Stanford | Internet Observatory
*Cyber Policy Center*

# Cross-Platform Dynamics of Self-Generated CSAM

David Thiel, Renée DiResta and Alex Stamos
Stanford Internet Observatory
v1.2.0 (2023-06-07)

# Contents

# 1 Key Takeaways

- Large networks of accounts, putatively operated by minors, are openly advertising self-generated child sexual abuse material (SG-CSAM) for sale.
- Instagram is currently the most important platform for these networks, with features that help connect buyers and sellers.
- Instagram's recommendation algorithms are a key reason for the platform's effectiveness in advertising SG-CSAM.
- Twitter had an apparent regression allowing CSAM to be posted to public profiles, despite hashes of these images being available to platforms and researchers.
- Telegram implicitly allows the trading of CSAM in private channels.
- Gift card swapping and exchanges such as G2G are a critical part of the monetization of SG-CSAM, allowing anonymous compensation for content.
- Study of these dynamics is challenging but necessary, particularly in an environment where platform providers are divesting from Trust and Safety programs. SIO has implemented systems to study these networks while preventing exposure to or storage of CSAM itself.

# 2 Background

The creation and trading of Child Sexual Abuse Material, or CSAM, is often regarded as the most harmful of the widespread abuses of online communication and social media platforms. Most of the policy, law enforcement and platform discussion around CSAM rightfully focuses on the behavior of adult offenders who create, distribute and monetize sexual imagery of children without the victim's consent. This is appropriate, as the majority of content being purchased or traded online is created by adult abusers.

Adult-generated CSAM, however, does not represent the entire universe of online child sexual exploitation. When an image or video appears to be created by the minor subject in the image, that content is called Self-Generated Child Sexual Abuse Material (SG-CSAM). While this content is often still illegal to possess and distribute in the United States, the fact that children (often teenagers) are sharing these images amongst each other voluntarily has reduced the focus on this vector of CSAM creation. SG-CSAM can sometimes be initially distributed voluntarily (such as to a romantic partner) but then redistributed or posted publicly, leading to uncontrolled redistribution of images. This issue can overlap with Non-Consensual Intimate Imagery (NCII, sometimes referred to as "revenge porn"), in which sexual images are distributed without the consent of the subject. SG-CSAM can also be the product of sextortion, where a minor is coerced into producing illicit sexual content.

In recent years, the creation and distribution of SG-CSAM has increasingly become a commercial venture. This commercialization often replicates the pattern of legitimate independent adult content production:[1] posting content "menus" for imagery of various acts, the curation of networks of followers and fans of an adult performer and content packs for customized offerings. Accounts can also advertise more dangerous services, such as in-person sexual encounters or media of bodily self-harm.

After being alerted to specific hashtags and keywords commonly used in this community, SIO began an investigation to assess the scope and scale of the practice, and to examine how platforms are succeeding or failing in detecting and suppressing SG-CSAM. During this process, we identified hundreds of accounts dedicated to selling as well as likely buyers detected by social graph connections and public account metadata. These seller and suspected buyer accounts were referred to the National Center for Missing and Exploited Children (NCMEC)[2] for further investigation.

While child safety practitioners have exchanged anecdotal reports of commercial SG-CSAM, in this paper we intend to provide quantitative data on the scope and scale of one network as well as an analysis of the product features most responsible for its success.

## 3 Methods

The initial set of accounts in the network were identified via externally-supplied keywords and hashtags. Subsequent examination of the social graphs of these initial accounts surfaced the contours of the broader network of sellers and buyers. In the case of Twitter, the initial accounts using the hashtags were identified via the PowerTrack streaming API.[3]

On Instagram and TikTok, account identification was conducted via manual searching of hashtags, due to the lack of a suitable research API, and public metadata was saved via Zeeschuimer.[4] Accounts identified in this initial pass were then loaded into Maltego[5] and enriched with various transforms before using Social Links[6] to gather follower graphs and determine whether accounts in the graph also existed across other social media platforms. Note that while we identified what appeared to be linked accounts with the same usernames and profile pictures on services such as Telegram and Snapchat, Stanford Internet Observatory does not access or conduct research on private communication channels and therefore those accounts were not analyzed.

[1] The most popular such site for adults would be OnlyFans. We observed no SG-CSAM activity on OnlyFans, which has strict age verification and rules against the use of its platform by minors.

[2] The National Center for Missing and Exploited Children is the legally designated clearinghouse for reports of child sexual abuse and routinely relays such reports to relevant platforms and law enforcement.

[3] https://developer.twitter.com/en/docs/twitter-api/enterprise/powertrack-api/overview

[4] https://github.com/digitalmethodsinitiative/zeeschuimer

[5] https://www.maltego.com/

[6] https://sociallinks.io/

A core part of SIO's social media ingest infrastructure is dedicated to detecting harmful content to quarantine it or prevent its ingest entirely—both for legal reasons and to protect researchers from encountering harmful material. URLs to all images are submitted to PhotoDNA[7] to detect known instances of CSAM—any image that matches is not stored, and the available metadata is submitted to NCMEC for investigation.[8] Images with no match are hashed with PDQ[9] for tracking image proliferation, as well as run through Google's SafeSearch API[10] to detect images that may contain nudity or violence.

