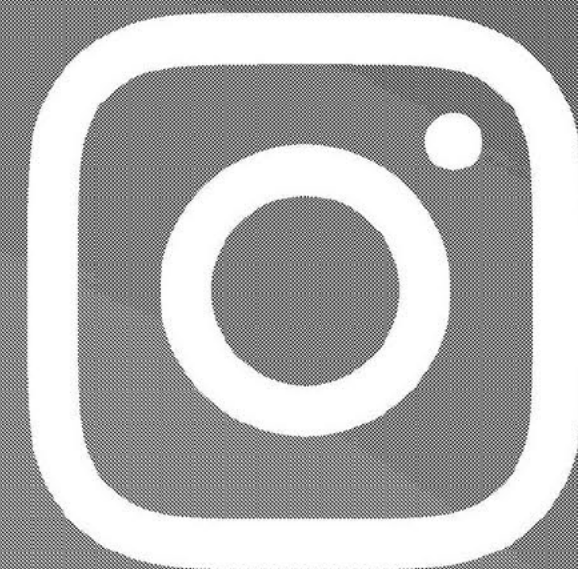


Well-Being



Child Safety on IG

Actor behaviors and opportunities

Hitomi Hayashi-Branson
Product Research
[REDACTED]@instagram.com

Daniel [REDACTED]
Data Scientist
[REDACTED]@fb.com

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY'S EYES ONLY)

META3047MDL-031-00192305
METANMAG-002-01771928

PLTF EX-00992_0001



GOAL

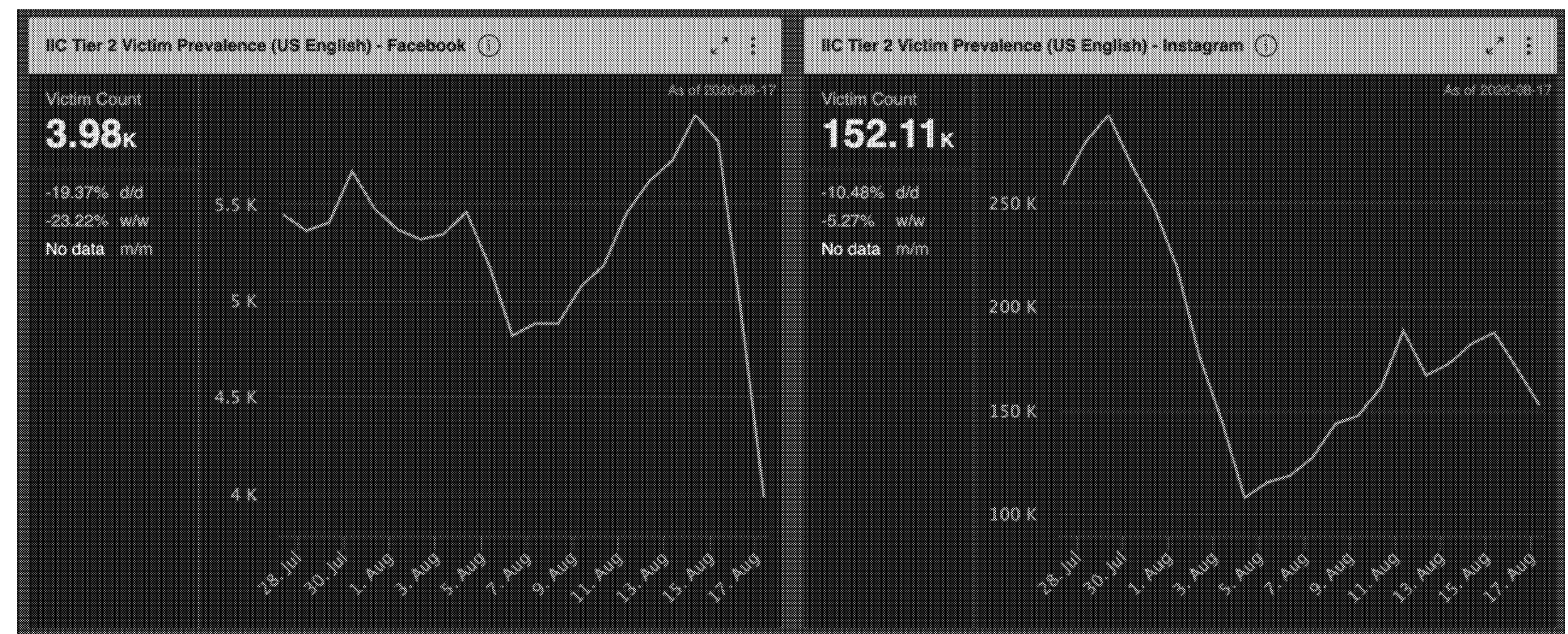
To summarize and present early data explorations on bad actors who seek to interact inappropriately with or harm minors on Instagram.

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]

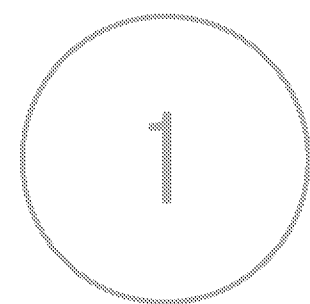
Child safety landscape - Instagram: https://docs.google.com/document/d/1qxutBnVxxt_ggf4rwPQ1EQcrlLWPrWuc-YvI5yMt1qg/edit
Gaps in Child Safety Detection and Enforcement: <https://docs.google.com/document/d/17EN8Qps-GAogmuUOB3LGr0Kpj49UKFxLfvmxDFnYjuM/edit>
All Roads Lead to Improving Age Prediction: <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

OUTLINE

1 Overview of actors and actor behaviors

2 Deep dive into actor behaviors - Examples and data

3 Opportunities



THEME

Overview of actors and actor behaviors

WHO ARE LIKELY ACTORS AND TARGETS?

Actors

- Accounts with a high proportion of CEI or MS
- Accounts with a history of CEI, MS, IIC
- Accounts associated with accounts with a history/high proportion of IIC, CEI, or MS (e.g., followers, users who interact with posts, use that IIC accounts send messages to)
- Fake accounts (e.g., adults posing as teens to distribute content)

Note: Some actors, especially those who don't post inappropriate content or interact with minors, may not have violations or may not be picked up by our current classifier.

Targets

- Minors
- Females
- Teen accounts
- Vulnerable populations (e.g., LGBTQ youth - particularly gay or questioning boys, users who are part of support groups)

Protecting youth from bad actors on Instagram - How it happens, what we're doing about it, and opportunities: <https://fb.workplace.com/notes/hitomi-hayashi-branson/protecting-youth-from-bad-actors-on-instagram-how-it-happens-what-were-doing-abo/615052212729301/>

ACTOR INTENT

There are a range of actors intents. The below taxonomic categories were developed by the online Child Safety team and reviewed by external experts (POC: John [REDACTED]).

Malicious

- 1 Preferential
Motivation is based on an inherent and underlying sexual interest in children
- 2 Situational
Those who take advantage of situations and opportunities that present to engage with CEI.
- 3 Commercially driven
Facilitate child sexual abuse for the purpose of financial gain.

Non-malicious

- 4 Unintended harm
People who may not mean to cause harm to the child depicted in the CEI share.
- 5 Situational - Risky
Users who share and consume a lot of porn, through which they come into sexualised contact with CEI.
- 6 Minor non-exploitative
Teens who are engaging in developmentally normative behavior,

ACTOR BEHAVIORS - WHAT DOES THIS LOOK LIKE ON INSTAGRAM?

PASSIVE/LOWER EFFORT BEHAVIORS

These behaviors are more behind the scenes. They are likely more prevalent than higher effort behaviors (see next slide), less directly harmful to targets.

1 Seeking/Discovering

4 Stalking

2 Viewing

5 Commenting

3 Collecting

6 Networking

ACTIVE/HIGHER EFFORT BEHAVIORS

The below are key interactions that actors have with targets. These behaviors range from low to high prevalence and are directly harmful to targets.

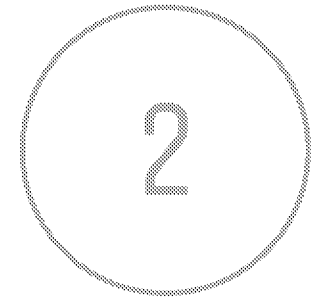
1 Contacting

4 Distributing content

2 Developing relationships

5 Scamming

3 Creating/posting content



THEME

Deep dive into actor behaviors - Examples and data

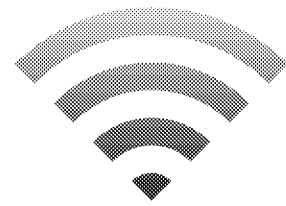
Note: Examples provided in the following slide were previously escalated for review and were removed if policy violating and ignored if not violating.

PASSIVE/LOWER EFFORT BEHAVIORS

- 1 Seeking/Discovering
- 2 Viewing
- 3 Collecting
- 4 Stalking
- 5 Commenting
- 6 Networking

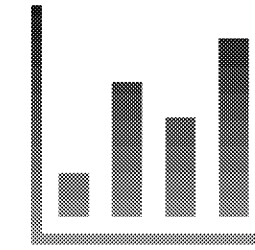
SEEKING/DISCOVERING

Searching for inappropriate content and targets. Discovering targets (e.g., via suggestions, hashtags, Explore).



- Searching for inappropriate content.*
- Searching for accounts with inappropriate or illegal content (CEI and MS).
- Searching for targets.
- Requesting images from other users/actors.
- Using hashtags to discover accounts who are sharing CEI.

*How prevalent is this content?



