

AMENDED Exhibit 210

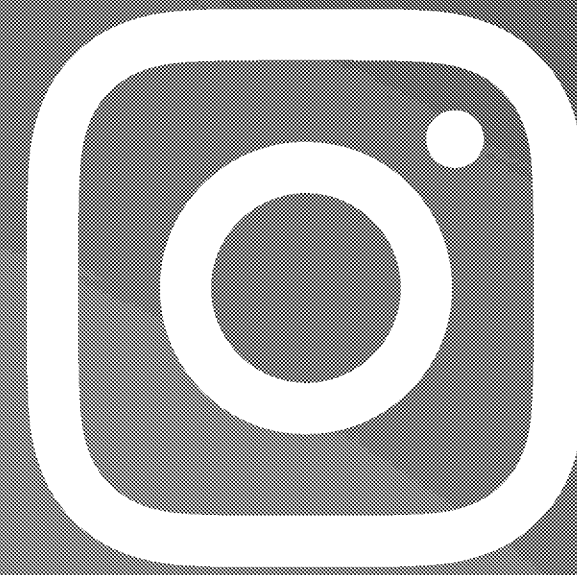
PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

Case No.: 4:22-md-03047-YGR

MDL No. 3047

In Re: Social Media Adolescent Addiction/Personal Injury Products Liability Litigation

Well-Being



Child Safety on IG

Actor behaviors and opportunities

Hitomi Hayashi-Branson
Product Research
hitomih@instagram.com

Daniel Alpert
Data Scientist
dalp@fb.com

GOAL

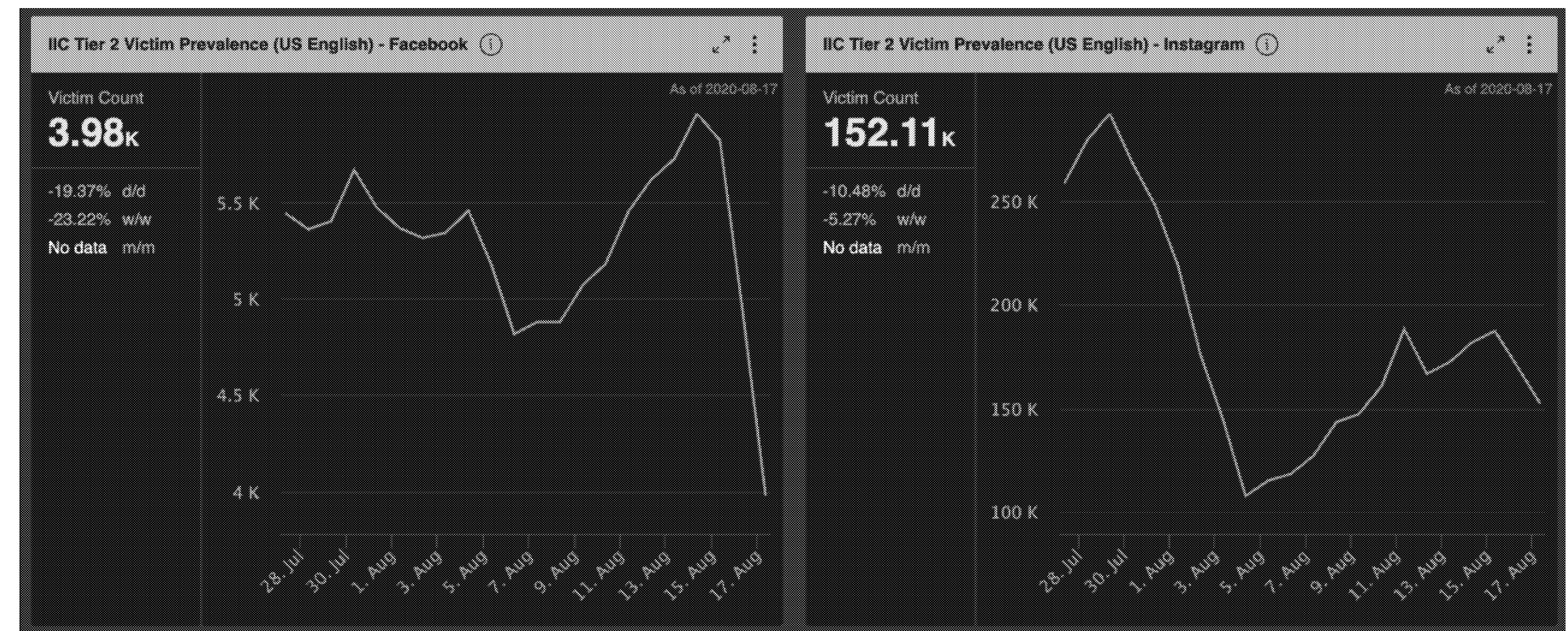
To summarize and present early data explorations on bad actors who seek to interact inappropriately with or harm minors on Instagram.

TRIGGER WARNING - This deck includes images that illustrate bad actors behaviors. These include (but are not limited to) IIC, CEI, and MS.

CONTEXT

A recent exploration into IIC on IGDM, revealed that there are **~38x as many victims on IG as on FB**. This is mostly due to there being many more minors on IG messaging with adults. Despite this, **CEI deletes on IG are ~21x lower and IIC auto-enforcing on IG is 90x lower** compared to on FB.

A unique challenge that Instagram faces is understanding **new and unique surfaces** (e.g., hashtag pages) that are likely vectors of abuse which we have yet to explore deeply.