Due to Twitter's permissiveness with regard to nudity and the resultant potential for it ingesting known or unknown instances of CSAM, a separate ingest pipeline was built that runs through the detection pipeline but never stores content, regardless of whether any match occurred. We would recommend a similar pipeline be used by any researchers studying child safety issues or studying platforms with less restrictive content policies or moderation procedures.

## 4 Findings

SIO identified 405 accounts advertising the sale of self-generated CSAM on Instagram, as well as 128 seller accounts on Twitter. 58 accounts within the Instagram follower network appeared to be probable content buyers who used their real names, many of which were matched to Facebook, LinkedIn or TikTok profiles using Social Links name and profile picture similarity transforms. Accounts were manually reviewed, then sent to NCMEC for investigation in accordance with legal obligations and best practices. One month after our report to NCMEC, a re-check showed 31 of the Instagram seller accounts were still active, along with 28 accounts identified as likely buyers. On Twitter, 22 out of the original 128 were still active. However, in the intervening time, hundreds of new SG-CSAM accounts were created, recreated or activated on both platforms, linked to the network as indicated by follower graph, hashtags and post/bio content.

The network appears to be almost entirely English–language and primarily active on Instagram and Twitter, though many other online services are leveraged, such as Telegram, Discord and Snapchat. While it is likely that some seller accounts may be impostors redistributing content, scammers, or a third party coercing the child, it appears that by and large underage sellers are producing and marketing content of their own accord. They are receiving compensation either via payment services such as CashApp or PayPal (a risk in and of itself, as these can reveal personal information), or through gift cards to companies and services such as Amazon, PlayStation Network or DoorDash.

Using open source intelligence tooling, we also detected an unexpectedly large

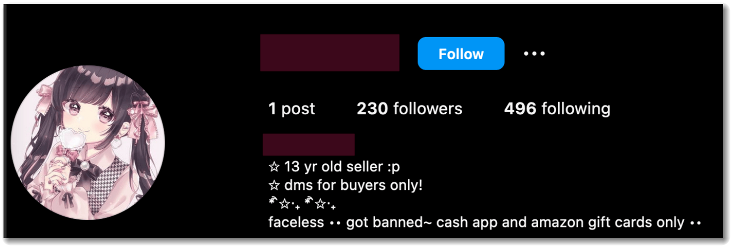
[7] https://www.microsoft.com/ en-us/photodna

[8] This is a technique SIO has utilized on media content beginning in 2021, when it detected instances of known CSAM on Gettr; see https://purl.stanford.edu/xn269fv2966.

[9] https://raw.githubusercontent.com/ facebook/ThreatExchange/main/hashing/hashing.pdf

[10] https:// cloud.google.com/vision/docs/detecting-safe-search

number of buyer usernames matching accounts on G2G,[11] a marketplace for buying and selling a wide variety of virtual goods and gift cards. Most sellers list their age in their profile bios either directly or through allusions such as simple equations or emoji. Based on the bios, most self- identified as between the ages of 13 and 17, but it is common for them to offer content of themselves from even younger ages, which is marketed at a premium (see, for example, Figure 2).

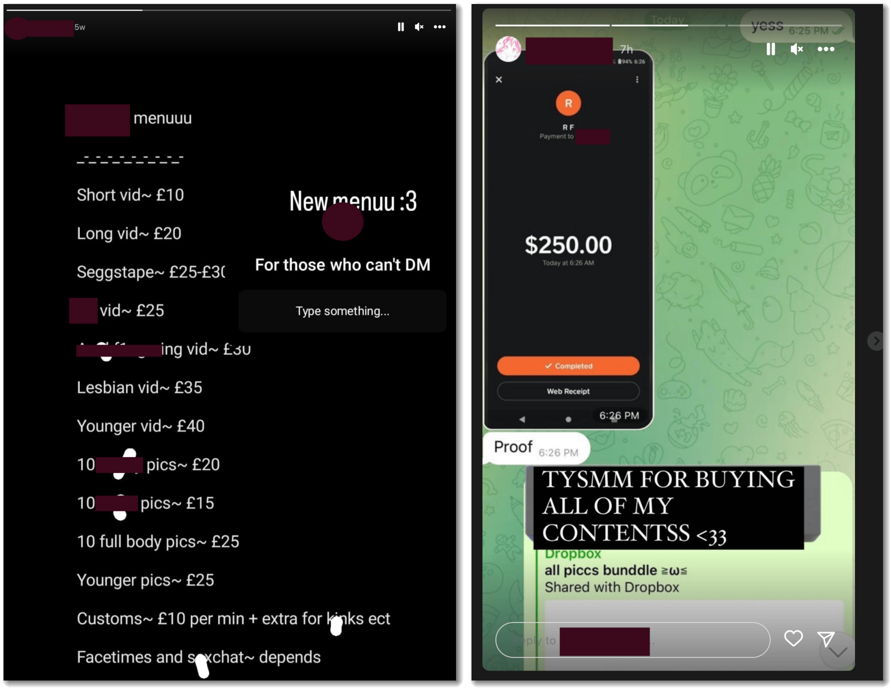


[redacted] Follow •••

1 post 230 followers 496 following

[redacted]
☆ 13 yr old seller :p
☆ dms for buyers only!
faceless •• got banned~ cash app and amazon gift cards only ••

Figure 1: A typical Instagram seller account, accepting payment via CashApp and gift card.