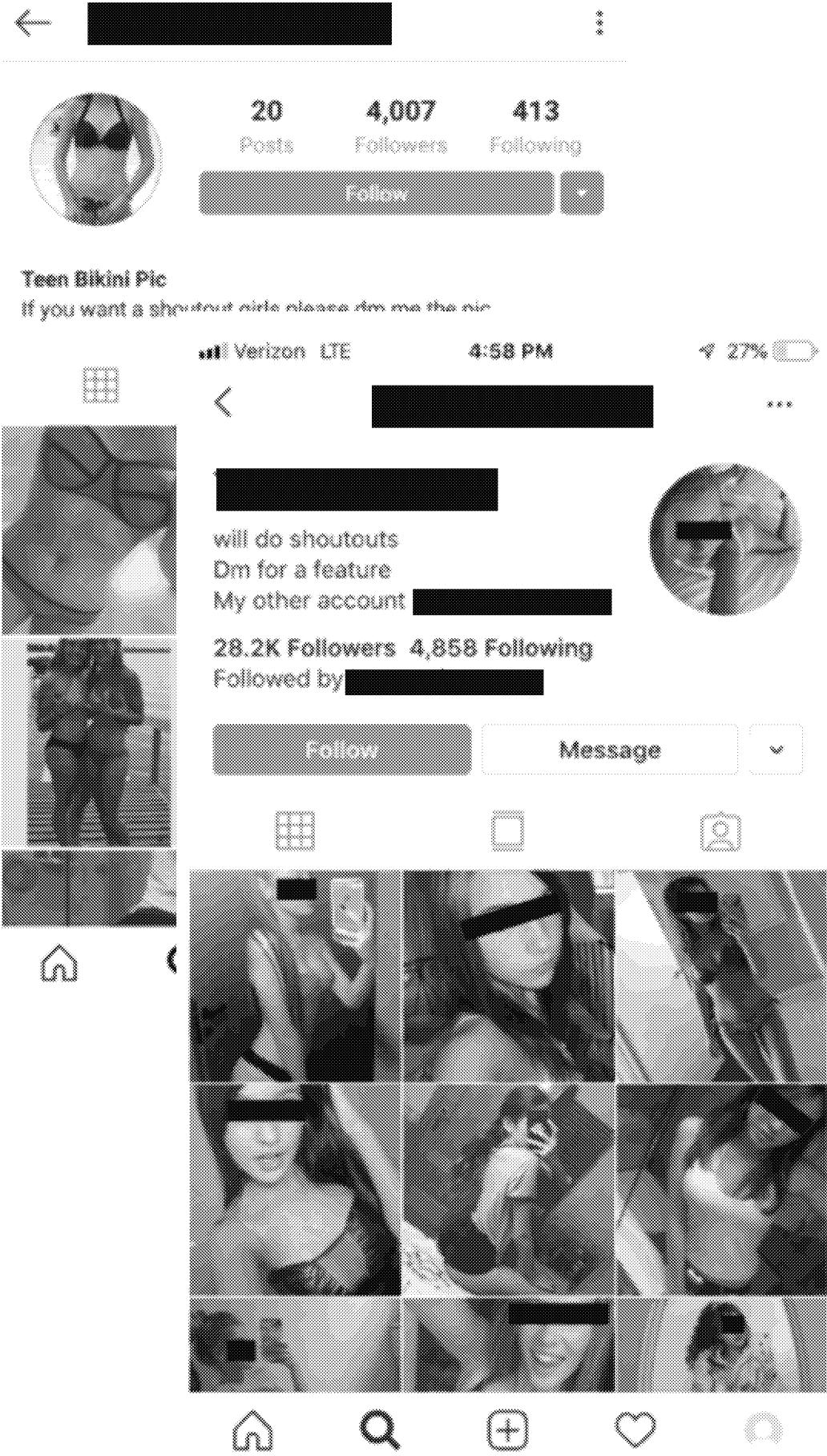
10.8K

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.

Hashtags: https://docs.google.com/spreadsheets/d/1lfZQ_qzambtDfBQFP1ecGLWnUK0shNh-jkC3qCmMd3s/edit#gid=893326739