Announcing IIC Tier 2 Prevalence Measurement on IG Direct: <https://fb.workplace.com/groups/135709743752034/permalink/624917758164561/>

TL;DR

1

Actors' intent range from malicious to non-malicious. Malicious intent includes those motivated by underlying sexual interest in children, those would take advantage of situations the engage in CEI, and those that are commercially driven.

2

Likely targets are minors, women, teen accounts, and vulnerable populations. Women appear to be more highly targeted by actors. Vulnerable populations that may be at risk include: LGBTQ youth (particularly young boys), and users who may struggle with mental health issues.

3

Passive/Low effort behaviors include: Seeking/discovering, viewing, and collecting inappropriate content, stalking targets, commenting inappropriately, and networking with other actors.

4

Active/High effort behaviors include: Contacting and developing relationship with targets, creating/posting and distributing content, and scamming targets.

5

We are only at the beginning of our journey. We have a long road ahead of understanding, sizing, detecting, filling gaps in policy, developing solutions, and ensuring consistent and appropriate enforcement.

DATA ON THE SIZE OF THE PROBLEM AND GAPS

38X MORE

victims on IG compared to FB.

21X FEWER

CEI deletes compared to FB.

90X FEWER

IIC auto-enforcing compared to FB.

DATA ON ACTOR BEHAVIORS

993K

actors had ≥ 5 impressions on content flagged for CES over a 14 day period.

133K had ≥ 10 impressions and 2.6K had ≥ 50 impressions.

10.8K

users following CEI-related hashtags identified by CS investigators. Examples include: #gaykid, #sexykid, #13yearoldboy, #gaykidz.

We've also identified 40 additional hashtags that appeared on at least 10 pieces of CEI content in the last week.

55

actors left ≥ 5 comments across 10.5K CES posts over the past 7 days.

The most comments left by an actor was 67 comments.

DATA ON ACTOR BEHAVIORS

2.2K

actors removed for IIC in last 14d.
These accounts follow 375K teen
accounts on IG.*

Note: This may be double
counting teens who are followed
by multiple violating accounts.

27.6K

threads between actors, ultimately
flagged for CEI, MS, IIC, and a teen
over a 7 day period.

3.9K threads had over 20
messages over 28 days.

~500 actors (flagged over a 7 period), who had
threads with at least 10 teens over the last month.
~200 actors had >20 message threads with at
least 5 teens over the last month.

5-7.5K

pieces of CEI actioned content
produced daily by 3-5K actors.

1-2.5K pieces of MS actioned
content produced daily by
800-2K actors.

OPPORTUNITIES

Note: Opportunities in green are those I believe would be primarily owned by the IG Child Safety team.

- 1 **Understand and size the problem** - Explore actors (e.g., their intent, strategies), targets (e.g., who is at risk, how do they respond), interactions (e.g., how do they start, evolve, end), and difference across groups (e.g., countries). Size the problem by key behaviors and interactions.

- 2 **Understand risk across IG surfaces to enable surface-level interventions** - Explore how actors target and interact with minors across different surfaces. Pay particular attention to high risk and ephemeral surfaces - Live, Reels, Story. Develop ways to prevent actors across surfaces, implement strategies to enforce blockers and penalties. Also seek to understand virality on IG (e.g., on Reels) and how this may increase risk.

- 3 **Identify likely bad actors** - Explore ways to improve our ability to identify bad actors on IG across different surfaces (e.g., improving age prediction, understanding more nuanced behavioral signals). [Primarily CI]

- 4 **Improve detection** - Improve detection based on our understanding of the problem and actors on IG. Explore how we can detect less overt or disguised abuses (e.g., text overlaid on Stories image). Continue to improve detection as our understanding grows and as actors evolve to continue evading our detection system. Given E2E encryption, explore non-content based signals that we can use to define and detect violations. [Primarily CI]

Gaps in Child Safety Detection and Enforcement: <https://docs.google.com/document/d/17EN8Qps-GAogmuUOB3LGr0Kpj49UKFxLfvmxDFnYjuM/edit>

- 5 **Assess where we are (compared to Facebook)** - Assess where we are for measurement, detection, reporting, safeguards, enforcement, and product solutions across surfaces. For all, explore the applicability of existing solutions and adopt as appropriate or develop new solutions.

- 6 **Develop product solutions** based on a deeper understanding of this problem on Instagram. Identify and fill limitations and gaps in current solutions. Draw insights and solutions from FB as appropriate. For current and proposed product solutions, refer to "Child Safety Landscape" (cited below).

- 7 **Ensure correct actions for new/niche surfaces** - Challenges that Instagram faces are largely due to new surfaces that we have less experience with. E.g., Despite efforts to calibrate FB classifiers to IG, CEI deletes on IG are ~21x lower and IIC auto-enforcing on IG is 90x lower compared to on FB.

- 8 **Ensure consistency in actions we take across violating content and across platforms.** In a summary of Gaps in Child Safety Detection, investigator found inconsistencies in enforcement and escalations. Additionally, enforcement of viral content in Stories was found to be inconsistent across platforms. [Primarily CI/CO]

Child safety landscape - Instagram: https://docs.google.com/document/d/1qxutBnVxxt_ggf4rwPQ1EQcrILWPrWuc-Yvl5yMt1qg/edit
 Gaps in Child Safety Detection and Enforcement: <https://docs.google.com/document/d/17EN8Qps-GAogmuUOB3LGr0Kpj49UKFxLfvmxDFnYjuM/edit>
 All Roads Lead to Improving Age Prediction: <https://fb.workplace.com/groups/1411637962558841/permalink/1429502370772400/>
 Protecting youth from bad actors on Instagram: https://fb.workplace.com/permalink.php?story_fbid=638544290380093&id=100026735829495

