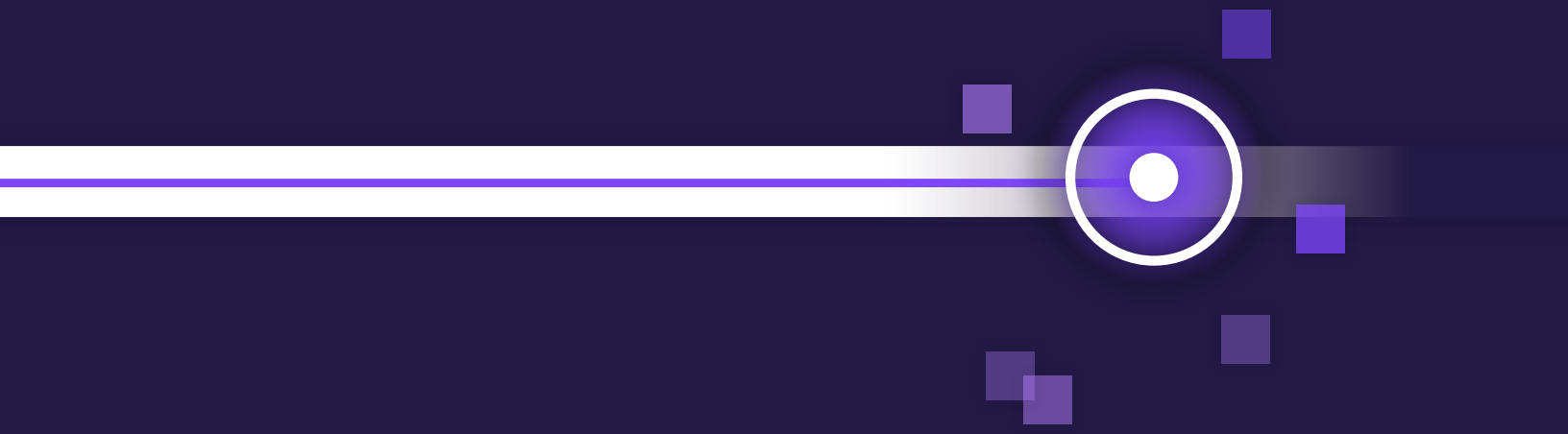
UNDERSTANDING AND MITIGATING
RISKS TO YOUR PLATFORM

Emerging Online Trends in Child Sexual Abuse

2023



safer BUILT BY THORN

CONTENTS

Safeguarding Children

In 2022, tech companies reported 88.3 million files of child sexual abuse material (CSAM) to the National Center for Missing and Exploited Children.

That's an average of 1.7 million CSAM files reported *per week*.

These numbers illuminate the alarming scale of CSAM. But they also reveal the critical role content-hosting platforms play in addressing it. Reports from tech companies constitute the majority received by NCMEC.

With the rise of user-generated content, the spread of CSAM accelerated. **The reality is, any platform with an upload button has an urgent need to address CSAM at scale.** Predators are also grooming and sextorting children online using self-generated CSAM as blackmail. Often, we're surprised to find CSAM and child exploitation spreading on platforms we use every day. But we can no longer ignore these threats.

Child predators today are more brazen than ever online—*because they're getting away with it.*

Eliminating CSAM from the web requires a focused and coordinated approach. We believe tech companies are key partners. Detection is the first step, and we're committed to empowering the tech industry with tools and resources to combat child sexual abuse at scale.

Clear trends emerged from our latest research on youth's attitudes, behaviors and actual experiences with online sexual threats. Captured in this report, these insights point to actionable steps tech companies can take to mitigate risk on their platforms.

Progress is being made: 236 companies submitted CyberTipline reports to NCMEC in 2022. That's an increase of 36% since 2020.

But we have a long way to go. This is about safeguarding our children. It's also about protecting your business and your users. With the right tools at our fingertips, together we can build a safer internet, one where every child is free to simply be a kid.

