State Interest in Content Governance *Through* Platforms

By Simran Agarwal
Doctoral Candidate
LabEx ICCA, Université Sorbonne Paris Nord

In India, content on media and technology platforms is, on the one hand, regulated by the state architecture like legislative acts, executive orders, and judicial judgements; and on the other, moderated and informed by platform policies, design, and algorithms. This dual approach to the issue of content governance is central to scholarly work in the field of platform governance as proposed by Tarleton Gillespie, principle researcher at Microsoft Research. However, this landscape of governance is marred with diverse interests and imbalances of power, which are reflected in various regulatory actions.

This article suggests that policy discussions should consider the value of studying content governance by the state, *through* platforms. This approach steps away from the specific functionalities of private platform policies, and sheds light on the insidious regulation of content by the state. Here, we look at the example of how the Indian state employs legislative instruments and executive actions that compel platforms to undertake privatised content regulation.

Governance *through Platforms*

In order to understand how the state achieves content regulation *through* platforms, this piece intends to introduce the reader to the Indian regulatory framework that governs public speech and, simultaneously, take an analytical lens to see how the same regulatory milieu is used by the state to gain control and prompt privatised regulation of content *through* platforms.

The Indian regulatory framework consists of a wide array of legal and statutory measures that came into existence through colonial era precedents and constitutional deliberations. These legislative acts and codes define, control, and place accountability on the limits of acceptable speech in the offline and online arena.

The Indian constitution places limits upon freedom of expression under the sub-clause 19 (2)(a) which lists unacceptable, albeit vaguely defined, speech as against "interests of the sovereignty and integrity of India, the security of the State, friendly relations with foreign States, public order, decency or morality or in relation to contempt of court." These constitutional limitations underly subsequent legal restrictions for online and offline speech in India. ~~constitutional limits~~ The **Indian Penal Code (1860)** (IPC) further criminalises seditious, defamatory, or libellous content**.** The IPC is the official criminal code which was devised during the British colonisation of India and has been evoked to **hold individuals on platforms liable** for speech that falls in above categories. Interestingly, the criminal status of these offences has since been revoked in the United Kingdom but continue to be in practice in India.

Speech is also controlled through limits on collective assembly and exchange in the physical and digital spaces through the **Section 144 of the Indian Criminal Procedure Code (1973)**. This section, another colonial legacy, intends to control rioting and public gathering. Since then, its interpretation has expanded to include control of assembly and exchange between individuals in an online community or groups on platforms. The Indian state, and its enforcement bodies, have enacted Section 144 to **assign liability upon platform users for checking the circulation of unlawful online content.** This is visible in cases where WhatsApp group administrators were deemed responsible for undesirable content shared by the members of their group.

Furthermore, online speech is controlled through access blocking or internet shutdowns authorised through the **Temporary Suspension of Telecom Services (Public Emergency or Public Safety) Rules (2017)**. These rules fall under the Indian Telegraph Act (1885), another colonial era law, which bares clauses (Section 5) granting the state powers to seize and intercept messages shared through operators that offer wireless telecom services. The above rules are enacted to **compel internet service providers (ISPs) to shut down internet access in a geographical region**. While the rules presuppose public emergency or safety for application, India has the highest number of internet shutdowns, totalling 674 cases, in the last decade, with the majority enacted during political instability and public protests against the state. The exploitation of power is further accentuated by the fact that ISPs risk losing their operating license in case of non-compliance, and that the state refuses to maintain a transparent log of these shutdowns.

Relatedly, the state can block access to specific online speech on the basis of the **Indian Technology Act (2000)** (IT Act). This law and its subsequent amendments cover a large number of intermediaries involved in the digital technology ecosystem which can be directed by the state to takedown specific content. Particularly, **Section 69 of the IT Act** (added through an amendment in 2008) blocks public access, upon request from the government, to content that is against the sovereignty and integrity of India, defence of India, security of the State, friendly relations with foreign States or public order or for preventing incitement to the commission of any cognizable offence. The state regularly invokes Section 69A of the IT Act, in order to prompt two forms of content moderation actions by platforms **– 1) taking down of websites and specific pieces of content, and 2) deplatformization of users and applications within geographical boundaries**. Further, the state mandates that content takedown requests from the government be kept confidential. Interestingly, this particular section of the law was upheld by the Supreme Court of India, even while another section (66A) that criminalised 'grossly offensive' and 'menacing' speech of individual users online was revoked in a landmark judgement in 2015. In this case, the judiciary identified Section 66 A as unconstitutional as it infringed the fundamental right of free speech but upheld Section 69A despite the widespread blocking power it allows the state. Moreover, despite this striking down, Section 66A has been repeatedly used by the Indian police to file cases relating to online content.

The same IT Act includes various provisions related to Intermediary Liability in **Section 79** (amended several times). This section evolved along the principles of Section 230 of the Communication Decency Act of the USA to grant immunity or 'safe harbour' to intermediaries over matters of content. However, the Indian state has repeatedly reserved the right to amend rules of liability exemption, protocols for content takedowns, and penalties in the absence of observing "due diligence". The law uses the term 'due diligence' and non-compliance to any of the guidelines under this section which can result in imprisonment or loss of immunity for intermediaries. However, the state interprets the contours of 'due diligence' to **mandate proactive monitoring and algorithmic moderation of content, expeditious take down of content, amend their design policies to enable traceability, and relinquish control over content to the state**.

