

facebook

Product Policy Forum

April 9, 2019

TOPICS: Gendered Cursing, Non-Sexual Adult Nudity, Delisting Terrorist Entities

Agenda

- 1 RECOMMENDATION: Gendered Cursing
- 2 RECOMMENDATION: Non-Sexual Adult Nudity
- 3 HEADS UP: Delisting Terrorist Entities

RECOMMENDATION:
Gendered Cursing

Gendered Cursing

Overview

Issue: We do not remove targeted cursing (e.g., cunt, bitch) directed at public figures, but we do remove derogatory sexual activity claims (e.g., whore, slut). Research and external engagement tells us that female-gendered targeting cursing feels especially intense to the person targeted. We want to ensure that all users feel safe on our site; however, restrictions on profanity may result in over-enforcement on political discussion, heated debate, and colloquial language.

Summary to Date:

- Met with 16 external stakeholders
- Convened 6 working group meetings and several 1:1 consultations

Recommendation for discussion: Define and remove female-gendered cursing as higher-intensity attacks

Gendered Cursing

Status Quo

Remove:

- Derogatory sexual activity claims for everyone
- Statements of intent or advocating to engage in sexual activity for all private individuals and for public figures when they are *purposefully exposed* to the attack
- Purposeful exposure is when the content is on the target's real account or page or when they are tagged or mentioned in the content

Targeted Cursing:

- Targeted cursing refers to profane terms that are used to attack an individual
- We remove targeted cursing aimed at minors, regardless of who reports it
- We remove targeted cursing aimed at private adults, when the target reports it to us
- We also remove Complex Objects dedicated to targeted cursing aimed at any private individual or at minor public figures

We want to protect public figures from the worst types of attacks while allowing discussion of people who are in the public sphere. As such, under our current policy, we remove certain attacks when the public figure is “purposefully exposed” to content attacking them (if, for example, the public figure is tagged in the content), which heightens the harassment.

Gendered Cursing

Examples

ALLOW ✓



Under our status quo policies, neither of these examples would be removed since Hillary Clinton and Kim Kardashian are public figures.

Gendered Cursing

Examples

DELETE X



Here, however, we'd remove the content under our policies because Jacob Sartorius is a public figure, but a minor, and, as previously mentioned, we do remove derogatory sexual claims even when they're about public figures like Teresa May.

Gendered Cursing

Product Policy Research Findings

- **Gendered harassment is a serious online problem.** A 2015 UN Women's report estimates that 3 out of 4 women will be harassed online. The most susceptible women are:
Young (18-24), public figures (e.g., celebrities, journalists), and people who are engaged politically.
- **Online harassment or aggression, includes targeted cursing when aimed as a gender-related insult and additional experiences** of high volume of offenses (too many to manage), excessive exposure (tagging), and severity of the act (e.g., death or physical threats).
Curse words are sensitive to contextual and cultural influences; context is vital to understand the intent of use. Four of the most harmful aggression types on Instagram include: threats, disrespects, betrayals, and unwanted contacts.
- In the context of setting harassment reporting rules, internal research suggests that **reporting include transparency, empathic communication, and assurance of privacy for those reporting.**

Gendered Cursing

Option 1: Treat the word “*cunt*” (and country-specific equivalents) as a slur

Pros:

- Consistent regionally
- Removes some toxic and harmful attacks

Cons:

- Cunt differs from other slurs and is not always used in an attack context (i.e., it’s often used self-referentially or in benign ways)
- Does not address other high-intensity attacks, including other gender-based attacks where relevant

Gendered Cursing

Option 2: Define and remove female-gendered cursing attacks [Recommendation for discussion]

- **Female-Gendered Cursing Attacks Definition:** Profane female-gendered terms for genitalia (e.g., cunt, pussy) and female-gendered profane animal terms (e.g., bitch)
- Remove for all minors and private individuals, remove for public figures with purposeful exposure
- Remove unless used as term of endearment or benignly, with burden of clarity on user (e.g. I love you, you bitch)

