

# Let's think global: Social Media Commission – A Federated Model for Governance

*Kamesh Shekar<sup>1</sup>*

## Abstract

The issues of hate speech, disinformation, misinformation, and polarisation on one side, and arbitrary exercise of power on public discourse and information on the other side, is what I coin as a broken social media problem. The urge to fix social media has mulled governments and agencies to think in the lines of regulating these entities. As a response to it, to combat governmental regulation, social media platforms had spin-off self-regulation models. But over-reliance on the government (as in India and the UK) or firm (as in self-regulation or Facebook's oversight board) for implementing consumer-centric governance is inadequate. So then, what could be the democratically legitimate global structure for social media governance? This paper tries to answer this question by proposing a federated model of governance called the Social Media Commission (SMC). The proposed structure will have international, country-level, district, or regional-level bodies, working in synergy and following a bottom-up approach. This paper will also discuss the federation structure of the SMC advised by input, throughput, and output legitimacy where SMC will have a system, which will be (i) responsive to user concerns through having active participation and representation of the users (input legitimacy), (ii) governed through robust accountability and process transparency mandates (throughput legitimacy), (iii) effective to deliver better consumer outcomes (output legitimacy).

**Submission to:** GigaNet Symposium, Internet Governance Forum 2021

---

<sup>1</sup>The author is a tech policy researcher. He is currently pursuing PGP in Public Policy from the Takshashila Institution. Views are personal and do not represent any organisations. The author can be reached at [kameshsshekar@gmail.com](mailto:kameshsshekar@gmail.com)

# Contents

<b>1. Introduction</b>	<b>2</b>
<b>1.1. A glance at Various Governance Frameworks</b>	<b>2</b>
1.1.1. Government Intervention: Case of India's IT Rules 2021	2
1.1.2. Social Media Intervention: Facebook Oversight Board	3
1.1.3. Association Intervention: A Case of IAMAI Voluntary Code During 2019 Indian Lok Sabha Elections (Lower house)	4
<b>1.2. Democratic Legitimacy Lens</b>	<b>5</b>
<b>2. Social Media Commission Model (SMC) – Structure</b>	<b>6</b>
2.1. Social Media Commission – International Level (SMC-IL)	6
2.2. Social Media Commission – National Level (SMC-NL)	7
2.3. Social Media Commission – Regional Level (SMC-RL)	8
2.4. Conduct Committee	8
<b>3. Social Media Commission Model (SMC) – Functions</b>	<b>10</b>
<b>3.1. Framing and Auditing of Universal Principle</b>	<b>10</b>
3.1.1. Framing Sets of Universal Principles	10
3.1.2. Auditing Universal Principles	15
<b>3.2. Advisory and Advocacy Role to the State</b>	<b>16</b>
3.2.1. Advisory Role to the Monist States	16
3.2.2. Advocacy Role to the Dualist States	16
<b>3.3. Aiding, Supervision and Accreditation Role to the Social Media Platforms</b>	<b>17</b>
<b>3.4. Grievance Redressal Role to Consumers/Entities</b>	<b>18</b>
<b>4. Potential Funding Options for Social Media Commission</b>	<b>19</b>
<b>5. Conclusion</b>	<b>21</b>
<b>Annexure: How Does SMCs Model Work? – Process Flow</b>	<b>23</b>
<b>Bibliography</b>	<b>24</b>

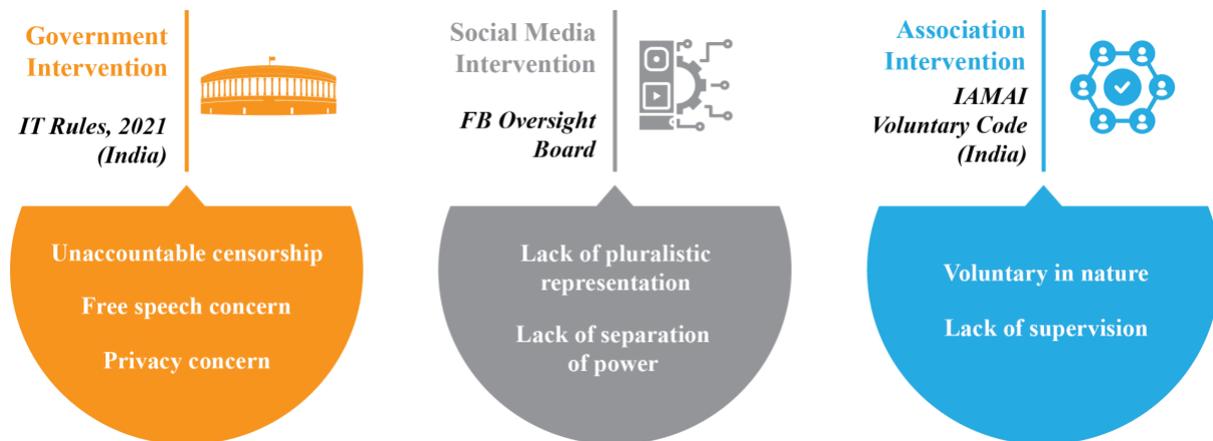
# 1. Introduction

Globally about 56.6% (as of July 2021) (Global Social Media Stats, 2021) of the population actively use various social media platforms to produce tons of information and network across geographical borders. While the internet and the emergence of social media platforms have positively impacted lives of people in terms of connectivity, freedom of speech, mobilising people for a greater cause (for instance, Arab Spring), crisis management etc., it has also opened up various problems. The issues of hate speech, disinformation, misinformation, and polarisation on one side, and arbitrary exercise of power on public discourse and information on the other side, is what I coin as a broken social media problem. Policy actors have taken various initiatives to fix the social media problem, as discussed below.

## 1.1. A glance at Various Governance Frameworks

Since the Cambridge Analytica fallout, various policy actors have recognised the urge to fix the social media problem, including policy experts, the government, and the general public. But over-reliance on the government (as in India and the UK) or social media platforms (as in self-regulation or Facebook's oversight board) for implementing consumer-centric governance is inadequate. This section maps the issues with the existing governance frameworks.

**Figure 1: Governance Framework By Various Stakeholders**



### 1.1.1. Government Intervention: Case of India's IT Rules 2021

The window for social media regulation opened up in India in 2018 when the Indian Ministry of Electronics and Information Technology (MEITY) put out draft intermediary guidelines (The Information Technology [Intermediaries Guidelines (Amendment) Rules] 2018) for public comments. After that, the matter was left untouched until 2021, when political will reopened the issue, and the MEITY recently notified the IT Rules, 2021. Compared to the 2018 draft guidelines, many things were newly added in the IT Rules, 2021, such as digital media and online curated content rules, which brought little to no chance for policy actors to contribute.

As the IT Rules, 2021 were passed as an executive order<sup>2</sup>, policy commentators and advocacy groups have already started pointing out many flaws in the IT Rules, 2021, seeking a more democratic route through legislation. While there are few positive elements to the IT Rules, 2021, negative aspects of the rules outweigh the same. One of the most debated aspects of the IT Rules, 2021 pertaining to social media platforms is part II Rule 4(2), which mandates significant social media intermediaries<sup>3</sup> to identify the first originator of a particular message. While there can be various ways to trace the first originator, in practicality, most of the methods break the end-end encryption (Grover, Rajwade, & Katira, 2021), which gives backdoor entry to every conversation of the consumers (directly or in-kind) to the government (on request), the private entities, and others such as hackers etc. This move stands against some of the universal best practices/principles such as data minimisation, proportionality, privacy by design etc. and some universal human rights and values such as freedom of expression, right to be forgotten, right to privacy etc.

Another contentious aspect of the IT Rules, 2021 (pertaining to social media) is part II Rule 3(1)(d), the government and court can order social media platforms to take down any content which they determine to be unlawful and has to be executed within 36 hours from the notice. As there is no administrative process statute in India, like in the case of the US (Goyal & Sane, 2021), the process and way in which takedown notice is issued can be discretionary, the rationale behind the removal of content, whether it's for collective consumer welfare, political reasons or wrongdoing cannot be determined. Therefore, this can have a chilling effect on freedom of expression. On similar lines, Rule 4(4) of part II mandates significant social media platforms to use technology-based measures to identify unlawful content to take it down. Scholars have pointed out that this can lead to over moderation by the platforms to secure their immunity under section 79<sup>4</sup> of IT Act, 2000 (Devdasan, 2021).

While this is the state of government intervention in India, the UK (Draft Online Safety Bill), Germany (NetDG Law), and other countries<sup>5</sup> have also taken a similar route of regulating social media platforms without weighing the unintended consequences of these actions.

### **1.1.2. Social Media Intervention: Facebook Oversight Board**

Facebook follows its community guidelines as a basis for performing hard moderation<sup>6</sup> of third-party content. The community guidelines followed by Facebook values 'voice' with a limitation of veracity, safety, privacy etc. (Bickert, 2019) and follows counter speech doctrine when the content is hateful but doesn't necessarily violate its community guidelines (Bartlett & Krasodonski-Jones, 2015). While the community guidelines did throw some light onto the process of hard moderation, which remained a black box previously, still from a legitimacy standpoint, making/revising the community guidelines lacked consumer representation and participation, which led to a lack of system countenance. Therefore, due to the negative press

---

<sup>2</sup> As per the section 87 of IT Act, 2000, the central government of India can make rules (delegated regulation) under IT Act, 2000.

<sup>3</sup> Government of India has notified 5 million users are the threshold to determine significant social media intermediary

<sup>4</sup> Provides safe harbor

<sup>5</sup> Various countries such as the US, Australia, New Zealand, Canada, and Japan (including UK and India) are mulling toward breaking end-to-end encryption (at least in the case of Facebook) to enhance online safety (Collins, 2020).