ACCOUNTS SEXUALIZING TEENS



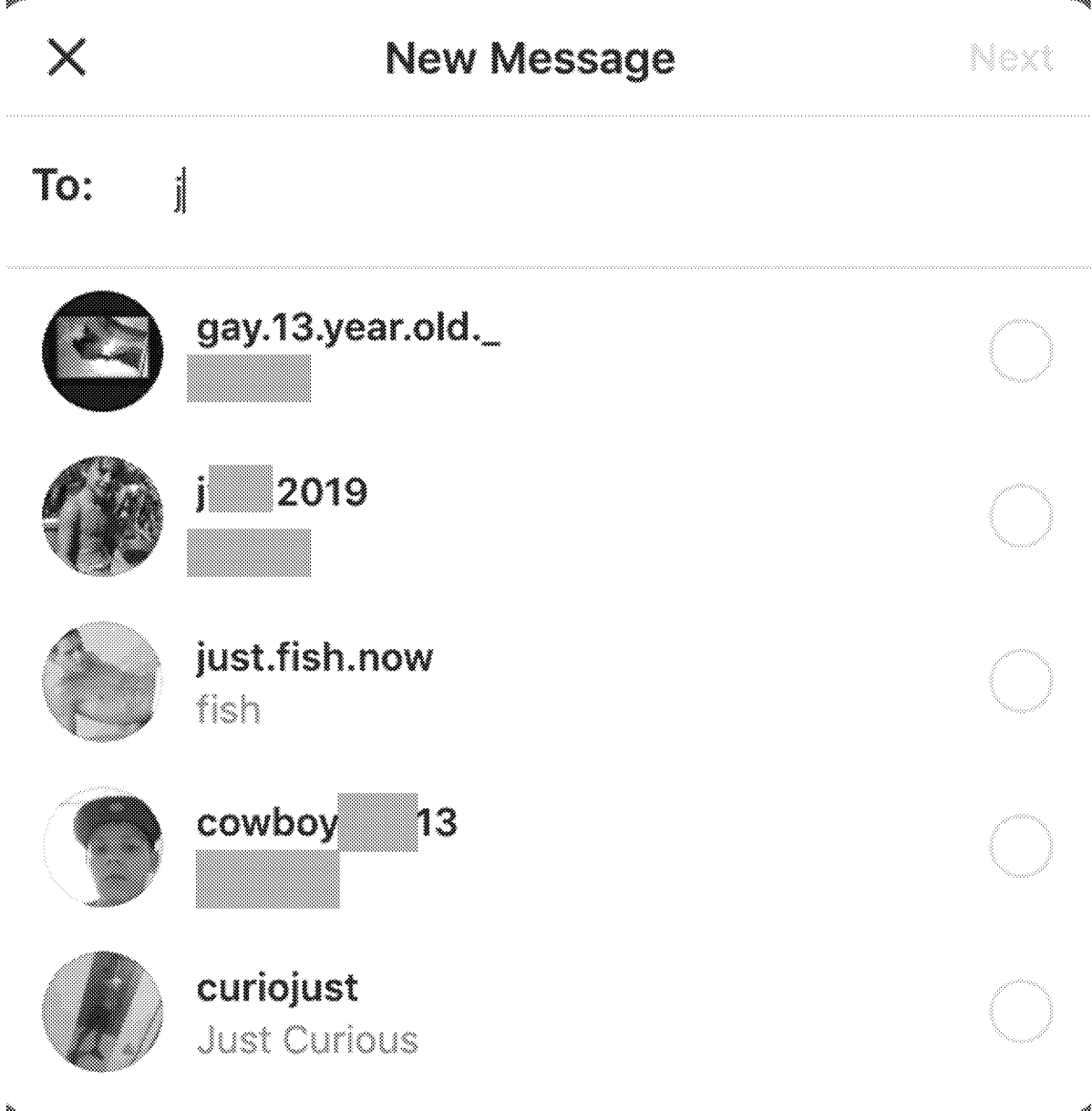
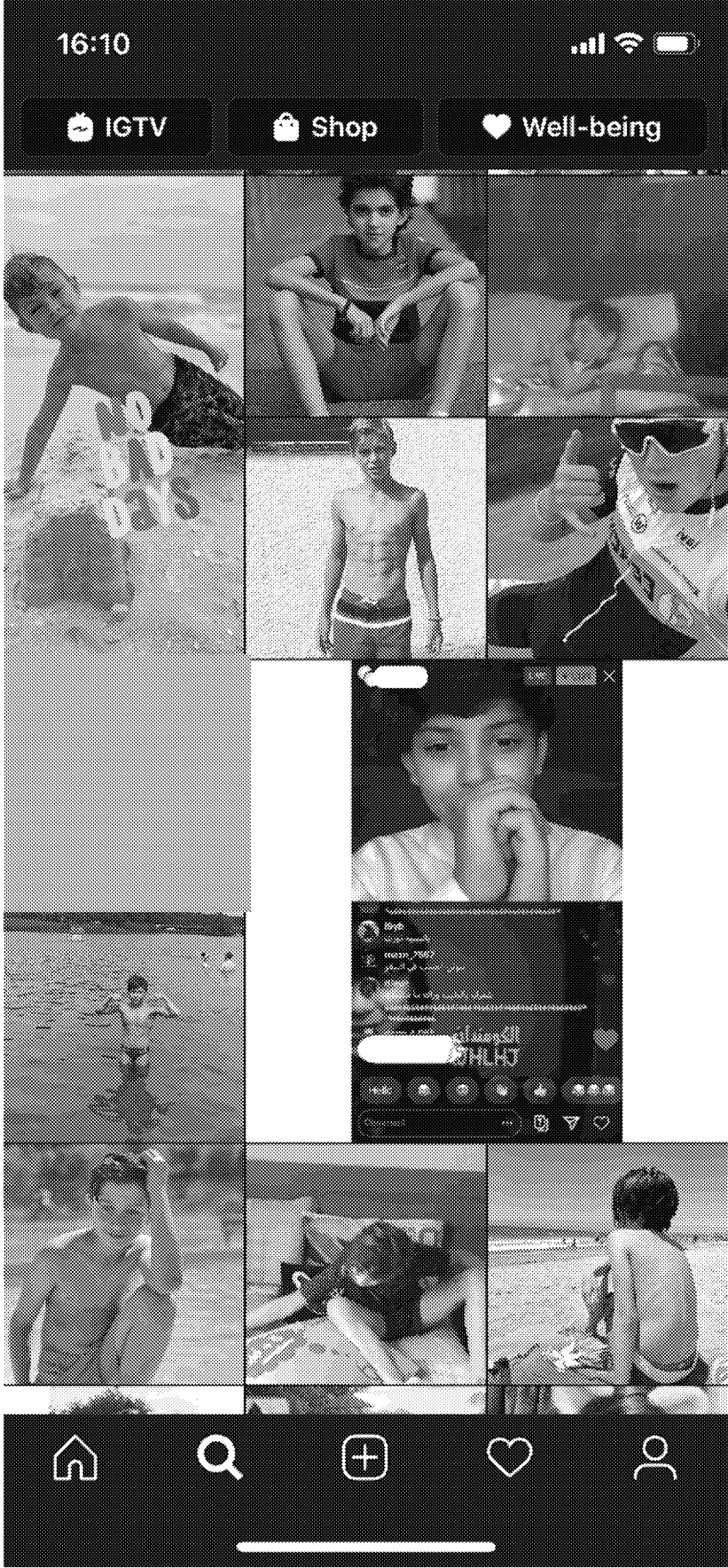
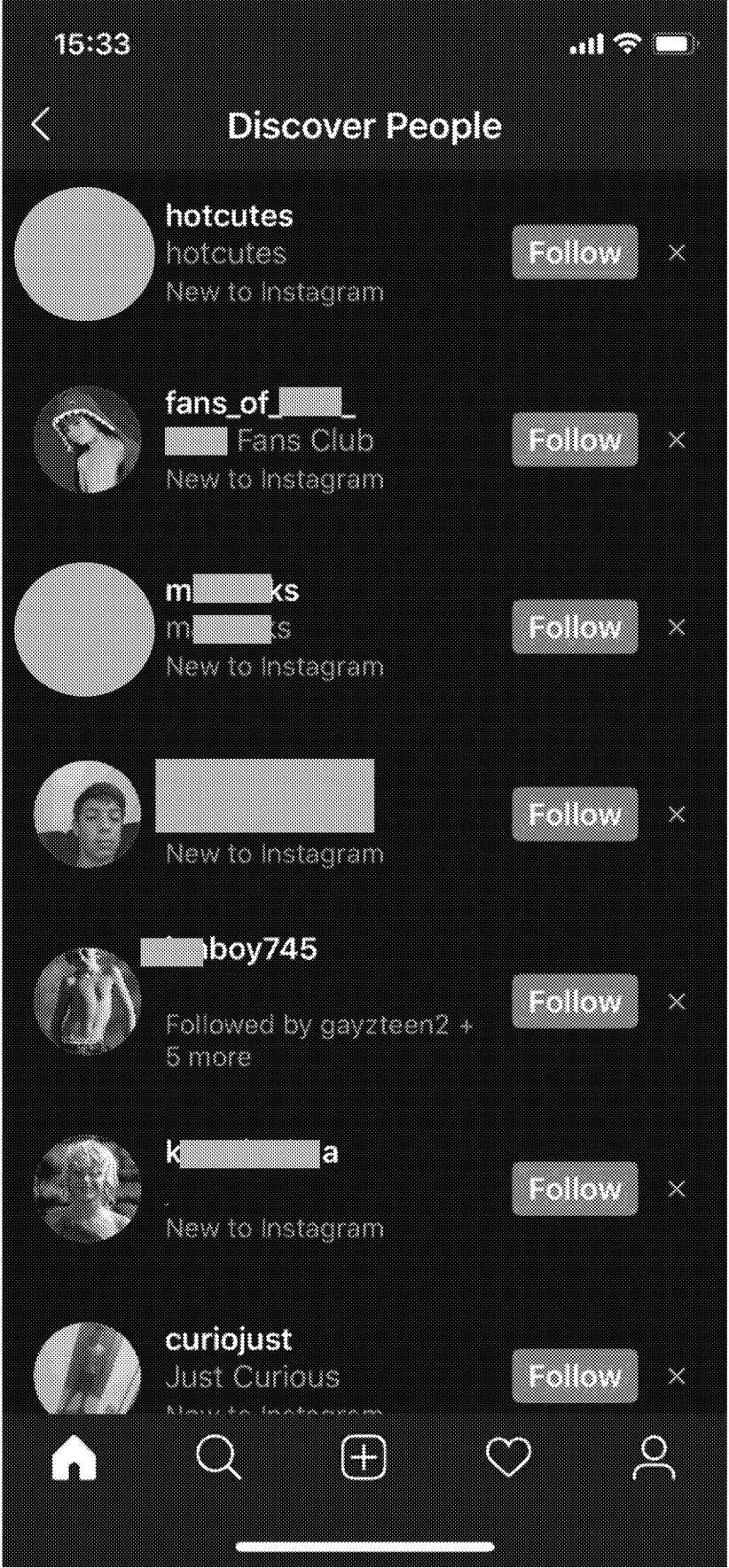
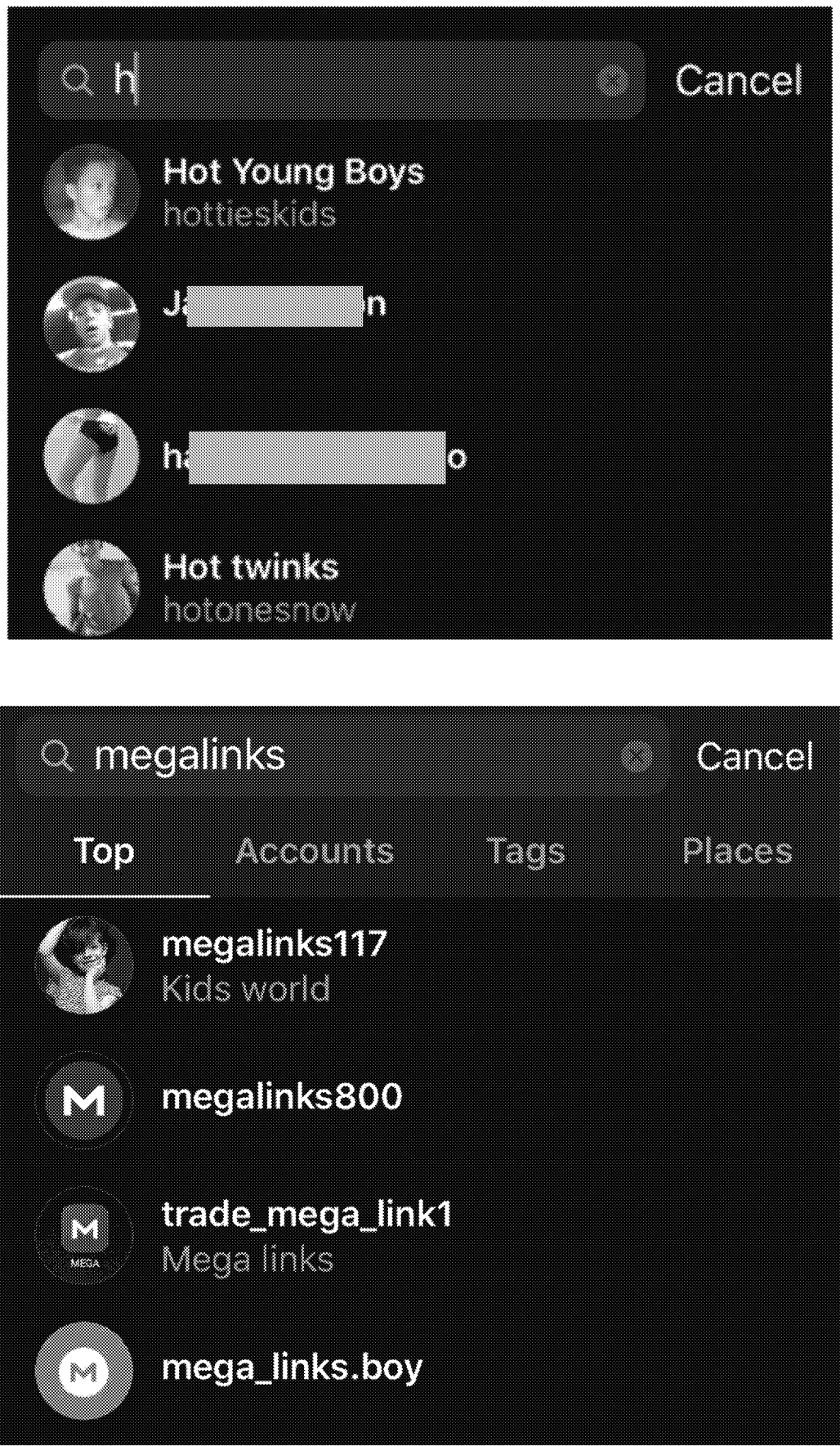
Pocher – Minor Sexualization (German Influencers Escalating CEI/Minor Sexualization Assessment): <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/933444310420909/>
CEI on IG Profile - <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/698909797207696/>

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY’S EYES ONLY)

META3047MDL-031-00192324
METANMAG-002-01771947

SEARCHING AND DISCOVERING

As part of Project Gumshoe, CO and OS investigators collaborated with CI to create a test account that simulated the bad actor discoverability process. Below are some recommendations and suggestions for this account.*

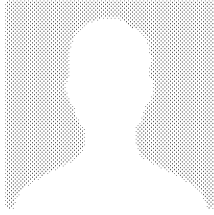


*How/when are we suggesting these accounts? Is this something we also surface for related interests as well (e.g., when searching for adult porn).

IG Test User: Minor Sexualization/CEI Discoverability: https://docs.google.com/presentation/d/14AMujK_1E1GZ9qemsEW0FtYyTmE9oVBUoi7ZaGRkgeg/edit#slide=id.p

SOLICITING IMAGES

Lower risk if soliciting from other actors. Higher risk if soliciting from targets (see below slides on Active/Higher effort behaviors). This behavior is illegal and reportable to NCMEC. It's common for actors to pose as teens.



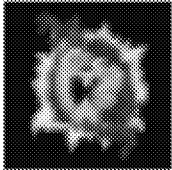
u[redacted]s

Mar 5, 2019 7:03pm

18009433798154504

Hmu for 🍷 #hmugirl #dmmegirls👉👈

#dmsexting



taz[redacted]

[redacted]

💖💖👉👈if any one have ANY vids or pic sets please send it thru telegram my username is t[redacted]2.