Pros:

- Acknowledges expert views that female-gendered cursing is more intense than other forms of cursing
- Protects all genders equally from highest-intensity cursing
- Expands protections to all public figures
- Provides most protections to private adults and all minors
- Consistent application regionally when terms are used in an attack

Cons:

- Does not address lower-intensity cursing, which, coupled with behaviors and other intensifiers, increases intensity of attack
- Assumes that gender-based attacks are the highest-intensity cursing attacks universally, which may not be the case

Gendered Cursing

Option 3: Define and remove gendered cursing

- **Gendered Cursing Definition:** Profane terms or phrases that are *not* used as exaggeration in speech, but are used to attack individuals by using terms for genitalia to refer to a person (e.g., bitch, cock)
- Remove for all minors, private adults (self-reporting required), and for public adults with purposeful exposure

Pros:

- Removes toxic and harmful attacks
- Protects all genders equally
- Expands protections to all public figures
- Provides most protections to private adults and all minors
- Consistent regionally

Cons:

- Possible over-enforcement for terms used to target males that aren't usually thought to be hateful
- Greater burden of reporting on private adults vs. public adults, consequently offering greater protection to public adults
- Removes broad types of cursing that may not be the highest-intensity cursing attacks
- Possible over-enforcement on discourse around public figures, which may limit political criticism

Gendered Cursing

Summary of Option 2 & Option 3

	Examples	All Minors	Private Adult	Public Figure Adult w/o Purposeful Exposure	Public Figure Adult with Purposeful Exposure	Page/Group /Event/IG account dedicated to abuse
Option 2 (Rec): <i>Female Gendered Cursing Attacks</i>	Cunt, Bitch, Pussy	Remove	Remove	Allow	Remove	Remove
Option 3: <i>Gendered Cursing Attacks</i>	Cunt, Dick, Cock	Remove	Remove with name/face match	Allow	Remove	Remove

Gendered Cursing

Option 4: Remove high-intensity cursing through marketized lists

- Define high-intensity cursing taking local context into consideration, allowing us to remove gendered cursing in countries/regions where attacks are most harmful, while removing other high-intensity cursing attacks in countries/regions where gendered cursing is not the most harmful attack
- Remove high-intensity cursing, treat other targeted cursing per status quo

Pros:

- Cultural sensitivity - allows us to remove the worst attacks in different regions
- Takes local context into account

Cons:

- Difficult to operationalize by requiring maintenance and vetting of lists
- Minimal difference in outcome despite operational challenges

Gendered Cursing

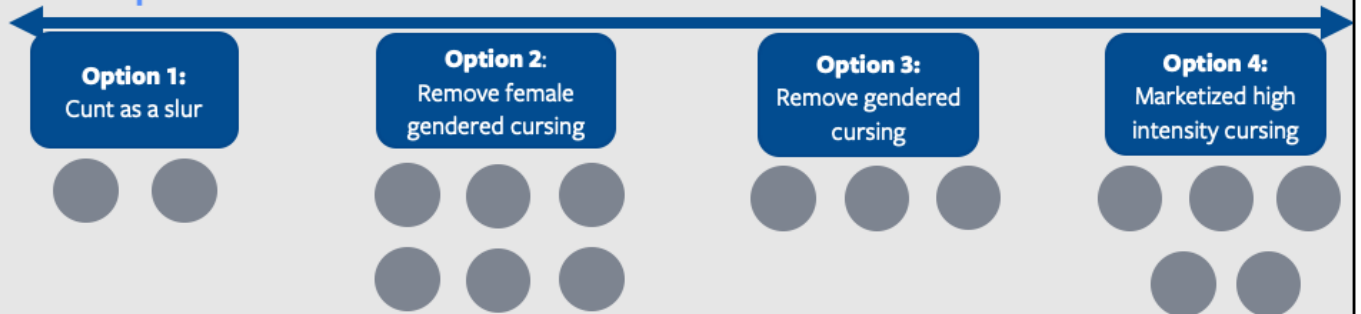
External Outreach



We spoke to 16 experts globally, including anthropological linguists, cognitive linguists, women's rights organizations, and safety organizations.

