

FACEBOOK & TATA INSTITUTE OF SOCIAL SCIENCES
COMMUNITY STANDARDS ROUNDTABLE CONVERSATIONS
7th October 2018
Bengaluru

Venue: The Lalit, Bangalore

Time: 2.30 - 4.30 pm

Social media, and more generally the internet, creates flows of information that continuously remain in a flux. Facebook not only creates a cyberspace for individuals to perform identities the offline world may not allow, and can even sustain a political movement. However, with this freedom, also comes the requirement to understand when content is emancipatory, and when it is oppressive. In order to build a community that tolerates and celebrates diverse views, people need to feel safe, especially those from marginalised communities. The extensive community standards that Facebook has formulated with feedback from users and experts in the field of technology and public safety, stand testament to the fact that in order to create global communities, sensitivity to cultural contexts need to be cultivated. However, considering the diversity that exists on Facebook, this can prove to be a monumental task. While content reports are evaluated in over 40 different languages on Facebook, there is always scope for community leaders from various political positions to contribute to understanding this cultural context. The roundtable discussion attempts to bring together diverse viewpoints of individuals who have expertise in the various forms of social critique. Whether this critique is through a Feminist perspective, an Ambedkarite view, or from the position of tribals, it attempts to articulate the concerns of a socially conscious social media user.

This note attempts to identify some of the relevant concerns from the existing Facebook Community Standards that may be opened for discussion and improved for more sensitivity. The complete community standards document is appended here. Click this link:

<https://www.facebook.com/communitystandards/>

The themes that need to be addressed are listed below as bullet points and the reference to the community standards document is provided in the tables. The areas that need further discussion and contextualization based on personal and social community experiences are articulated as questions in italics. This will be the framework within which the roundtable is organised, and the questions seek to speak to the experiences (both personal and professional) of the participants. An important aspect that holds true for all the categories is that local cultural contexts are clarified during the discussion. The document only serves as a loose framework, and participants are encouraged to focus on or bring up their own concerns, in case they are missing here.

We would appreciate it if each participant can clearly decide on which themes they would like to focus on. Since the time for the Round Table is short, it is going to be helpful to address one or two themes per participant. You may write to us after reading this note. We will structure the round table interventions accordingly.

Broad Themes to be Covered:

1. Bullying

Part III. 12. <u>“We define "attack" as violent or dehumanising speech, statements of inferiority, or calls for exclusion or segregation.”</u>
“Sometimes people share content containing someone else's hate speech for the purpose of raising awareness or educating others. Similarly, in some cases, words or terms that might otherwise breach our standards are used self-referentially or in an empowering way. When this is the case, we allow the content, but we expect people to clearly indicate their intent, which helps us better understand why they shared it. Where the intention is unclear, we may remove the content.”, <i>ibid.</i>
“Protected Categories included here are race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity and serious disease or disability”, <i>ibid.</i>
Part I. 1. <u>“Any content created for the express purpose of outing an individual as a member of a designated and recognisable at-risk group”</u>
“Exposure of vulnerable individuals' identities without their permission”, <i>ibid.</i>

- Personal attacks based on physical, mental or moral values (like calling someone, especially those from marginalised groups (or those from ‘protected characteristics’), “hideous”, “retarded” or “slutty”):
 - ***How does the term ‘personal attack’ not include some vicious ways of attacking someone verbally?***
- Outing people to be from marginalised groups - based in protected characteristics.
 - ***When does an individual not want to be identified as such?***
- Dehumanising speech: For example, comparison to animals or inanimate objects as a form of insult.
 - ***In what way does gender play a role here?***
 - ***There may be many ways in which marginalised gender groups may be talked of in a dehumanising manner. What kind of language is used here that escapes notice?***
- ***Bullying related to disabilities***

2. Sexual exploitation of adults

Part II. 8. In an effort to create space for this conversation while promoting a safe environment, we remove content that depicts, threatens or promotes sexual violence, sexual assault or sexual exploitation, while also allowing space for victims to share their experiences. We remove content that displays, advocates for or coordinates sexual acts with non-consenting parties or commercial sexual services. We do this to avoid facilitating transactions that may involve trafficking, coercion and non-consensual sexual acts. "Sexual services" include prostitution, escort services, sexual massages and filmed sexual activity.

- Threats of sexual violence
 - *In what ways is fear created on social media and how can it be mitigated?*
 - *What are the forms in which feedback on this may be given to Facebook and are there changes that might be desirable?*
- Language of harassment
 - *What are the subtle ways in which individuals are harassed?*

3. Harassment

Part II. 10. Our Harassment Policy applies to both public and private individuals because we want to prevent unwanted or malicious contact on the platform. Context and intent matter, and we allow people to share and re-share posts if it is clear that something was shared in order to condemn or draw attention to harassment. In addition to reporting such behaviour and content, we encourage people to use tools available on Facebook to help protect against it.

“Repeatedly contact a single person despite that person's clear desire and action to prevent that contact”, *ibid.*

“Claims that a victim of a violent tragedy is lying about being a victim, acting/pretending to be a victim of a verified event, or otherwise is paid or employed to mislead people about their role in the event when sent directly to a survivor and/or immediate family member of a survivor or victim.”, *ibid.*

- Repeated contacts from unwanted individuals
- Accusations about lying about the event of harassment
 - *The victim is accused of lying about being harassed. This may be done in very diverse kinds of ways. Thoughts on this.*
 - *What are the ways in which such contact might be blocked?*

4. Mental health

Part II. 6. “In an effort to promote a safe environment on Facebook, we remove content that

encourages suicide or self-injury, including real-time depictions that might lead others to engage in similar behaviour. Self-injury is defined as the intentional and direct injuring of the body, including self-mutilation and eating disorders. We want Facebook to be a space where people can share their experiences, raise awareness about these issues and support each other through difficult experiences, and so we allow people to discuss suicide and self-injury. We encourage people to offer and to seek support from one another in connection with these difficult topics.”

