



Public Comment Appendix for

2022-012-IG-MR

Case number

Case description

In March 2022, an Instagram account describing itself as a platform for Dalit perspectives posted a video from India showing a woman being assaulted by a group of men. **The woman’s face is not visible. The text accompanying the video states that a “tribal woman” was sexually assaulted and harassed by a group of men in public, and that the video went viral.** The account has around 30,000 followers, mostly located in India.

The content was reported by another Instagram user for sexual solicitation and sent **for human review. Human reviewers determined that the content violated Meta’s Adult Sexual Exploitation policy. Under this policy, Meta removes content “that depicts, threatens or promotes sexual violence, sexual assault or sexual exploitation.”** Following additional human review, **Meta issued a newsworthiness allowance, restored the content and placed a warning screen on the video alerting users it may contain violent or graphic content.** The warning screen prevents users under the age of 18 from viewing the content and requires all other users to click through the screen to view the video. A newsworthiness allowance permits content **on Meta’s platforms that might otherwise violate** its policies where the content is newsworthy and keeping it visible is in the public interest. It can only be applied by specialist teams within Meta, and not by human reviewers who review content at scale.

Meta referred this case to the Board, stating that it demonstrates the challenge in **striking “the appropriate balance between allowing content that condemns sexual exploitation and the harm in allowing visual depictions of sexual harassment to remain on [its] platforms.”** It states that the content was initially removed for violating the Adult Sexual Exploitation policy for depicting non-consensual sexual touching, and that **“because of the graphic and harmful nature of this kind of depiction, the policy does not allow this kind of content to be shared in a condemning context.”** It says it has only allowed such content **“in limited circumstances, upon escalation, and on a case-by-case basis.”**

The Board would appreciate public comments that address:

- Whether Meta’s policies and enforcement practices, including its newsworthiness allowance, appropriately balance protecting users from potentially harmful content and allowing users to raise awareness.
- Insights into the socio-political context affecting the treatment of Dalit and Adivasi individuals and communities, in particular women. These insights may address any relevant power dynamics, practices of physical and social segregation and discrimination, and how existing hierarchies may be reproduced digitally.
- The role of social media in raising awareness of and condemning sexual exploitation and other harmful acts against people from marginalized communities in India.
- The harm caused by allowing visual depictions of sexual harassment to **remain on Meta’s platforms, even if the victims cannot be recognized or identified.**
- The potential for visual depictions of violence against marginalized groups to contribute to an increase in such violence, even when shared in a condemning context.

In its decisions, the Board can issue policy recommendations to Meta. While recommendations are not binding, Meta must respond to them within 60 days. As such, the Board welcomes public comments proposing recommendations that are relevant to this case.



Public Comment Appendix for

2022-012-IG-MR

Case number

The Oversight Board is committed to bringing diverse perspectives from third parties into the case review process. To that end, the Oversight Board has established a public comment process.

Public comments respond to case descriptions based on the information provided to the Board by users and Facebook as part of the appeals process. These case descriptions are posted before panels begin deliberation to provide time for public comment. **As such, case descriptions reflect neither the Board's assessment of a case, nor the full array of policy issues that a panel might consider to be implicated by each case.**

To protect the privacy and security of commenters, comments are only viewed by the Oversight Board and as detailed in the [Operational Privacy Notice](#). All commenters included in this appendix gave consent to the Oversight Board to publish their comments. For commenters who did not consent to attribute their comments publicly, names have been redacted. To withdraw your comment, please email contact@osbadmin.com.

To reflect the wide range of views on cases, the Oversight Board has included all comments received except those clearly irrelevant, abusive or disrespectful of the human and fundamental rights of any person or group of persons and therefore violating the [Terms for Public Comment](#). Inclusion of a comment in this appendix is not an endorsement by the Oversight Board of the views expressed in the comment. The Oversight Board is committed to transparency and this appendix is meant to accurately reflect the input we received.