John Starr
VP of Strategic Impact
Thorn

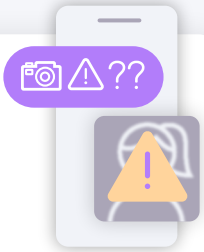
HOW SG-CSAM IMPACTS LIVES

Riley's Story

Thorn's research directly with youth indicates a sustained rise in **self-generated sexual imagery**. Riley's hypothetical story paints a picture of common ways this content spreads online, with devastating impacts.

Meet Rylie, a shy 12-year-old. She gets As, plays soccer, and just got her first phone.

Jake, a handsome 8th grader, begins texting her. Soon, he pressures her to send him an explicit photo.



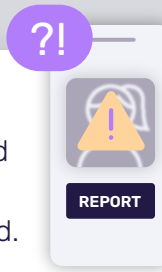
He shares it with a friend, who shares it in a group chat. It quickly spreads throughout school.

Eventually her image is shared on the open web. Messages swirl at school calling her a slut.

Riley's devastated and humiliated, but she can't bring herself to tell her parents.

Riley becomes withdrawn and depressed. Her parents don't know what's wrong.

She desperately tries to figure out how to report the image and get platforms to take it down. But she can't contain the spread.



She wants to just hide from the world and ***MAKE IT STOP.***

Self-Generated Child Sexual Abuse Material is on the Rise

DEFINITION

SG-CSAM

Explicit imagery of a child that appears to have been taken by the child in the image. It can result from consensual or coercive acts. Kids often refer to consensual experiences as “sexting” or “sharing nudes.”

Young people are increasingly taking and sharing sexual images of themselves.

Over four consecutive years, Thorn monitored the perspectives of 9- to 17-year-olds. The research identified a sustained increase in SG-CSAM.[1] From 2021 to 2022 alone, the Internet Watch Foundation also noted a 9% rise.[2] Self-produced sexual content often results from pressure, whether from peers or bad actors, and can quickly spread to the open web. Self-generated material - either consensually or coercively produced - now represents more than 25% of the circulated identified cases known to NCMEC.[3]

CONTRIBUTING FACTORS

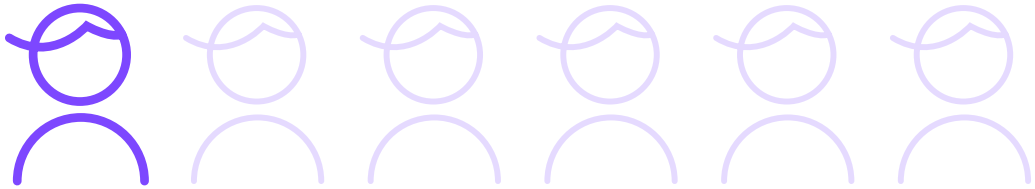
A surge in sextortion is one driver. Children are groomed or coerced into taking sexual images and videos of themselves and supplying them to online predators. In 2022, the FBI received more than 7,000 reports related to the online sextortion of minors.[4]

In addition to these high-risk pathways, Thorn’s research shows youth increasingly see sexting or sharing nudes as normal behavior. This includes sharing another child’s SG-CSAM without permission. The data suggests boys and Hispanic/Latino youth are at higher risk. Today, minors view sharing explicit images of themselves with adults they only know online as common as they do with minors they only know online.[1]

1 Self-Generated Child Sexual Abuse Material: Youth’s Attitudes and Experiences in 2021, Thorn, 2022
2 Internet Watch Foundation Annual Report, 2022
3 NCMEC CyberTipLine 2022 Report
4 US Department of Justice, 2022

1 in 6

minors has shared their own SG-CSAM[1]



43%

of minors who shared their own SG-CSAM did so with someone they didn't know offline[1]

What's the role of platforms?

THIS NEW CSAM CAN'T BE FOUND WITH HASHING & MATCHING.