Facebook makes it clear that simply talking about mental health issues is not a problem.

Content is subject to removal if it encourages or glorifies the following:

- Self-injury
- Suicide
- Eating disorders
- Trigger warnings
 - *Is the design of the website conducive for this according to you?*
 - *In what unexpected ways do conversations (even those that may be considered educational or for social awareness) about mental health be triggers for some people?*
 - *Do complaints meet with swift and adequate responses?*

5. Issues particular to minors

Part II. 7. “When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children (NCMEC), in compliance with applicable law. We know that sometimes people share nude images of their own children with good intentions; however, we generally remove these images because of the potential for abuse by others and to help avoid the possibility of other people reusing or misappropriating the images.”

- **Fetishization of minors:** Individuals may post content about their own children with good intentions, however, because there is scope for misuse, such content may be removed.
 - *In what ways does fetishization of minors remain hidden?*
- Solicitation over social media
 - *How does one articulate this idea of keeping children safe from sexual solicitation?*
 - *How might this be operationalised? Is it possible?*

6. Graphic content

Part III. 14. “Our Nudity Policies have become more nuanced over time. We understand

that nudity can be shared for a variety of reasons, including as a form of protest, to raise awareness about a cause or for educational or medical reasons. Where such intent is clear, we make allowances for the content. For example, while we restrict some images of female breasts that include the nipple, we allow other images, including those depicting acts of protest, women actively engaged in breastfeeding and photos of post-mastectomy scarring.”

- Related to death or dismemberment
- Related to any kind of sexual activity
 - *What kind of conversations about sexual activity are policed?*
 - *Is this policing gendered such that marginalised sexualities are more severely punished?*
 - *How is feedback on these grounds received? Are there narratives on this that might illuminate ways forward?*

7. Presence of false news and misinformation

Part IV. 18. “We want to help people stay informed without stifling productive public discourse. There is also a fine line between false news and satire or opinion. For these reasons, we don't remove false news from Facebook, but instead significantly reduce its distribution by showing it lower in the News Feed.”

Part I. 1. “Misinformation that contributes to imminent violence or physical harm”

- Policy rationale to keep false news and satire separate.
- Misinformation that can cause harm.
 - *What are the ways in which communities can identify false news?*
 - *What are the politics of false news? What do you think of the dichotomy of false and truthful news?*
 - *What are the ways in which a platform might deal with this better?*

8. Violence, threat and crime

Part I. 2. “Any association of three or more people that is organised under a name, sign or symbol and that has an ideology, statements or physical actions that attack individuals based on [protected] characteristics.”

“In an effort to prevent and disrupt real-world harm, we do not allow any organisations or individuals that are engaged in the following to have a presence on Facebook:

- Terrorist activity
- Organised hate
- Mass or serial murder
- Human trafficking

● Organised violence or criminal activity
We also remove content that expresses support or praise for groups, leaders or individuals involved in these activities.”, *ibid*.

- Terrorist organisations
- Hate crime organisations, especially those based in religious, racial and caste identities
 - *Is there hate in the glorification of certain religious thoughts and races, especially in the Indian context and Hindutva?*
 - *How do we see hate speech in this context?*

9. Intellectual property rights

Part V. 20. “You own all of the content and information that you post on Facebook, and you control how it is shared through your privacy and application settings. However, before sharing content on Facebook, please make sure that you have the right to do so. We ask that you respect other people's copyrights, trademarks and other legal rights. We are committed to helping people and organisations promote and protect their intellectual property rights.”

- *How can the IPR of marginalised individuals protected?*
- *What are the fears in relation to IPR in relation to Facebook?*

10. Privacy and confidentiality

Part II. 11. “Except in limited cases of newsworthiness, content claimed or confirmed to come from a hacked source, regardless of whether the affected person is a public figure or a private individual.”

“Privacy and the protection of personal information are fundamentally important values for Facebook. We work hard to keep your account secure and safeguard your personal information in order to protect you from potential physical or financial harm. You should not post personal or confidential information about others without first getting their consent. We also provide people with ways to report imagery that they believe to be in breach of their privacy rights.”, *Ibid*.

- Outing private information of an individual (government IDs, address, bank account information, private phone numbers, passwords) except when for newsworthiness.
- Identity theft on the basis of above information
- Impersonation of real individuals or corporations
- Artificially increasing likes for financial gain
- Creating inauthentic profiles of real individuals

Additional terms that may be discussed

- **Newsworthiness:** At a lot of points in the standards, the idea is that certain content may be allowed (for example, nudity, violence, usage of slurs) if there is newsworthiness in the content.
 - *What, in a variety of different contexts, does ‘newsworthiness’ then means?*
- **Public figures:** This term is again used to mark certain content as being allowed. For example, sexually graphic content is not otherwise allowed, but it is if images of ‘public figures’ are photoshopped, it may be allowed.
 - *Who are public figures? On social media, the term ‘public figures’ becomes very loose.*
 - *The policy on bullying does NOT apply to public figures who have a ‘large audience’. Thoughts on this.*

**Note prepared by the TISS team (Chinar Mehta, Lakshmi Lingam and Shilpa Phadke)
6th October 2018.**