

From: Diego San Martin [REDACTED]@meta.com
Sent: 3/17/2023 4:34:26 PM
To: Gabriel [REDACTED] [REDACTED]@meta.com
Subject: Re: [tasks] T148254533: # **Discovery Intelligence Ops (DIOps): Child Sexual Exploitation and Nudity Proactive Sweep**

T148254533 - # **Discovery Intelligence Ops (DIOps): Child Sexual Exploitation and Nudity Proactive Sweep**

RECENT ACTIVITY

Diego San Martin commented:
Thanks, Rishi!

Charita - For the accounts that were not escalated to PRESC for review but you still found creepy/implicitly sexualizing can you please escalate to the Child Safety taskforce via this contact form - <https://www.internalfb.com/butterfly/form/862300394563344?>

If you need anything else reviewed please also use that form. Thank you!

[Comment](#) · [Close](#) · [Unsubscribe](#) · [Assign](#) · [Add Subscribers](#) · [Tag](#) · [Set Priority](#) · [Star](#)

STATUS	Open
PRIORITY	None
PROGRESS	In Progress
SIZE	None
OWNER	Charita [REDACTED]
CREATED	Mar 16 at 11:42 PM by Charita [REDACTED]
DESCRIPTION	Synced with an internal group post: https://fb.workplace.com/groups/725214898469503/... Discovery Intelligence Ops (DIOps): Child Sexual Exploitation and Nudity Proactive Sweep TL;DR IG DIOps proactively reviewed IG surfaces by assessing the search results of 16 Queries, Tags and Emojis (e.g. tweenxstars, #cutegirls and Little 🍷🍷). 50+ suspicious accounts were found, and we observed patterns indicating the existence of a much larger net work of similar accounts. Working with the Inauthentic Behavior Team for behavioral insights regarding suspected networks. Follow-up with DiscoS on the discoverability of Non-Recommendable accounts. Coordinate with the Child Safety Team to optimize our approach to tackling CEI-specific discoverability risks. Overview On February 1st, YouTuber Natly Denise posted a 30-minute video titled "Instagram Pizza", in which she identified Instagram accounts that allegedly promoted and sold child pornography [Source (https://www.youtube.com/watch?v=TTC8W0FL-QI)]. Pizza and cancer emojis (e.g. 'Little 🍷🍷' or 'petite 🍷🍷'), known as secretive communication tactics by Pedophiles [Source (https://www.dailymail.co.uk/.../Paedophiles-using-cheese-pizza...)], were found in bios for a majority of these accounts. The YouTuber also mentioned accounts present on the platform with computer-generated images of "child-like faces with womanly bodies". Instagram recommen



ended some of these malicious accounts. In response to these concerns, IG DIOps started an investigation into these claims.

Potential Risks

DIOps investigated the claims made by Natly Denise in an effort to ensure the following violation types are not present on IG Search Platform:

****Child Exploitation Imagery**** - Users leveraging IG to share inappropriate imagery of children.

Unwanted Interactions - Pathway for malicious and unsolicited interactions with minors.

Nudity and Sexual Activity - Implicit and/or explicit Sexual Activity and Nudity Imagery of minors and adults on the platform.

Methodology & Findings

Keywords for Queries were derived from the original YouTube video of Natly Denise and sweeps were conducted to gauge risk on the following surfaces: TypeAhead, Top Tab, Accounts Tab and Tags Tab on Instagram.

Among 50+ *suspicious accounts mentioned* in the video, we observed the following patterns:

Accounts shared media of scantily clad children's pictures of which some are AI-manipulated or simply stolen from other accounts.

Classified as not "real" pictures but CGI-animated (eg. 17841457894122934) or in case of organically derived: they seem to be just allegedly 'stolen' from other users (eg. 17841457443294988, 17841457303684351).

Accounts having '3d' in their IG Username (eg. 'claire_3dwaifu', 'lola_unreal3d') indicative of 'Computer Generated Imagery'.

Telegram accounts were found, possibly facilitating purchases of CGI media of child-like images in sexually suggestive positions.

We observed concerning patterns which might indicate the existence of a network of malicious accounts:

IG Users 'evebloom3d', 'lolawaifu3d', 'allnatural_3dbabes', 'unrealcandies_cgi' sharing Child Sexual Exploitation Imagery, having 31k+ Common Followers.

Total Followers of 'evebloom3d', 'lolawaifu3d', 'allnatural_3dbabes', 'unrealcandies_cgi' were 83k+.

We flagged 116 IGIDs; 101 Non-Recommendable posts of Sexually Explicit and Suggestive Imagery with Insufficient Coverings and 15 posts violating Community Guidelines for Child Safety.