Gendered Cursing

Snapshot of External Outreach



Gendered Cursing

Next Steps

- Fine tune definition for targeted cursing and female gendered cursing
- Clarify operational guidelines for endearment and benign carve-outs
- Explore where low-intensity cursing attacks are coupled with behaviors and/or other intensifiers that increase the intensity of the attack, similar to high-intensity cursing (e.g., *“you good for shit, stupid fucking idiot”*)
- Enforcement criteria for reports not stemming from target, when on target’s profile (e.g., Marcy writes “Happy Birthday Bitch” to Anna, and John reports)
- Launch

RECOMMENDATION: Non-Sexual Adult Nudity

Note: This policy recommendation did not result in consensus at the Product Policy Forum and has not been adopted. The subject matter experts leading this effort will reconvene internal and external working groups to do additional analysis and research.

Non-Sexual Adult Nudity

Overview

Issue: We do not allow adult nudity except in limited circumstances related to breastfeeding and health. We have increasingly made exceptions for cultural nudity; however, these exceptions are difficult to define and operationalize and we do not want to open the platform to sexualized nudity.

Working Groups

- 4 cross-functional working groups
- 17 conversations with external stakeholders

Recommendation for discussion

- Continue to remove uncovered female nipples per status quo policy
- Add further carve-outs for cultural nudity and pregnancy nudity
- Clarify operational guidelines for non-binary nudity enforcement

Non-Sexual Adult Nudity

Status Quo

REMOVE

- Real images of nudity
- Nudity is defined as images of uncovered nipples of those who identify as female as well as visible genitalia

EXCEPT

- Uncovered female nipples in the context of: breastfeeding; mother and baby after-birth; mastectomy/gender confirmation surgery; female breast exposure in breast health context or in the act of protest
- Visible genitalia and/or anus in medical/health context

When we talk about content shared in the context of health, we mean: birth-giving and after-birth moments (including both natural vaginal delivery and cesarean section), gender confirmation operation and genitalia self-examination for cancer or disease prevention/assessment.

WARNING:
Mature Content

Non-Sexual Adult Nudity

Cultural Nudity



Under our status quo policy, all of these images would come down as violating.

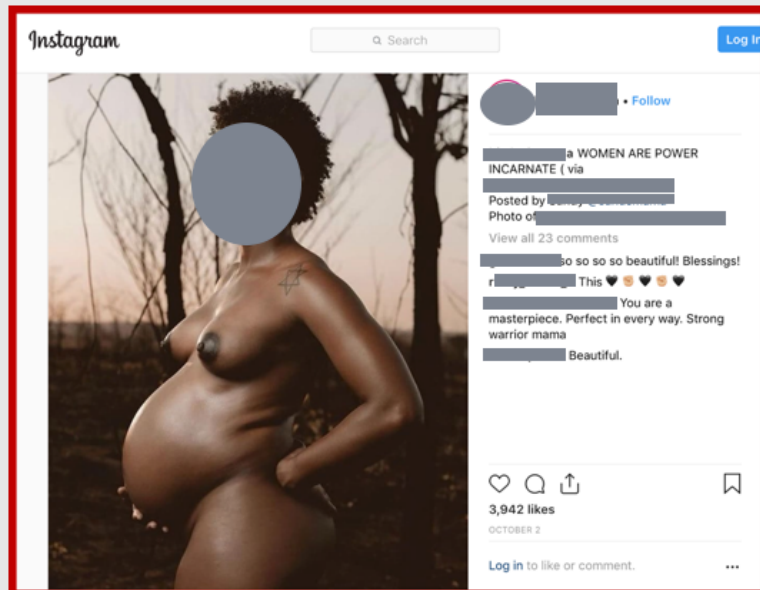
- Top right - an article published by the Brazilian newspaper Folha de S. Paulo containing a picture taken by one of the most renowned Brazilian photographers — Sebastião Salgado — depicting naked indigenous people
- Left hand side - Indoni is a South African women's heritage group that aims to increase awareness of local culture. They have protested outside Facebook and Google offices claiming that "imperialist censorship."
- Right bottom – Photo of an Indian politician meeting Jain naked monks. Jainism is a religion in India, and some monks who adhere to the region do not wear clothes.

