

# **LEGAL LOCK JOURNAL**

## **2583-0384**

---

**VOLUME 4 || ISSUE 3**

---

**2025**

This Article is brought to you for “free” and “open access” by the Legal Lock Journal. It has been accepted for inclusion in the Journal after due review.

---

To submit your Manuscript for Publication at Legal Lock Journal, kindly email your Manuscript at [legallockjournal@gmail.com](mailto:legallockjournal@gmail.com).

## CONTENT MODERATION IN THE DIGITAL AGE: STRIKING A BALANCE BETWEEN RIGHTS AND RESPONSIBILITIES UNDER THE INDIAN CONSTITUTIONAL FRAMEWORK

Sowbaakhya Y<sup>1</sup>, Angelin Subiksha<sup>2</sup> & R.S.Krishna Kanth<sup>3</sup>

### Abstract

India is witnessing a boom in technological advancement. Along with that boon comes the bane of threats connected to it. While certain problems associated with technological advancement such as job displacement are well-known, other concerns such as technologies that intervene in people's daily lives and acts as an impediment in exercising their rights remain un-renowned. Content moderation is one such development. The practice of scrutinising user-posted content on Social Media Intermediary (SMI) platforms to ensure it complies with the company's standards, policies, and guidelines is known as content moderation. Any content that is found to violate their policies—such as offensive, dangerous, unlawful, defamatory, etc., may result in the suspension or removal of the user account or post. The content moderation can be done through manual process, automated techniques, machine learning such as AI methods. India enacted the Information Technology (Intermediaries Guidelines) Rules, 2011 that governed the content moderation later it was amended as the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. It prima facie appears to be an effective method of surveilling posts that are unethical, and has no serious detrimental effects. But when examined meticulously it is evident that it fosters an atmosphere that curtails basic liberties protected by articles 19(1)(a) and 21, respectively, such the freedom of speech and expression and life. Therefore, even if it is vital and effective in today's high-tech culture, a serious note of this contravention needs to be taken into consideration. The Constitution is the fundamental/supreme law of the land and no infringement of it should be tolerated. But in some exceptional cases it becomes necessary to accord the recently passed laws, which keep up with technological advancements, a modicum of importance. It does not imply that the constitution must be disregarded, rather, if there appears to be any inconsistency, a harmonic construction between

---

<sup>1</sup> The author is a student of law at Government Law College, Vellore.

<sup>2</sup> The Co-author is a student of law at Government Law College, Vellore.

<sup>3</sup> The Co-author is a student of law at Government Law College, Vellore.

the new act and the constitution must be made. This research paper aims to highlight the ways in which content moderation and the Intermediary Rules, 2021 is effective and helpful and analyse if its implementation in any way breaches the rights given under the constitution and if it violates the constitutionally guaranteed rights, the research paper suggests neutralised solutions that helps in maintaining the checks and balances that prevent one from dominating another.

**Keywords**

Content Moderation, Social Media Intermediary, Intermediary Rules, Constitution, Fundamental Rights.

**Introduction**

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 is a bold move on part of India for moderating content and the public accountability of the Social Media intermediaries. Though this regulation brings private sector liabilities there is a thin line between these liabilities and the state restrictions. The Constitution of India has guaranteed freedom of speech and expression as a fundamental right in Article 19(1)(a) it includes one's right to express their opinions in writing, posting pictures or in any other forms by using social media intermediaries as a medium of communication. Further, UDHR assures the right to freedom of speech and expression under Article 19.

Regulations affecting online intermediaries must be extremely carefully studied in light of their potential impact on users' freedom of speech due to the complexities of the relationship between governments, its people, and private social media platforms. Ultimately in this paper we would discuss the intermediary rules and how by balancing the rights of the people and the responsibilities of the state there is hope that checks and balances could be maintained by adopting a few changes and would be able to clear the impediments in exercising the people's Constitutional Rights.