Actions Taken

Manually demoted 101 Insufficient Coverings Media content in violation of Instagram Recommendation Quality Guidelines.

Investigation in Centra in order to look at possible Connections between the 50+ suspicious accounts.

Escalated a request to the PRESC Team for deletion of 24 suspicious accounts (26 were automatically disabled over time) and 15 posts in violation of the Community Standard Guidelines.

Next Steps

Coordinate efforts with the Inauthentic Behavior Team to analyze and discover potential malicious networks (T148055563 (t148055563)).

Work to seek advice from the Child Safety Team on how to improve future investigations.

Continue to monitor Search for harmful trends and proactively investigate potential problem areas.

cc: Andrew [REDACTED] Sayed Otaru, Dhiraj [REDACTED] Santhosh [REDACTED] Jean [REDACTED] Robin [REDACTED] Ben [REDACTED] Ashley [REDACTED] Alex [REDACTED] Allen [REDACTED] Cory [REDACTED] Marina [REDACTED] Julie [REDACTED] Paulo [REDACTED]

Child Sexual Exploitation and Nudity sweep sheet (<https://docs.google.com/.../1hpxwHa-3JKyBR-WZ2DKXAD-rV-Q.../edit>)

TAGS

oncall oncallteam-ig-diops-search-integrity-response groupsynced ig_diops

SUBSCRIBERS

Malia [REDACTED] Diego San Martin Gabriel [REDACTED] Anthony [REDACTED] Rory [REDACTED] Wayne [REDACTED] Hudson [REDACTED] Rob [REDACTED] (PDO) Luca [REDACTED] Rachel [REDACTED] Harmanjit [REDACTED] Charita [REDACTED] Samuel [REDACTED] Sefa [REDACTED] Cory [REDACTED] Rishi [REDACTED] Kevin [REDACTED]

COMMENT HISTORY

Diego San Martin

Thanks, Rishi!

Charita - For the accounts that were not escalated to PRESC for review but you still found creepy/implicitly sexualizing can you please escalate to the Child Safety taskforce via this contact form - <https://www.internalfb.com/butterfly/form/862300394563344?>

If you need anything else reviewed please also use that form. Thank you!

Mar 17 at 4:34 PM

Anthony [REDACTED]

Wayne, Malia [REDACTED] (tagged above) and Luca [REDACTED] would have the most context on profile and network detection capabilities for CEI. We did have a network detection pipeline built for CEI networks on FB and I want to say that Luca extended it to IG in early 2022 but him or Malia would need to confirm

Mar 17 at 4:33 PM

Jyoti [REDACTED]

#thanks Charita for leading the proactive sweeps on this topic and sharing this note. It is very insightful. The insights you shared would be helpful to followup with different teams who could help with mitigating these to an extent.

Mar 17 at 3:28 AM

Rishi [REDACTED]

Charita [REDACTED] You can escalate through here:

https://www.internalfb.com/.../Safe.../child_safety_escalations/(https://www.internalfb.com/.../Safe.../child_safety_escalations/)

Diego San Martin or Sefa [REDACTED] may also have some guidance here.

Mar 17 at 1:44 AM

Wayne [REDACTED]

Thanks for driving this proactive sweep, Charita [REDACTED] Having a toddler daughter, this hits extra hard.

Thinking through existing steady state levers and how we can move fast here to mitigate risk:

1/ We do have an existing CEI search inform module. We should consider adding these emoji terms to empty SERP or CEI triggering if we do not have evidence the terms are inherently violative. (cc Cory [REDACTED] for input here)

2/ IGPR has precedence for filtering out accounts based on profile elements. Harmanjit [REDACTED] and Hudson [REDACTED] know more details about how to set this up. If these are known secret terms, we can seek to find/enforce accounts with such terms in the bios plus block them using IGID for specific queries.

3/ I noticed you used the term "network." Tagging Anthony [REDACTED] as I recall he previously worked on a system that would identify networks of bad actors. Anthony, any POCs you can point us to?

Given this was initiated by an external impetus, I'd like to see where we can expedite short term mechanisms to remove these accounts and/or reduce their viz.

Longer term, do we have a CEI problem area POC we can reach out to, to better understand their long term plans?

Mar 17 at 12:37 AM

Charita [REDACTED]

Rishi [REDACTED] Hi Rishi, we are not in touch with the i3 investigation teams, but are open to share our findings and work collaboratively to protect Children and the safety of our users

Mar 17 at 12:30 AM

Rishi [REDACTED]

Thanks for doing this, Charita. Are you in touch with the i3 investigative teams on these findings?

Gabriel Rory Malia FYI on potential signals for MaCSA

cc Wayne [REDACTED] on account patterns to be potentially used in Search

Mer 16 at 11:59 PM

Customize your notifications at [Email Settings](#)