Translate

Mar 16, 2020 10:31am



i[redacted]a • Follow

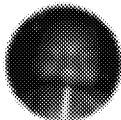
...



ix[redacted]a DM me I am 15

#gay12yeasrolds #gay14yearoldsold
#gay15teen #gaykik

1w



n[redacted] Dm me I'm 14 bottom looking for daddy or bbc

1w Reply

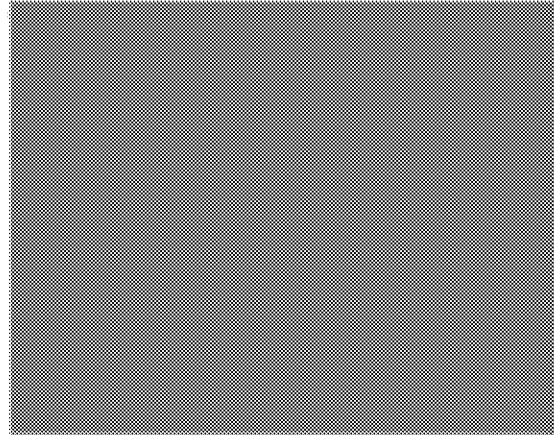
View replies (2)



m[redacted]k Dm dropbox trade

1w Reply

View replies (2)



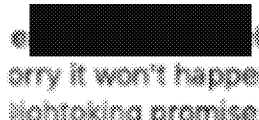
e[redacted]cx • Follow

...

e[redacted]cx Send me omething horny asf girls and maybe oys... but mostly girls...bi btw 15 ears old 🍷👉👈

hornyyyyyyxxx #teendm dmteenguris #dm #Pokemon Chelsea grin #hardcore

97w



e[redacted]ore_cx I'm orry it won't happen again ijohtoking promise :)

97w Reply



ike[redacted]y 🍷

97w Reply

IIC on IG Proactive Risk Investigation: <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/728865210878821/>

Risk Assessment: https://fb-my.sharepoint.com/:p:/r/personal/jjanssen_fb_com/_layouts/15/guestaccess.aspx?e=TKayCd&share=ETZ50yZSD_VLrLhxldFwjbwBQqWTEnzPvznF9Re84DXMfA

Project Gumshoe x Discoverability Archetype: https://docs.google.com/presentation/d/1CbqG91o6ZwYWI1_uPd-gJZI3LaR6uYovwKgUQCZZ_Ro/edit#slide=id.p; https://docs.google.com/document/d/1kSHkQyKvTiKfxga_gTupEbtpnctj4aSg5zDff-u6wHM/edit#heading=h.l9v3n5lembo2

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY’S EYES ONLY)

META3047MDL-031-00192326
METANMAG-002-01771949

HASHTAGS

Using Hashtags to find (and also share) content.

childporn

#childpornporn

4 posts

#childporn4l

24 posts

#childpornst

2 posts

#childpornoc

23 posts

#childpornag

15 posts

#childporngr

16 posts

#childporncollection

3 posts

#childpornisgood

1 post

@jb_teens_101

0y 23d 0h 0m



@jb_teens_101

#teen #jailbait #jbprime #primejailbait

0 Likes

i[redacted] a DM me I am 15

#gay12yeasrolds #gay14yearoldsold
#gay15teen #gaykik

#gay11s #gay101
#gays11concept #gayorepost12
#gay13yearsold #gay14y
#gay14year #gayteen

I Love You My Kids Friends
And Kids Sex Group Join Me

#Kids #Fun #Sex #group

My Girlfriend Kiss Time And Kids Sex Group Add Me My Friends

#Sex #Kids #Love

Kids Sexy Girls
Porn On My Group Kids Club

#Sexy #Cute #Girls





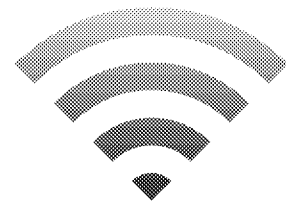
HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY’S EYES ONLY)

IG Test User: Minor Sexualization/CEI Discoverability: https://docs.google.com/presentation/d/14AMujK_1E1GZ9qemsEW0FtYyTmE9oVBUoi7ZaGRkgeg/edit#slide=id.p
Hashtags: https://docs.google.com/spreadsheets/d/1lfZQ_qzambtDfBQFP1ecGLWnUK0shNh-jkC3qCmMd3s/edit#gid=893326739
IIC on IG Proactive Risk Investigation: <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/728865210878821/>
Post on use of hashtags: <https://fb.workplace.com/groups/946056792456490/permalink/1063709817357853/>

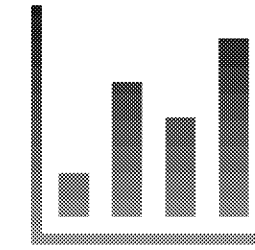
META3047MDL-031-00192327
METANMAG-002-01771950
PLTF EX-00992_0023

VIEWING (+ LIGHT INTERACTIONS)

View inappropriate content on accounts, Feed, Explore. May also include liking inappropriate content or following accounts.



- Viewing inappropriate content.
- Viewing content from accounts with inappropriate or illegal content (CEI and MS).
- Viewing content from targets.
- Liking inappropriate content
- Liking targets' content
- Following accounts



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.

COLLECTING

Collecting/Saving inappropriate images.



- Saving inappropriate images.
- Saving images from accounts with inappropriate or illegal content (CEI & MS).
- Saving images from accounts belonging to targets.

STALKING

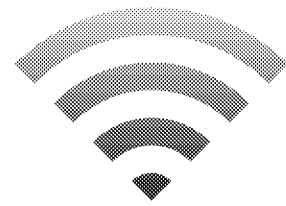
Stalking minors by viewing their content. Most evident in Stories given that users are provided this information.



- Watching stories
- Following
- (Persistent) Liking

INAPPROPRIATE COMMENTING

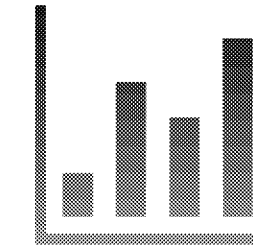
Commenting inappropriately on content which may or may not be appropriate.



Commenting inappropriately on:

- inappropriate content
- content on accounts with inappropriate or illegal content (CEI and MS)
- targets' content

Liking inappropriate comments.



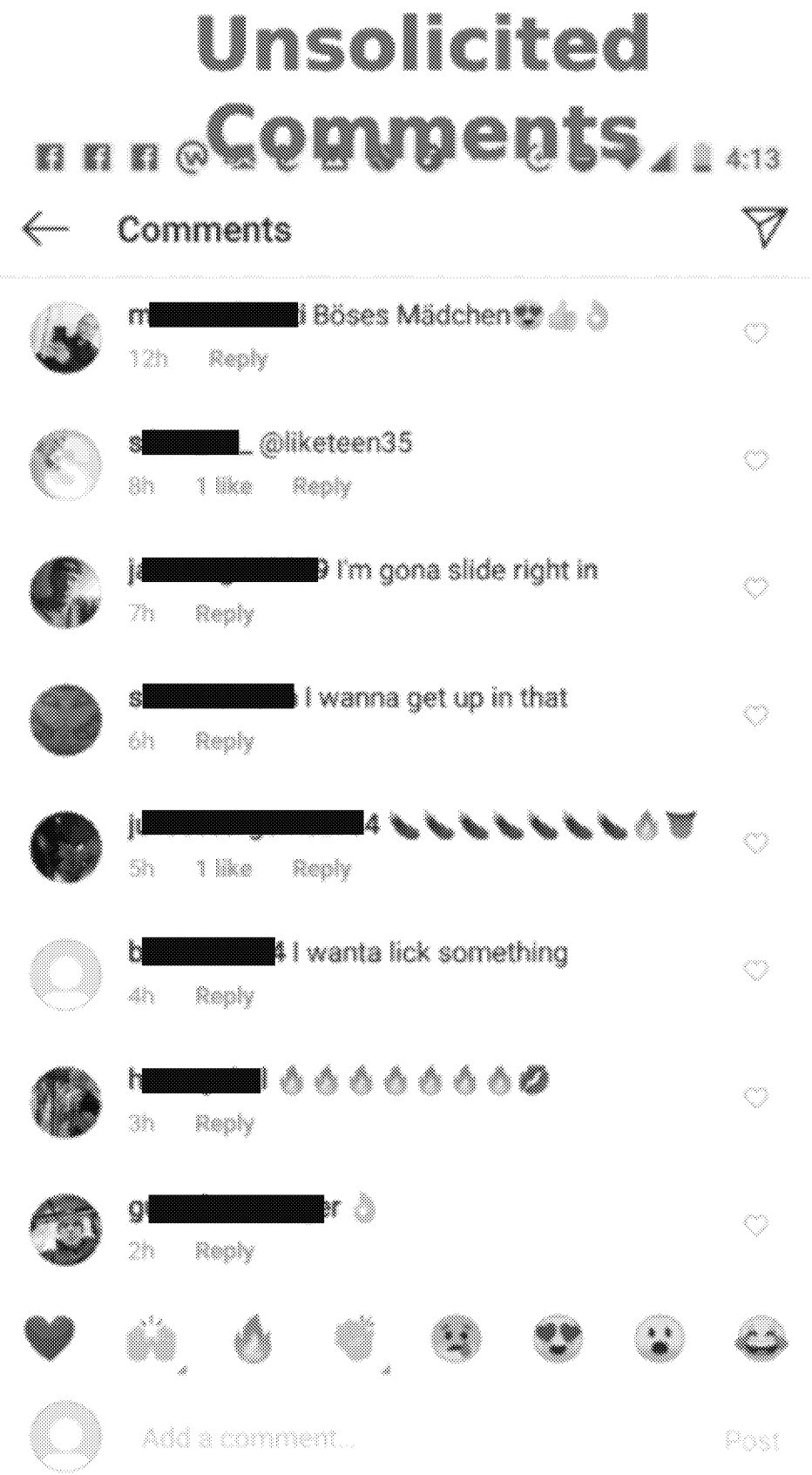
55

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

The most comments left by an
actor was 67 comments.

Improve the minor sexualisation classifier on IG: This enqueues comments for CO review that explicitly sexualise minors and is trained on previous violating posts. For more on proactive detection and product solutions, refer to: "Child safety landscape - Instagram" (POC: Vaishnavi J)

COMMENTS



IIC on IG Proactive Risk Investigation: <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/728865210878821/>
Risk Assessment: https://fb-my.sharepoint.com/:p:r/personal/jjanssen_fb_com/_layouts/15/guestaccess.aspx?e=TKayCd&share=ETZ50yZSD_VLrLhxldFwjbwBQqWTEnzPvznF9Re84DXMfA
Pocher – Minor Sexualization (German Influencers Escalating CEI/Minor Sexualization Assessment): <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/933444310420909/>

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY’S EYES ONLY)

META3047MDL-031-00192332
METANMAG-002-01771955

NETWORKING

Seeks to connect with other actors to observe, discuss, share, view relevant content etc.