Because these self-generated images and videos often represent new CSAM, hashing and matching won't detect them. **Platforms must use a CSAM Classifier to identify this content.**

CASE STUDY | flickr

Flickr Uses Safer's CSAM Image Classifier to Detect New CSAM

THE CHALLENGE

- Millions of photos are uploaded to Flickr every day, and with those comes a responsibility to keep the platform safe.
- Flickr sought to expand detection to include unknown CSAM — material that's never been hashed or seen, or is altered or computer generated.

THE SOLUTION

- In 2021, Flickr deployed Safer's CSAM Image Classifier, a machine learning classification model trained to detect new and previously unknown CSAM.
- **Flickr's Trust & Safety team used the Classifier to increase their efficiency and detect images they likely wouldn't have discovered otherwise**, except through user reports.

THE RESULTS

- One Classifier hit led to the discovery of 2,000 previously unknown images of CSAM and a law enforcement investigation.
- Flickr added the new hashes to their SaferList, thereby sharing them with other Safer customers. As a result, another Safer user found new CSAM on their platform from that series.

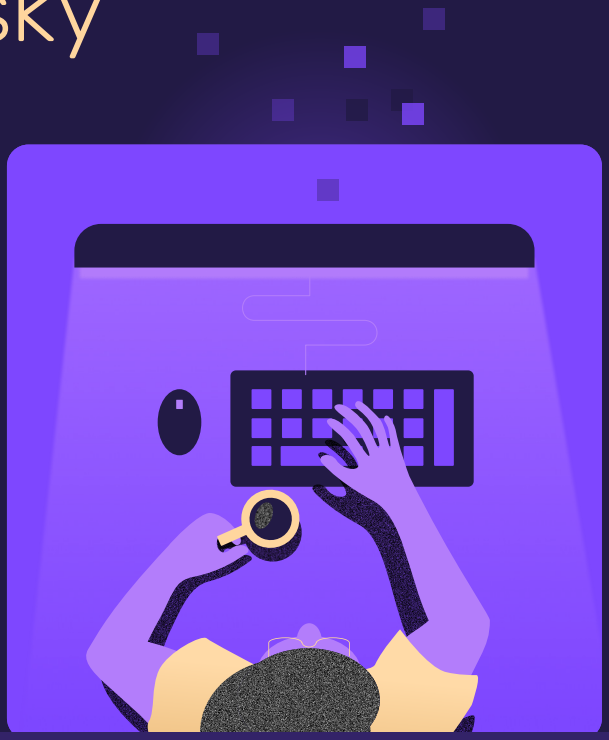
IN 2022

14,537	8,692	34,176
Classifier hits	Hash matches	Files reported to NCMEC

"We don't have a million bodies to throw at this problem, so having the right tooling is really important for us. [Thorn's technology] enables us to be proactive. We're not just keeping our platform safe, we're protecting it and making it a safe, equitable platform."

JACE POMALES, Trust and Safety Manager, Flickr

Youth are Having Risky Online Interactions With Adults



DEFINITION

Online Grooming

The intentional use of the internet to manipulate and/or coerce someone into participating in sexually explicit interactions or exchanges.

Child predators grew bolder over the past year.

ENTICEMENT

From 2021 to 2022, NCMEC saw an astounding 82% increase in reports of online enticement of children for sexual acts.[5] Also in 2021, Thorn surveyed 1,200 youth (aged 9-17) about their experiences with online flirting and their responses to threats of grooming and abuse. Two in 5 youth reported being approached by someone they thought was attempting to "befriend and manipulate" them.[6] It's logical to expect these numbers to increase as generative AI tools allow for scaling grooming attempts.