Public Comment Appendix for

2022-012-IG-MR

Case number

10

Number of Comments

Regional Breakdown

1	3	3	0
Asia Pacific & Oceania	Central & South Asia	Europe	Latin America & Caribbean
0	0	3	
Middle East and North Africa	Sub-Saharan Africa	United States & Canada	

2022-012-IG-MR

Case number

PC-10791

Public comment number

United States and Canada

Region

Withheld

Commenter's first name

Withheld

Commenter's last name

English

Commenter's preferred language

Withheld

Organization

Yes

Response on behalf of organization

Short summary provided by the commenter

In case after case, it has been demonstrated that Facebook wrongfully removes posts made by individuals to raise awareness or protest wrongdoing. Simultaneously, they insist upon leaving up posts that promote hatred, violence, division, and the dissemination of false information.

Full Comment

In case after case, it has been demonstrated that Facebook wrongfully removes posts made by individuals to raise awareness or protest wrongdoing. Simultaneously, they insist upon leaving up posts that promote hatred, violence, division, and the dissemination of false information. The complete lack of a human representative in email or chat based contact only exacerbates these conditions, and drives participants away from the platform.

Link to Attachment

No Attachment

2022-012-IG-MR

Case number

PC-10795

Public comment number

Europe

Region

Withheld

Commenter's first name

Withheld

Commenter's last name

English

Commenter's preferred language

Withheld

Organization

No

Response on behalf of organization

Short summary provided by the commenter

A woman/ man/a female or male being sexually assaulted is not newsworthiness! It is a criminal offence and not for public viewing. Meta needs to review what it deems acceptable!

Full Comment

A woman/ man/a female or male being sexually assaulted is not newsworthiness! It is a criminal offence and not for public viewing. Meta needs to review what it deems acceptable! This is a criminal offence and there are young/impressionable/disturbed people viewing that video who will copy and believe that is acceptable behaviour towards females in particular whom are already at risk and more frequent victims of domestic abuse/rape and murder. It is unacceptable to allow this content of the actual attack to remain viewable online. It is something that should be talked about in a preventative manner but the actual videoed attack should be removed out of respect and privacy for the victim/for any victims of sexual assault/murder/abuse. This video and those alike are to be shown in courts of law and nowhere else. Only when a victim themselves gives their personal permission/ consent and authority for their assault to be displayed online or on any public platform or media coverage should it them be shown. Otherwise it is a breach of the victims personal rightful privacy. And more stringent methods of allowing access to anything sexual/abusive online or elsewhere need to be in place for individuals under the age of 18. Simple box ticking to confirm over the age of 18 or 16 years of age in some cases is not protective or proactive enough in ensuring this type of violation behaviour does not persist. Victims consent and permission must be sought for public viewing of the victims assault/attack as the victim may well

be volatile and under the age of 18 years of age and have been easily coerced. We the public do not need to view individuals sexual assaults as they happened. We do not need negative sexually aggressive graphic abusive content visible and viewable to every single human being on this planet. Many of whom already have very disturbed messed up aggressive mindsets and whom think abuse/rape/assult/murder/theft is acceptable. Viewing victim abuse content makes individuals become desensitised.

[Link to Attachment](#)

No Attachment

2022-012-IG-MR PC-10802 Central and South Asia
Case number Public comment number Region

Shmyla Khan & Noor Waheed English
Commenter's first name Commenter's last name Commenter's preferred language

Digital Rights Foundation Yes
Organization Response on behalf of organization

Short summary provided by the commenter

It is recommended that in considering the facts of this case, special consideration and standards be developed for violence sexual content given the particular impact it has on the viewer, survivors of gender-based violence, and families of victims by developing specific and detailed standards for the violence allowed under the newsworthiness exception. Furthermore, special consideration should be given to the cultural context in which the violence happened and the stigma attached to survivors of sexual violence in that context. Lastly, it is recommended that newsworthy content that is ultimately removed should be archived by the platform in a secure, purpose-limited, and access-limited manner.

