



DigitalRightsFoundation
"KNOW YOUR RIGHTS"

DIGITAL RIGHTS FOUNDATION PUBLIC COMMENT(S) ON OVERSIGHT BOARD CASE 2023-002-IG-UA and 2023-005-IG-UA ON VIOLENCE AGAINST WOMEN (SWEDISH) (JOINT CASES)

Submission Author: Noor Waheed

Submission Date(s): March 23rd & May 4th 2023

Meta's Hate Speech Policy is taking too literal an approach in this case. Meta must consider the underlying reasoning why it has certain conditions in its hate speech policy such as "unqualified behavioral statements" levied against a particular sex or gender. The pressure here lies in the understanding of the expression "unqualified" in the case of male-female violence especially when there is a studied disproportionate amount of evidence that men perpetuate violence against both other men, women and gender minorities. The description of the post did not say "all men" simply "men". Even in cases that use "all men", the intention is often to shed light on the gender hierarchy in society rather than literally condemning all men as violent perpetrators. Additionally, Meta needs to be clear about why they have this requirement in their hate speech protocols in the first place, namely to counter harmful stereotypes especially towards vulnerable groups and gender minorities - so assumptions like "all trans women are frauds" or "all women are promiscuous" "all men from XYZ race are dangerous" are more likely to fall within the ambit of this "hate speech" requirement because it does tangible harm offline and poses real-life consequences by exacerbating already existing societal problems, application and interpretation of policies cannot be divorced from real-life harms. Any statements regarding any gender or sex should be assessed within the context.

Testimonials and the ability to share one's personal experiences is a fundamental aspect of freedom of expression, even if those experiences are negative, disturbing, graphic or troubling in nature. The ability to share testimonials also forms the basis of the appeal of social media. However, user-testimonials are obviously not taken under oath and there is no concept of perjury on social media. In the event that a testimony may be false, there are local legal mechanisms by which a court-order request of removal can be made to Meta. However, Meta should be cognizant that provided such testimony is not levied against public figures, it cannot be tried as defamation and that often instances where testimonies may be true but may lack the necessary threshold of evidence such mechanisms (such as defamation claims) can be abused by perpetrators in positions of power to silence survivors (as was seen in some cases of the MeToo movement). Additionally, a testimony may be true without having to bear the burden of being universal or exclusively representative.

According to key facts published by the World Health Organization in 2021, globally about 1 in 3 (30%) of women worldwide have been subjected to either physical and/or sexual intimate partner violence or non-partner sexual violence in their lifetime. Most of this violence is intimate partner violence. Worldwide, almost one third (27%) of women aged 15-49 years who have been in a relationship report that they have been subjected to some form of physical and/or sexual violence by their intimate partner.¹ Additionally, the US Department of Justice's Bureau of Justice Statistics reports in its Criminal Victimization, 2020 Supplemental Statistical Tables that regardless of the victims' sex, a greater percentage of violent incidents involved male offenders (79%) than female offenders (17%) or offenders of both sexes (3%).² Given the alarming prevalence of this phenomenon globally and that most violent crimes, including intimate partner violence towards all genders, are statistically largely perpetuated by men, it is irresponsible for Meta to render testimonials by women and gender minorities against men as "unsubstantiated". It also damages a survivor's credibility and deprives them of one of the foremost ways by which they can raise awareness of the situation, i.e on social media.

In the event Meta is concerned about an allowance for testimonials or anecdotal evidence being taken out of context to the detriment of the work of activists, journalists, human rights defenders and journalists and users themselves, they can institute an internal mechanism of post tagging that can allow posters to pre-tag their work or disclaim the content of the post as a testimonial or an anecdote. This would allow Meta to re-evaluate the content of the post in light of that context as opposed to a direct application of their various content regulation policies. (With obvious exceptions being in case of dangerous criminal conduct, misuse of the tagging feature etc. – in which case a high standard of transparency behind why an account is receiving a strike should be maintained). That this post was flagged multiple times shows a concerning lack of awareness on part of the human content moderation team at Meta. Moderators must also be mandatorily subjected to gender-sensitivity and GBV training that can help better distinguish the social and political nuances when making decisions involving user experiences around gender and sexuality. This will allow them to make more informed decisions on whether or not the context is helpful or harmful.

ADDITIONAL COMMENT (SWEDISH GBV CASE 2023-005-IG-UA CASE 2)

In light of the second case added to the deliberation of the Swedish Gender-Based Violence (GBV) case, the Digital Rights Foundation would like to add the following points for consideration along with its original submission. The term "man-hater" (both assigned or self-identified) is a broad generalization about men, however, it cannot be considered a form of discrimination or bullying in the context it was used in. The original post is in the context of an ongoing conversation around gender-based violence, of which men are statistically and

¹ World Health Organization, Violence Against Women Key Facts, (WHO, 2021)
<https://www.who.int/news-room/fact-sheets/detail/violence-against-women>

² Bureau of Justice Statistics, Criminal Victimization 2020 - Supplemental Statistics Tables (US Department of Justice, 2022)
<https://bjs.ojp.gov/library/publications/criminal-victimization-2020-supplemental-statistical-tables>

historically the largest perpetrators – not only against women but towards other men. It is imperative that Meta does not categorize this kind of generalization as a form of “discrimination” in the traditional sense. Discrimination standards exist in light of hierarchical systems of power based on gender, race, class, etc. and thus dependent on the context of the speaker and recipient of the comment. It is understandable that Meta may consider this a “double standard”, however given that such moderation has the potential to silence conversations around GBV a balance must be struck. While gender identity is a protected characteristic, broad comments on “men” do not have the same potential for harm as broad comments about women, gender minorities, sexuality, race, ethnicity and religion. In considering whether or not a post is harmful, Meta should look to other context clues, for example, any post calling for violence against men, bullying specific individuals or making racially disparaging comments in addition to attacking their gender can be removed.

*Read the Oversight Board’s full decision on this case [here](#)

**See all submitted Public Comments [here](#)