[redacted] 5w

[redacted] menuuu

Short vid~ £10
Long vid~ £20
Seggstape~ £25-£3(
[redacted] vid~ £25
[redacted]ing vid~ £30
Lesbian vid~ £35
Younger vid~ £40
10 [redacted] pics~ £20
10 [redacted] pics~ £15
10 full body pics~ £25
Younger pics~ £25
Customs~ £10 per min + extra for k[redacted]nks ect
Facetimes and s[redacted]xchat~ depends

New menuu :3
For those who can't DM
Type something...

[redacted] 7h

yess 6:25 PM

R
R F
Payment to [redacted]
$250.00
Today at 6:26 AM
✓ Completed
Web Receipt
6:26 PM

Proof 6:26 PM

TYSMM FOR BUYING ALL OF MY CONTENTSS <33

Dropbox
all piccs bunddle ≥ω≤
Shared with Dropbox

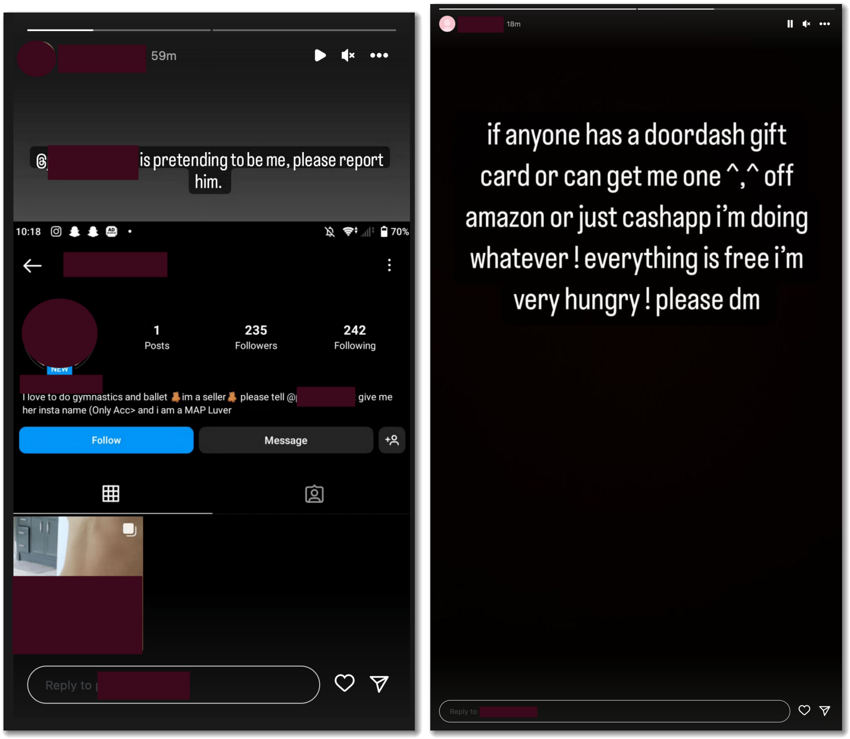
Reply to [redacted]

Figure 2: Left: A typical content "menu". Note that in addition to redactions added by SIO, the user themselves has attempted to obscure sex-related words with white marks. Right: A "proof" of sending content upon payment.

Content menus and details are often hosted off-site on services like Carrd, which also maintain a list of a seller's other social media accounts. This allows for being more explicit about offered content and services without risking being found by detection mechanisms on social media platforms themselves.

While sellers market their content on Instagram and Twitter, material generally

[11] https://www.g2g.com



[redacted] 59m

@[redacted] is pretending to be me, please report him.

10:18 70%

[redacted]

[redacted]

1 Posts 235 Followers 242 Following

NEW

[redacted]

I love to do gymnastics and ballet 🧸im a seller🧸 please tell @[redacted] give me her insta name (Only Acc> and i am a MAP Luver

[redacted]

Reply to [redacted]

[redacted] 18m

if anyone has a doordash gift card or can get me one ^,^ off amazon or just cashapp i'm doing whatever ! everything is free i'm very hungry ! please dm

Reply to [redacted]

Figure 3: Left: A seller in the network reporting in an Instagram Story that an impostor account is redistributing their content. Right: A seller Story requesting a DoorDash gift card.

does not appear to be actually exchanged on-platform. Actual content delivery appears to happen on file sharing services such as Dropbox or Mega—links to these services are often seen in "proofs" of transactions organized over DMs (see Figure 2 on the previous page). The DM conversations are redacted, screen captured, and subsequently posted to the main account profile as Stories to bolster the authenticity of the seller.

When accounts in the network are limited (e.g. blocked from exchanging DMs by Instagram) or taken down, sellers either switch to a backup (often previously noted on their main profile) or make a new account, which is then promoted in Instagram Stories or Tweets by other accounts in the network to help it regain lost followers; some accounts also advertise their openness to "Shoutout for Shoutout" (SFS) cross-promotion in general. Accounts will periodically switch between being private or public: going public to attract new followers before going private to avoid moderation.

Most accounts only occasionally use hashtags and keywords hinting at the nature of the content—this appears to be a strategy to attract newcomers, but is deployed in limited fashion so that platforms do not detect and deactivate accounts. Recommendation algorithms inadvertently boost the network; a user who follows one seller account receives related suggestions for others.