Full Comment

Meta has made considerable strides in allowing content to remain on its platforms (on a case-by-case basis), with due warning, for the purposes of news reporting and the general public interest that would otherwise violate its Community Standards (See: Colombia Protests case, Sudan Graphic Video case). However, it should be noted that the sufficiency of these measures has thus far not been tested against content depicting sexual violence. Whilst all forms of violence have tremendous potential for harm to those subjected to it, there is still a distinction between violence in general and sexual violence. The distinction between sexual violence and general violence is an important point for Meta to consider. Meta already has a precedent where it has made a distinction between different kinds of violence such as its separate guidelines for child sexual exploitation, abuse and nudity.

As traumatic as general graphic violence during an armed conflict or through police brutality is, sexual violence has greater potential for harm on its viewer, and has the added consequence of retraumatizing sexual assault survivors and depending on cultural context can further exacerbating the harm suffered from the attack. If such content is allowed to stay up due to newsworthiness, it needs to be censored beyond just the identity of the victim i.e. through diminishing the resolution of the content, **or blurring the act considerably. Meta should consider „únewsworthiness,À here as raising awareness that a particular act or phenomenon is occurring without necessarily fully depicting the act in an uncensored way. Also Meta,Àôs warning label may be accidentally bypassed by people without the requisite literacy to consent to viewing such content. Meta should also consider keeping a threshold i.e. at what level of detail does the depiction of graphic sexual violence become reductive and more harmful than serve the public good. Further, it should account for the fact that visual depiction of sexual violence has a different impact and should be considered at a different (higher) threshold than a verbal description of the same act. (See: Swedish journalist describing child sexual violence case)** Secondly, to account for the fact that the impact of violent content will be dependent on the cultural context and play out along intersectional lines. Dalit, **Adivsai and other members of „úScheduled Castes,À and „úScheduled Tribes,À in India are disproportionately subject to violence and discrimination. As reported by National Family Health Survey 2015-2016, women from Scheduled Tribes (Advisai or Indigenous Indians) and Scheduled Castes (Dalits) were more like to face sexual violence at higher rates, (7.8% and 7.3% respectively) than non-caste women (4.5%). Traditionally, women from these marginalized castes and communities are routinely failed by the criminal justice systems and rarely have access to institutional recourse. Sharing stories about their plight online has been helpful in raising awareness, and there have been instances where public outrage and social pressure was successful in expediting the criminal prosecution of cases and getting these women justice. Social media has been a powerful tool to raise a voice for those who have been silenced by the status quo and those whose conditions have been concealed by state apparatuses and have no other access to justice. However, videos of sexual exploitation can be exploitative in their own right. For example, the videos left up for the purpose of news reporting may result in sexually violent content on the platform proliferated in private group chats for voyeuristic purposes and shared **in a celebratory fashion as opposed to „únewsworthiness,À. This can also devolve into inadvertent pornography, or even a form of revenge porn, where the continuous sharing of the content can be used to humiliate the victim and cause second-hand trauma for others.** Additionally, the context of the video i.e. the location, faces of the perpetrators, clothing of the victim can all be used as identifiers, making the victims life worse by being doxxed even if the face of the victim is not visible. There is still a massive stigma for assault survivors in South Asia and a cultural proclivity towards victim**

blaming, especially in tribal areas. Oftentimes, due to social pressure alone, women from these castes/tribes are unable to pursue their cases to the fullest extent of the law and having this kind of content available even for the purposes of newsworthiness may prove to be damaging and more harmful to the victim. For those cases, Meta should also consider that even if the content is „únewsworthy,“ and is eligible to remain on the platform, requests from the victim or the victim,“s family to remove/ban the content should be given the highest priority and complied with accordingly, in an expedient manner. However, the rights of victim/survivor should not be made solely dependent on reporting for privacy violations as often marginalized communities lack the digital literacy and access to adequately use reporting mechanisms.

In cases content otherwise considered newsworthy or in the public interest is removed from the platform, Meta should keep a highly secured cache of such content that is accessible (on a case-by-case basis) primarily to law enforcement officials, researchers and journalists for the purposes of news reporting, data collection and use of evidence in criminal justice proceedings. Special measures should be taken to ensure that such content, especially dealing with sexual assault, is not copied, shared, or disseminated in any way by those accessing the database.