<sup>6</sup> Hard moderation refers to decisions taken by the platforms in terms of content such as takedown, flagging and account such as temporary moratorium/lock decisions, permanent banning etc. (Groove, Binns, & Katzenbach, 2020)

coverage and constant pushback from the users, employees, and governments globally on private speech laws, i.e., community guidelines (Klonick, 2018), Facebook decided to constitute an oversight board<sup>7</sup>. This independent body will hold Facebook accountable for its decisions on hard moderation by taking a final call on community speech within the perimeter of community guidelines.

While the Facebook oversight board stands to be the first practical attempt to shift the governance framework from the national angle to global, still it falls through the cracks. It has been two years since the incorporation of the board; at this moment, the Facebook oversight board crop up to be a symbolic intervention in place of substantive action (Napoli, 2021) where the oversight board's advice is not very well taken up<sup>8</sup>.

The making of the board did involve an amount of consultation with experts from various regions, and preliminary scanning of the board members was done through consumer suggestion (Klonick, 2020); still, the final selection was at the discretion of Facebook, which once again makes this process lack legitimacy in terms of separation of power, participatory process. In addition to this, the oversight board mandates stand within the perimeter of the community guidelines, which still layers away from consumer inputs and put together by Facebook. Over and above this, the Facebook oversight board doesn't account for the pluralistic aspect of the users as a single board can't cover diversity in users and contextual heterogeneity in the issues.

### **1.1.3. Association Intervention: A Case of IAMAI Voluntary Code During 2019 Indian Lok Sabha Elections (Lower house)**

During the 2019 Lok Sabha Elections, the Internet and Mobile Association of India (IAMAI)<sup>9</sup> produced a (Voluntary Code of Ethics by the Social Media platforms for the General Election 2019, 2019). This voluntary code ensured that the platforms gave due diligence to the Election Commission of India (ECI). In September 2019, ECI announced (PIB, 2019) that it would observe the IAMAI Voluntary Code of Ethics during all general elections. While the code has legitimacy as recognised by the ECI (fourth branch regulator), it is voluntary, where social media platforms are not held accountable for non-compliance. Therefore, this code also appears to be more symbolic in a place of substantive action.

This illuminates that gap within analysed intervention bring out sub-optimal outcomes and give birth to various unintended consequences. Besides, analysing the existing interventions also shows that each of the interventions lacks some crucial elements of policy and policymaking, making it less democratically sound and legitimate. Therefore, this paper will be using the below discussed democratic legitimacy framework to advise the social media commission model.

---

<sup>7</sup> Besides, making an independent appellate body was also floated through (Santa Clara Principles on Transparency and Accountability, 2018) and Noah Feldman memorandum (Srinivasan, 2021)

<sup>8</sup> <https://twitter.com/OversightBoard/status/1437434013153640455?s=20>

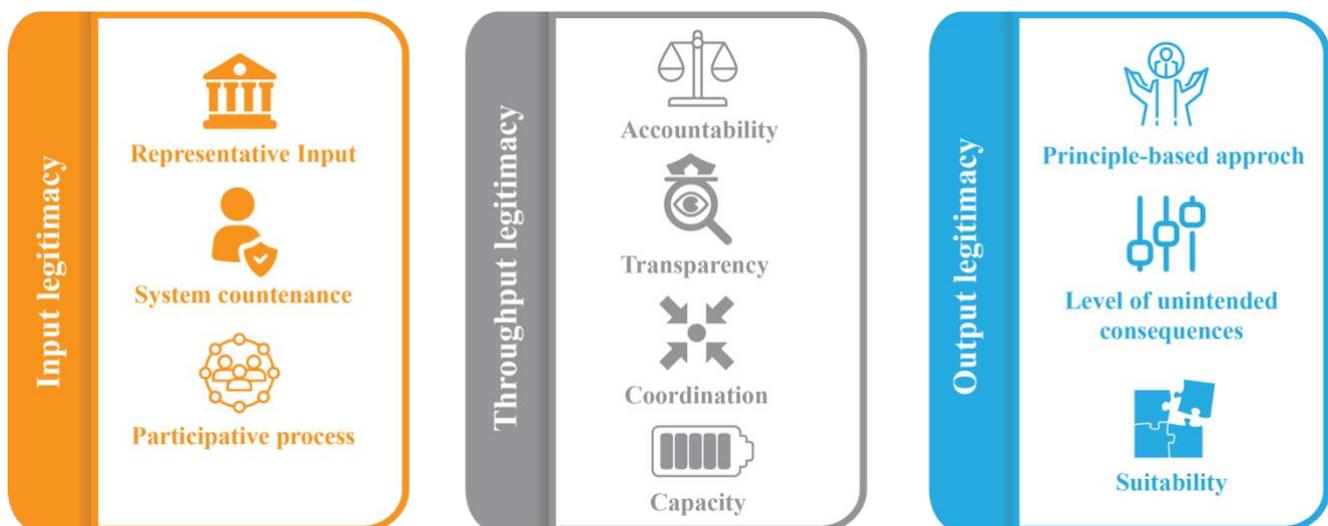
<sup>9</sup> “The Internet & Mobile Association of India (IAMAI) is a not-for-profit industry body registered under the Societies Registration Act, 1860. Its mandate is to expand and enhance the online and mobile value added services sectors. It is dedicated to presenting a unified voice of the businesses it represents to the government, investors, consumers and other stakeholders.” (IAMAI)

## 1.2. Democratic Legitimacy Lens

Democratic legitimacy is a political concept that determines how democratic the systems and functions of the institutions, practices and rules are. While there are various approaches towards studying democratic legitimacy, borrowing the idea from system theory, Schmidt proposes a robust threefold framework – Input, throughput, and output legitimacy (Schmidt, 2010). Schmidt's framework is a modification to Scharpf's taxonomies by adding the missing theorisation of throughput legitimacy. Dividing the various functions of the system, Schmidt's framework looks into the (i) institution and processes (input), (ii) accountability to practices (throughput), and (iii) framed regulations, rules and policies (output).

Therefore, input legitimacy tries to evaluate the responsiveness of institutions to consumer concerns through having active participation and representation of the consumers. Throughput legitimacy tries to analyse the quality of the governance structures. Finally, output legitimacy analyses effectiveness in delivering better consumer outcomes. For the purpose of this paper, specific features under each of the legitimacy, i.e., input, throughput and output legitimacy (refer to figure 2), are considered for advising the social media commission model. The features illustrated below are also extracted from the learning from existing governance frameworks through a meta-analytic literature review.

**Figure 2: Features of Input, Output and Throughput Legitimacy**



In section 2 of the paper, I will be discussing the social media commission model and provide contour to the federated structure advised by features of input legitimacy. In section 3 of the paper, I will outline the functions of the social media commissions within the federated structure. Annexure of the paper will discuss how the social media commission model will work in terms of processes.

## 2. Social Media Commission Model (SMC) – Structure

Social media commission is a multi-level federated global governance body that will bring in a uniform principle-based approach towards tackling the narrative harms<sup>10</sup>. The proposed structure will have international, country-level, district or state-level bodies working in synergy and following bottom-up approach. The social media commissions at the country-level and state-level should be registered as a not-for-profit organisation<sup>11</sup>. At the international level, it should be formed as an international coordination organisation, something similar to intergovernmental organisations like the United Nations. SMC must be an autonomous and independent commission, free from external inferences.

At each level, the commission will have various actors and stakeholders vested with different responsibilities. The following subsections will discuss the composition of the commissions in various levels and the constitution process in detail.

### 2.1. Social Media Commission – International Level (SMC-IL)

The top layer of federated SMC structure is the international level SMC, which would work in tandem with intergovernmental organisations such as United Nations (UN)<sup>12</sup>. SMC-IL shall comprise delegates representing all the SMC-NLs (refer to section 2.2) who choose to be a member. There shall be five delegates<sup>13</sup> from each SMC-NLs; alternative delegates shall be present in the absence of the actual delegates in the commission. The delegates and alternate delegates shall be notified by the SMC-NLs no later than three months from the formation of the commission. Delegates and alternate delegates can hold office for five years<sup>14</sup>, and six months prior to the term expiry, SMC-NLs have to notify incoming delegates. Countries that don't follow the SMC model or are not a member of SMC-IL can have representation in the SMC-IL on a context basis without voting rights.

Every delegate of the SMC-IL will have the voting right (one vote) to elect a chair and deputy chair of the commission through a secret ballot. The chair and deputy chair must have a special majority (two-third votes) in the secret ballot and can hold office for five years, starting from the election date. The deputy chair and chair can't be re-elected consecutively from the same SMC-NL to have diversity in representation.

Besides, as both state and non-state actors (like social media platforms) resort towards content moderation to tackle the issue of narrative harms, SMC-IL must form main committees to look into various facets of content moderation. SMC-IL shall have four main committees: algorithmic moderation committee, human moderator and welfare committee, lateral moderation committee, and state-mandated content moderation committee. Main committees must have at least one delegate from every member SMC-NLs and elect a chair and deputy chair in a similar fashion to the commission election.

---

<sup>10</sup> The term narrative harms refer to a potential for manipulation of information flows in society or an impact on public discourse. (Waghre & Ramprasad, 2021)

<sup>11</sup> It is worth considering setting up social media commissions as a statutory organisation in countries where there are no regulations.