**Content Moderation**

The process of examining and keeping an eye on user-generated content on websites to make sure it complies with rules and regulations is known as content moderation. The content moderation is frequently used by online platforms that largely rely on user-generated material, including social networking sites, e-commerce platforms, the sharing economy, online dating, and forums. The types of Content Moderation are given below.

1. Manual moderation
2. Automated moderation
3. Artificial Intelligence Moderation
4. Commercial Content Moderation

### **Background of IT (Intermediary Guidelines Digital Media Ethics Code) Rules, 2021**

The IT Act, 2000 was India's first step towards regulating internet intermediaries. It established new restrictions for entities that receive, store, or send communications or provide related services.<sup>4</sup> A significant shortcoming of the initial structure was the absence of appropriate safeguards for intermediary responsibility. This meant that the platforms may be held directly and strictly liable for some damaging information uploaded by its users.<sup>5</sup> The issues faced by the intermediaries were first recognised in the ruling *Shreya Singhal v. Union of India*<sup>6</sup>, wherein the judiciary acknowledged the pragmatic limitations of reviewing all user-generated content that are hosted by the intermediaries.

In 2008, the IT Act was modified to give intermediaries greater safe harbour protection for content for which they served as a conduit, under section 79 providing that they used the necessary "due diligence" against problematic content.<sup>7</sup> But a series of regulations were enacted in response to these reforms, giving the government more authority to regulate the flow of information online by requiring intermediaries to censor or remove user content.

Further strengthening these capabilities, the 2011 Intermediary Guidelines mandated that certain inappropriate information be removed within twenty-four to thirty-six hours after receiving a complaint.

### **IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021**

The word "Social Media Intermediary," who falls within the category of intermediaries, was included in the Rules. SMI are defined by the 2021 Rules as those who facilitate online connection between two or more users exclusively.

---

<sup>4</sup> Information Technology Act 2000, s 2(w).

<sup>5</sup> *Avnish Bajaj v State* 2008 SCC OnLine Del 688.

<sup>6</sup> *Shreya Singhal v Union of India* (2013) 12 SCC 73 [116]

<sup>7</sup> Information Technology Act 2000, ss 79(1), 79(2)(c).

SSMIs (Significant Social Media Intermediaries) are defined as intermediaries having more registered users than a certain notified threshold.<sup>8</sup>

The 2011 Rules intended to provide more clarity and structure to the intermediary provisions given in the IT Act, 2000. However, it was superseded by the Intermediary Rules, 2021. Some of the prominent features of the Rules are discussed below.

The 2011 rules already stipulated the due diligence that the intermediaries had to perform, but the new regulations strengthened their effectiveness by making some amendments. Rule 3 enumerates the information that intermediaries are not permitted to host, display, upload, alter, transmit, publish, update, or share. Additionally, the term 'grossly harmful' which was criticised for being ambiguous has been removed. Rule 3(b)(ix) and 3(b)(ix) have been added, along with two clauses that prohibit the intermediaries from carrying out the aforementioned activities if the information contains any software viruses or other computer code that renders the computer resource unusable or if the information contains patently false information intended to deceive anyone for financial gain.

One of the key requirements of the Intermediary Rules is the setting up of a Grievance Redressal Mechanism. A Grievance Officer must be appointed by the intermediaries, who will be responsible for addressing user complaints regarding objectionable content, fake news, hate speech, and other violations of the law. The grievance officer's name, his contact information, and the procedure for filing a complaint must be made public. When a complaint is received, it must be acknowledged by the Grievance Officer within 24 hours, resolved within 15 days, and given a distinct tracking number for future reference.

The following additional due diligence is required to be followed by SSMIs as per Rule 4(1):

1. the appointment of a resident grievance officer
2. a nodal person for law enforcement agency coordination, and
3. A chief compliance officer, all of whom must reside in India to make sure adherence to the Act and Regulations.