Link to Attachment

[PC-10802](#)

2022-012-IG-MR	PC-10805	Central and South Asia
Case number	Public comment number	Region

Yasal	Munim	English
Commenter's first name	Commenter's last name	Commenter's preferred language

Media Matters for Democracy	Yes
Organization	Response on behalf of organization

Short summary provided by the commenter

The comment discusses how restoring the video essentially violates the exceptions **laid out in Meta's newsworthiness allowance and compromises the victim's physical safety** by putting them in broader exposure, and fail to take into account the overall safety situation of women in South Asia, where they face harassment in their daily lives. Circulation of such videos are not only triggering for survivors of harassment, but play a role in normalizing crimes against women, especially those belonging to minority or vulnerable communities.

Full Comment

The sensitive nature of the case under discussion warrants a two-dimensional **analysis in light of Meta's policies on newsworthiness and content depicting adult sexual exploitation**. Firstly, restoring a video showing a woman being subjected to an assault of sexual nature publicly by a group of men, to a significant extent, **retires its newsworthiness, for the video's earlier availability on Instagram** before being taken down is very likely to have allowed sufficient time for its widespread circulation on the platform. Given the potential of such sensitive content for virality, as has been demonstrated in similar cases before, the video would not have been exclusive to Instagram by the time it was removed and probably exported to other social media platforms, not to mention news reports (if any) that generally ensue viral content depicting abuse on the internet, amplifying incidents of abuse **in the course of reporting and documentation**. Therefore, **"newsworthiness allowance"** being cited as the reason to restore this video necessitates a protracted analysis to determine whether or not the decision really serves public interest in accordance with the **exceptions as laid out in Meta's policy on newsworthy content**,

especially when the platform has already drawn attention to the incident. Second, although there has been a continuing debate on what exactly constitutes potentially harmful content, which should be timely and effectively regulated by the platform on which it is uploaded and circulated, the video at hand clearly presents certain markers on the spectrum of content that requires immediate attention in terms of **action by the platform**. Given Meta's Adult Sexual Exploitation policies, "content that depicts, threatens or promotes sexual violence, sexual assault or sexual exploitation" is liable for removal or a warning screen on the platform (which may not be very effective in case of underage users signing up with emails of their parents, siblings, etc, or using their accounts). Additionally, it has been categorically stated in Meta's policies that despite possessing newsworthiness, content that "presents a risk of harm, such as physical, emotional and financial harm, or a direct threat to public safety" will be removed. The video under discussion explicitly exhibits these qualifiers that further endorse its removal. Going by Meta's approach to managing content depicting adult sexual exploitation, the platform prioritizes protection of victims and survivors, and so content inadvertently forcing them to relive the trauma, fetching it broader exposure, and **putting their lives at risk clearly violates the Meta's community standards**. That is, however, not to say that the measure will ensure the removal of such posts from the internet altogether. The content can, moreover, prove to be triggering for other women who have faced sexual harassment, which can lead to their exit from the social media platform. Even measures such as hiding the face of the woman is not enough because it is the reality of many women, and circulation of such videos play a big part in normalizing violence against women. According to India's National Crime Records Bureau, crimes against women increased by 26.35% from 2016 to 2021. The top cases were of kidnappings and abduction, rapes, domestic violence, dowry deaths and assaults. The numbers show how prevalent crimes against women is in the society and circulation of videos on sexual assault prove to make online spaces unsafe for women, who already engage with social media platforms in a limited capacity due to familial pressures. In this case, it is highly important to **gauge the possible adverse impact of the video's amplification on the victim and their community**. The pivotal role of social media in bringing to light incidents of abuse, which may not always be picked up by mainstream media, cannot be denied **but, at the same time, the position of vulnerability that social media's broad exposure places particular subjects in contributes heavily towards exacerbating the situation in some cases**. In South Asia, marginalized communities such as Dalits and Adivasis in India (as cited in this case) continue to face oppression and violence in various forms, including rampant hate crimes and societal prejudice, which largely influence and motivate the attacks against them and leads to their discrimination in job markets. Where social media platforms have helped bring these incidents to global awareness conversations taking place online, allowing visuals that depict them being subjected to violence to remain on the platform often escalates chances

for increased hostility against the victims as, at the end of the day, the fact remains that majority of these individuals will have to return to the same circumstances due to ruthless social segregation and widespread discrimination.