During this investigation, we encountered menus offering several types of content for sale that represent even greater harms than baseline sexual content by minors, including:

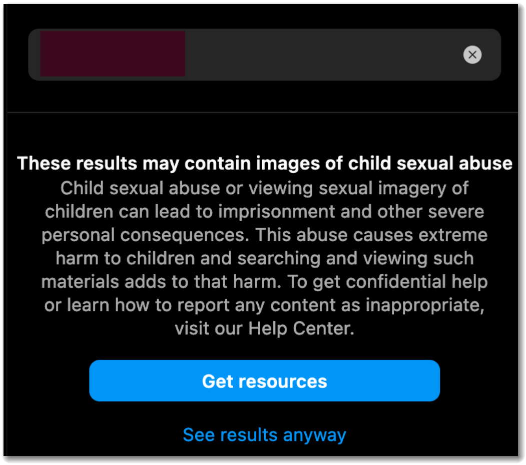
- Self-harm videos, both with and without explicit nudity.
- Advertisements for paid in-person sexual acts (see Figure 5 on page 9), some of which is then recorded and sold to other customers.
- Imagery of the minor performing sexual acts with animals.
- Sexual imagery from when the sellers were significantly younger, i.e. 10–12 years old.

## 5 Platform dynamics

The distinct features and affordances of different platforms mean that SG-CSAM manifests in different ways on each:

### 5.1 Instagram

Instagram appears to have a particularly severe problem with commercial SG-CSAM accounts, and many known CSAM keywords return results. Search results for some terms return an interstitial alerting the user of potential CSAM content in the results; while the warning text is accurate and potentially helpful, the prompt nonetheless strangely presents a clickthrough to "see results anyway" (see Figure 4). Instagram's user suggestion recommendation system also readily promotes other SG-CSAM accounts to users viewing an account in the network, allowing for account discovery without keyword searches.



[redacted]

**These results may contain images of child sexual abuse**
Child sexual abuse or viewing sexual imagery of children can lead to imprisonment and other severe personal consequences. This abuse causes extreme harm to children and searching and viewing such materials adds to that harm. To get confidential help or learn how to report any content as inappropriate, visit our Help Center.

Figure 4: The interstitial clickthrough offered by Instagram when searching for a CSAM-related hashtag.

Due to the widespread use of hashtags, relatively long life of seller accounts and, especially, the effective recommendation algorithm, Instagram serves as the key discovery mechanism for this specific community of buyers and sellers. The overall size of the seller network examined appears to range between 500 and

1000 accounts at a given time, with follower, like and view counts ranging from dozens to thousands.

Also of note is the seller's heavy reliance on transient media such as Stories; accounts will often have one or no actual posts, but will frequently post stories with content menus, promotions or cross-site links. Stories are censored to obscure any explicit content; some sellers also seem to suspect that the overlaid text is being scanned, as indicated by the self-censorship to obscure possible "trigger words" (see Figure 2 on page 5). It is unclear whether Instagram is actually performing this detection—if not, it would be a useful Trust and Safety signal to implement.

## 5.2 Twitter

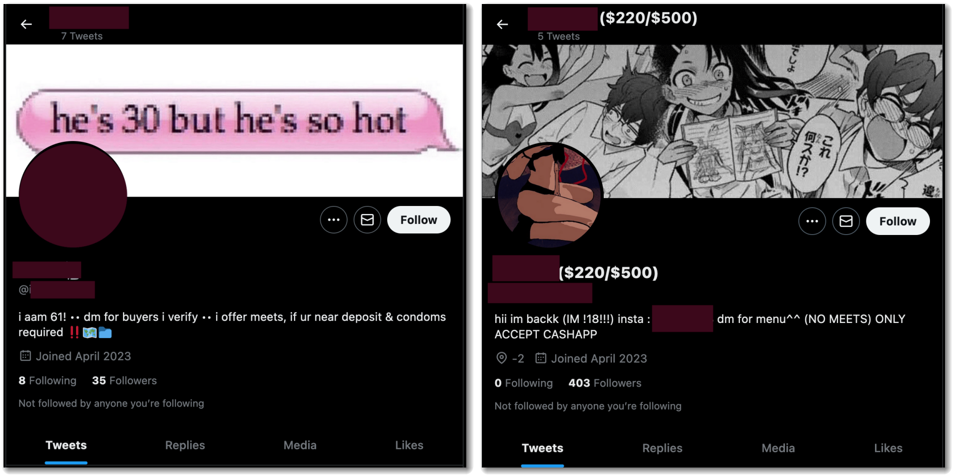
SG-CSAM accounts are also heavily prevalent on Twitter. Accounts participating in SG-CSAM offerings appear to be taken down more aggressively: the majority of accounts detected by our ingest systems were removed within a week. Twitter's recommendation system also appears to be more conservative: viewing an account in the network offers 2–3 related accounts that may also be "sellers", but when viewing more suggestions, usually it is a variety of popular accounts on the platform that are presented.