One of the things that we've heard internally and externally is that

we shouldn't remove this stuff from Facebook as it's often included in classroom textbooks and is part of a country's culture.

Non-Sexual Adult Nudity

Pregnancy Photos



This post would come down as well because it depicts nudity, but none of our existing carve-outs apply.

We've received feedback from parenting and maternal groups who don't think we should be taking this content down.

Non-Sexual Adult Nudity

Product Policy Research Findings on Non-Binary, Gender Fluid Nudity

There is limited internal and external research, as a result it's unknown how many non-binary people exist in the population or on the platforms.

Internal and external research finds non-binary and trans people use social media to seek and engage with community and expressions of their body, including imagery, are part of this community's needs and conversations.

Internal research found trans users experience high levels of harassment on Facebook; the most reported are verbal abuse offline, comments, upsetting content, private messages, and being outed.

Implications for policy:

- Restricting the way this community exists on the platforms, may further marginalize this community and mitigate online conversations that determine how gender is categorized more broadly across societies.
- Development and implementation should consider ways to mitigate harassment for this population.

Non-Sexual Adult Nudity

Option 1 - Status Quo

REMOVE

- Real images of nudity.
- Nudity is defined to include images of uncovered female nipples & visible genitalia.

Pros:

- Easily explicable
- Limits the sharing of pornographic content

Cons:

- Removes possible socially acceptable forms of nudity
- Inconsistent enforcement

Non-Sexual Adult Nudity

Option 2 – Age-gate and/or interstitial non-sexual uncovered nipples

- Interstitial/ age-gate content that includes uncovered visible nipples but does not include sexually suggestive pose nor visible genitalia. (Remove when sexually suggestive pose present)

Pros:

- Would eliminate carve-outs/exceptions
- Equitable treatment across genders
- Allows for more socially acceptable nudity
- Interstitial may protect vulnerable groups

Cons:

- Opens the platform up to more nudity
- An interstitial may imply negative judgment on the content
- Content without context would still be removed
- Tooling changes required

Non-Sexual Adult Nudity

Option 3 – Carve-outs [Recommendation for discussion]

- Interstitial and age-gate content that meets indicators of cultural/indigenous nudity.
- Interstitial and age-gate content when shared in explicit pregnancy/motherhood context.
- Continue to remove content that depicts uncovered female nipples and doesn't meet one of carve-outs.
- Clarifying operational guidelines for non-binary nudity enforcement.

Cons:

Pros:

- | | |
|--|---|
| • Easily explicable | • May not have context necessary to confirm consent (of person photographed / depicted) |
| • Limits the sharing of pornographic content | • Carve outs open up possibility of false positives |
| • Allows for more socially acceptable nudity | • Defining cultural nudity is challenging |
| | • Slippery slope - context can be subjective |
| | • Tooling changes required |

Non-Sexual Adult Nudity

Next Steps

- Finalize the tooling and actions in our review tool for both Facebook and Instagram
- Launch TBD
- Quality test in second half of the year to determine consistency of enforcement

Discussion

Question: The examples of cultural / indigenous nudity represented here are from Sub-Saharan Africa, India, etc. What about places like the Nordics? How are we defining cultural and indigenous”?