- Following other actors.
- Following accounts that other actors follow.
- Messaging/Group messaging with other actors.
- Replying to other actors in comments.
- Sharing content/accounts with other actors.
- Tagging/mentioning other actors.
- Directing others off-platform to interact.

SHARING CONTENT & CONNECTING

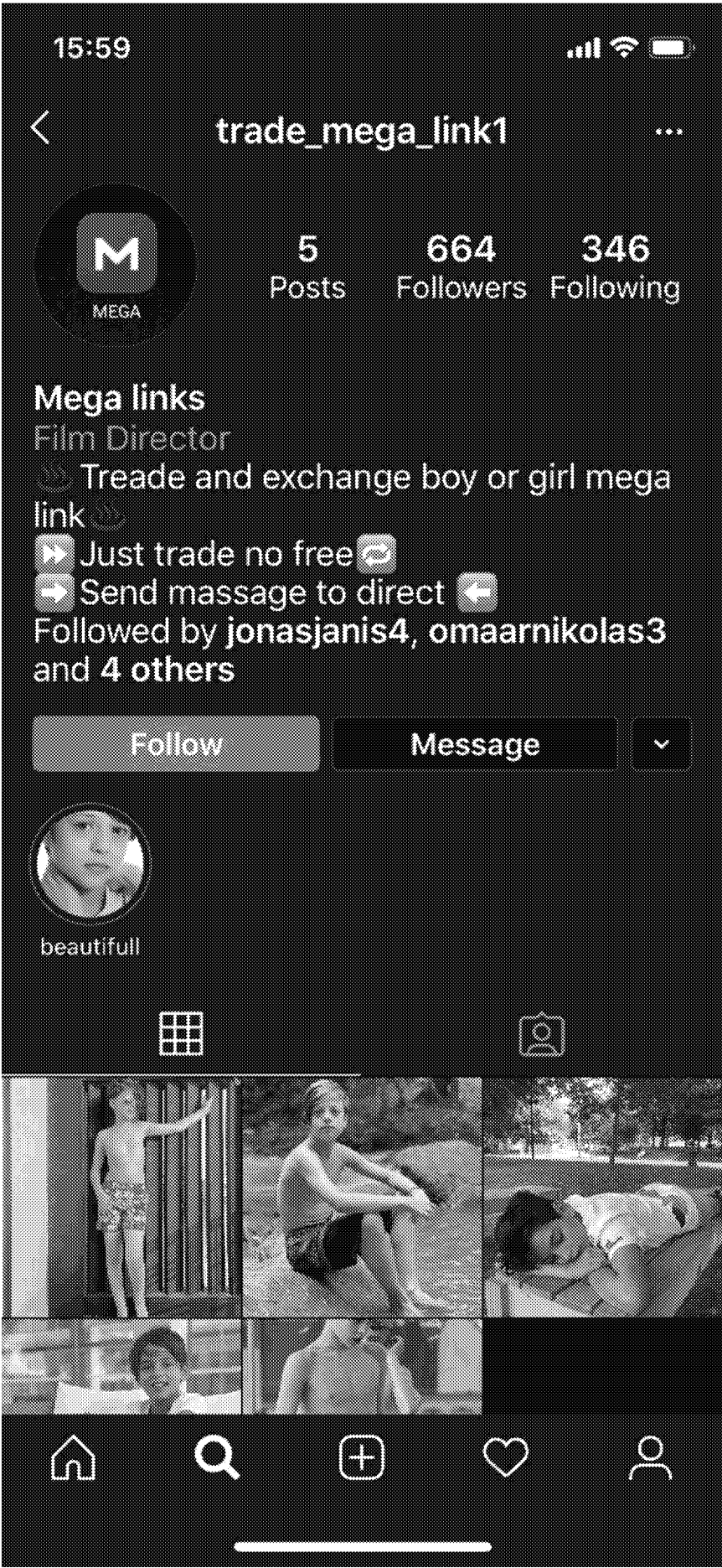
██████████98 - 2020/06/03 2:35:05pm PDT
Traded my way till I collected 100 gigs of it.

██████████ - 2020/06/03 2:35:11pm PDT
Now I don't trade anymore, only selling.

██████████72347 - 2020/06/03 2:35:53pm PDT
So you dont use the deep web?

██████████ - 2020/06/03 2:36:08pm PDT
I used to, but not anymore, I don't have to.

██████████347 - 2020/06/03 2:36:44pm PDT
Yeah fair enough



rob: Links??? Translate Mar 8, 2020 12:06pm

The Bat: @ABatr telegram. is my Translate Apr 9, 2020 11:54am

The Bat: @.tazz I'll send you a message here shortly. Translate Apr 9, 2020 11:54am

tazz: @.thebat add me two if you want tazz Translate Apr 9, 2020 11:53am

The Bat: @robe telegram. I will add you on Translate Apr 9, 2020 11:48am

Evan: @.tazz Could you still send it to me? Translate Apr 3, 2020 9:09am

#gay11s #gay101
#gays11concept #gayorepost12
#gay13yearsold #gay14y
#gay14year #gayteen
Translate



5 Like(s)

b██████████8
17865142114425753
I'm 14 with a big dick. I respond only for pic
Translate
Jul 1, 2019 9:07am

w██████████
17849911234478700
Hey want to trade dick Pic 10 to 16 boys dm me I'm horny
Translate
Jul 1, 2019 5:30am

ay██████████0
17907254986317389
I have young gay14 sex video dm me

Project Gumshoe: CEI and Child Roleplay Accounts: https://docs.google.com/presentation/d/1COI_rtS863w9wotx49EF3nH5blrX42Wc8C8MAekEa38/edit#slide=id.g6fa958b600_0_155
Project Gumshoe x Discoverability Archetype: https://docs.google.com/presentation/d/1CbqG91o6ZwYWI1_uPd-gJZI3LaR6uYovwKgUQCZZ_Ro/edit#slide=id.p; https://docs.google.com/document/d/1kSHkQyKvTiKfxga_gTupEbtpnctj4aSg5zDff-u6wHM/edit#heading=h.l9v3n5lembo2

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY’S EYES ONLY)

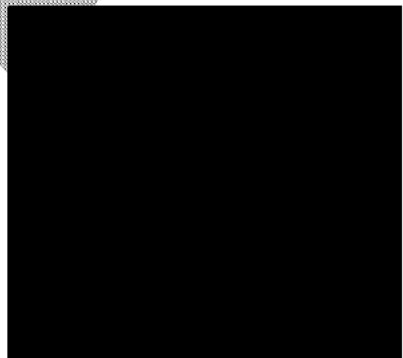
META3047MDL-031-00192334
METANMAG-002-01771957

USER EXPERIENCE

Seeking

“We had an account called 'High school girls robotics team' and we started to get some older men following us and watching us. I realized that they probably just searched 'high school' or 'girls' to find us.”

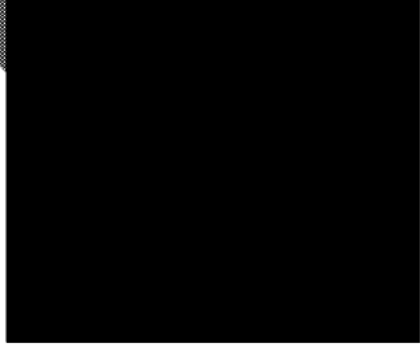
TEEN, WOMAN



Commenting

“She posted a picture in a bikini and there were all these older men making inappropriate comments and asking when she would post more pictures.”

*MOM SPEAKING ON DAUGHTER'S/
INFLUENCER'S EXPERIENCE*



Stalking

“I’ll see a random person watching my Story and when I look into them I realize it’s an old man with kids and a wife. What business does he have watching a young girl’s Story?”

TEEN WOMAN

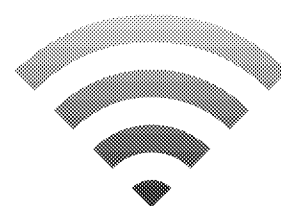


ACTIVE/HIGHER EFFORT BEHAVIORS

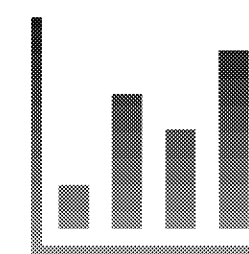
- 1 Contacting
- 2 Developing relationships
- 3 Creating/Posting content
- 4 Distributing content
- 5 Scamming

CONTACTING

This captures the first step to building a connection with targets which may evolve to more interactions.



- Following targets.
- Reaching out with an initial message - Many times not violating (e.g., "Hi"), but can be outright sexual (e.g., eggplant emoji).
- Commenting on target's content (to gain their attention or initiate conversation).
- Posing as a minor (e.g., in picture, bio, conversations).



