

From: Sayed Otaru [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=]
Sent: 4/14/2023 10:45:49 PM
To: Rong []@meta.com]
Subject: Re: [A/C Priv] Top Reels Integrity Review for US Teens

From: Rong []<[]@meta.com>
Sent: Thursday, April 13, 2023 12:04 PM
To: Max Eulenstein <[]@meta.com>; Tomas Kabbabe <[]@meta.com>; Blake []<[]@meta.com>; Aameek []<[]@meta.com>; Nikita []<[]@meta.com>; Sayed Otaru <[]@meta.com>; Stephanie []<[]@meta.com>; Andrew []<[]@meta.com>; Sharon []<[]@meta.com>; Adam []<[]@meta.com>; Haotian []<[]@meta.com>; Zak []<[]@meta.com>; Melissa []<[]@meta.com>; Kate []<[]@meta.com>; Mike []<[]@meta.com>; Dhiraj []<[]@meta.com>
Cc: Ben []<[]@meta.com>; Karim []<[]@meta.com>; Ethan []<[]@meta.com>; Jordi []<[]@meta.com>; Eric []<[]@meta.com>; Erion []<[]@meta.com>; Melissa []<[]@meta.com>; Ming []<[]@meta.com>; Shreekar []<[]@meta.com>; Alice []<[]@meta.com>
Subject: [A/C Priv] Top Reels Integrity Review for US Teens

Hi Max, Relevance, and IGWB Leads,

We wanted to share some (rather concerning) early data insights and next steps, as we looked into recommendability and media quality of US Teens’ Top recommended Reels.

Context

- Anecdotal evidence indicates that Teens see higher levels of non-rec content. An ex-Meta employee recently shared feedback that his teenage step-daughter’s Reels feed is highly offensive with content including: “people chugging alcohol and vomiting, blatantly racist jokes, pedophilia jokes, videos of people dying in car crashes, etc.”.
- There is a non-zero chance that Tiktok gets banned in US, and there will be a flush of US teen viewers into IG. We want to make sure the Top Impression Reels that represents IG are healthy and thus leave a safe reputation.
- The IFR equivalent study showed that teen and adults consume drastically different media. Hence, there seems to be a knowledge gap in terms of what media teen users consume.
- Apple age gating risk (the threat to increase IG’s age rating from 12+ due to sexually suggestive Reels) urges IG to provide a safe experience for US cold start, which tends to be dominated by popular top impression media.

Early data (March 2023)

- **11% of US Teen’s Weekly Top 200 Reels** on Reels Tab/Chaining was non-recommendable.
 - These 200 Reels made up 3% of US Teens’ Reels consumption on Reels Tab/Chaining.
 - Here’s a media fiddler of some non-rec examples within the top 200 Reels for US Teen on Reels tab. To provide some examples: the top 1 Reel by VPV used the N word and depicted gun (regulated goods), the 9th Reel by VPV depicted violence and physical bullying, the 13th involved sexually suggestive positioning, and the 36th was disturbing.
 - This prevalence number is very high in comparison with similar segments:



	Country	Teen/Adult	Surface	Metric	Prevalence	Implication
1	US	Teen	Reels Tab	Top 200 Content-level	11.0%	
2	US	Teen	Connected Feed	Top 200 Content-level	2.5%	[Row 1 v.s. 2]: US teens choose to follow relatively safe media and accounts, yet we recommend more offensive content.
3	US	Adult	Reels Tab	Top 200 Content-level	4.0%	[Row 1 v.s. 3]: Top 200 Reels was a lot less safe for US teens compared to US adults
4	US	Teen	Reels Tab	Overall Impression-level	5.9%	[Row 1 v.s. 4]: For US teens, top Reels are more dangerous than overall Reels
5	US	Adult	Reels Tab	Overall Impression-level	5.1%	
6	Global	Teen	Reels Tab	Overall Impression-level	4.2%	
7	Global	Adult	Reels Tab	Overall Impression-level	4.4%	

- **Why did we not find out how dangerous US Teen Top Reels were earlier?**
 - We did not pay enough attention to country variance. We have global *teen* impression level prevalence measurement (row 6), which is in fact lower than global *adult* impression level prevalence (row 7). However, from the rows 4-7 in the table above, we see that US prevalence is higher than the rest of the world, and US teen prevalence even higher than US adults.
- **Why does our current enforcement leak so much non-recommendable content to the top for US Teens?**
 - The current label-based enforcement set up doesn't cover top US Teen consumed media because they are so different from the global top N, yet our human review of viral content (VCE) is set up to protect global top N. Only 3 of the US Teens top 200 made it to the global top 100, and 13 made it to the global top 200.
 - The Sensitive Content Control (SCC) setting for teens doesn't help because only 7% US teen WAU receive stronger enforcement through SCC. For reference, on a global scale, 11.2% teen WAU are under SCC - high enforcement setting.
 - Moreover, the feature to "default new teen users to high SCC enforcement" generally has a low coverage because (1) the feature uses a U16 age threshold, (2) it's determined on stated age rather than predicted age, and (3) it only applies to new users who registered after August 2022.

Next Steps

1. Immediate Actions

- We opened a SEV3 [S335572](#) and are actively working on it. (Done Apr 12)
 - We will escalate to SEV2 if the US teen top content prevalence exceeds 15%.
- We moved over some existing Viral Content Enforcement resources to focus on popular content among US Teens, though at the risk of reducing some enforcement for the general population at the midst of a non-rec prevalence surge on Reels Tab. (Done Apr 13)
- We will set up measurement for US Teen Top N and impression-level prevalence country breakdown. (ETA: Apr 21)
- We are working with PCO to understand media quality beyond the Recommendation Guidelines. (ETA: Apr 26)

2. The following items need follow up to give recommendations. We will circle back by the listed ETAs.

- We will follow up on discussion for team to take an additional goal for US Teen top content non-rec prevalence. (POC Rong [REDACTED] ETA Apr 26)
- a. We will assess labeling capacity to review media with more than a certain number of recommended impressions, and we may request additional rater resources accordingly. (POC Rong [REDACTED] ETA Apr 26)
- b. We plan to review enforcement model performance targeting US teen recommended media and upsample for US teens if needed. (POC Ming [REDACTED] ETA Apr 26)
- c. We plan to partner with IGWB to tune teen-specific enforcements through [Age Gating](#). (POC Ming [REDACTED] ETA Apr 21)
- d. We will recommend to corresponding POCs to add teen samples in standard experiment reviews across Relevance. (POC Rong [REDACTED] ETA Apr 21)