Answer: We are trying to focus on historic, social and cultural norms that exist across different groups of people, whether that’s religious groups, racial groups, social groups. In this vein, we’re speaking with academics and experts. We are also working with local teams to explore and fully understand the scope of cultural nudity. One of the examples that came up in conversation with local teams is ice bathing in Finland. We want to define these terms clearly and objectively so that something like sunbathing in Florida wouldn’t count as cultural nudity. Also worth noting that we want to move away from using

the word “indigenous” because it’s something that is only applicable in some countries and we want these carveouts to be global in scope. We are instead focused on cultural nudity.

Comment: You walked through our status quo policy, which includes carveouts for things like breastfeeding and cancer awareness. I worry that allowing indigenous or cultural nudity will be difficult to defend (and won’t be all encompassing) because it is difficult to define as already mentioned.

Comment: And what about the issue of consent / agency? What about the case of Amanda Todd where she posted a video online exposing her breasts, and ultimately ended up killing herself. This wasn’t on our platform, but does this policy raise that risk? There are a lot of nuances here that we want to be sure we have thought through fully.

Response: On the question of age, this policy is purely focused on adult nudity. When it is difficult to identify if the person in the image is an adult or minor, we always err on the side of assuming the person is a minor and providing enhanced protections. On the question on consent, this is a big one. We are working closely with our operations and product teams to explore different tools that might help with the issue of consent when the content is self-reported.

Comment: The discussion makes it clear to me that we aren’t ready to move forward with adoption the proposed recommendation. We’ll continue to apply our newsworthy policy where we have additional context available to ensure that consent isn’t an issue. And we’ll go back to the drawing board to understand if there’s more we can do in the realm of cultural nudity.

HEADS-UP:

Delisting Terrorist Entities

Delisting Terrorist Entities

Overview

Issue: We want to devise a framework that strikes the right balance between keeping terrorist entities off the platform and considering entities for removal from our list of terrorist organizations when they have disavowed ties to terrorism. However, it will be difficult to assess whether an entity has genuinely disavowed terror-motivated violence and could lead people to believe that we are relaxing enforcement against terrorism.

Goals:

- Create cross-functional working groups to explore options for delisting terrorist entities.
- Define clear criteria that will inform our assessment on when to remove an entity from our internal lists.

Delisting Terrorist Entities

Status Quo - Designation Criteria

We designate when:

- The entity is on the US State Department's list of Foreign Terrorist Organizations, OR
- Meets all of the following **internal** designation criteria:
 1. A non-state actor,
 2. Use of violence against civilians or governments, AND
 3. In pursuit of a political, religious or ideological aim

Delisting Terrorist Entities

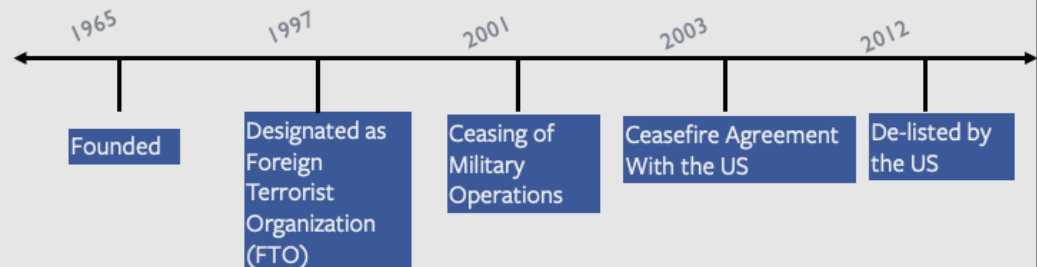
Relevant Case Studies



Gulbuddin Hekmatyar/Hezbi Islami



People's Mujahedin of Iran



These case studies will be helpful in informing the work we do here – are the signals identified here the right ones to be looking at when delisting individuals and entities? What else can/should we think about? One of the things we want to think about, for example, is the timing of events – how long is long enough to establish disavowal.

Delisting Terrorist Entities

Next Steps

- Convene cross-functional working groups to discuss different policy options
- Engage with external stakeholders and our research teams to benefit from prior thought on these issues

Working Groups scheduled:

- APAC Working Group
- NA / EMEA Working Group

facebook