<sup>12</sup> Especially with one of the subsidiary organs of general assembly i.e., Human Rights council (refer footnote 18)

<sup>13</sup> Like Article 9(2) of United Nation Charter

<sup>14</sup> This doesn't mean the delegates are not part of national SMC anymore

In addition to this, with the advice of the delegates of main committees, the chair (of the respective main committee) can form part-time expert sub-groups to support the matters discussed by the main committee regarding framing and reviewing the uniform principles. The expert sub-group members should be subject matter and policy experts who would advise the commission through their expertise (this enhances the capacity of SMC-IL – element of throughput legitimacy). Besides, the expert sub-group should closely work with the in-house research team formed under SMC-NL.

SMC-IL should also have an adjudication wing set up. The adjudicators will be non-permanent members of the commission, where the commission will assign a bench according to the cases and context of the issue. SMC-NLs (through automatic route too – refer to section 3.4), consumers and social media platforms can raise their grievance at a global scale to the adjudicatory wing. The adjudicatory wing should also work in tandem with the International Court of Justice. The commission must also form an appellate body to whom the SMC-NLs, consumers and social media platforms can raise their concerns related to the decision of the adjudication wing.

To support the various functions and procedures of SMC-IL, the general secretariat can be formed by the commission headed by the general secretary.

## **2.2. Social Media Commission – National Level (SMC-NL)**

The next level below in federated structure is establishing a national level social media commission that would work within the respective country. The members of SMC-NL must be selected, sufficing one of the crucial elements of throughput legitimacy, i.e., capacity to pull the task. Therefore, SMC-NL shall comprise members who are stakeholders. There shall be four representatives from social media platforms, one member each from small<sup>15</sup>, medium and large market sized platforms and one member representing indigenous platforms<sup>16</sup>. There shall be four representatives from the national government from relevant ministries and departments depending upon the country. There shall be four representatives from the state government, one member from each geography, i.e., east, west, south, and north, depending on the country. In addition to the government, it is also essential to have one representation from parliament and various standing and joint committees<sup>17</sup>. There shall be two representatives from civil society organisations, two lawyers, one academician, two representatives from media, i.e., one from national and regional, one representative from advertising agencies, one representative from minority groups. Every member of the SMC-NL must be a legal citizen of the respective country. Each member can hold office for five years and have voting rights to select chair and deputy chair of the SMC-NL. Therefore, as SMC-NL will have representation across the board, it considers the multistakeholder aspect of the problem to be tackled.

SMC-NL shall have two main committees which would aid the commission in two different functions, i.e., (a) state advisory and advocacy function (refer to section 3.2) (b) social media supervisor and accreditation function (refer to section 3.3). Besides, SMC-NL shall also

---

<sup>15</sup> For this purpose of this paper definition of ‘Small-sized platforms’ also includes start-ups

<sup>16</sup> Indigenous platform must have their root of origin in the respective country

<sup>17</sup> This will differ according to the system that the country follows i.e., parliamentary system or presidential system

constitute a grievance redressal wing that takes the complaints at the national level (refer to section 3.4).

## **Participatory Process**

Main committees must aid SMC-NL in developing guidelines, standard operating procedures, and other necessary tools and documents per uniform principles prescribed by SMC-IL through stakeholder consultation. The process of conducting and collating the responses from consultation must be one of the in-house research wing's mandates which would also closely work with the expert sub-groups (if formed) of SMC-IL.

Besides, the research wing should also review and collect all inference reports of prevalence-based complaint audits submitted by SMC-RL to aid the universal principle audit function (refer to section 3.1) conducted by SMC-IL.

### **2.3. Social Media Commission – Regional Level (SMC-RL)**

The final component to the federated structure is the social media commission that operates at the regional level<sup>18</sup>. SMC-RL will be the second tier in the grievance redressal role for consumers/entities (refer to section 3.4). The members shall be representatives with domain expertise, and they can form sub-commissions depending upon the cases. SMC-RL shall also have an Interactive Voice Response (IVR) wing for navigating the grievance management system in terms of finding the appropriate portal for lodging complaints.

Besides, the SMC-RL shall work in tandem with SMC-NL by providing inference reports based on the prevalence-based compliant audit.

### **2.4. Conduct Committee**

One of the major concerns with the Facebook oversight board was the lack of consumer involvement in the system because the formation process only had a borderline representation of consumers, and the final discretion to choose the board members was that of Facebook. Alternatively, in this model, a conduct committee shall be formed in respective countries by the consumers through nomination and voting, which would, in turn, constitute the SMC-NL and SMC-RL. This voting and nomination process can be aided by the Human Rights Council<sup>19</sup> of the UN and the respective country's election authority.

The conduct committee shall comprise experts with domain expertise such as content moderation laws/policies and narrative harms, universal human rights and values advocates, technology experts, legal experts. The conduct committee can hold office for five years have not more than ten members<sup>20</sup>.

---

<sup>18</sup> This can be at the state level or county level depending upon the respective country and population

<sup>19</sup> Human Right Council is appropriate because within its broad ambit this council closely works on topics related to content moderation to preserve freedom of expression and opinion

<sup>20</sup> Subjected to the size of the country and population

## **Representative Input**

The conduct committee shall appoint the members of the SMC-NL. In addition, consumers can also nominate members of the commission (like the Facebook oversight board). The conduct committee can also appoint members of the SMC-RL.

Post constituting the conduct committee through voting, the committee should take the nomination from the consumers for the members under categories like civil society organisation, minority groups, lawyers, academicians for SMC-NL (refer to section 2.2). In addition, the committee should also reach out to social media platforms, government and parliament seeking representation.

Therefore, the SMC model ticks the box of respective input as the consumers will be represented as part of both the SMC-NL and SMC-RL formation process through the conduct committee.

## **System Countenance**

Another major issue with the Facebook oversight board is the lack of a reliable mechanism for holding the board's actions accountable<sup>21</sup>. Besides, Facebook itself has formulated internal oversight board's regulations through its bylaws. Alternatively, in this SMC model, the conduct committee should be entrusted to constitute an internal regulation for SMC-NL and SMC-RL, which would provide procedural clarity to the commissions and process transparency to the consumers. Further, the committee should also monitor the actions of SMC-NL and SMC-RL and hold them accountable.

Therefore, as the entire system of SMC-NL and SMC-RL, starting from the constitution to forming internal regulation and monitoring, is done by the conduct committee (formed by the consumers), it advances the trust ratio, which in turn ticks the box of system countenance.

## **Transparency and accountability**

One of the critical concerns with the existing governance structure, be it Facebook oversight board or government intervention, is the lack of accountability. While a Trust was formed in the case of the Facebook oversight board to hold the board accountable to its ideals such as independent operation without Facebook's inference, still accountability is suboptimal. Besides, the process and functions of the existing governance structures are less transparent, especially in the case of government interventions. Therefore, this model will ensure that two important elements of throughput legitimacy, i.e., transparency and accountability are intact, through accrediting conduct committee the responsibility to ensure that the process and functions of SMC-NL and SMC-RL are transparent and accountable to them.

---

<sup>21</sup> While Oversight Board Trust is formed as an independent body their role is very limited as of now

### 3. Social Media Commission Model (SMC) – Functions

Every commission, i.e., SMC-IL, SMC-NL and SMC-RL, has a unique function to perform within the federated structure of the social media commission model. All the three structures of this model will work in synergy as their functions are interdependent. Besides, the model also follows a bottom-up approach where SMC-RL is where ground-level information is collected and crunched, and inference is moved upwards in the funnel to SMC-NL, which further aggregates and transfers it to SMC-IL. In addition, the dispute management function also follows a similar structure. This section will discuss various functions of social media commissions, how they are performed, and the role of each committee formed under different commissions.

#### 3.1. Framing and Auditing of Universal Principles

##### 3.1.1. Framing Sets of Universal Principles

SMC-IL frames the universal principles with the help of main committees to aid various forms of content moderation functions constituted by both state and non-state actors to tackle narrative harm. There shall be four sets of universal principles depending upon the form of content moderation, i.e., automation, human moderation, peer-to-peer moderation, and state moderation. Each main committee has a respective aspect/form of content moderation to deliberate upon and develop a set of principles that will be finally debated by the SMC-IL and voted for implementation. Below I will discuss the preamble of four main committees.

##### **Algorithmic moderation committee**

With mounting pressure over the platforms from the government<sup>22</sup> and consumers to tackle narrative harms<sup>23</sup>, the platforms have resorted to algorithmic content moderation<sup>24</sup> to tackle the scale problem<sup>25</sup>. While there is extensive debate over the scale problem, algorithmic moderation of content will not fade away due to its timeliness, and platforms use them for hard and soft moderation. The use of algorithmic systems for hard moderation exacerbated exponentially during the pandemic as the human moderators were sent home. Therefore, to unlock these technologies' positive capabilities (rather than completely taking an aggressive regulatory stand), the main committee on algorithmic moderation must come with principles that can be embedded within these technological processes for delivering better consumer welfare. Below are some of the issues highlighted through various research. These inferences must act as a preamble for the main committee on algorithmic moderation.

**Hard moderation:** The platforms use various forms of technological measures like word filters, automated hash-matching, geo-blocking, content ID, and other predictive machine learning tools for detecting unlawful content like child sexual images, pornography, dis/misinformation etc. Platforms take decisions related to the detected content and account of the individuals using human moderators or through means of algorithms themselves. Besides,

---

<sup>22</sup> India's IT Rules 2021 mandate significant social media intermediaries to deploy technological measures such as automated tools to proactively identify unlawful content (rule 4)

<sup>23</sup> Refer to footnote 7

<sup>24</sup> One form of commercial or platform moderation

<sup>25</sup> Platforms state that it would be hard to perform content moderation without the help of automation because quantity of content is produced is enormous

they perform hard moderation at both ex-ante (to an extent) and ex-post levels. While these technological measures have their merits, we increasingly see content falling through the crack due to false negatives and getting stuck or taken down due to false positives. This is problematic as false negatives can have real-world implications; for instance, despite efforts taken by the social media platforms to flag fake information on election integrity, the U.S. Capitol was attacked on 6 January 2021 due to the spread of dis/misinformation.