[Link to Attachment](#)

No Attachment

2022-012-IG-MR

Case number

PC-10806

Public comment number

Central and South Asia

Region

Merrin

Commenter's first name

Ashraf

Commenter's last name

English

Commenter's preferred language

It for Change

Organization

Yes

Response on behalf of organization

Short summary provided by the commenter

In Indian context, in assessing whether a content qualifies as newsworthy, the long-standing injustice faced by dalit and tribal community and the power imbalance between dominant social groups and marginalized majority should be duly considered. However, this must be balanced with the rights of those who fall at the intersection of caste and other disadvantaged identities, especially gender. Allowing newsworthiness exception to the impugned video instrumentalizes the traumatic experience of the woman, especially if sharing of the video was non-consensual, thereby denying her dignity and privacy. The Board must also examine the basis for the decision of human reviewers and their knowledge of the socio-political context of India.

Full Comment

Dalits are victims of the caste system, a defining feature of Indian society that assigns individuals a ranked social identity at birth. The representation of dalits and dalit issues in the Indian mainstream media is abysmally low. The advent of social media platforms (SMPs) was, therefore, a liberating moment for the dalits as it enabled them to voice out and mobilize their opinion, and to report and condemn violence and atrocities committed against them, sometimes even through posting of images and videos of such acts. Thus, today, SMPs have become crucial sites for dalits to participate in public discourse and to overcome their socio-historical **exclusion and discrimination. In India, one's caste identity affects one's exercise of personal rights, choices, freedom, dignity, access to capital and effective political participation.** Despite constitutional protections that seek to put an end to caste

discrimination and remedy the historical injustice done to dalits, a large majority of them still languish at the lowest socio-economic levels and continue to face brutal forms of violence which reinforces power relations in the caste system. Adivasis (indigenous communities), nomadic, and semi-nomadic tribes, largely lying outside the folds of the caste system, constitute another significant marginalized group in India. While SMPs have created alternate media spaces for the dalits and other marginalized communities, its impact to counter the mainstream media narrative and to make discourses inclusive have so far been limited due to the pervasive online casteist hate-speech, abuse and harassment and the discriminatory practices of the platforms in prioritizing speech of caste-privileged individuals and groups. In this scenario, the newsworthiness allowance of Meta can be of significant help to share images/videos/texts that spotlight the persistent discrimination, exclusion and violence that dalits and other marginalized communities face, which may otherwise be in violation of its content policies. Therefore, in the context of India, the longstanding injustice faced by these communities and the continuing power imbalance between the dominant social groups and the marginalized majority should be given due consideration while assessing whether a content qualifies as newsworthy. But in making this assessment, it is also vital to take into account the rights of those who fall at the intersection of caste and other disadvantaged identities, especially gender. Caste and tribal identity intersects with gender to produce unique sites of discrimination and violence. This should alert us to viewing acts of discrimination and violence against women from a single axis of either caste/tribe or gender. Caste cannot be delinked from patriarchy, as the entire edifice of caste system is built on endogamy which is maintained through a tight regulation of female sexuality and labour. Within caste societies in India, manhood is defined both by the degree of control men exercise over women and the degree of passivity of the women. Hence, humiliating women of another caste is perceived as degrading the manhood of those castes for their inability to protect their women. Further, as much as Dalit women have been targets of violence by the dominant castes, they are also oppressed by men within their caste, thanks to the pervasive influence of Brahminical patriarchy. Further, the experience of dalit and adivasi women should not be conflated because of the differences in their social locations and the multi-layered abuse that the latter experience. The video in question depicting a tribal woman being subject to assault and non-consensual sexual touching by a group of men **prima facie violates Meta's adult sexual exploitation** policy. The video was purportedly shared by the Dalit instagram handle to condemn such violence against women from marginalized communities. While it is important to report and condemn such acts of violence against people from marginalized communities, the gender implications of sharing the video cannot be overlooked. First, it is not clear if the video was shared with the consent of the woman who was attacked. Non-consensual sharing of the video, **even if the woman's face is blurred**, denies her right to bodily integrity, and her agency to take control of the narrative

