Safety Centre

Online child protection

Overview

Across our family of apps, we take a comprehensive approach to child safety that includes zero tolerance policies prohibiting child exploitation, cutting-edge technology to prevent, detect, remove and report policy violations, and victim resources and support. We also collaborate with industry child safety experts and civil society around the world to fight the online exploitation of children because our commitment to safeguarding children extends beyond our apps to the broader Internet.

Our efforts

At Facebook, our work on child safety has spanned over a decade. Our industry-leading efforts to combat child exploitation follow a three-pronged approach: prevent this abhorrent harm in the first place; detect, remove and report exploitative activity that escapes these efforts; and work with experts and authorities to keep children safe. We apply this approach across both the public spaces of our platform, such as Pages, Groups and profiles, as well as on our private messaging services, such as Messenger. In addition to zero tolerance policies and cutting-edge safety technology, we make it easy for people to report potential harm, and we use technology to prioritise and swiftly respond to these reports. We have specially trained teams with backgrounds in law enforcement, online safety, analytics and forensic investigations review potentially violating content and report findings to the National Center for Missing and Exploited Children (NCMEC). To understand more about the specifics of our approach to safety on WhatsApp, go here and on Instagram, go here.

Safety Centre

Understanding offenders

Understanding how and why people share child exploitative content is critically important to deploying effective comprehensive solutions to combat it. To help with that understanding, we conducted an in-depth analysis of the illegal child exploitative content that we reported to the National Center for Missing and Exploited Children from Facebook and Instagram during two representative months, and used what we learned to deploy new tools and launch new programmes targeted at reducing the sharing of this abhorrent content. Here's what we found:

- More than 90% of this content was the same as or visually similar to previously reported content.
- Copies of just six videos were responsible for more than half of the child exploitative content that we reported in that time period.
- While this data indicates that the number of pieces of content does not equal the number of victims, and that the same content, potentially slightly altered, is being shared repeatedly, one victim of this horrible crime is one too many.

The fact that only a few pieces of content were responsible for many reports suggests that a greater understanding of intent could help us to prevent this revictimisation. With that in mind, we worked with leading experts on child exploitation, including NCMEC, to develop a research-backed taxonomy to categorise a person's apparent intent in sharing this content. Based on this taxonomy, we evaluated 150 accounts that we reported to NCMEC for uploading CSAM (child sexual abuse material) in July and August of 2020 and January 2021, and we estimate that more than 75% did not exhibit malicious intent (i.e. did not intend to harm a child). Instead, these accounts appeared to share for other reasons, such as outrage or poor humour. While this study represents our best understanding, these findings should not be considered a precise measure of the child safety ecosystem. Our work to understand intent is ongoing.

Based on our initial findings, we are developing some targeted solutions, including new tools and policies, to reduce the sharing of this type of content. We've started by testing two new tools: one aimed at potentially malicious searches for this content and another aimed at sharing of this content for reasons other than to harm a child. The first intervention is a pop-up that is shown to people who initiate searches on Facebook using terms associated with child exploitation. The pop-up offers offender diversion resources from child protection organisations and shares information about the consequences of viewing illegal content. The second is a safety alert designed for people who have shared viral memes of child exploitative content, informing them of the

reason, resharing this content is illegal and revictimises the child and that they should formally report it instead.

Our work to understand intent is ongoing, and based on our findings, we'll continue to develop targeted solutions, including new tools and policies to reduce the sharing of this type of content across both our public platforms and private messaging.

Our policies

Our Community Standards outline what is and is not allowed on Facebook. We do not allow content that sexually exploits or endangers children. When we become aware of apparent child sexual exploitation, we report it to NCMEC in compliance with applicable law. We also work with external experts, including the Facebook Safety Advisory Board, to discuss and improve our policies and enforcement around online safety issues, especially with regard to children. We know that sometimes people share nude images of their own children with good intentions; however, we generally remove these images because of the potential for abuse by others and to help avoid the possibility of other people reusing or misappropriating the images. Read our Community Standards on child nudity and sexual exploitation of children here.

User reporting

If you see a post on Facebook that you think violates our Community Standards, you can report it by using the "Report" link that appears when you tap on the downwards arrow or "..." next to a post. On all major surfaces, you can additionally tell us that troubling content involves a child. To learn more about how to report a photo or video, please visit the Help Centre. You can also report a conversation on Facebook by tapping the name of the person you're having the conversation with and selecting Something's Wrong. To learn more, click here.

When you report a post on Facebook, our specially trained teams determine whether to take action on the post. This team works 24 hours a day, seven days a week around the globe. We don't include any information about the person who filed the report when we contact the person who posted the reported content.

Safety Centre

Tools and technology

Technology is our business, and we use an array of industry-leading tools and technologies to fight child sexual exploitation. We use these along all three prongs of our approach to child protection – prevention, detection and response – and in ways tailored to public spaces such as Pages, Groups and profiles, as well as to our private messaging services.

For example, to prevent harm, we've designed many of our features to remind minors who they're sharing with and to limit interactions with strangers. This includes protecting sensitive information – including minors' contact information, school or birthday – from appearing in search to a public audience. Additionally, we take steps to remind minors that they should only accept friend requests from people they know, and we do not allow unconnected adults to message minors.

Amongst the detection technologies that we use are photo matching technologies that help us detect, remove and report the sharing of images and videos that exploit children. These photo matching technologies create a unique digital signature of an image (known as a "hash"), which is then compared against a database containing signatures (hashes) of previously identified illegal images to find copies of the same image. We use these technologies across our public surfaces, as well as on unencrypted information available to us on our private messaging services, including user reports. We also run these technologies on links from other Internet sites shared on our apps and their associated content to detect known child exploitation housed elsewhere on the Internet. Not only does this help keep our platforms safer, but it also helps keep the broader Internet safer, as all violating content is reported to NCMEC.