ISOLATING VICTIMS

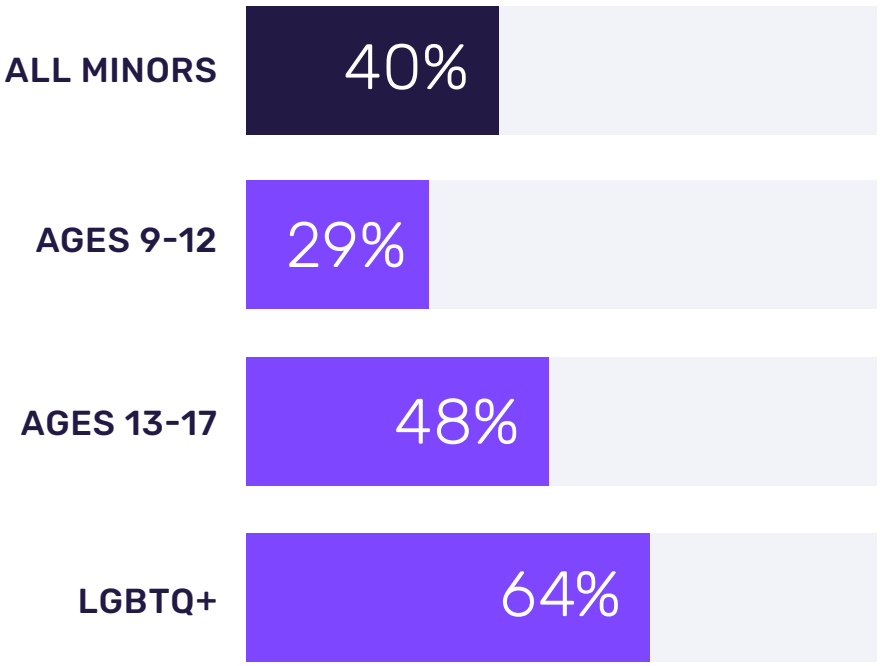
Online offenders use proven tactics. In the same way that offline abusers build rapport and then isolate their victims prior to hands-on abuse, so too do online perpetrators. After meeting minors in public forums, they move victims across platforms to increase both their own security and the victim's isolation. Thorn's research showed 2 in 3 minors reported being asked by someone they met online to move from a public forum to a private conversation on a different platform.[5]

MAINTAINING CONTACT

Even if made to feel uncomfortable, a minor may not cut off contact. Nearly one quarter of kids stayed in contact with someone online who made them uncomfortable, with LGBTQ+ youth more than twice as likely to be in this position.[5]

5 NCMEC CyberTipLine 2022 Report
6 Online Grooming: Examining Risky Encounters Amid Everyday Digital Socialization, Thorn, April 2022

Percentage of minors who have received a cold solicitation online[6]



1 in 7

minors are asked for nudes by a stranger online daily or weekly[6]

1 in 5

minors reported having an online sexual interaction with someone they believed to be an adult[7]

What's the role of platforms?

These numbers stress the urgent need for relevant and scalable interventions. By infusing a **safety-by-design** approach (see page 14), companies can eliminate harms before they occur.

CONSIDER ADVISORY SERVICES

Thorn's team offers guidance on child safety policies, intervention and prevention strategies, and safety-by-design, along with workshops that help platforms hear firsthand from youth.

7 Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting and Blocking in 2021, Thorn, February 2023

Youth Rely More on Built-In Safety Tools Than Offline Help

TYPES OF SAFETY TOOLS

Block Report Mute

Safety tools, like blocking and reporting, remain a child's first choice for addressing bad actors online.

Often, safety tools provided by platforms are the only place a child will disclose a harmful interaction. Thus, even as platforms build prevention strategies, they must optimize their current features.