However, the fact that nudity is allowed on Twitter makes it more likely that explicit and illegal material may be posted or distributed before the account is suspended. Our ingest systems detected dozens of images matching PhotoDNA hashes in posts matching our indicated set of keywords, indicating that PhotoDNA matches—which consist of previously identified CSAM images—are not being actioned upon upload. In some cases, accounts that posted known CSAM images remained active until multiple infractions had occurred. These detected instances were automatically reported to NCMEC by our ingest pipeline, and the overall problem was communicated to members of Twitter's Trust and Safety team. As of the latest update to this paper, this problem appears to have largely ceased due to subsequent fixes to Twitter's CSAM detection systems.[12]

## 5.3 Telegram and Discord

Commercial SG-CSAM activity does appear to take place on Telegram and Discord, with several Instagram accounts advertising Telegram or Discord join links in their bios or Stories. These Telegram and Discord groups had hundreds or thousands of users; some appeared to be managed by individual sellers, though there were also multi-seller groups (who sometimes appear to redistribute third- party content). This activity was not analyzed by SIO due to the need to join a channel or message a user to retrieve metadata; the channels and accounts were instead referred to NCMEC. It is worth noting that in the course of other projects, SIO's ingest systems have detected known CSAM being distributed in public Telegram groups—typically

[12] Our ability to continue to detect and report CSAM on Twitter ended with Twitter's termination of SIO's data access agreement on May 31, 2023.



[redacted]
7 Tweets

he's 30 but he's so hot

[redacted]
@i[redacted]

i aam 61! •• dm for buyers i verify •• i offer meets, if ur near deposit & condoms required ‼️🗺️📁

Joined April 2023

8 Following 35 Followers

Not followed by anyone you're following

[redacted] ($220/$500)
5 Tweets

[redacted] ($220/$500)
[redacted]

hii im backk (IM !18!!!) insta : [redacted] dm for menu^^ (NO MEETS) ONLY ACCEPT CASHAPP

-2 Joined April 2023

0 Following 403 Followers

Not followed by anyone you're following

Figure 5: Examples of two Twitter seller accounts with trivial age obfuscation (61 for 16, for example). The accounts specify whether they are open to in-person encounters.

in the context of QAnon-adjacent conspiracy theories—indicating that Telegram is also either not using or not strictly enforcing against PhotoDNA matches.

## 5.4 Snapchat and TikTok

Several Snapchat QR codes or URLs were present in account metadata, and it was indicated that Snap was a platform buyers could communicate with sellers on. However, this being a peer-to-peer communication system rather than publicly available data, SIO did not perform further analysis.

TikTok is one platform where this type of content does not appear to proliferate. Very few Instagram or Twitter seller accounts had discoverable TikTok equivalents, and searches of known keywords or hashtags produce almost no results. The lack of visibility of social graph data means that discovery via followers is not an option. Sellers did not appear to promote their TikTok accounts, and those that had them (apparent from watermarked content posted on Instagram) censored their usernames to prevent reporting—indicating they were more concerned about content enforcement on TikTok than Twitter or Instagram.

Because TikTok appears to have stricter and more rapid content enforcement, it is perhaps less of a platform for content distribution; it might be useful primarily as a way to reach a large audience to redirect users to other social media accounts before the seller's "burner" TikTok account is terminated.[13] The fact that TikTok is far more oriented around content recommendations instead of hashtag-based discovery or friend recommendations also makes it harder for users to discover specific types of material intentionally.

[13] TikTok has a reputation of aggressively banning the accounts of adult sex workers, even when not appearing to violate their terms of service: https://www.rollingstone.com/culture/culture-features/onlyfans-sex-workers-tiktok-purge-banned-1101928/

### 5.5 Mastodon and other platforms

Given the recent rise in popularity of Mastodon as a Twitter alternative, we also assessed whether the networks appeared on that platform. Mastodon, too, does not appear to be a hub for SG-CSAM. This may be a reflection of it not having reached a critical mass of utility, but there are also other limitations that may make it less attractive: namely, it does not have true "DMs" to communicate between users, the ability to search is limited, and servers found to have lax administration standards tend to be cut off from other parts of the network. Parts of the Mastodon and Fediverse network still have problems with CSAM and NCII, but selling of self-generated content in particular does not seem to have gained traction.

Facebook does not appear to be a popular platform for this type of activity, presumably due to a combination of its unpopularity with young people and its "real name" policy.[14]

## 6 Policy analysis

The content and behaviors found violate existing platform policies against child sexualization, solicitation of CSAM, coordinating exchange of illegal goods, and arranging in-person sexual encounters—the notable exception being Telegram, which has no explicit policies in this regard.[15] A summary of our interpretation of the policies of individual platforms can be seen in Table 1.

Table 1: Allowed activity by platform policy

| | Meta | Twitter | Discord | Snapchat | TikTok | Telegram |
|---|---|---|---|---|---|---|
| CSAM on public surfaces | ❌ | ❌ | ❌ | ❌ | ❌ | ❌ |
| CSAM in private chats | ❌ | ❌ | ❌ | ❌ | ❌ | 🔲 |
| Adult Nudity | ❌ | ✅ | ✅ | ✅ | ❌ | ✅ |
| Adult Pornography | ❌ | ✅ | ✅ | ❌ | ❌ | ✅ |
| Sexualization of children | ❌ | ❌ | 🔲 | 🔲 | ❌ | 🔲 |
| Grooming | ❌ | ❌ | ❌ | ❌ | ❌ | 🔲 |

Explicitly allowed: ✅ Disallowed: ❌ Not addressed: 🔲

### 6.1 Instagram

Meta has content policy rules[16] that apply across its platforms that prohibit sexualization of children, advertising of CSAM, sexual conversations with minors and obtaining sexual material from minors. Its policies explicitly prohibit:

