprudence from a human rights perspective to provide a reliable standard for regulators. It would be beneficial to use international standards such as the '*The Johannesburg Principles on National Security, Freedom of Expression and Access to Information'*, provide that "no restriction on freedom of expression or information on the ground of national security may be imposed unless the government can demonstrate that the restriction is prescribed by law and is necessary in a democratic society to protect a legitimate national security interest. *The burden of demonstrating the validity of the restriction rests with the government*."[62] Principle 2 states that "*a restriction sought to be justified on the ground of national security is not legitimate if its genuine purpose or demonstrable effect is to protect interests unrelated to national security, including, for example, to protect a government from embarrassment or exposure of wrongdoing, or to conceal information about the functioning of its public institutions, or to entrench a particular ideology, or to suppress industrial unrest*."[63]

62. In France, an interim mission report submitted to the French Secretary of State for Digital Affairs titled "*Creating a French framework to make social media platforms more accountable: acting in France with a European vision*" suggested that any model of regulation should:

> "*(1) respect the wide range of social network models, which are particularly diverse, (2) impose a principle of transparency and systematic inclusion of civil society, (3) aim for a minimum level of intervention in accordance with the principles of necessity and proportionality and (4) refer to the courts for the characterisation of the lawfulness of individual content*."[64]

63. In its report titled '*Watching the watchmen Content moderation, governance, and freedom of expression'*, the organization 'Article 19' found that current proposals for online content regulation "effectively demand that companies police human communications and decide what speech is 'illegal' or 'harmful'. This is deeply problematic as *only the courts can determine illegality* and *different types of content may well call for different types of regulation*; the solutions used to deal with child-abuse

[62] The Johannesburg Principles on National Security, Freedom of Expression and Access to Information, Principle 1(d).
[63] The Johannesburg Principles on National Security, Principle 2(b).
[64] "Creating a French framework to make social media platforms more accountable: Acting in France with a European vision," Submitted to the French Secretary of State for Digital Affairs, May 2019, https://www.numerique.gouv.fr/uploads/Regulation-of-social-networks_Mission-report_ENG.pdf

material may not be appropriate to deal with disinformation or copyright."[65] The report laid down the following guiding principles for online content moderation:

> "*1. States should refrain from unnecessary regulation of online content moderation.*
> *2. Overarching principles of any regulatory framework must be transparency, accountability, and the protection of human rights.*
> *3. Conditional immunity from liability for third-party content must be maintained, but its scope and notice and action procedures must be clarified.*
> *4. General monitoring of content must continue to be prohibited.*
> *5. Any regulatory framework must be strictly limited in scope. Regulation should focus on illegal rather than 'legal but harmful' content. Private-messaging services and news organisations should be out of scope. Measures should not have extraterritorial application.*
> *6. Obligations under any regulatory scheme must be clearly defined. These include, in particular, transparency obligations and internal due-process obligations.*
> *7. Any regulator must be independent in both law and practice.*
> *8. Any regulatory framework must be proportionate.*
> *9. Any regulatory framework must provide access to effective remedies.*
> *10. Large platforms should be required to unbundle their hosting and content-curation functions and ensure they are interoperable with other services*."[66]

64. Lastly, self-regulation models on content moderation need to be strengthened through the development of strong incentive structures and co-creation on part of the government. The Government's current approach towards Service Providers and Social Media Companies is a "carrot-and-stick" approach with platforms negotiating with the Government in order to avoid blanket bans that have either been threatened or imposed. The Government would benefit from a tiered approach towards content moderation, with some types of content, such as child pornography, requiring direct intervention and other types of content, such as "fake news", being subject to self-regulation where platforms are incentivized to invest in local fact-checkers. Other incentive structures that address structural issues can include requiring more investment in local content reviewers, both in terms of context and local languages. Companies can be required to meet transparency

[65] "Watching the watchmen: Content moderation, governance, and freedom of expression," Article 19, 2021, https://www.article19.org/wp-content/uploads/2021/12/Watching-the-watchmen_FINAL_8-Dec.pdf, pg. 8.
[66] "Watching the watchmen Content moderation", pg. 26-37.

requirements regarding content moderation and provide adequate appeal mechanisms to users.