YOUTH'S REPORTING BEHAVIORS

Thorn's research found that 84% of minors who faced an online sexual interaction used a built-in safety tool to respond. In fact, youth are 2.5x more likely to use these tools than to seek help offline, from say parents or caregivers. Anonymity plays a key role — 75% said privacy makes a tool more appealing to use. Of the tools, minors prefer blocking over reporting. And yet, they indicate these actions ultimately do little to stop a problematic user.[7]

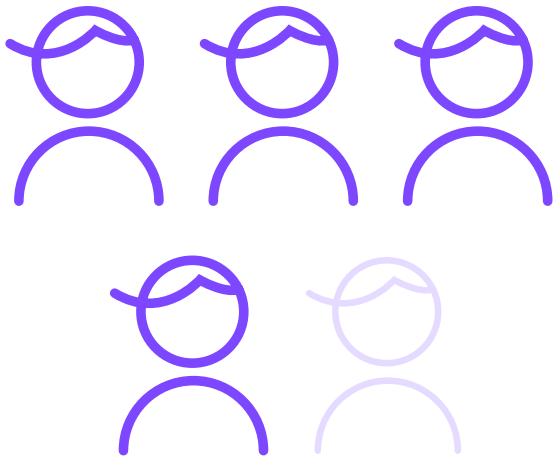
INSUFFICIENT AND INCONSISTENT TOOLS

Today, platforms' child safety systems and policies have not kept pace with the risks youth face online. After blocking or reporting threatening users, 50% of minors continue to be recontacted, leading to potentially recurring harm. Not surprisingly, many minors think an app will do nothing in response to reported abuse.[6]

Adding challenges for youth, responsive actions vary from platform to platform. On one platform, blocking users disables them from viewing a child's videos or engaging with their content. On another, it results in removal of the bad actor's likes and comments but still allows that user to see the child's social activities.

4 in 5

minors want platforms to provide more information on how to stay safe from risky online sexual interactions[7]



50%

of minors experience recontacts from problematic users after blocking or reporting them[7]

What's the role of platforms?

Even as platforms build prevention strategies, they must improve their reporting features based on youth's real behaviors and experiences. A safety-by-design approach (see next page) supports both initiatives.

Safety by Design

DEFINITION

Safety by Design

An approach to the design of online products and services that focuses on user safety and rights to minimize online harm by anticipating, detecting and eliminating threats before they occur.

Safety by Design encourages thoughtful development: Rather than retrofitting safeguards after an issue has occurred, technology companies should strive to minimize threats and harms to children throughout the development process.

This proactive approach incorporates three principles:

1 TECHNOLOGY PLATFORM RESPONSIBILITY

Children shouldn't be asked to defend themselves. From the outset, teams should understand online threats, anticipate where and how they might occur, and address them in the design and engineering of products.

2 EMPOWER YOUTH

Tools should be built to serve the best interests of youth. Blocking and reporting remain minors' preferred response to bad actors, yet 50% report being recontacted. As their first line of defense, these features must enable children to promote their own safety.

3 YOUTH-INFORMED DESIGN

Incorporate youth's perspectives in the design process to enhance their protective tools. Platforms can gain feedback on tools and features directly from youth through Thorn's NoFiltr Youth Innovation Council workshops.

These three steps advocate for safety built into each stage of the design process, from ideation through production. Teams must ask themselves: "What are the risks to the child if we build our product this way?" And, "How do we design-out areas of harm to ensure it's safe for kids before we release it to the world?"

TikTok

"The input we received from Thorn/NoFiltr's Youth Innovation Council has been invaluable to our team, as we continue to enhance our policies and tools with teens' feedback at the forefront."

KRISTELLE LAVALLEE COLLINS, TikTok's Youth Safety & Wellbeing Policy Lead, North America

Conclusion

The internet provides a world of wonder for kids, and today's innovative platforms collectively create a vibrant and thriving online ecosystem. At the same time, predators continue to exploit these spaces to access and abuse children.

With the right policies in place, and products and features designed with child safety in mind, companies can protect their businesses, their users and our children.

Equip your Trust & Safety team with the right tools

- Use hashing and matching technology to detect known CSAM and ensure users aren't exposed to abuse content. Safer offers the largest database in the world of known hashes.
- Use machine learning classification models, like Safer's CSAM Classifier, to detect new and previously unknown content, like self-generated CSAM.