On the other hand, false positives can hamper freedom of expression and opinion. For instance, Facebook took down the award-winning image of a naked girl fleeing napalm bombs captured during the Vietnam War. This incident illuminates that the algorithmic content moderation techniques can't differentiate the contextual meaning of the content; after receiving negative feedback, Facebook reinstated the image. Therefore, the main committee must develop principles that would strike a proper balance between tackling the harmful content and preserving freedom of expression/opinion subjected to context.

**Soft moderation:** When we see and read posts on social media, we don't realise that the content which appears on our screen is not coincidental. Social media platforms collect information (input data) on our preferences, behaviour, relationships and so on, to develop an algorithmic system and recommend content (output) that will help them retain our attention for longer. These addictive recommendation algorithms also provide a safe harbour to the consumers who produce unlawful content. While platforms are extensively moderating the harmful content through hard moderation techniques, it does little in tackling the issue at the root by reducing the supply of harmful content by tackling the demand caused for it through these addictive recommender algorithms. Therefore, to have a holistic approach to algorithmic content moderation, the main committee must also consider the soft moderation problem. Suggested principles must draw a line between paternalism (in the form of recommendation systems etc.) and consumer harm caused due to it.

Besides, another major fall through the crack happens in the form of borderline content. The borderline content is a grey area content which doesn't entirely fall within the watertight categories of harmful and harmless content, depending upon the socio-cultural-legal context of the jurisdiction. For instance, nudity might be completely normal within the legal and social fabric of a nation, whereas in some other nations, it would be considered a violation. Currently, platforms define and tackle the borderline content using a combination of algorithmic hard and soft moderation techniques at their discretion. This ad-hoc and disjointed way of handling borderline content causes various repercussions in society. Therefore, the goal of SMC-IL must be to harmonise these differences by coming with a uniform definition and principles for tackling the borderline content. To that means, the main committee on algorithmic moderation must aid SMC-IL by finding ways to develop universal principles without comprising any nation's social and legal fabric.

Thus, the algorithmic moderation committee must come with hard and soft moderations principles by following the procedures mentioned in section 2.1.

### **Human moderator and welfare committee**

Understanding that technology can't be used entirely for content moderation purposes (as they lack context and interpretation), social media platforms deploy human moderators to patch the gap. The human moderators do both ex-ante moderation for some high-stake con-

-tent and ex-post moderation for content flagged by the users. While it is essential to have a pair of human eyes to evaluate the content, there are various problems with this process, as discussed below.

**Human moderator welfare:** It has been noted that there are various levels of human moderators within the system. Some work as contract workers under the payroll of outsourced companies (within which there are hierarchical systems), and others work as direct employees of the platforms. While direct employees work under the social media's payroll, others ply within the silos under pseudo designation with the strict non-discourse requirement. This secrecy provides room for the social media platforms to mistreat and provide a suboptimal work environment for the contract-based human moderators. In addition, the human moderators are constantly exposed to psychologically tolling content daily, which disturbs their mental health and physical health (to an extent). There are some key humanitarian concerns spotted through research and reporting in regard to using human moderators and their nature of work: (a) there is inadequate mental health rehabilitation facilities, (b) contract-based human moderators are compensated less (Newton, 2019) (c) human moderators are treated like a machine (Ruckenstein & Turunen, 2019) (d) human moderators are forced to work in secrecy (Banerjee, 2021) which brings out more repercussions.

Besides, human moderators were usually hired from developing countries where labour is cheap; until recently, Facebook hired human moderators from the USA and other first world countries due to the need for more human moderators amidst mounting concerns on content that violates the community guidelines. This highlights that most of the outsourced contractors work under sub-optimal across the globe.

Therefore, the human moderator and welfare committee should mitigate these concerns through principles that would eventually push social media platforms to ensure welfare to the human moderators directly through themselves or indirectly via contractors.

**Arbitrary moderation guidelines:** Starting from broad standards for content moderation, the social media platforms narrowed their content moderation guidelines to rule-based prescription, as the human moderators spread across the globe with diverse backgrounds (Klonick, 2018). While rule-based prescriptions try to bring uniformity, it drains the human element from content moderation as human moderators are provided less room for deliberation, analysis, and discretion based on context and time. Besides, it has been noted that the content moderation guidelines prescribed to the human moderator (which is frequently subjected to changes) by the platforms are different from the external guidelines produced for user consumption. Using these guidelines, the human moderators pass professional judgement over the content and are trained to weed out cognitive biases while making the same. It has been noted that human moderators are pushed and trained to have the highest accuracy level, i.e., judgement in congruence with rule-based content moderation guidelines. While it is essential to have rule-based content moderation guidelines, currently, the framing and operationalisation of the same happen in clandestine, and platforms exercise arbitrary power in this process without any principles backing to their decision.

Therefore, the human moderator and welfare committee should deliberate on these issues and list out principles that would support the framing and operationalisation of rule-based content moderation. In addition, these principles should also bring in stability and act as a prescription to the platforms to follow democratic procedures, be it in-house human moderation or outsourced.

Thus, the human moderator and welfare committee must come with a set of principles by following the procedures mentioned in section 2.1, ensuring welfare to human moderators and clarity to their work (i.e., principles aiding moderation guidelines).

### **Lateral moderation committee**

While a traditional moderation set-up involves two actors at different levels, i.e., the state/platform and user account/content, in lateral moderation, set-up breaks this code by enabling peer-to-peer moderation. Both governments and platforms have orchestrated peer-to-peer moderation to patch the capacity issue, but they cause various implications, as discussed below.

**Content reporting system:** Social media platforms have a feature where users can report content that they find offensive or against platform norms. In Facebook, the report content feature provides users with choices for flagging the content, which helps the platform prioritise content and channel it to respective portals for resolution. Twitter recently instituted a feature called birdwatch - a separate platform that collates and presents user reported misleading content on a majority basis. While non-state actors are extensively moving towards content reporting systems to bridge the capacity issue, states have recognised this feature and recommended it through law. For instance, India's IT Rules 2021 encourage the content reporting system to enable users to report unlawful content for takedown.

While this system is not direct moderation, it acts as the first step in content moderation where platforms send these user requests for human moderator reviews. While this feature does help in tackling the capacity issue, without a concrete principle-based framework that aids the platforms as well as the users to utilise this feature wisely is problematic. Therefore, the lateral moderation committee must develop principles that would act as a backbone for content reporting systems and any other lateral moderation system instituted by the social media platforms in the future.

**Peer-to-peer reporting system:** Like platforms, some governments across the globe institute various peer-to-peer reporting systems. While these systems again act as the first step towards moderation where the government can direct platforms to act upon the reported content, the language and operationalisation of these systems ultimately end up causing lateral surveillance<sup>26</sup>. For instance, recently Indian Cyber Crime Coordination Centre under the Indian Ministry of Home Affairs launched the cyber volunteers programme, which seeks citizens to report unlawful activities on the internet and social media. Therefore, the lateral moderation committee must deliberate on the issues related to peer-to-peer reporting systems and develop a set of principles that would act as a standard for designing these systems weeding out the threat of surveillance and other unintended consequences.

Thus, the lateral moderation committee must develop principles for both non-state and state actors to aid their development of lateral moderation systems.

---

<sup>26</sup> Lateral surveillance is a phenomenon where citizens 'watch over others' i.e., other citizens through monitoring their actions.

## State-mandated moderation committee

Social media is a deliberative medium. The nation's mood towards some significant problems deliberated on social media directly impacts government actions; hence the state is sceptical about the content on social media platforms. Amongst the other mechanisms, state (a) enforces social media platforms to monitor for unlawful content proactively (b) orders takedown of content. These state moderation mechanisms are enforced in an ad-hoc manner without any concrete principle-based framework to evaluate. The state moderation committee must deliberate on the below-discussed problems related to state enforcement mechanisms to develop principles that would aid the state and become the pedestal for SMC-NL's state engagement (refer to section 3.2) in terms of advocacy/advisory role.

**Proactive measures:** Many countries have moved towards a proactive mechanism as part of their social media regulations - governments direct social media platforms to monitor unlawful content on their platforms actively. For instance, in India, IT Rules 2021 (for details, refer to section 1.1.1) directs significant social media platforms<sup>27</sup> to deploy technological tools like automation to identify unlawful content on their platforms. These measures provide a lot of room for social media platforms (a) to determine illegal content (b) might also indulge over moderation to protect themselves from legal implications (c) to trade-off between content. Therefore, be it any form of moderation (human or algorithmic), it is essential to ringfence the extent to which social media platforms can determine the legality of the content. Government should move towards principle-based approach toward proactive measures where SMC-NL (refer to section 3.3) would aid the social media processes to align with the principles through producing operational guidelines. SMC-NL must also advocate and advise the states to move from the current discretionary mechanism towards principle-based legal determination of content.

Thus, the state-moderation committee must develop principles that form the base for the legal determination of content.

**Notice-and-action procedure:** Most of the governments across the globe issue content takedown orders to which social media platforms must bind. For instance, according to the Indian IT Rules 2021, the government can order (through court order or by a designated government agency) the platforms to take down content they deem unlawful within 36 hours of notification. It has been noted that these takedown orders are issued in secrecy without any accountability and transparency. While the other subset of principles discussed above will aid legal determination of content, the process and practices behind taken down orders must follow certain principles. This is important because the number of content takedown orders has increased over time; for instance, in India, about 6000 takedown orders were sent by the government to the platforms. Therefore, the state moderation committee must develop principles that would make the state-mandated content takedown process and practice more procedurally fair, democratic, accountable, and transparent.