2.2K

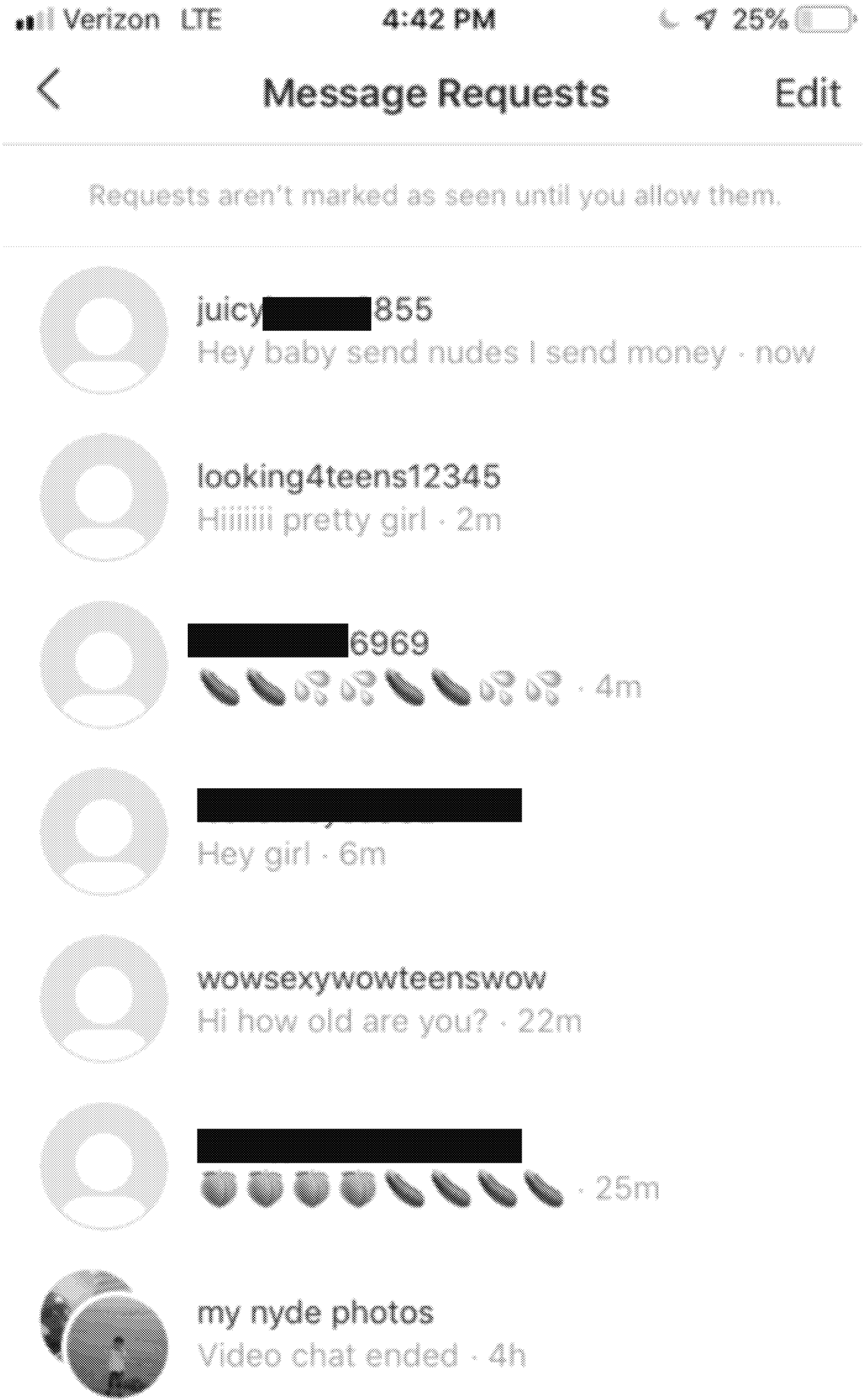
actors removed for IIC in last 14d. These accounts follow 375K teen accounts on IG.*

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

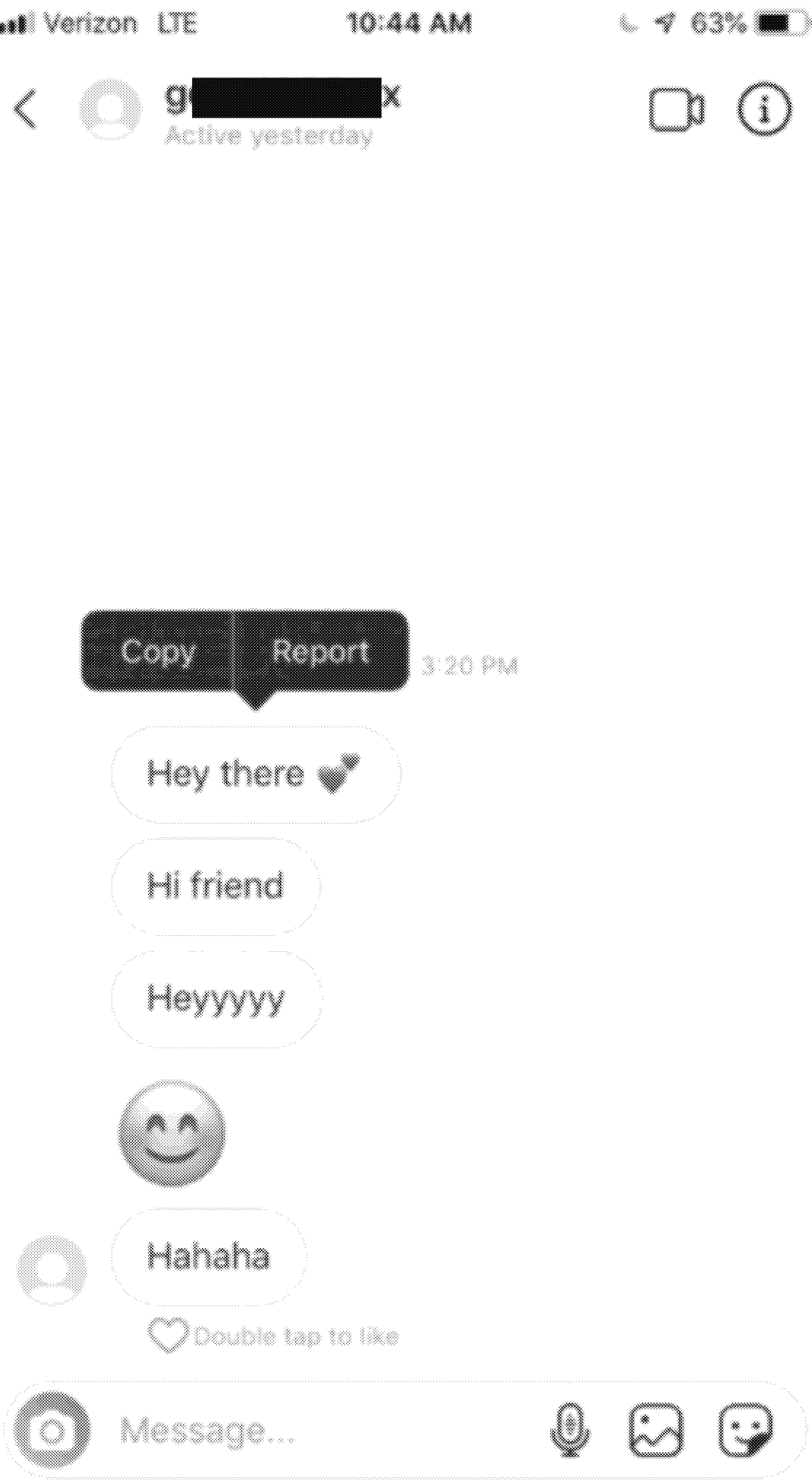
Disrupt the ability for predators to find targets on IG: FB disrupts the ability for predators to discover minors (using People You May Know) to prevent unconnected adults from discovering or initiating conversations with minors, **Instagram has no similar logic for Accounts You May Follow.** We're in the process of understanding minor discoverability on selected surfaces on IG (e.g. search, suggested accounts). For more on proactive detection and product solutions, refer to: "Child safety landscape - Instagram" (POC: Vaishnavi J)

MESSAGES

Many messages are not explicit (e.g., "Hey girl") and are not included in our IIC measurement. The IIC T2 classifier does attempt to find and auto-disable these type of users.



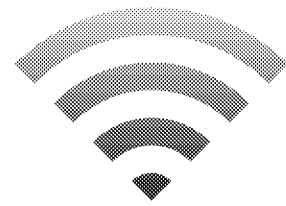
Unsolicited DMs



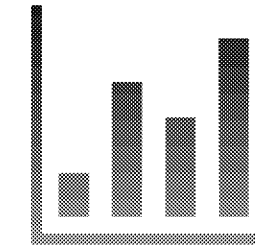
Pocher – Minor Sexualization (German Influencers Escalating CEI/Minor Sexualization Assessment): <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/933444310420909/>

DEVELOPING A RELATIONSHIP

These are people who are trying to build inappropriate relationships with minors. Primarily happens in DMs.



- Prolonged conversations with targets.
- Inappropriate conversations (e.g., Sex talk)
- Directing targets to meet offline.
- Directing targets to other secure platforms, messaging apps, or file-sharing platforms.
- Inappropriate requests (e.g., Enticement - Asking the target to send images or produce content for them. This is illegal in the US)
- Moving to RTC (live video call).
- Deleting messages (to remove evidence)



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.

Improve IIC text classifiers on FB and IG: Child Safety CI is working to develop a text model across both FB and IG that will capture conversations with minors where someone solicits images from them. For more on proactive detection and product solutions, refer to: "Child safety landscape - Instagram" (POC: Vaishnavi J).

TRUST BUILDING AND INAPPROPRIATE REQUESTS

Actors may ask to meet up in person or may ask for inappropriate images.

What do you look like baby?

Ill send a pic if you do

I'll send a dick pic and you send a pic of your ass ok?

And we can both screenshot

Deal baby?

I love you too my husband

Im gonna find away to come see u next month probably during the week when ur mom works every day til 5

Ok

Because i need to release some stress and im going to pound ur pussy and ass when i get there

Mar 10, 2019 6:31am

17980680856201794

1 Previous Review; 0 Note

Content in Review

@ [redacted]

Because im older then u plus yes i do like girls ur age and i don't need ur parents knowing that im talking to especially with

Hide Rotate

Responsible

0 Posts

0 Followers

0 Following

@ [redacted]

Reporter

48 Posts

107 Followers

132 Following

@ [redacted] angel

13 years old

Taken by: @ [redacted]

Pfp by: me

HIGHLY CONFIDENTIAL (COMPETITOR)

HIGHLY CONFIDENTIAL (ATTORNEY’S EYES ONLY)

IIC on IG Proactive Risk Investigation: <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/728865210878821/>

META3047MDL-031-00192340

METANMAG-002-01771963

36

PLTF EX-00992_0036

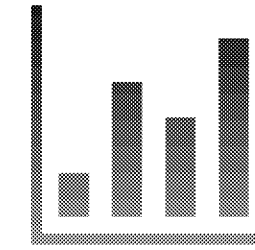
CREATING/POSTING CONTENT

Actors who post inappropriate content of minors. May include original content, content they take from others, and manipulated or edited content. Actors may benefit or profit from creating/sharing exploitive material.



- Posting inappropriate content (original or created).
- Creating an account with CEI or MS content.
- Directing users to accounts with inappropriate content.
- Directing other off-platform or to an email for more information or content.
- CEI/MS content paired with currency symbols (\$, usd, £, Rs, etc.).
- Links for payment in bio.

*Actors may pose as a minor (e.g., picture or bio)



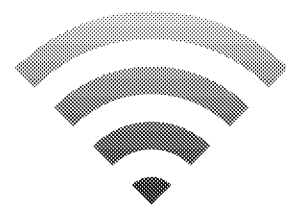
5-7.5K

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.