In addition to photo matching technology that detects known images, we're using artificial intelligence and machine learning to proactively detect child nudity and previously unknown and new child exploitative content, as well as inappropriate interactions with children, sometimes referred to as "grooming".

Across our private messaging services, we also use highly predictive behavioural signals, information from user reports, and content and signals from our public spaces to prevent, detect and respond to harm to children. For example, on Messenger we recently launched a feature that educates people under the age of 18 to be cautious when interacting with an adult who they may not know and empowers them to take action before responding to a message.

Safety Centre

Direct.

Transparency and accountability

At Facebook, we strive to be open and proactive in the way we safeguard users' privacy, security and access to information online. We've published biannual transparency reports since 2013. We also release a quarterly Community Standards enforcement report, which includes data on action that we take against violating content on Facebook, Messenger and Instagram. We believe that increased transparency tends to lead to increased accountability and responsibility over time, and publishing this information will push us to improve more quickly too.

In our Community Standards enforcement report, you will find information on the amount of content that we removed for violating our child nudity and exploitation policies as well as information on trends and prevalence. You can read our most recent enforcement report here.

Building the child protection ecosystem

Child protection requires a global and comprehensive response from industry, law enforcement, government, civil society and families, which is why we are committed to working with child safety stakeholders to build and support the child safety ecosystem. We also collaborate across the industry through organisations such as the Technology Coalition - Fighting Child Sexual Abuse, an industry association dedicated solely to eradicating the sexual exploitation of children online.

Partnering with experts

We work closely with our Safety Advisory Board, which is comprised of leading online safety charities, as well as over 400 safety experts and NGOs from around the world, including specialists in combating child sexual exploitation and aiding its victims. Our efforts include developing industry best practices, building and sharing technology to fight online child exploitation, and supporting victim services, amongst other things.

Safety Centre

- **Open sourcing PDQ-TMKF**

  In 2019, we developed two open source photo and video matching technologies that detect identical and nearly identical photos and videos. Known as PDQ and TMK+PDQF, these technologies are part of a suite of tools that we use at Facebook to detect harmful content. These technologies allow industry partners, developers and charities to more easily identify abusive content, share hashes (or digital fingerprints) of different types of harmful content, and allow hash-sharing systems to communicate with each other, making the systems that much more powerful.

- **Aselo: Bringing modern technology to child helplines**

  Since 2020, we have been supporting TechMatters' project, Aselo: Bringing modern technology to child helplines, to develop new technology to support child helplines around the world and make them more accessible to children in crisis. TechMatters is building a customisable, open source contact centre platform that allows children and young people to get in touch with helplines via voice, SMS, web chat, WhatsApp and Facebook Messenger if they come across child abuse material.

- **Rebuilding NCMEC's case management tool**

  Since 2019, we have been working with NCMEC to update their case management tool to ensure that international law enforcement can receive, triage, manage and organise the CyberTipline reports as part of the global response to child sexual exploitation.

- **Hosting child safety hackathons**

  Since 2016, we have hosted regular child safety hackathons with NGOs, including the National Center for Missing and Exploited Children and Thorn, to bring together engineers, data scientists and designers from across the industry, as well as non-profit partners, to code and prototype projects focused on making the Internet a safer place for children.

In 2020, we launched a pilot with the UK Government, Internet Watch Foundation and the NSPCC for minors to report sexually explicit, self-generated images through a partnership with the National Center for Missing and Exploited Children. The reporting channel, Report Remove, will be launched formally in Spring 2021.

- Amplifying AMBER Alerts

  We launched AMBER Alerts on Facebook in 2015 to help families and authorities successfully recover missing children and have since expanded the programme to over 20 countries. People in a designated search area where local law enforcement has activated an AMBER Alert will see the alert in their News Feed. The alert includes a photo of the missing child, a description, the location of the abduction, and any other pertinent, available information. People can share the alert with friends to spread awareness, tapping into an organic desire to help. We know that the chances of finding a missing child increase when more people are on the lookout, especially in the critical first hours. Our goal is to help get these alerts out quickly to the people who are in the best position to help.

Partnering with industry

Fifteen years ago, the Technology Coalition was formed when industry leaders came together to fight online child sexual exploitation and abuse (CSEA). In 2020, Facebook joined Google, Microsoft and 15 other member companies of the Technology Coalition to launch ProjectProtect – a plan to combat online child sexual abuse. This includes a renewed commitment and investment from the Technology Coalition, expanding its scope and impact to protect kids online and guide its work for years to come. ProjectProtect focuses on five key areas: tech innovation, collective action, independent research, information and knowledge sharing, and transparency and accountability. Read more about the recent efforts of the Technology Coalition.

From our partners

demonstrated their leadership and willingness to be a proactive leader in the fight to keep the Internet safer for everyone. While the amount of content taken down may be surprising, it is a reminder that the sexual exploitation of children is a global problem that demands a multi-faceted global solution."

Michelle C DeLaune, NCMEC Chief Operating Officer

Child safety news

11 February 2020: Our commitment to keeping people safe

11 June 2020: Industry effort to fight child exploitation online

11 August 2020: Community standards enforcement report, August 2020

20 December 2020: Changes to Facebook messaging services in Europe

1 August 2019: Open sourcing photo and video matching technology to make the Internet safer

24 May 2018: Using technical solutions to keep children safe

24 October 2018: New technology to fight child exploitation

Family of apps child safety initiatives

How WhatsApp fights child exploitation

Keeping Instagram a safe and supportive space

Messenger safety and privacy

Safety Centre

Our family of companies:

About

Create Page

Careers

Create ad

Developers

Privacy

Safety Centre

Terms

Help

Meta © 2022

English (UK)