of her experience. This constitutes a violation of the constitutionally guaranteed right to privacy and dignity of the woman. Second, the circulation of video of sexual assault can have a dehumanizing effect on the victim and subject her to voyeuristic gaze and pleasure of the viewers, whereby she is reduced to an object to be judged by them. The easy and ubiquitous presence of such graphic content on social media can also numb the audience response to such acts of violence. Third, scholars have argued that circulation of such graphic videos depicting violence against real persons can have the effect of emboldening perpetrators, and inducing fear in future victims. This can lead to the reinforcement of existing gender and social hierarchies that subordinate the position of the tribal woman in this case. Fourth, permitting newsworthiness allowance for the impugned video has the effect of instrumentalizing the traumatic experience of the woman (especially if the video was shared non-consensually), thereby denying her dignity and privacy. If the intention of the Instagram handle was to raise awareness and condemn violence against marginalized communities, it could have been done by narrating the **incident without sharing the video, which is allowed by Meta's Adult Sexual Exploitation Policy**. Due to the above reasons, the content should be denied the benefit of newsworthiness allowance completely. Finally, we recommend an **examination of the basis for the decision of Meta's human reviewers to adjudge the impugned content as newsworthy**, and their knowledge of the deeply discriminatory structures of caste and patriarchy in India. Dalits are victims of the caste system, a defining feature of Indian society that assigns individuals a ranked social identity at birth. The representation of dalits and dalit issues in the Indian mainstream media is abysmally low. The advent of social media platforms (SMPs) was, therefore, a liberating moment for the dalits as it enabled them to voice out and mobilize their opinion, and to report and condemn violence and atrocities committed against them, sometimes even through posting of images and videos of such acts. Thus, today, SMPs have become crucial sites for dalits to participate in public discourse and to overcome their socio-historical exclusion and **discrimination. In India, one's caste identity affects one's exercise of personal rights, choices, freedom, dignity, access to capital and effective political participation**. Despite constitutional protections that seek to put an end to caste discrimination and remedy the historical injustice done to dalits, a large majority of them still languish at the lowest socio-economic levels and continue to face brutal forms of violence which reinforces power relations in the caste system. Adivasis (indigenous communities), nomadic, and semi-nomadic tribes, largely lying outside the folds of the caste system, constitute another significant marginalized group in India. While SMPs have created alternate media spaces for the dalits and other marginalized communities, its impact to counter the mainstream media narrative and to make discourses inclusive have so far been limited due to the pervasive online casteist hate-speech, abuse and harassment and the discriminatory practices of the platforms in prioritizing speech of caste-privileged individuals and groups. In