As given in Rule 4(2) an SSMI, which mainly offers messaging services, needs to make it possible to identify the originator of content on its platform within India. A court order or an order from the appropriate authority under the Information Technology Act may insist on

---

<sup>8</sup>PRS Legislative research, (25 February 2021) 'The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021', <<https://prsindia.org/billtrack/the-information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021>>

this. Orders of this type would be issued for certain reasons, such as the prevention, detection, and investigation of specific offenses, such as those connected to public order, national security, and sexual assault. But if the originator could be located using less invasive methods, such requests would not be granted.

In order to identify the (i) content that shows child sexual abuse and rape; or (ii) information that is exactly the same as the information that was previously restricted by a court or government order, SSIMs should adopt technology-based techniques as per Rule 4(4) subject to the provisos given there under.

Due diligence of intermediaries are to be specified in the service agreements or policies of that site regarding class of content which aren't authorised to post or share if they are posted, they would be taken down within thirty-six hours upon the issuance of order by government or court. SSIM assists agencies that enforce law and retains restricted content and related records for ninety days and also provides GRM for affected users. The Rules prescribe online publishers such as online papers, news portals, aggregators and agencies and curated videos, a three tier mechanism. They are:

1. Self- regulation by publishers.
2. Self-regulation by publishers associations.
3. Oversight by the Union government.

Other intermediaries who will not come under the definition of SSIM can also be asked by the MeitY to comply with all the obligations mentioned in Rule 4 for the reasons that must be recorded in writing.

### **Problems associated with the Impugned Rules and its Amendments**

The 2021 Rules provides a Grievance Redressal Mechanism under which, by establishing a Grievance Appellate Committee or committees, which are constituted by an executive. The Union Government will take over as the arbiter of what constitutes appropriate speech on the internet, rather than an independent judicial or regulatory agency.<sup>9</sup> The government may encourage social media intermediaries to censor any speech that it deems objectionable. When it comes to granting orders for content removal or access denial, there is insufficient judicial scrutiny. Online platforms become political battlefields where

---

<sup>9</sup> Tejasi Panjiar, Prateek Waghre, 'A public brief on the IT Amendment Rules, 2022 a.k.a 'how the government is trying to moderate online speech'', (Internet Freedom Foundation, 10 November 2022), <<https://internetfreedom.in/public-brief-on-the-it-amendment-rules-2022/>>

any government agency with a formal order can demand the removal of whatever content it thinks appropriate without going through a court process.

There are questions about the independence of the Grievance Appellate Committee and the capacity of the government to sway decisions regarding content moderation in an opaque way because there are no facts available about how the GAC(s) operate or choose appeals for review.

In less than 72 hours after being requested by any government agency or law enforcement, the guidelines direct intermediaries to provide information under their possession for identity verification or prevention, detection, investigation, or prosecution. The right to privacy of all users that is enshrined as fundamental right in Article 21 of the Constitution of India is undermined by these rules by disclosing users' identities to government agencies, in the digital sphere.

The transparency of the regulations' criterion for identifying content deemed to be fake or objectionable by committees or the government is lacking. Following these guidelines could have a chilling effect on free speech and deter individuals from expressing diverse opinions.

The 2021 framework offers guidelines and a supervisory system to control the material for Online Publishers. The 2021 Rules establish guidelines for content control for online news and current affairs publishers including curators of audio-visual material (podcasts, series and films). The publishers of print media regulated under the Press Council Act of 1978, the Press and Registration of Books Act of 1867 etc, enjoy more freedom in expressing their views than the online publishers of news and current affairs as there is intervention of the government in the latter.

The 2021 Rules further limit the types of content that intermediate platform users are allowed to create, upload, and share. These include content that is invasive of another's privacy, insulting towards gender, and knowingly and deliberately communicating any information that is obviously false or misleading but could be reasonably interpreted as fact. These grounds are ambiguous and overly restrictive.

SSMIs if mandated by an order of court or competent authority must permit the identification of the information's original source or originator when it comes to issues

pertaining to India's sovereignty, public order, friendly relations with other countries, incitement to commit crimes, sexually explicit or child abuse material, prevention, prosecution, detection, investigation, or punishment.