DISTRIBUTING

Actors who distribute inappropriate material. They may do this by sharing, sending, using hashtags, mentioning or tagging others.



- Sharing inappropriate content, accounts, links, or suspicious hashtags (gaykid, sexyteens, cuteboys, etc.) in DM (1-1 and group).
- Directing others to content (tagging, mentioning, posting, Stories).
- Directing other off-platform or to an email for more information or content.
- Using hashtags make content discoverable.

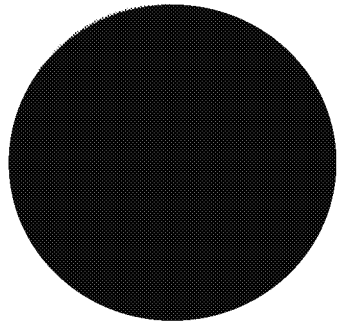
Bring the NCMEC reported signal to IG: this signals to product teams users who have shared CEI but not been disabled (currently running on Rooms and Messenger Kids). For more on proactive detection and product solutions, refer to: "Child safety landscape - Instagram" (POC: Vaishnavi J).

CREATING/USE A RANGE ON CONTENT AND ENGAGING ACTORS



JCR Extracted Text

Creation Time: 2020/05/27 11:28am PDT
Message:
What would you do with her? 🍑
Copy Translate
Product Type: STORY
Impression Count: 0



stealingboyspvtpicsfb Follow ...

61 posts 19 followers 35 following

Stealing Kids Personal FB Pics
Some are worth jerking off to! Some of them might not be so cute but that's because I steal kids personal Facebook photos regardless of their looks

m ██████████ 88 DON'T LOOK at my STORY,if you don't want to M A S T U R B A T E! 🍑🍑



r ██████████ d Follow ...

0 posts 5 followers 109 following

Videos for boys
If you want to have sex videos for young boys and girls, enter private lessons and before that send your picture and write to us your age to respond



Name: tommy 🍑 Loves: mommy, milk, being nakey, sucking items Hates: mean people, being alone, being in the cold

WSJ Child Role Play Incident Review - <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/693831037715572/>

Project Gumshoe x Discoverability Archetype: https://docs.google.com/presentation/d/1CbqG91o6ZwYWI1_uPd-gJZI3LaR6uYovwKgUQCZZ_Ro/edit#slide=id.p

IG Test User: Minor Sexualization/CEI Discoverability: https://docs.google.com/presentation/d/14AMujK_1E1GZ9qemsEW0FtYyTmE9oVBUoi7ZaGRkgeg/edit#slide=id.p

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY’S EYES ONLY)

META3047MDL-031-00192343
METANMAG-002-01771966

SENDING ACTORS OFF PLATFORM, ASKING FOR PAYMENT

2020/06/03 8:09:13am PDT

I don't send free shit, all those files are videos

2020/06/03 8:09:43am PDT

Ok

2020/06/03 8:11:51am PDT

if you change your mind for a video, here is telegram @
Moi400987

2020/06/03 8:12:03am PDT

Nah bye

ul [redacted] m (Ulrike) 10893954208 – 2019/02/24
5:17:13pm PST

Heey baby !!! Keep an eye on My Secrete NUDES Now.
Press the Website link on @_danteyaden543 Bio

rob

Links???

Translate

Mar 8, 2020 12:06pm

The Bat

@ABatr telegram.

Translate

Apr 9, 2020 11:54am

The Bat

@tazz I'll send you a message here shortly.

Translate

Apr 9, 2020 11:54am

tazz

@thebat if you want tazz add me two

Translate

Apr 9, 2020 11:53am

The Bat

@robe telegram.

Translate

Apr 9, 2020 11:48am

Evan

@tazz Could you still send it to me?

Translate

Apr 3, 2020 9:09am



Selling Dropbox links hmu #Dropbox #gaydropboxlinks #dropboxxx

2020/05/30 1:32:01am PDT

Steam Gift Card 20 USD - Buy cheaper on G2A.COM
buy this gift card, and send me the code they send you upon purchase. Then as soon as you give me the code I'll send you a Megalink with her set. <https://www.g2a.com/en/steam-gift-card-20-usd-steam-key-global-i10000000258091>
 <https://www.g2a.com/en/steam-gift-card-20-usd-steam-key-global-i10000000258091>

May 29, 2020 10:16am
18035675614260174
IG Post

Come to my inbox if you want to buy videos, dm me

Translate

COME TO MY INBOX IF YOU WANT TO BUY VIDEOS, DM ME

tazz

Mar 11, 2020 5:23am

IG Post

For special pics join me on telegram.

Username ta:

Translate

30 Like(s)

 J. [redacted] 5_
18075677974104247
Dm 14
Translate
Aug 25, 2019 6:53pm

 x. [redacted] 03
17854678885521041
I'm 16 into young
Translate
Aug 24, 2019 5:11pm

 G. [redacted] e
17921285677316780
Dm I'm 13
Translate
Aug 20, 2019 11:13pm

 K. [redacted] n
17876197609430181
@d. [redacted] 0 dm
Translate
Aug 24, 2019 4:29pm

 s. [redacted] a
17844743542592492
8-16 dm me
Translate
Aug 18, 2019 9:25am

 [redacted] ❤️
17969718133303510
Teen bottom here, tops dm me! 🌈
Translate

#gay12yearolds
#gay11yearolds #gay13yearold
Translate



168 Like(s)

 b. [redacted] e
18038365852051197
Dm me daddies I'm horny
asf with hot ass looking for a
bf I'm 13
Translate
Feb 27, 2019 3:21pm

 J. [redacted] e
18033655738071271
12
Translate
Feb 26, 2019 11:35am

 t. [redacted] 3
18018534568123592
@j. [redacted] fe.11 dm me
rn
Translate
Mar 1, 2019 12:48am

 b. [redacted]
@ [redacted] e.11 dm
Translate

#menhairy #instaboy #boom
#lgbtq #gaypride #lolihentai
#ecchi hentai #gayfetish
#spicyhentaimeemes
#waifuwednesday #bhfyp
Translate



68 Like(s)

 F. [redacted]
18001265089215258
وخيرك
Translate
May 20, 2019 11:21am

 [redacted]
18038487157140219
@ca. [redacted] i have much
links gay videos if you want
send to me
Translate
May 20, 2019 8:24am

 j. [redacted] s
18062253115068663
Nice
Translate
May 20, 2019 7:09am

#gay11s #gay101
#gays11concept #gayorepost12
#gay13yearsold #gay14y
#gay14year #gayteen
Translate



5 Like(s)

 b. [redacted] 8
17865142114425753
I'm 14 with a big dick. I
respond only for pic
Translate
Jul 1, 2019 9:07am

 w. [redacted]
17849911234478700
Hey want to trade dick Pic
10 to 16 boys dm me I'm
horny
Translate
Jul 1, 2019 5:30am

 a. [redacted]
17907254986317389
I have young gay14 sex
video dm me
Translate

 d. [redacted]

Project Gumshoe: CEI and Child Roleplay Accounts: https://docs.google.com/presentation/d/1COI_rtS863w9wotx49EF3nH5blrX42Wc8C8MAekEa38/edit#slide=id.g6fa958b600_0_155

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL (ATTORNEY'S EYES ONLY)

META3047MDL-031-00192345
METANMAG-002-01771968

PLTF EX-00992_0041

SCAMMING

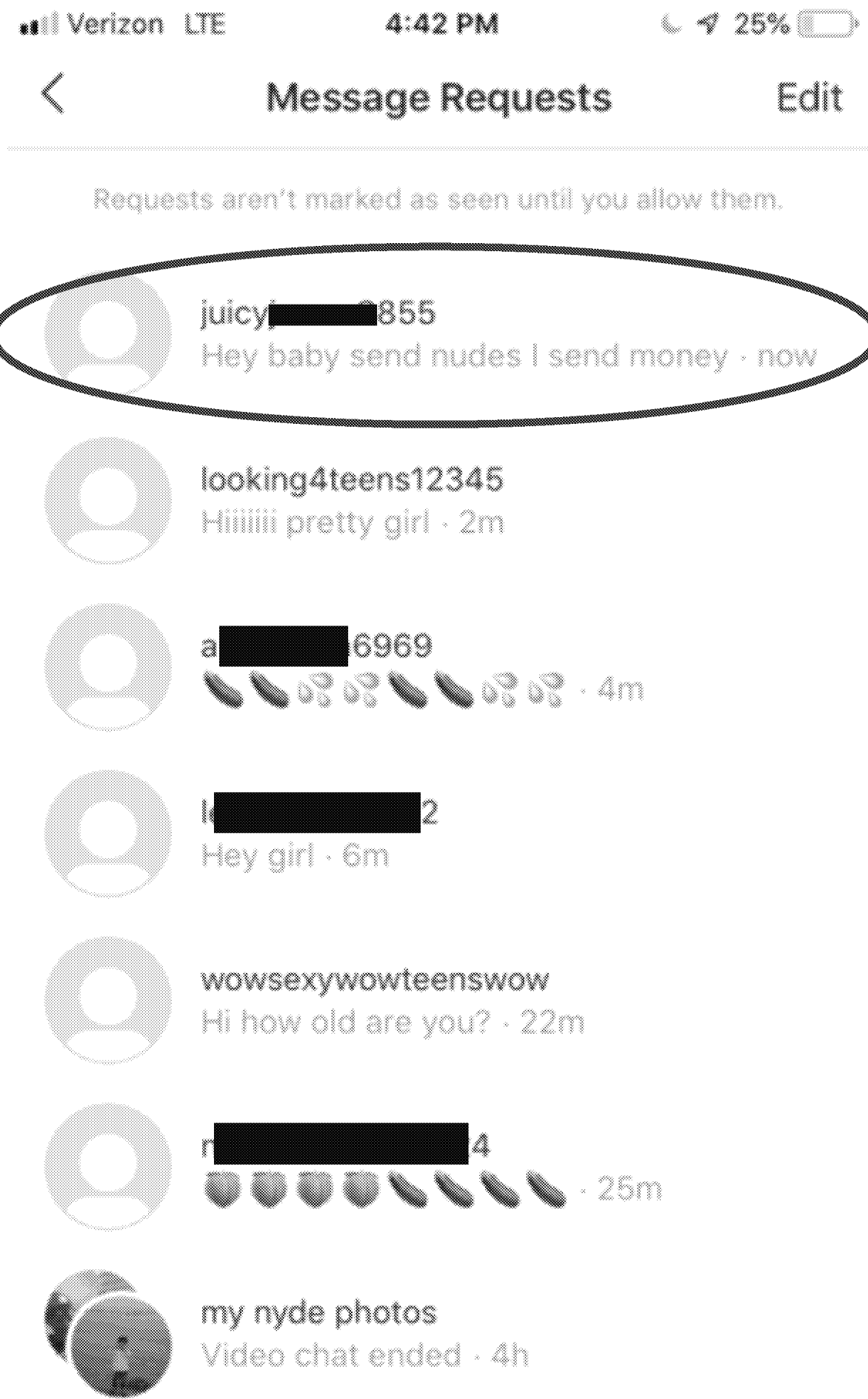
These people target users to scam them out of money or personal information/content. Most interactions happen in DM. Primary tactics include using lies and promises to draw in and entice targets.