this scenario, the newsworthiness allowance of Meta can be of significant help to share images/videos/texts that spotlight the persistent discrimination, exclusion and violence that dalits and other marginalized communities face, which may otherwise be in violation of its content policies. Therefore, in the context of India, the longstanding injustice faced by these communities and the continuing power imbalance between the dominant social groups and the marginalized majority should be given due consideration while assessing whether a content qualifies as newsworthy. But in making this assessment, it is also vital to take into account the rights of those who fall at the intersection of caste and other disadvantaged identities, especially gender. Caste and tribal identity intersects with gender to produce unique sites of discrimination and violence. This should alert us to viewing acts of discrimination and violence against women from a single axis of either caste/tribe or gender. Caste cannot be delinked from patriarchy, as the entire edifice of caste system is built on endogamy which is maintained through a tight regulation of female sexuality and labour. Within caste societies in India, manhood is defined both by the degree of control men exercise over women and the degree of passivity of the women. Hence, humiliating women of another caste is perceived as degrading the manhood of those castes for their inability to protect their women. Further, as much as Dalit women have been targets of violence by the dominant castes, they are also oppressed by men within their caste, thanks to the pervasive influence of Brahminical patriarchy. Further, the experience of dalit and adivasi women should not be conflated because of the differences in their social locations and the multi-layered abuse that the latter experience. The video in question depicting a tribal woman being subject to assault and non-consensual sexual **touching by a group of men prima facie violates Meta's adult sexual exploitation** policy. The video was purportedly shared by the Dalit instagram handle to condemn such violence against women from marginalized communities. While it is important to report and condemn such acts of violence against people from marginalized communities, the gender implications of sharing the video cannot be overlooked. First, it is not clear if the video was shared with the consent of the woman who was attacked. **Non-consensual sharing of the video, even if the woman's face is blurred,** denies her right to bodily integrity, and her agency to take control of the narrative of her experience. This constitutes a violation of the constitutionally guaranteed right to privacy and dignity of the woman. Second, the circulation of video of sexual assault can have a dehumanizing effect on the victim and subject her to voyeuristic gaze and pleasure of the viewers, whereby she is reduced to an object to be judged by them. The easy and ubiquitous presence of such graphic content on social media can also numb the audience response to such acts of violence. Third, scholars have argued that circulation of such graphic videos depicting violence against real persons can have the effect of emboldening perpetrators, and inducing fear in future victims. This can lead to the reinforcement of existing gender and social hierarchies that subordinate the position of the tribal woman in this case. Fourth,

permitting newsworthiness allowance for the impugned video has the effect of instrumentalizing the traumatic experience of the woman (especially if the video was shared non-consensually), thereby denying her dignity and privacy. If the intention of the instagram handle was to raise awareness and condemn violence against marginalized communities, it could have been done by narrating the **incident without sharing the video, which is allowed by Meta's Adult Sexual Exploitation Policy**. Due to the above reasons, the content should be denied the benefit of newsworthiness allowance completely. Finally, we recommend an **examination of the basis for the decision of Meta's human reviewers to adjudge the impugned content as newsworthy, and their knowledge of the deeply discriminatory structures of caste and patriarchy in India.**

Link to Attachment

[PC-10806](#)

2022-012-IG-MR

Case number

PC-10807

Public comment number

Europe

Region

Withheld

Commenter's first name

Withheld

Commenter's last name

English

Commenter's preferred language

Withheld

Organization

Yes

Response on behalf of organization

Short summary provided by the commenter

Caste discrimination is a serious issue across South Asia severely affecting the rights of Dalits online and offline. We would like to direct the Oversight Boards attention to the publication 'Caste hate-speech' - The report examines caste-hate speech with a particular emphasis on digital media. It is based on extensive research and offers numerous real-life examples of caste-hate speech and its relation to caste-based hate crime. This report would inform the board with crucial background knowledge. The report can be downloaded here <https://idsn.org/portfolio-items/caste-hate-speech-addressing-hate-speech-based-on-work-and-descent/>

Full Comment

Please read the report here <https://idsn.org/portfolio-items/caste-hate-speech-addressing-hate-speech-based-on-work-and-descent/>

Link to Attachment

Received & Reviewed, Not Publishable.

2022-012-IG-MR

Case number

PC-10808

Public comment number

Asia Pacific and Oceania

Region

Ellen

Commenter's first name

Kusuma

Commenter's last name

English

Commenter's preferred language

SAFEnet

Organization

Yes

Response on behalf of organization

Short summary provided by the commenter

As a social media company, Meta is responsible for developing internal policies that respect human rights and conduct thorough content due diligence. Besides, meta must set standard limits on content that cannot be uploaded and provide remediation mechanisms for victims affected by the content. Achieving more equitable moderation of social media content for marginalized groups is urgent and critical; therefore, the platform features allowing the visibility of content moderation should be considered, as well as where and when content moderation should and should not be viewed.