The table below summarises section 3.1.1 by listing a set of principles to be formed by different main committees of SMC-IL. Thus, the function of the SMC-IL illuminates that the elements under output legitimacy are embedded within the model where it (a) follows principle-based approach (b) tries to reduce unintended consequences through ringfencing the process and practices of content moderation within the ambit of principles proposed

---

<sup>27</sup> Refer to footnote 3

and (b) tries to shift from current neutral international principles to more tailored, nuanced, and suitable principles for various nuanced content moderation practices and process.

**Table 1: Set of principles to be formulated by SMC-IL**

Main committee	Set of principles
Algorithmic moderation committee	<ul style="list-style-type: none"> <li>● Principles for hard moderation (tech-based)</li> <li>● Principles for soft moderation (tech-based)</li> </ul>
Human moderator and welfare committee	<ul style="list-style-type: none"> <li>● Welfare principles for human moderator</li> <li>● Principles for human moderation</li> </ul>
Lateral moderation committee	<ul style="list-style-type: none"> <li>● Principles for developing content reporting systems</li> <li>● Principles for developing peer-to-peer reporting systems</li> </ul>
State-mandated moderation committee	<ul style="list-style-type: none"> <li>● Base principles for proactive measures</li> <li>● Principles for notice-and-action procedure</li> </ul>

### 3.1.2. Auditing Universal Principles

One of the critiques of any technology regulations is the delay in the policymaking process and no update over time which may render it obsolete given the fast pace of technological advancement. This shows that even though the technology is advancing exponentially, the policy discourse around the same is still barely scraping the trails. Therefore, as the scope and pace of development of social media platforms are unprecedented, the model suggests that irrespective of framing the universal principles, as a mandate, main committees have to review/update those principles over a period of time in an adequate interval to keep up with the pace.

This model proposes a more representative and bottom-up approach for auditing principles by analysing disputes' prevalence. A feedback mechanism based on analysing the prevalence of dispute received by the SMC-RL (refer to section 3.4) is one way of identifying gaps in universal principles. For instance, Facebook reinstated the award-winning image of a naked girl fleeing napalm bombs during the Vietnam War after receiving negative feedback following its takedown. While Facebook received this feedback through newspapers and civic movements, this event still shows that aggregating grievances can provide feedback on policies and actions (Shekar, 2021).

If all the complaints received by SMC-RL were to be analysed for prevalence, patterns would emerge to reflect where the mechanism has formed a void or lacks seamlessness. These complaints could flag the proportion of problems in different sets of principles and signal any new technological trends that aren't covered by principles or made obsolete. Once this audit of dispute is conducted and patterns emerge through data crunching, the problem areas become more apparent and easier to solve. Newer technologies can then be studied, understood, and principles can be formed/modified accordingly. Besides, recurring problems can be addressed with alternative principles, which might, in turn, ease the blockages in redressal mechanisms.

Therefore, within the federated system, SMC-RLs<sup>28</sup> of the respective countries must funnel the inferences from the grievances received during a stipulated time to SMC-NLs (to research wing) as a report. In turn, SMC-NLs must collate the inferences across the country, find patterns, and present it at the principles audit meet under different main committees of SMC-IL. This way, the model tries to keep the process of auditing universal principles participatory, which is one of the elements of input legitimacy.

### **3.2. Advisory and Advocacy Role to the State**

The principles framed through the SMC-IL must be considered as international obligations that the state must strive to follow. In the current structure of international laws, the state can ratify international obligations by signing treaties, but there are no national structures that would aid the state in translating the international principles to the domestic context such that it is seamless for the state to consider. Therefore, this model proposes that SMC-NL must perform the advisory and advocacy role to the state depending upon the relationship theory (as discussed below) followed by the state, i.e., monism and dualism.

#### **3.2.1. Advisory Role to the Monist States**

In monist countries, the international law becomes the rule of the land with the state signing the international treaties as a commitment to the international principles and obligation, directly without incorporating it into the domestic legislation. Under this system, domestic legislation becomes subordinate; therefore, ICC statute becomes enforceable within the national boundaries and adjudicated by the national courts. There are about 16 countries, i.e., the US, Chile, Austria, South Africa etc., which are monist states. In all the 16 countries, in some form or the other, it is mandated for the state to go for legislative approval before committing to international obligations for the nation.

Therefore, here the role of one of the main committees of SMC-NL<sup>29</sup>, i.e., state advisory committee, will be crucial as they will have to work with the state by helping them introduce the international principles<sup>30</sup> farmed by SMC-IL in legislation. Besides, the international principles must be optimised to the domestic context such that the state motion gets legislative approval seamlessly. After ratifying the international treaty, the states face various roadblocks in implementing the international obligation within the domestic legal order. Here again, the committee will have to aid the states in their implementation process through thorough advice.

#### **3.2.2. Advocacy Role to the Dualist States**

In dualist countries, international laws and domestic laws are two different legal frameworks, where domestic laws precede and form the rule of the land. Here just the ratification of international obligation and principles is not enough, where the domestic laws must explicitly incorporate the same within the statute. Most commonwealth countries are dualist states and

---

<sup>28</sup> Though the grievances will be escalated to SMC-NL and SMC-IL within the function proposed in section 3.4, still the grievance inference report must come from SMC-RL, as they are the primary consumer touch point for grievance redressal.

<sup>29</sup> The SMC-NL will strive to make sure that the state is convinced to join the SMC model of social media governance.

<sup>30</sup> (a) Principles for developing peer-to-peer reporting systems (b) Base principles for proactive measures (c) Principles for notice-and-action procedure.

follow various ways to incorporate treaties within the domestic legal systems. But in some dualist countries, signing a treaty is ineffective as they are not obliged to follow the same. For instance, in India, only very few international principles and treaties are considered within the domestic system. Besides, antinomy is noticed to be very high in these countries. Therefore, in the case of a dualist nation, the state advocacy committee of SMC-NL must find ways to advocate the international principles within the policy discourse such that it is well taken by the state and legislation. The committee must understand the state's perspective<sup>31</sup> and accordingly plan for the advocacy strategy. Besides, the committee must also consider the perspective of stakeholders such as platforms (as a collective and coordinated action – an element of throughput legitimacy) and consumers (referring to inferences from disputes submitted by the SMC-RL<sup>32</sup>) as part of their submissions to the state.

Therefore, the state advisory/advocacy committee of SMC-NL will try to enhance the state capacity (one of the elements of throughput legitimacy) to understand the problem and direct them towards principles<sup>33</sup> framed by the SMC-IL when the regulations are made.

### 3.3. Aiding, Supervision and Accreditation Role to the Social Media Platforms

In addition to the advisory and advocacy role to the state, SMC-NL must also perform the below-discussed roles via social media supervisor and accreditation committee to ensure social media platforms align their functions, features and practices to the principles<sup>34</sup> framed by SMC-IL.

**Aiding & supervision role:** the committee must aid the social media platforms to implement principles within their procedures and processes by forming various operational guidelines, SOPs, awareness programs, and private consultations. These aiding materials prepared by the committee must be tailored according to the domestic context and socio-economic-political fabric that the platforms are working within for seamless implementation. Besides, moving forward, it is also crucial for the committee to supervise procedures and processes of social media platforms to see if they are aligned with the principles through mechanising various soft enforcement measures like “name and shame”, temporary/permanent cease of representation in the commission etc. The enforcement measures must be evaluated and implemented according to the nature and size of the non-compliance.

**Accreditation role:** In addition to the direct supervision and enforcement, the committee must also institute market mechanisms such as (a) audit of content moderation features used by social media platforms (b) market for principles-based accreditation, which should enable a competitive edge for platforms. While an independent auditing agency must perform the audit, the committee must perform the accreditation<sup>35</sup> process at a nominal cost based on defined principles. The accreditation process must have a well-laid process and procedure that balances transparency and safeguards to protect intellectual and proprietary information. Besides, the

---

<sup>31</sup> SMC-NL will have representation of state and legislation (refer to section 2.2)

<sup>32</sup> As the state might not be following the SMC-IL principles, the SMC-NL must utilise the report for best possible ways.

<sup>33</sup> States can also refer to the other sets of international principles while making legislations

<sup>34</sup> (a) Principles for hard moderation (b) Principles for soft moderation (c) Welfare principles for human moderator (d) principles for human moderation (e) principles for developing content reporting systems.