Under the Section 79 of the IT Act, the state recently introduced the **Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.** These rules were enacted through executive action and not passed through legislative procedures and are being contested by several associations in India. These rules allow the Indian state to lay down exemption from liabilities so long as the **platforms comply with blocking and takedown orders, reveal first originator (first sender of a message) information to state, and strictly self-regulate in ways prescribed by the rules,** thus binding platforms to toe the line. The enaction of these rules without adequate legislative procedures highlights the state's attempts at expanding its control over platforms.

Increasing control over Content Governance

In addition to using legislative instruments to assign culpability, enforce action, and coerce platforms to undertake moderation through automated and transparent means, the state undertakes executive decisions to gain further control over decision making.

First, it includes the Indian policy attempts to curb specific content online, such as misinformation and hate speech, by pushing platforms to undertake more action through automation, literacy, and design. This was evident in the state pushing WhatsApp to revise their policies to curb the spread of misinformation in India. The pressure from Indian ministries led to WhatsApp changing its design policy of labelling forwarded messages with the 'Forward' tag and capping the number of times a message can be forwarded by the same user to five in India. This change was implemented worldwide after a year.

Second, the Indian state, particularly the Ministry of Home Affairs launched a citizen flagging programme named **Cyber Crime Volunteer Programme (2021)** under which the state appoints citizen volunteers to surveil and flag undesirable content on platforms to the government. This programme **endows selected citizens with unfettered power to surveil platforms and flag content** that state vaguely defines as 'unlawful' and 'anti-national'. The implication here is that content flagged by volunteering citizens allows the state to alternatively use **coercive police action to take down of user content on platforms**.

Third, the Indian state is a **signatory to the international statement on End-to End Encryption and Public Safety** involving the following states - UK, Australia, Canada, India, Japan, New Zealand and the United States. This statement guarantees the state backdoor access to encrypted content under the implication of public safety. Here, the Indian government **demands to gain access to encrypted content through backdoor methods** under the presupposition of public safety. This intention is also reflected in the state's demand for traceability on WhatsApp which has resulted in Facebook suing the Indian government. Further, the state cyber security and monitoring body has directed all Virtual Private Network (VPN) providers in India to store, and share upon request, consumer data**.**

Finally, the state also regulates content through **state sponsorship, financial and political, to domestic platforms**. These include homegrown alternatives such as Koo (micro blogging platform launched in 2020) and ShareChat (non- English social media and messaging platform launched in 2015). These platforms have received political support in the form of important political leaders joining the platforms since their launch. State sponsorship also came in the form of monetary support when Koo won the state launched innovation competition in line with the state visions of "Self-Reliant India" and "Digital India". Further, the state support to these platforms could also be seen as an outcome of tussle with international platforms over content decisions (such as Twitter and Facebook). Thus, financial and political support to alternatives **threaten the reach of international players and suggests a degree of state influence over content moderation decisions**.

These governance mechanisms have allowed the state to effectively and erratically regulate content in the offline and online space. The above discussion throws light on how the state employs legislative instruments to coerce platforms into taking action, as well as executive instruments to extend its control over content governance in India. This raises the question of underlying intentions and interests when these legislative and executive measures are implemented.

Why does the Indian state govern *through* platforms?

One could make the argument that legislation and policies are efficient tools for ensuring adequate corporate governance and are needed to curb violence facilitated by online extremism, misinformation and platforms' control over the public sphere. However, the ambiguous remit and uneven implementation of these statutory and executive governing mechanisms creates the ground to assess the motivations, accountability, and interests of the state in content governance. It initiates the question of 'Why does the state govern through platforms?' and leads us to some conclusions and considerations of the state acting in its own interests.

The Indian state uses laws and executive powers as an apparatus to achieve the **interests of sovereign power, such as avoiding criticism and dissent of hegemony, and ceding power to transnational corporations**. This claim can be supported by the instances of the state compelling internet service providers to shut down internet access. Despite the economic and reputational losses of such actions, the state acts through excessive and legal control. Further, it shapes and prompts private moderation of content, so that the visible governing of content appears to be done by the platforms. This allows the state to **act with impunity and limited accountability towards platforms, citizens, and itself**. This again is evident from the state avoiding and transferring the responsibility of maintaining transparency to platforms, while at the same time legally binding them to keep content takedown requests from the state confidential. Finally, these actions of the state reinforce the ruling government's dominant ideology and **perform the discursive function of presenting the state as upholding national security and public order,** thereby allowing it to appear as a powerful sovereign acting in the public interest through legal safeguards against platforms.

Therefore, the discussions in this article present an important analytical frame of 'governance *through* platforms'. It aims to distinguish and emphasise the ways in which the excessive state power over content is achieved. This approach focuses on the instruments used by the state as well as its underlying interests and can offer valuable insight into the dynamic interaction of power between states and platforms in India, and elsewhere.