65. Self-regulation models have been successful in some cases. Article 19 notes in its report[67] that the European Commission found in its evaluation of the "EU Code of Conduct on Countering Illegal Hate Speech" that "IT companies are now assessing 89% of flagged content within 24 hours and 72% of the content deemed to be illegal hate speech is removed, compared to 40% and 28% respectively when the Code was first launched in 2016."[68] It also found that in France "the speed of deployment and progress made during the last 12 months by an operator such as Facebook show the benefits of capitalising on this self-regulatory approach already being used by the platforms, by expanding and legitimising it."[69]

66. Meta's (formally Facebook) Independent Oversight Board is another model for self-regulation and example of the implementation of the *UN Guiding Principles*. Formed in 2020, the Board consists of 40 members, including lawyers, activists, academics and policy experts, often referred to as the "supreme court" of Meta/Facebook, whose judgments on content moderation are binding on the company.[70] The Board uses the international human rights framework as a basis for its judgments, creating caselaw and precedent for the company.[71] The Board's judgments have resulted in changes in policy and practice. The Board's decisions have also resulted in a ripple effect for other platforms; for instance, Twitter has adopted the Board's jurisprudence with regards to misinformation on Covid-19 and the vaccine,[72] that have employed standards and rulings issued by the Board.

Recommendations

67. It is recommended to the Honourable High Court that:
    a. The current framework being used by the Government to approach content moderation is fundamentally flawed and thus the 2021 Rules should be immediately denotified and that the Government reengage with the question of

[67] "Watching the watchmen," pg. 19.
[68] "Countering illegal hate speech online – EU Code of Conduct ensures swift response," *European Commission*, February 4, 2019, https://ec.europa.eu/commission/presscorner/detail/en/IP_19_805.
[69] "Creating a French framework to make social media platforms more accountable," pg. 11.
[70] "Oversight Board Charter," Meta Oversight Board, September 2019,https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf.
[71] "Bylaws," Meta Oversight Board, November 2021, https://about.fb.com/wp-content/uploads/2020/01/Bylaws_v6.pdf.
[72] Davey Alba, "Twitter Permanently Suspends Marjorie Taylor Greene's Account," The New York Times, January 2, 2022,https://www.nytimes.com/2022/01/02/technology/marjorie-taylor-greene-twitter.html.

online content moderation and regulation of social media through a completely new framework that places human rights standards at the center.

b. Restrictions on online content should be the exception, not the norm. This requires that unlawful content should be narrowly-defined in line with the tripartite tests required under Article 19 of the ICCPR. The categories of unlawful content need to be narrowed, removing vague language on morality and national security, and replaced by precise definitions.
c. Heed be paid to the human rights implications of Section 37 of PECA and for it to be struck down given the wide and unfeterred powers it accords for content moderation.
d. Decisions of cotent moderation should be taken by an independant and competent judicial body by divesting the PTA of its powers to make determinations on freedom of expression and access to information.
e. A new, tiered framework for content moderation should be co-created through a multi-stakeholder consultation process that is led by civil society, digital rights organisations, journalists/media and other civic stakeholders.
f. Different content moderation mechanisms be developed to deal with various cateogries of content such as child pornography, hate speech, incitement to violence. While more focus on self-regulation be adopted for content such as misinformation, disinformation, defamation, etc.
g. Transparency requirements be placed on the PTA regarding its regulation of content; currently, there is no way of knowing what content has been removed and the reasons behind the content restriction. It is recommended that any Rules proposed by the Government should contain provisions compelling the PTA to maintain a publicly available registry of content (URLS, applications, platforms) it removes along with the reasoned order mandating removal. In case the order is overturned and content is restored, the date of restoration and period of removal should be updated in the index.
h. More focus should be placed on long-term interventions such as user awareness and digital literacy as some issues cannot be solved solely through regulation. For instance, hate speech exists in other mediums which have been around for a while, such as electronic and print media and laws have failed to adequately address the issue.