This is the most contentious part of the guidelines, as stated in Rule 4(2), essentially cracking the end-to-end encryption used by services like Signal and WhatsApp to identify the information's original source. The Indian government has put safeguards in place to ensure that this provision is only used as a last resort and that an order won't be issued when there are less intrusive ways to identify the original author. Additionally, intermediaries would only need to reveal their identity; they would not be obliged to reveal the communications' content or any other information about the sender. This is being contested by Facebook and Whatsapp in the cases *Facebook Inc vs Union of India*<sup>10</sup>, *WhatsApp LLC v. Union of India*<sup>11</sup> and its submission before the High Court of Delhi that it will cease its operation in India if told to break encryption made headlines. This rule infringes internet users' basic right to privacy and casts doubt on the ruling made by the Honourable Apex Court in *K.S. Puttaswamy vs Union of India*<sup>12</sup>. Another controversial point is, as per the rules, if the original originator is based outside of India, the originator within India will be deemed to be the first originator. Recently, the Supreme Court has transferred all the petitions challenging IT Rules 2021 to the Delhi High Court.<sup>13</sup>

The impugned Rules established a fact-check unit under Rule 3(1)(b)(v). The rules require the SMI to take "reasonable steps" to ensure that they did not host any content about the "business of the central government" that the Union government's fact-checking unit assessed to be "fake, false, or misleading." This gives the Union government an arbitrary power to have control over the things being shared about it and by using this power the Union government may curtail the Freedom of Speech of people. The establishment of the fact-check unit was challenged before the Bombay High Court in *Kunal Kamra vs Union of India*<sup>14</sup> and the Divisional Bench of the Court delivered a split verdict on the case. Later, Justice Chandurkar rejected the plea to stay. Aggrieved by the decision, the petitioners approached the Apex Court and the Supreme Court bench, finally stayed the Union government's notification establishing the fact-check unit.

---

<sup>10</sup>*Facebook Inc vs Union of India (WP(C) 7281/2021)*

<sup>11</sup>*WhatsApp LLC v. Union of India (WP(C) 7284/2021)*

<sup>12</sup>*K.S. Puttaswamy v. Union of India, (2017) 10 SCC 1*

<sup>13</sup> TP (C) 100-105/2021

<sup>14</sup>*Kunal Kamra v. Union of India (SLP(C) No. 6871-6873/2024)*



**Suggestions**

- A. Strike a balance between hindering the spread of dangerous content and defending the fundamental rights.
- B. Make sure that there is accountability and transparency in the standards and procedures of content moderation in fact-checking and grievance redressal. Mandate transparency reports from intermediaries that list the quantity of government requests for user data and content removals.
- C. Look into strategies to locate the source/ originator of a problematic content without completely jeopardising encryption.
- D. In order to guarantee fairness and balance, include stakeholders in the development and assessment of the guidelines, such as representatives from the industry, civil society organisations, and legal experts.
- E. Consider less intrusive tactics like age-gating, content warnings, or restricted visibility for content deemed objectionable instead of content takedowns.
- F. Establish independent fact-checking organisations with diverse representation to reduce the lack of transparency and ensure unbiased evaluations. Separate budget should be allotted for the Fact check unit to manage its expenses.
- G. Ranging the liabilities of the intermediaries through the legislation would help SMIs and SSIMs to be at ease as they are given immunity through the IT, 2000 parent act.

**Conclusion**

Crucial measures for content regulation in India's digital landscape have been established by the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. But there are still issues that need to be resolved concerning fact-checking, government control, privacy, and grievance redressal procedures. Resolving these difficulties calls for independent fact-checking efforts, improved privacy and freedom of speech protections, a clearer understanding of government functions, and improved grievance mechanisms. By prioritising transparency, accountability, and other related aspects, India can effectively traverse the intricacies of content moderation and establish a dynamic online environment that protects user safety, freedom of speech as well as enable the government to discharge its duties.