- Requesting personal information in DM.
- Offers (e.g., I'll be your sugar daddy) paired with currency symbols (\$, usd, £, Rs, etc.).
- Threats (e.g., I have your personal images) paired with currency symbols (\$, usd, £, Rs, etc.).
- Directing targets off-platform or to an email.

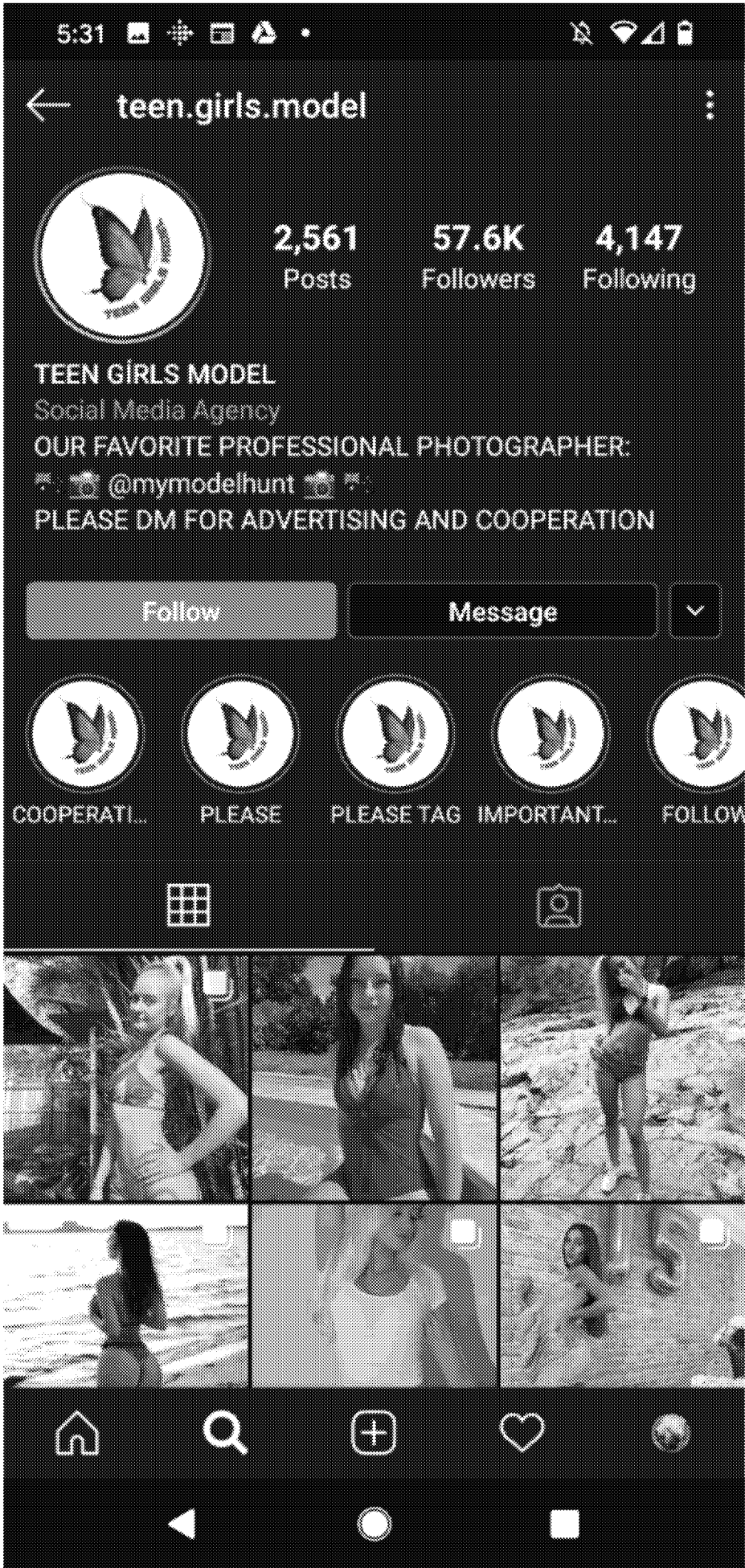
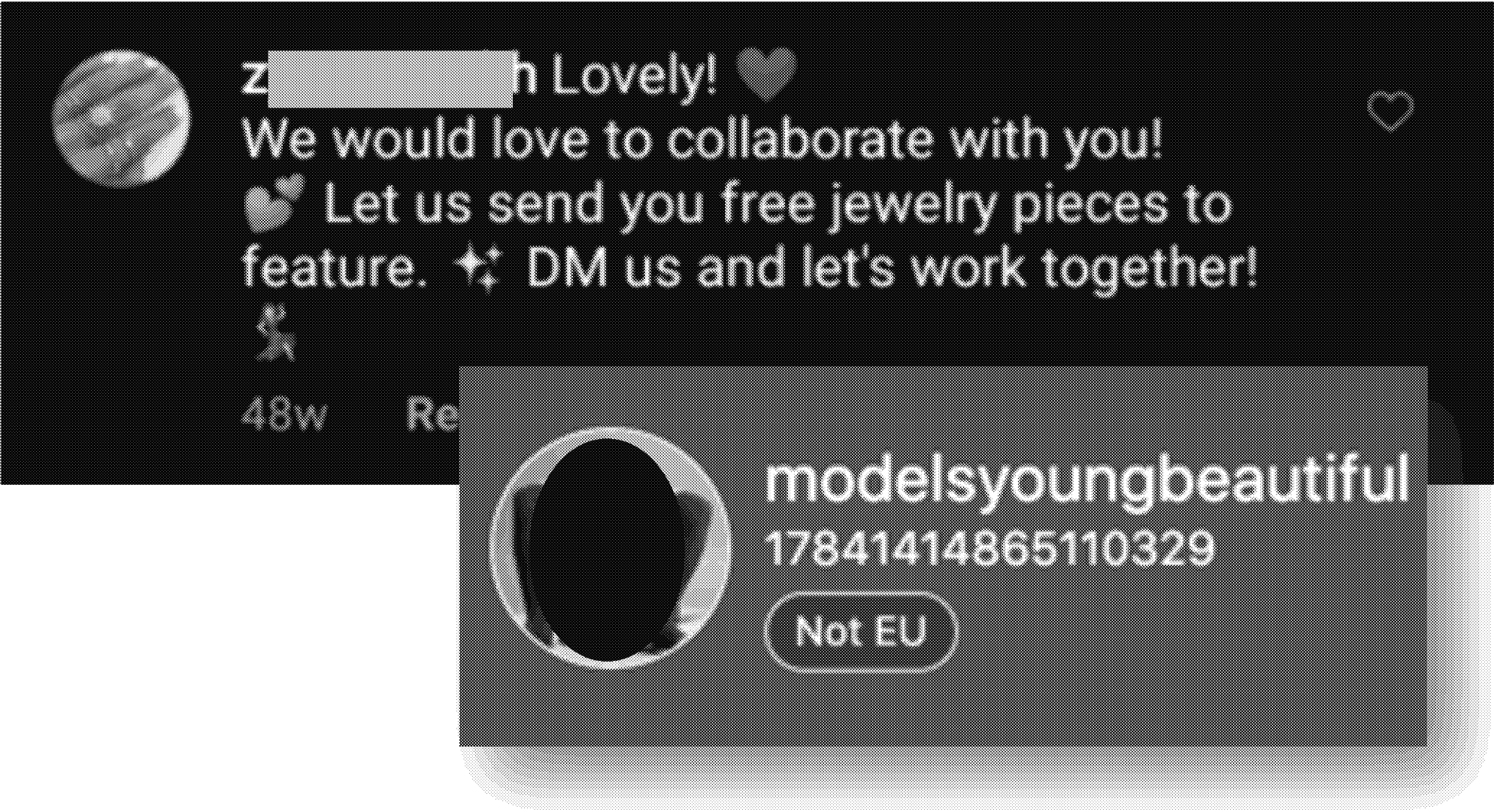
MANIPULATION THROUGH OFFERS AND PROMISES

Comments used to entice targets may motivate them to DM the actor first. How can we better understand how interactions move across surfaces?



Fake model scouts tricking UK girls into sharing explicit photos

Leading model agencies say calls from worried parents are increasing as scammers use social media to exploit children



IIC on IG Proactive Risk Investigation: <https://fb.workplace.com/groups/CIPCOIncidentReview/permalink/728865210878821/>

USER EXPERIENCES

Posting

Contact/Building trust

“I’ll get a random DM from strangers. They won’t seem creepy at first. They just want to talk, but then they might start asking for more, like what school I go to or they might ask for pictures.”

TEEN, WOMAN

“I’ve had friends whose pictures were taken and used to create fake profiles [with sexual undertones]. On Facebook, someone took my pictures and put it on a page called 'Cute boys and girls'. They asked people to comment on it and linked my account.”

TEEN, WOMAN

Scams

“A friend of mine was contacted by a guy who said that he wanted to get to know her and wanted to take care of her. He wanted to send her money but needed her back account. She gave it to him and he stole \$5k from her.”

ADULT MAN SPEAKING ON TEEN WOMAN'S EXPERIENCE

Scams

“They know that teens are always looking for ways to make money, so they take advantage of that to scam you.”

TEEN MAN



THEME

Opportunities

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_ggf4rwPQ1EQcrlLWPrWuc-YvI5yMt1qg/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