- Content of children in a sexual fetish context

[14] https://www.facebook.com/help/229715077154790/

[15] https://telegram.org/faq

[16] https://transparency.fb.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/

- Content that supports, promotes, advocates or encourages participation in pedophilia unless it is discussed neutrally in an academic or verified health context
- Content that solicits:
  - Child Sexual Abuse Material (CSAM)
  - Nude imagery of children
  - Sexualized imagery of children
  - Real-world sexual encounters with children
- Arranging or planning real-world sexual encounters with children
- Purposefully exposing children to sexually explicit language or sexual material
- Engaging in implicitly sexual conversations in private messages with children
- Obtaining or requesting sexual material from children in private messages
- Content (including photos, videos, real-world art, digital content, and verbal depictions) that sexualizes children

These policies are comprehensive and should effectively apply to the sexualized, non-nude content that is used as a standard advertisement for SG-CSAM sellers as well as a broad array of sexualized content featuring minors. Instagram's role as the key platform in our investigation is likely not due to a lack of policies, but ineffective enforcement.

## 6.2 Twitter

Twitter's child sexual exploitation policy[17] prohibits much of the same content, along with links to third-party sites that contain CSAM. However, its policy of permitting adult nudity and depictions of sexual behavior[18] in posts make removing explicit imagery a more manual process than on platforms such as Instagram.

Twitter explicitly bans:

- Visual depictions of a child engaging in sexually explicit or sexually suggestive acts;
- Links to third-party sites that host child sexual exploitation material
- Recruiting, advertising or expressing an interest in a commercial sex act involving a child, or in harboring and/or transporting a child for sexual purposes
- Trying to obtain sexually explicit media from a child or trying to engage a child in sexual activity through blackmail or other incentives
- Promoting or normalizing sexual attraction to minors as a form of identity or sexual orientation

[17] https://help.twitter.com/en/rules-and-policies/sexual-exploitation-policy
[18] https://help.twitter.com/en/rules-and-policies/media-policy

## 6.3 Discord

Discord's policies[19] are unique in their focus on the behavior not just of adults but of the minors who comprise a large percentage of their user base. Discord explicitly prohibits users under 18 from engaging "in sexual conduct or any conduct that puts your online or physical safety at risk", a broad prohibition against not only the image and video-based content we focus on in this report but text and audio-based sexual conduct.

Like Twitter, Discord allows adult nudity in channels marked 18+, and simple searches show hundreds of Discord servers dedicated to sexual content. This likely complicates Discord's enforcement efforts against the exchange of nude material from post-pubescent minors (B2 in the Tech Coalition's Industry Classification System[20]).

Discord specifically prohibits:

- Solicitation or sexual conduct between adults and minors
- Making sexual content available to minors (although how server admins are supposed to enforce this is not defined)
- Distribution of non-consensual intimate imagery
- Any content that "depicts, promotes, or attempts to normalize child sexual abuse."

## 6.4 Snapchat

Snap's Community Guidelines[21] explicitly mention self-generated nudity of minors with specific guidance:[22]

> "Never post, save, send, forward, distribute, or ask for nude or sexually explicit content involving anyone under the age of 18 (this includes sending or saving such images of yourself)."

It further prohibits:

- Promoting, distributing, or sharing pornographic content.
- Commercial activities that relate to pornography or sexual interactions
- Activity that involves sexual exploitation or abuse of a minor, including sharing child sexual exploitation or abuse imagery, grooming, or sexual extortion (sextortion).

[19] https://discord.com/guidelines
[20] https://paragonn-cdn.nyc3.cdn.digitaloceanspaces.com/technologycoalition.org/uploads/Tech_-Coalition_Industry_Classification_System.pdf
[21] https://values.snap.com/privacy/transparency/community-guidelines
[22] https://values.snap.com/privacy/transparency/community-guidelines/sexual-content

## 6.5 TikTok

TikTok has detailed Community Guidelines[23] outlining its policy and child safety measures, prohibiting:

> "...content that may put young people at risk of exploitation, or psychological, physical, or developmental harm. This includes child sexual abuse material (CSAM), youth abuse, bullying, dangerous activities and challenges, exposure to overtly mature themes, and consumption of alcohol, tobacco, drugs, or regulated substances."

It also limits discovery of content created by users under 16 (insofar as they can detect a user's real age), restricts direct messages to those 16 and older, and restricts the use of livestreaming features to users 18 and older.[24] CSAM and grooming are explicitly mentioned, notably including digitally created content:

> "We do not allow youth exploitation and abuse, including child sexual abuse material (CSAM), nudity, grooming, sextortion, solicitation, pedophilia, and physical or psychological abuse of young people. This includes content that is real, fictional, digitally created, and shown in fine art or objects."

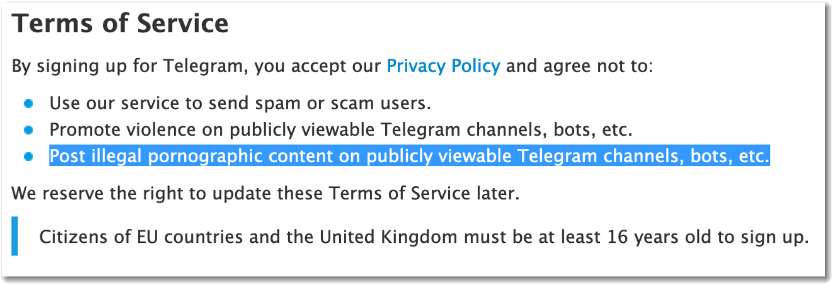
Also notable are TikTok's links to global and local support resources to those potentially seeking out CSAM,[25] though it is unclear under what contexts this information might surface to a user during actual use of the app.