Full Comment

Sexual violence against Dalit women in India and the dissemination of information and videos on the internet demonstrate that social media is a massive and challenging platform to control. For example, there was the spread of rape content against Steubenville High School students in 2013. The video went viral on Instagram, YouTube, and school emails (Social Media Victims, 2022). Although the content can be reported manually, Meta or another platform impression provider may be appointed to review the content. However, even though the victim's face was not recognized and the social media platform had already set an age limit for the audience, the video continued to circulate. These cases highlight how Meta lacks appropriate, contextual, and standard content filters. For example, the omission of a Dalit woman being raped can have a massive and widespread impact on one ethnic/group, resulting in the demonstration of violent attitudes toward marginalized groups seen as prevalent and appropriate. As in the Steubenville High

School rape case in 2013, allowing content to spread will result in how social media encourages the rape culture. Rape is justified, and the victim is considered to have contributed to the rape, for example, for clothing reasons. The impact on other victims is shame and hesitation to report cases of violence that have happened to them as the construction of a rape culture that perpetuates victim blaming. Talking about raising awareness of humanitarian issues, we can look at the case of Mahsa Amini. Mahsa Amini is a 22-year-old woman who died after being arrested by the police for not following the Iranian government's rules for wearing the headscarf. In Iran, protests and large-scale actions have been called in response to Mahsa's death. Of course, the problem does not end in Iran. This case is receiving international attention. Social media has also become a means of disseminating information about Mahsa's case. A search on Instagram for the hashtag #mahsaamini yielded 312 thousand posts, with #MahsaAmini yielding up to 114 thousand posts. After all, this is just a rough search, and there may be content that includes this hashtag unrelated to the Mahsa Amini case or content that is related to Mahsa Amini but does not include the hashtag in the post. However, we can use this rough data as a reference to see that social media is more than just a place to connect with family and friends; it also plays a role in information dissemination. Studies show that the visibility of content moderation increases the accountability of social media moderation controllers and addresses concerns around unfair content removal or account suspension. Achieving more equitable moderation of social media content for marginalized groups is urgent and critical; therefore, the platform features allowing the visibility of content moderation should be considered, as well as where and when content moderation should and should not be viewed (Thach et al., 2022). According to Thach et al. (2022), ideal content moderation allows users from marginalized communities to point out inconsistency and unfair treatment regarding content removal or content suspension. It's also worth noting to holistically improve the automated content moderation system with AI and machine learning algorithms to avoid disproportionately targeting marginalized groups. In any way possible, Meta should be more transparent about the deployment, operation, and training for the AI-operated content moderation. Human intervention, on the other hand, is required to assist the entire process. Meta should investigate the safeguards and scrutinize the role of third-party service providers, such as contractors, in risk assessment, as in human rights risks in the moderation system. Continuous systemic reviews are required to prevent third-party employees from posing human rights risks. Back to the case of sexual violence against women in Dalits, various efforts have been made to oppose all forms of violence against women. Allowing this content becomes double-edged; on the one hand, it can help spread information and awareness that sexual violence against women from marginalized communities still exists. However, it has the potential to escalate sexual violence against marginalized communities. At its core, Meta, as a social media company, is responsible for developing internal policies that respect

human rights and conduct thorough content due diligence. Besides, Meta must set standard limits on content that cannot be uploaded and provide remediation mechanisms for victims affected by it. Such remediation mechanisms could allow users to contact local sexual assault services for assistance by placing a sign on violent content **through extensively expanding Meta's counterspeech Search Redirect Program** to a feature that can reach out to marginalized groups in need of assistance. It is also possible for individuals to apply for the removal of content whose rights have been violated. To conclude, Meta must consider the upcoming risk to victims of sexual violence in Dalits and the rights of victims of sexual violence in general, even if the victims cannot be identified. In addition, the omission of such content creates an obligation for Meta to develop a common transnational understanding that sexual violence should not be normalized, let alone tolerated, through their platform. Moderating content and allowing such violent content shouldn't be based only on newsworthiness, as someone's or certain group's traumatic experience shouldn't be only look at as "content". The objective of the restoring the content shouldn't stop at the "newsworthiness" but whether or not the content could add values and contribute to the betterment of the marginalized communities themselves, and be "part of the voice" from the communities.

Link to Attachment

[PC-10808](#)



Public Comment Appendix for

2022-012-IG-MR

Case number

End of public comments