Review your child safety policies and enforcement

- Empower youth by giving them greater control of their own safety online through built-in tools that incorporate their real-life experiences.
- Leverage consultation services—like Thorn's—on child safety policies and enforcement, on-platform intervention, and prevention strategies.
- Have clear procedures for reporting CSAM discovered on your platform. Safer's reporting tool supports you in sending quality reports to NCMEC or RCMP.

Take a Safety-by-Design approach when building new products and features

- Build safety into the entire development process to minimize harms before they occur.
- Incorporate advice from Thorn's experts on product-policy considerations when developing a new product or feature.
- Control your security and scale CSAM detection while delivering the data privacy your users expect. Safer integrates directly with your tech stack.

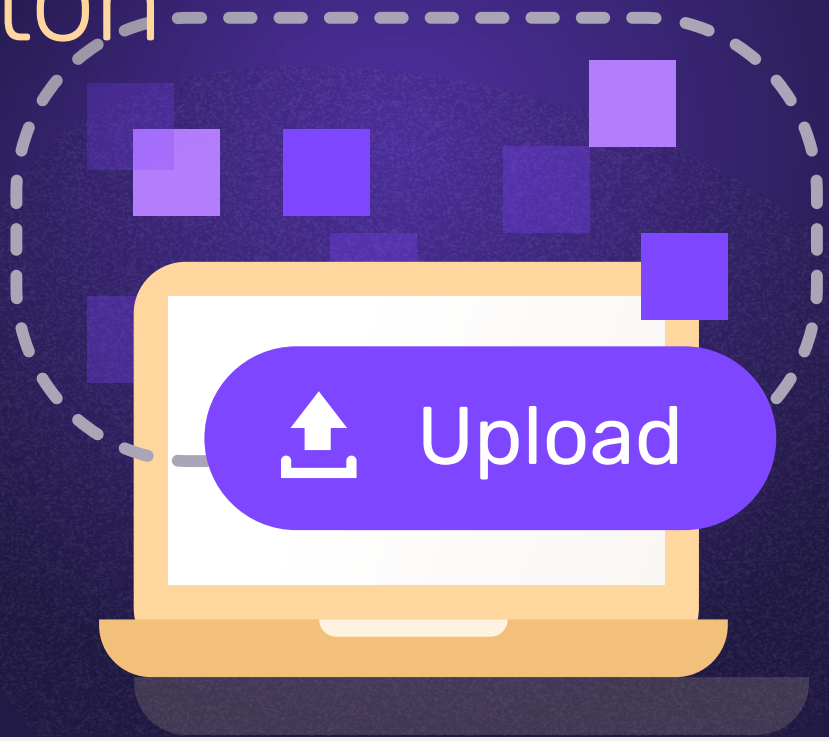
The astounding scale and acceleration of CSAM and SG-CSAM, among many other abuses, evidences our need to act now. As a united digital community, we can, and must, meet this challenge. In doing so, we will forge an internet that not only protects our children but empowers them to be kids.

Ready to help build a safer internet?

To learn more about Safer, contact us.
For consulting, email info@wearethorn.org.

CSAM Detection for Any Platform with an Upload Button

An all-in-one solution built by experts in child safety technology



safer BUILT BY THORN

Comprehensive CSAM Detection

Detect known and new CSAM with the largest database of hashes (32+ million) and machine learning image and video Classifiers.

Secure, On-Premises Deployment

Keep data secure and maintain user privacy while accessing the CSAM detection tools you need, with self-hosted deployment.

Cross-Platform Hash Sharing

Leverage hashes contributed by other Safer customers and help improve cross-platform intelligence to diminish the viral spread of CSAM.

CONTACT US TO SCHEDULE A DISCOVERY CALL TODAY

safer.io/contact