## 6.6 Telegram

Telegram's Terms of Service[26] (see Figure 6 on the following page) states that posting illegal pornographic content is not allowed on publicly viewable channels, implicitly allowing CSAM on its platform, provided it is shared in private groups or direct messages.[27] It further states that "All Telegram chats and group chats are private amongst their participants. We do not process any requests related to them,"[28] presumably even if reported by a user. They further state that "To this day, we have disclosed 0 bytes of user data to third parties, including governments."

As noted in Section 5.3, Telegram has also been observed by SIO as failing to perform even basic content enforcement on public channels, with instances of known CSAM being detected and reported by our ingest systems. Given that use of PhotoDNA requires reporting material to NCMEC, which would violate their principle of not providing data to third parties, Telegram may not be using it at all.

[23] https://www.tiktok.com/community-guidelines/en/
[24] https://www.tiktok.com/community-guidelines/en/safety-civility/#4
[25] https://www.tiktok.com/safety/en-us/prevent-csam/
[26] https://telegram.org/tos/terms-of-service-for-telegram
[27] Note that private chats and channels, contrary to popular perception, are not end-to-end encrypted; see https://www.kaspersky.com/blog/telegram-why-nobody-uses-secret-chats/46889/.
[28] https://telegram.org/faq



### Terms of Service

By signing up for Telegram, you accept our Privacy Policy and agree not to:

- Use our service to send spam or scam users.
- Promote violence on publicly viewable Telegram channels, bots, etc.
- Post illegal pornographic content on publicly viewable Telegram channels, bots, etc.

We reserve the right to update these Terms of Service later.

> Citizens of EU countries and the United Kingdom must be at least 16 years old to sign up.

Figure 6: A screenshot of an except of Telegram's Terms of Service, as of May 11, 2023.

## 6.7 G2G

G2G is something of an outlier, in that it is acting as a de-facto payment provider rather than a networking or messaging service. Its safety guidelines[29] primarily center around scams, and a broad declaration that illegal activities are not allowed.[30] It allows payment with a broad range of e-wallets and cryptocurrencies,[31] with little in the way of identity verification measures for buyers. As a platform, it has relatively little data to work with to detect SG-CSAM-related activity; however, it is notable as a potential source of signal for other platforms, as well as a platform that may be useful in investigations to determine identity of buyers.

# 7 Recommendations for platforms

**Better proactive investigations and heavier enforcement on keywords and hashtags:** The proliferation of these accounts, particularly on platforms such as Twitter and Instagram, and the recurring patterns common to many accounts in the network (i.e., mentions of "menu", certain emoji in bio, obvious hashtags involving variations on "pedo"), indicate a general lack of resources devoted to detecting SG-CSAM and associated commercial activity. SG-CSAM distribution rings could easily have been detected and acted upon by internal investigation teams, and there are existing policies in place that cover this type of content, but accounts often stay active for months. Reactive enforcement based on user reports is inadequate: investigative reporters regularly find activity that platforms have missed, simply by actively looking rather than relying on reports. Platforms should have a well-staffed internal team responsible for the detection and content enforcement of SG-CSAM, with proactive discovery mechanisms at their disposal.

**Signal sharing across platforms:** Given the breadth of platforms used by the SG-CSAM ecosystem, information sharing between platforms and services on the changing characteristics, hashtags, keywords and advertising tactics could improve detection and enforcement mechanisms on an industry-wide level.

[29] https://support.g2g.com/support/solutions/articles/5000001431
[30] https://support.g2g.com/support/solutions/articles/5000861666-prohibition-of-illegal-activities
[31] https://www.g2g.com/payment-guide/

**Models to detect accounts that behave like buyers:** Apparent buyers had relatively distinctive positions in the network graph, and presumably have distinctive behaviors (following seller accounts, messaging underage users, receiving links to fileshare platforms). Platforms should build classifiers based off of these buyer accounts to detect them in the future and report to law enforcement when there is reason to suspect illegal activity.

**Better age detection mechanisms:** Sellers use fairly trivial and predictable methods of obfuscating their real age, in addition to keywords and hashtags that imply being underage. CSAM detection logic and classifiers should take these signals into account.

**Detection of gift card-related transactions:** Transactions involving gift cards are particularly likely to be used by unbanked younger sellers or in exchange of illicit goods. Similar to how various companies have mechanisms to detect credit card numbers in data where they should not appear, mechanisms to detect exchange of gift card redemption codes should be possible to implement as a signal that a financial transaction may be taking place.

**Reevaluation of recommendation systems:** It is very likely that part of the reason that Instagram and Twitter were platforms of choice for this network is because their recommendation systems are extremely efficient at suggesting similar accounts to follow, as a growth tactic. Any medium that is going to rely heavily on recommending content to other users—whether those be posts, accounts "Reels" or "Stories"—needs to be implemented in such a way that does not recommend SG-CSAM producers to buyers or vice versa.