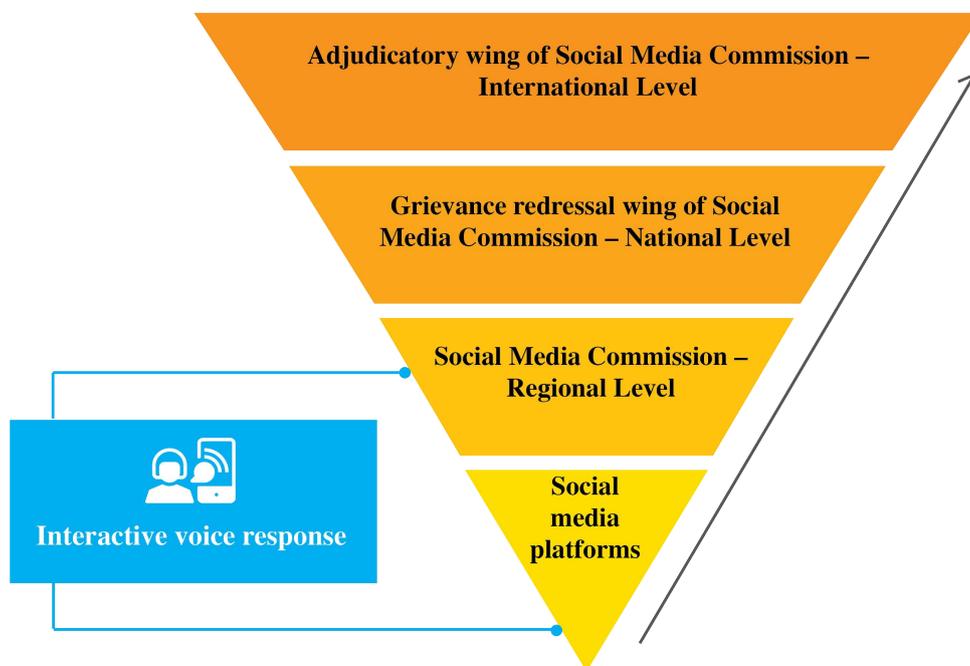
<sup>35</sup> Social media supervisor and accreditation committee can provide global-level accreditation to big social media platforms (if they opt for) who operate in multiple countries to avoid redundancy.

accreditation process must be aspirational such that it pushes the social media platforms toward performing better on the user outcome aspect.

### 3.4. Grievance Redressal Role to Consumers/Entities

Currently, we don't have a well-functioning and agile grievance redressal mechanism that would take complaints on social media platforms for various actions. For instance, when social media platforms constantly remove my post from their platform stating that my content is against the norms of the platform, I don't have a place to complain against this action if my content is not violating the platform norms. While legislation in some countries mandates some form of dispute management systems (like India's IT Rules 2021), this model proposes a calibrated hierarchical grievance management system (refer to figure 3) that is federated, coordinated and agility proofed.

**Figure 3: Grievance management system**



**Interactive voice response:** The step zero in the grievance redressal process is to find the designated portal for lodging a dispute. In many cases, navigating the grievance management system for both consumers and entities is arduous, making it difficult to reach the designated portal. Therefore, the zero-step of the proposed grievance management system should involve Interactive Voice Response (IVR). This automated voice response system through call will navigate consumers (or entities) to reach step 1, i.e. to the social media platform's grievance redressal system or to step 2, i.e. reaching out to SMC-RL if step 1 is already complete.

**Social media platform's grievance redressal system:** Social media platforms have various grievance redressal mechanisms, which are also mandated through legislation (as discussed) in some countries. Therefore, step one of the proposed grievance management system is to get the dispute redressed by reporting it to the respective social media platform.

**Social media commission - regional level:** Step 2 of the proposed grievance management system is to reach out to SMC-RL if step 1 fails or the respective social media platform doesn't have a grievance redressal system. The functions of the SMC-RL doesn't override the judiciary system of the country; instead, it is advised to work in tandem with district and regional courts. Where, in case of failure of mediation, the SMC-RL can move towards litigation.

Besides, various SMC-RLs within a country must work in tandem through integrating various functions and databases, wherein the case of a similar form of complaint on a particular incident coming from across the country should be aggregated and moved to step 3 with all the details of the complaint.

**Grievance redressal wing of SMC-NL:** In case of a national-level dispute, step 3 will be initiated. The grievance redressal wing of the SMC-NL will take up the aggregated disputes escalated by SMC-RL. Again, the functions of the grievance redressal wing will not override the powers of the country's judiciary system; instead, the wing should work in tandem with courts. Where, in case of failure of mediation, the SMC-NL can move towards litigation.

Besides, grievance redressal wings of various SMC-NLs must work in a coordinated manner (similar to SMC-RLs of the country). If similar complaints on particular incidents are lodged across the globe, they must be aggregated and moved to step 4 with all the complaint details.

**Adjudicatory wing of SMC-IL:** Step 4 of the grievance redressal system will be initiated if there is an international level dispute on a particular aspect escalated by the consortium of SMC-NLs. In addition, any dispute between SMC-NLs on any matter related to social media platforms and content moderation can be mediated through SMC-IL.

#### **4. Potential Funding Options for Social Media Commission**

The Social media commissions would need a steady and sustainable flow of revenue for keeping its operations (a) seamless, (b) financially and functionally independent (c) safe from social media platforms and state capture. To achieve these goals, this model proposes various facets of revenue streams for the commission at different levels, which let the operations move without dependence on one source of income.

**Social media platforms contribution:** Social media platforms must contribute to social media commissions at different levels. The proportion of the contribution can be determined through various factors such as reach, active user base, income level etc. In addition, there shall be no restriction on social media platforms to contribute more. Besides, they shall also pay a nominal fee to the SMC-NL for securing accreditation

**Human Rights Council contribution:** As discussed above, SMC-IL must collaborate and coordinate with the Human Rights Council (HRC) – a subsidiary organ of the UN general assembly. HRC is funded through regular (as part of UN budget) and voluntary routes for running its programmes and operations. As functions of social media commission would be a subset of the HRC's mandates performed in a more nuanced manner, the model proposes seeking funding from HRC or UN directly. As the commissions start expanding their scope within the larger ambit of securing the digital public sphere, the model suggests the commission fosters more collaborations with the UN and its organisations.

**State contribution:** In the case of monist states, the state shall compensate the SMC-NL with a nominal fee for the advisory service provided by one of the main committees in terms of aligning the domestic legislation with international obligations. Besides, SMC-NL must try to strike an MOU or agreement with the state (in both monist and dualist countries) to avail various economic benefits like subsidies, grants etc.

The SMC-NL must seek the states to consider a financial contribution under corporate social responsibility or any other tax subsidies (if there is legislation for the same).

**Philanthropy contribution:** Amongst the other funding sources, philanthropy funds stand to be a significant contribution to UN funds. For instance, it has been recorded that Bill and Gates foundation’s contribution stands within the highest funders of UN list. Similarly, the commission should venture to secure funding from philanthropy.

**Plaintiff and defendant contribution:** SMC-RL shall levy a nominal fee for providing dispute resolution to both plaintiff and defendant. This nominal fee should be an operational cost in addition to a small margin.

**Other means and ways:** The model suggest that the commission shall be provided independence to decide other means and ways to generate revenue without hampering the goals discussed above in this section.

**Table 2: Mapping potential funding to the commissions**

<b>Funding options</b>	<b>Recipient</b>	<b>Beneficiary</b>
<b>Social media platforms contribution</b>		
Social media platforms – operating at the international level	SMC-IL	SMC-IL
Social media platforms – operating at national-level	SMC-NL	SMC-NL SMC-RL Conduct Committee
Nominal fees for accreditation	SMC-NL	SMC-NL
<b>Human rights council/UN contribution</b>		
UN contribution – For setting up SMC-NL (at least starting from 193 member countries)	Conduct committee	Conduct Committee
Human rights council/UN	SMC-IL	SMC-IL SMC-NL SMC-RL
<b>State contribution</b>		
Tax subsidies, grants etc.	SMC-NL	SMC-NL SMC-RL Conduct Committee

Nominal fees (in the case of monist states)	SMC-NL	SMC-NL
CSR funding (through state-route discussed in this session)	SMC-NL SMC-IL	SMC-IL SMC-NL SMC-RL
<b>Philanthropy contribution</b>		
Philanthropist at the global/national/regional scale	SMC-IL SMC-NL	SMC-IL SMC-NL SMC-RL
<b>Plaintiff and defendant contribution</b>		
Plaintiff and defendant nominal fee	SMC-RL	SMC-RL
<b>Other means and ways</b>		
Various autonomous revenue generation stream	SMC-IL SMC-NL SMC-RL	SMC-IL SMC-NL SMC-RL

## 5. Conclusion

Globally we are witnessing tremendous growth in digitalisation efforts, where footprints of social media use are seen across the borders and amongst various strata of people. While this development has positively impacted society by enabling innovation in business models, providing better service, and helping in crisis management and connectivity, it has also created various new sets of issues and gaps in the regulatory perimeter. In deciding what interventions would best solve the identified gaps in the current system, policymakers, various industrial associations, and social media platforms have started framing many new governance structures, mainly to tackle narrative harms. This paper has showcased that the existing governance structure lacks democratic legitimacy and proposes a federated model of governance called the Social Media Commission (SMC). The proposed structure has international-, national-, regional-level commissions with unique roles and will work in synergy following a bottom-up approach.

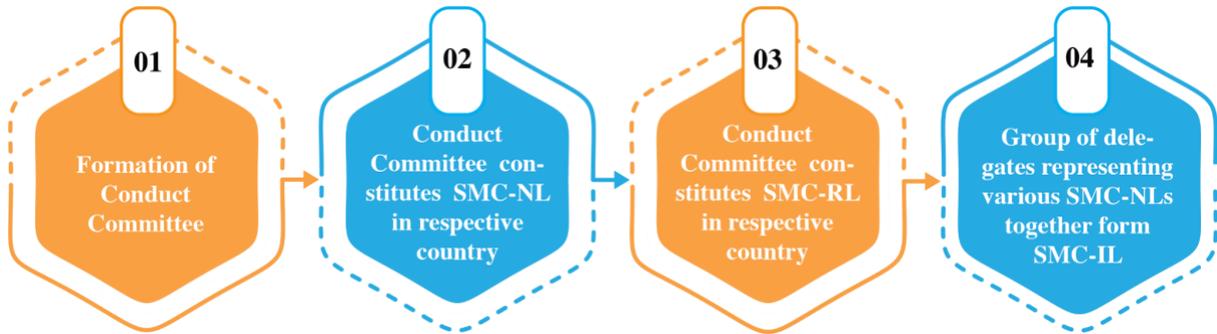
The paper also discussed the federation structure of the SMC advised by input, throughput, and output legitimacy. The proposed system in this paper is (i) responsive to user concerns through having active participation and representation of the users (input legitimacy), (ii) governed through robust accountability and process transparency mandates (throughput legitimacy), (iii) adequate to deliver better consumer outcomes (output legitimacy).