**Education for sellers:** When an account is identified as selling SG-CSAM, disabling the account should be accompanied by messaging to the seller to attempt to discourage recidivism. This messaging might include:

- The fact that this content is widely illegal and can result in prosecution; being a minor does not prevent legal consequences.
- That it increases the user's risk for being a victim of sextortion, stalking and other harms.
- That producing this content helps prop up a CSAM industry where content is created against people's will—the content they sell may be traded to gain access to this.
- Links to supportive resources, such as NCMEC's Take It Down which can help prevent content from proliferating.

## 8 Ethical and Safety Considerations for Researchers

Academic and civil society research teams looking to have a positive impact against online child sexual exploitation (OCSE) have to navigate a thicket of complicated equities, including:

- Treating victims ethically and empathetically, even if they are participating in their own abuse

- Fulfilling legal requirements around the handling and reporting of CSAM
- Meeting organizational requirements around human subject research, if appropriate
- Protecting researchers from legal risk and potential emotional harms related to dealing with OCSE
- Creating trustworthy research that serves our goal of reducing the harm of online abuses
- Mitigating short-term harms via reporting illegal activity

The SIO team took these steps to fulfill these goals:

**Avoiding direct user interaction:** During this and other child safety projects, the SIO team has avoided any interactions with individual victims or offenders. The purpose of our research is to understand the network behaviors, not understand the individual motivations.

**Automatic scanning of high-risk media:** The SIO has deployed PhotoDNA scanning into all our standard media ingest pipelines to prevent the possibility of known CSAM being ingested or viewed by SIO researchers. If our detection system is triggered, the metadata of the image is encrypted and stored in a location only available to two senior SIO staff members (both of whom have worked on child safety investigations professionally) and the original image deleted before it touches persistent storage. This also triggers an automatic report to the NCMEC CyberTipline. SIO is currently integrating the new Take It Down[32] hash bank into our pipeline.

**Reporting to NCMEC:** Beyond the automatic reporting of known CSAM described above, SIO reached out to brief NCMEC analysts on our findings and provided them with full Maltego graphs of the networks of buyers and sellers we discovered. Our goal was to help NCMEC coordinate the various relevant law enforcement agencies as they investigated and took action on our reports.

**Briefing relevant platforms and coordinating bodies:** The SIO team also briefed the child safety teams at several platforms, including Instagram and Twitter. While NCMEC has the primary responsibility for relaying CyberTipline reports, such referrals from NCMEC are often aimed at specific offenders and not at the broader problem. In our briefings, we provided recommendations on the product and operational changes we believe could reduce the size and effectiveness of the SG-CSAM ecosystem. We also worked closely with Twitter to troubleshoot the unwanted appearance of known, hashed CSAM on public profiles.

We recommend that academic and other research groups performing work on Twitter data deploy PhotoDNA or an equivalent technology on any data ingest pipelines. This suggestion also holds for Telegram. Academic research groups are invited to contact the authors for assistance or implementation details.

[32] https://takeitdown.ncmec.org/

# 9 Conclusions

The social and technical causes and conditions that lead to the production of SG-CSAM are varied,[33] with mitigation requiring technical countermeasures but also education, social services and support—as well as preventing the circumstances in which minors might feel coerced or compelled to exchange SG-CSAM for money. Even when material is truly self-produced and distributed intentionally, minors do not have the ability to meaningfully consent to the implications of having widely distributed explicit material and the other harms for which it puts them at risk. Future work by qualified groups could include surveys or interviews of the producers of SG-CSAM, to understand how many producers have been forced into this ecosystem and what drove those who choose to sell underage sexual material to do so.

While the primary platforms identified as having significant SG-CSAM activity were Instagram and Twitter, a wide cross-section of the industry is connected to this ecosystem—some of which we could not analyze using open-source methods. An industry-wide initiative to limit production, discovery, advertisement and distribution of SG-CSAM is needed—not just with social media platforms, but with file sharing services, merchants and payment providers. Given the highly multi-platform nature of the problem, this will require a better knowledge sharing of the shifting nature of the SG-CSAM production networks, countermeasures and methods for identifying buyers.

This work also demonstrated a weakness the authors have noticed in the global child safety framework. While there are some examples of law enforcement and child-safety centers working on SG-CSAM as a major source of illegal material, the ease with which we found and explored this network, with no special data access or investigatory powers raises questions about the effectiveness of the current enforcement regime. Further work is necessary on whether the correct statutory and law enforcement framework exists to deal with this issue.

With as much time and energy that platforms and service providers have spent policing and deplatforming legal, adult sex workers, the lack of attention to commercial SG-CSAM and the apparent difficulty some platforms have in controlling it was unexpected and unfortunate. This is an issue needs enough resources devoted to it such that it can be identified and rapidly triaged—proactively, instead of based on user reports. We hope that this report will aid the industry in doing so, and will continue to to partner with technology and child safety organizations to further research and recommend countermeasures.

[33] https://www.thorn.org/blog/youth-continue-sharing-sg-csam/

Stanford | Internet Observatory
*Cyber Policy Center*

*The Stanford Internet Observatory is a cross-disciplinary program of research, teaching and policy engagement for the study of abuse in current information technologies, with a focus on social media. The Stanford Internet Observatory was founded in 2019 to research the misuse of the internet to cause harm, formulate technical and policy responses, and teach the next generation how to avoid the mistakes of the past.*

Stanford | Internet Observatory
*Cyber Policy Center*