While the paper tries to provide an operational model for governing content moderation function in terms of social media content, various other facets of issues emerge from the other functions of social media platforms. Some of the immediate concerns which need attention are surveillance capitalism, lack of interoperability and portability, consequences of network effects, privacy concerns, security concerns etc. As the social media commission model scales

up, it would be ideal for the commissions to expand the scope into solving these problems (and any other) by utilising the federated structure.

## Annexure: How Does SMCs Model Work? – Process Flow

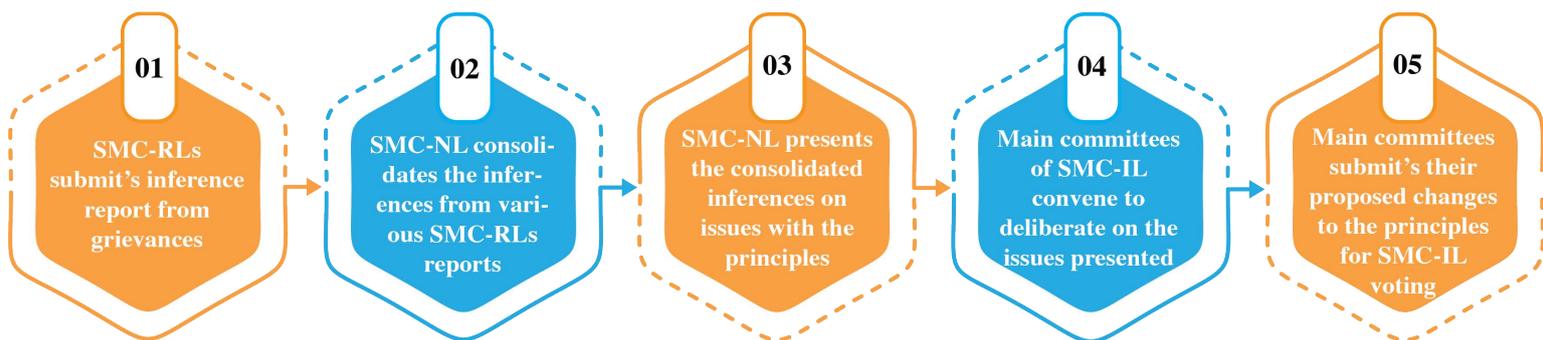
**Figure 4: Setting up various SMCs**



**Figure 5: Framing universal principles**



**Figure 6: Auditing universal principles**



## Bibliography

A rights-respecting model of online content regulation by platforms. (2018). Retrieved from Global Partners Digital: <https://www.gp-digital.org/publication/a-rights-respecting-model-of-online-content-regulation-by-platforms/>

Adler, S. (2018). Post No Evil. Retrieved from Radiolab: <https://www.wnycstudios.org/podcasts/radiolab/articles/post-no-evil>

Alexander, J. (2019). YouTube claims its crackdown on borderline content is actually working. Retrieved from The Verge: <https://www.theverge.com/2019/12/3/20992018/youtube-borderline-content-recommendation-algorithm-news-authoritative-sources>

Ananny, M., & Crawford, K. (2016). Seeing without Knowing: Limitations of the transparency ideals and its applications to algorithmic accountability. *new media & society*, 1-17.

Bal, H. S. (2021). Paranoia about digital coverage led ministers to propose media clampdown, monitoring “negative influencers”. Retrieved from The Caravan: <https://caravanmagazine.in/politics/paranoia-about-digital-coverage-led-gom-propose-media-clampdown-monitoring-negative-influencers>

Banerjee, P. (2021). Facebook’s content moderators demand for an end to culture of ‘fear and secrecy’. Retrieved from Livemint: <https://www.livemint.com/technology/tech-news/facebooks-content-moderators-demand-for-an-end-to-culture-of-fear-and-secrecy-11626998967857.html>

Bartlett, J., & Krasodonski-Jones, A. (2015). counter-speech examining content that challenges extremism online. Retrieved from Demos: <http://1i7vb53afs9ypj6n31joq8ch.wpengine.netdna-cdn.com/wp-content/uploads/2015/10/Counter-speech.pdf>

Bhardwaj, D. (2021). 6k social media content takedown orders this year. Retrieved from Hindustan Times: <https://www.hindustantimes.com/india-news/6k-social-media-content-takedown-orders-this-year-101623014539309.html>

Bickert, M. (2019). Updating the Values That Inform Our Community Standards. Retrieved from FB News: <https://about.fb.com/news/2019/09/updating-the-values-that-inform-our-community-standards/>

Bulck, H. V., Puppis, M., Donders, K., & Audenhove, L. V. (n.d.). *The palgrave Handbook of Methods for Media Policy Research*. Palgrave macmillan.

Coleman, K. (2021). Introducing Birdwatch, a community-based approach to misinformation. Retrieved from Twitter Blog: [https://blog.twitter.com/en\\_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation](https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation)

Collins, B. (2020). Mission Impossible: 7 Countries Tell Facebook To Break Encryption. Retrieved from Forbes: <https://www.forbes.com/sites/barrycollins/2020/10/11/mission-impossible-7-countries-tell-facebook-to-break-encryption/?sh=773da2cc3a5a>

Conference Report: Social media councils - From concept to reality. (2019). Retrieved from Stanford Global Digital Policy Incubator: <https://fsi.stanford.edu/content/social-media-councils-concept-reality-conference-report>

Confessore, N. (2018). Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. Retrieved from New York Times: <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>

Constine, J. (2018). Facebook will change algorithm to demote “borderline content” that almost violates policies. Retrieved from Techcrunch: <https://techcrunch.com/2018/11/15/facebook-borderline-content/>

Content & Jurisdiction Program. (2019). Retrieved from Internet & Jurisdiction Policy Network: <https://www.internetjurisdiction.net/uploads/pdfs/Papers/Content-Jurisdiction-Program-Operational-Approaches.pdf>

Devdasan, V. (2021). Intermediary Guidelines and the Digital Public Sphere: Balancing the Scales. Retrieved from Indian Constitutional Law and Philosophy: <https://indconlawphil.wordpress.com/tag/intermediary-liability/>

Docquir, P. F. (2019). The Social Media Council: Bringing Human Rights Standards to Content Moderation on Social Media. Retrieved from CIGI Online: <https://www.cigionline.org/articles/social-media-council-bringing-human-rights-standards-content-moderation-social-media/>

Douek, E. (2019). Facebook's "Oversight Board:" Move Fast With Stable Infrastructure and Humility.

Efficiency of the Human Rights Council – addressing financial and time constraints. (2019). Retrieved from Human Rights Council: <https://undocs.org/en/A/HRC/PRST/OS/13/1>

Ghosh, D. (2021). Are We Entering a New Era of Social Media Regulation? Retrieved from Harvard Business Review: <https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation>

Gillespie, T. (2018). Custodians of the Internet. Yale University Press.

Gillespie, T. (2020). Content moderation, AI, and the question of scale. Big Data & Society.

Gillespie, T. (2017). Regulation of and by platforms. SAGE Handbook of Social Media.

Global Social Media Stats. (2021). Retrieved from Data reportal: <https://datareportal.com/social-media-users>

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data & Society.

Goyal, T., & Sane, R. (2021). Towards better enforcement by regulatory agencies. Retrieved from Data Governance Network: <https://datagovernance.org/files/research/1612454004.pdf>  
Groove, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Sage Journal (Big data & society).

Grover, G., Rajwade, T., & Katira, D. (2021). The Ministry And The Trace: Subverting End-To-End Encryption. NUJS Law Review.

Haggart, B., & Keller, C. I. (2021). Democratic legitimacy in global platform governance. Retrieved from Telecommunications Policy: <https://www.sciencedirect.com/science/article/pii/S0308596121000562>

Heldt, A. (2020). Borderline speech: caught in a free speech limbo? Retrieved from Internet Policy Review: <https://policyreview.info/articles/news/borderline-speech-caught-free-speech-limbo/1510>

How does international law apply in a domestic legal system? (n.d.). Retrieved from The Peace and Justice Initiative: <https://www.peaceandjusticeinitiative.org/implementation-resources/dualist-and-monist>

Impact of the work of Special Procedures: Contribution to governmental and/or judicial processes. (n.d.). Retrieved from Human Rights Council: <https://www.ohchr.org/EN/HRBodies/SP/Pages/Contributiontogovernmentaljudicialprocesses.aspx>

Impact of the work of Special Procedures: Human rights standard setting. (n.d.). Retrieved from Human Rights Council: <https://www.ohchr.org/EN/HRBodies/SP/Pages/Humanrightsstandardsetting.aspx>

Impact of the work of Special Procedures: Legislative reform. (n.d.). Retrieved from Human Rights Council: <https://www.ohchr.org/EN/HRBodies/SP/Pages/Legislativeform.aspx>

Implementing the United Nations “Protect, Respect and Remedy” Framework. (2011). Retrieved from Human Rights Council: [https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr\\_en.pdf](https://www.ohchr.org/documents/publications/guidingprinciplesbusinessshr_en.pdf)

Information Technology (Intermediaries Guidelines) Rules, 2021. (2021). Retrieved from [https://www.meity.gov.in/writereaddata/files/Intermediary\\_Guidelines\\_and\\_Digital\\_Media\\_Ethics\\_Code\\_Rules-2021.pdf](https://www.meity.gov.in/writereaddata/files/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf)

International Covenant on Civil and Political Rights. (n.d.). Retrieved from UN Human Rights: <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>

Jain, M. (2021). Twitter’s ‘Soft Block’ Feature Is For Users Who Need Some Space From Annoying Followers. Retrieved from Medianama: <https://www.medianama.com/2021/10/223-twitter-feature-softblock-followers/>

Jørgensen, R. F. (2021). A Human Rights-Based Approach to Social Media Platforms. Retrieved from Berkley Forum: <https://berkleycenter.georgetown.edu/responses/a-human-rights-based-approach-to-social-media-platforms>

Kirkconnell-Kawana, L. (2021). Online Safety Bill: Five thoughts on its impact on journalism. Retrieved from LSE Blog: <https://blogs.lse.ac.uk/medialse/2021/06/03/online-safety-bill-five-thoughts-on-its-impact-on-journalism/>

Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. Harvard Law Review. Retrieved from Harvard Law Review.

Klonick, K. (2018). The New Governors: The People, Rules, And Processes Governing Online Speech. Harvard Law Review, 1589-1670.

Klonick, K. (2020). The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. Yale Law Journal, 2232-2605.

Koebler, J., & Cox, J. (2018). The Impossible Job: Inside Facebook's Struggle to Moderate Two Billion People. Retrieved from The Vice: <https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works>

Kyle, D. (2021). The battle for social media regulation: can international human rights bridge the governance gap in the digital space? Retrieved from Universal Rights Group: <https://www.universal-rights.org/universal-rights-group-nyc-2/the-battle-for-social-media-regulation-can-international-human-rights-bridge-the-governance-gap-in-the-digital-space/>

Latonero, M. (2020). Can Facebook's Oversight Board Win People's Trust? Retrieved from Harvard Business Review: <https://hbr.org/2020/01/can-facebooks-oversight-board-win-peoples-trust>

Malcolm, J. (2019). Fair Processes, Better Outcomes. Retrieved from Electronic Frontier Foundation: <https://www.eff.org/deeplinks/2016/09/fair-processes-better-outcomes>

Mcsherry, C. (2019). Social Media Councils: A Better Way Forward, Window Dressing, or Global Speech Police? Retrieved from Electronic Frontier Foundation: <https://www.eff.org/fr/deeplinks/2019/05/social-media-councils-better-way-firward-lipstick-pig-or-global-speech-police>

Mirchandani, M., Sahai, D., & Goel, O. (2018). Encouraging counter-speech by mapping the contours of hate speech on Facebook in India. Retrieved from ORF: [https://www.orfonline.org/research/encouraging-counter-speech-by-mapping-the-contours-of-hate-speech-on-facebook-in-india/#\\_edn8](https://www.orfonline.org/research/encouraging-counter-speech-by-mapping-the-contours-of-hate-speech-on-facebook-in-india/#_edn8)

Napoli, P. M. (2019). Social Media and the Public Interest. Columbia.

Napoli, P. M. (2021). The symbolic uses of platforms: The politics of platform governance in the United States. Journal of Digital Media & Policy, pp. 215–230.

Newton, C. (2019). The Trauma Floor: The secret lives of Facebook moderators in America. Retrieved from The Verge: <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>

Nover, S. (2021). What if Facebook is lying to its own Oversight Board? Retrieved from Quartz: <https://qz.com/2064919/what-if-facebook-is-lying-to-its-own-oversight-board/>  
OHCHR's Funding and Budget. (n.d.). Retrieved from Human Rights Council:  
<https://www.ohchr.org/en/aboutus/pages/fundingbudget.aspx>

PIB. (2019). "Voluntary Code of Ethics" by Social Media Platforms to be observed in the General Election to the Haryana & Maharashtra Legislative Assemblies and all future elections. Retrieved from Government of India:  
[https://www.meity.gov.in/writereaddata/files/Intermediary\\_Guidelines\\_and\\_Digital\\_Media\\_Ethics\\_Code\\_Rules-2021.pdf](https://www.meity.gov.in/writereaddata/files/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf)

Popiel, P., & Sang, Y. (2021). Platforms' Governance: Analyzing Digital Platforms' Policy Preferences. University of California Press. Retrieved from University of California Press.

Regulating social media and OTT services: Comparing rules from around the world. (n.d.). Retrieved from The Week: 2021

Reality Check Team. (2020). Social media: How do other governments regulate it? Retrieved from BBC: <https://www.bbc.com/news/technology-47135058>

Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. (n.d.). Retrieved from UN General Assembly: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement>

Roth, Y., & Achuthan, A. (2020). Building rules in public: Our approach to synthetic & manipulated media. Retrieved from Twitter Blog:  
[https://blog.twitter.com/en\\_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media](https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media)

Ruckenstein, M., & Turunen, L. L. (2019). Re-humanizing the platform: Content moderators and the logic of care. Sage Journal. (New Media and Society)

Santa Clara Principles on Transparency and Accountability. (2018). Retrieved from <https://santaclaraprinciples.org/>

Schmidt, V. (2010). Democracy and Legitimacy in the European Union Revisited: Input, Output and Throughput. KFG Working Paper Series.

Scott, M., & Isaac, M. (2016). Facebook Restores Iconic Vietnam War Photo It Censored for Nudity. Retrieved from New York Times:  
<https://www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photo-nudity.html>

Self-executing treaties. (n.d.). Retrieved from The Peace and Justice Initiative:  
<http://www.peaceandjusticeinitiative.org/implementation-resources/dualist-and-monist/self-executing-treaties>

Shadow Regulation. (n.d.). Retrieved from Electronic Frontier Foundation:  
<https://www.eff.org/issues/shadow-regulation>

Shekar, K. (2021). How to Fight Electoral Fake News on Social Media. Retrieved from Freedom Gazette: <https://freedomgazette.in/2021/02/how-to-fight-electoral-fake-news-on-social-media/>

Shekar, K. (2021). Principles, Not Government's Discretion, Can Fix Social Media. Retrieved from Freedom Gazette: <https://www.freedomgazette.in/2021/05/principles-not-governments-discretion-can-fix-social-media/>

Shendruk, A. (2018). Where does the UN get its money? A simple explanation of a complex system. Retrieved from The Quartz: <https://qz.com/1396994/where-does-the-un-get-its-money-a-simple-explanation-of-a-complex-system/>

Shendruk, A., & Rosenthal, Z. (2021). Funding the United Nations: What Impact Do U.S. Contributions Have on UN Agencies and Programs? Retrieved from Council on Foreign Relations: <https://www.cfr.org/article/funding-united-nations-what-impact-do-us-contributions-have-un-agencies-and-programs>

Sloss, D. (2011). Domestic Application of Treaties. Santa Clara Law Digital Commons. Special Procedures of the Human Rights Council. (n.d.). Retrieved from Human Rights Council: <https://www.ohchr.org/EN/HRBodies/SP/Pages/Welcomepage.aspx>

Srinivasan, S. (2021). What is Facebook's Oversight Board? Retrieved from The Hindu: <https://www.thehindu.com/sci-tech/technology/internet/explainer-what-is-facebooks-oversight-board/article34500035.ece>

Swaminathan, M., & Saluja, S. (2020). Essay: Watching Corona or Neighbours? - Introducing 'Lateral Surveillance' during COVID-19. Retrieved from CIS-India: <https://cis-india.org/internet-governance/blog/essay-watching-corona-or-neighbours-introducing-2018lateral-surveillance2019-during-covid201919>

The Information Technology [Intermediaries Guidelines (Amendment) Rules] 2018. (n.d.). Retrieved from Meity: [https://www.meity.gov.in/writereaddata/files/Draft\\_Intermediary\\_Amendment\\_24122018.pdf](https://www.meity.gov.in/writereaddata/files/Draft_Intermediary_Amendment_24122018.pdf)

The Social Media Councils: Consultation Paper. (2019). Retrieved from Article 19: <https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf>

The personal data and information privacy code bill, 2019. (n.d.). Retrieved from <https://saveourprivacy.in/bill>.

The United Nations System. (n.d.). Retrieved from UN: [https://www.un.org/en/pdfs/un\\_system\\_chart.pdf](https://www.un.org/en/pdfs/un_system_chart.pdf)

UN System. (n.d.). Retrieved from UN: <https://www.un.org/en/about-us/un-system>

United Nations Charter. (n.d.). Retrieved from United Nations: <https://www.un.org/en/about-us/un-charter/full-text>

Voluntary Code of Ethics by the Social Media platforms for the General Election 2019. (2019). Retrieved from Election Commission of India: <https://eci.gov.in/files/file/9468-voluntary-code-of-ethics-by-the-social-media-platforms-for-the-general-election-2019/>

Waghre, P., & Ramprasad, V. (2021). Governance of DCNs I: Categorisation of Harms. Takshashila Working paper 2021 - 01.

White House Holds Press Briefing: 26 April | NBC News. (n.d.). Retrieved from <https://youtu.be/3LDSzjWKzOg>

Wingfield, R. (2018). Content regulation laws threaten our freedom of expression. We need a new approach. Retrieved from Global partner digital: <https://www.gp-digital.org/content-regulation-laws-threaten-our-freedom-of-expression-we-need-a-new-approach/>

York, J. C., & Mcsherry, C. (2019). Content Moderation is Broken. Let Us Count the Ways. Retrieved from Electronic Frontier Foundation: <https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways>

Zuidijk, D. (2021). Misinformation Is Bigger Than Facebook, But Let's Start There. Retrieved from Bloomberg: <https://www.bloomberg.com/news/articles/2021-08-30/how-to-fix-covid-vaccine-misinformation-on-facebook-fb-social-media>