

Guide for tech companies considering supporting the “Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse”

February 2021



Engagement with the Voluntary Principles

Developed for industry, by industry, this informational guide is intended to assist tech companies considering operationalizing the “Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse” (Voluntary Principles), as appropriate to their platform or service.

Because the Voluntary Principles are designed to be flexible and non-prescriptive, this guide does not recommend a particular approach toward implementing any or all of the Principles. Rather, it provides an overview of operational, policy, and other practices that may be relevant, based on the experiences of leading companies involved in the development of the Principles.

We recognize and understand that companies differ from one another and may take unique, tailored approaches to building their online child safety program. We have written this document with the understanding that the practices should be treated as a starting point for internal discussions and are not intended to set or define an industry standard of care.

The Voluntary Principles

The Five Country Ministerial (US, UK, Australia, Canada, and New Zealand) launched the Voluntary Principles on 5 March 2020.

Developed in consultation with six technology companies, the Voluntary Principles:

- aim to provide a consistent and high-level framework to combat online child sexual exploitation and abuse (CSEA)
- are intended to drive collective action across the tech industry
- are intended to be flexible and acknowledge that every service is different, with a different risk profile, and
- will evolve over time.

Supporting the Voluntary Principles provides a way for companies to demonstrate their commitment to countering online CSEA.

You can find the Principles, and further context for each, [here](#).

Considering how to operationalize the Voluntary Principles

Companies of all sizes may wish to consider how the Voluntary Principles apply to their platforms and services, although the steps they might take will likely differ depending on the nature of their service and available resources. How a company applies the Voluntary Principles will also depend on technical requirements, available resources, and legal and privacy considerations, to name a few. The guide is not intended to be prescriptive or exhaustive but rather offers suggestions for practical action, along with illustrative examples of how other companies have approached these issues, and links for further information. Please note that legal restrictions may vary depending on the relevant jurisdiction, be subject to review, and impact products in different ways – always seek the advice of local counsel.

For companies who are members of the Technology Coalition (TC), an industry group dedicated to fighting online CSEA, more details are available in the TC Starter Kit, which gives detailed information about establishing or extending a child safety program. To join the TC, please [visit the TC’s website](#).

Note that throughout this guide we use the terms “CSEA” and “child sexual abuse material” (CSAM) – the latter refers to unlawful material.



Principles 1 & 2: Prevent child sexual abuse material (CSAM)


Principle 1: Companies seek to prevent known child sexual abuse material from being made available to users or accessible on their platforms and services, take appropriate action under their terms of service, and report to appropriate authorities.

Principle 2: Companies seek to identify and combat the dissemination of new child sexual abuse material via their platforms and services, take appropriate action under their terms of service, and report to appropriate authorities.

You may wish to consider:

- Ensuring your terms of service or other content policies explicitly prohibit uploading, posting, transmitting, sharing, accessing, or making CSAM available.
- Ensuring your terms of service or other content policies outline the consequences of posting or sharing CSAM, such as referral to the appropriate authorities and/or account termination.
- Providing a mechanism to report abuse on the platform or service (note that there may be legal restrictions on how CSAM may be reported).
- Providing consumers with (region-specific, where possible) information on ways to report CSAM to the appropriate authorities, and/or links to national reporting hotlines or the International Association of Internet Hotlines (INHOPE).
- Deploying automated tools to detect duplicates of CSAM photos and videos based on existing, known imagery, such as robust hash-matching or URL-blocking technologies.
- Deploying tools or features designed to prevent the creation of, interaction with, and dissemination of, CSAM.
- Incorporating relevant CSAM hash-sharing databases, and keyword and URL lists, such as The National Center for Missing & Exploited Children's hash database, the Internet Watch Foundation URL list or the Thorn Keyword Hub.
- Employing safety-enhancing technology, such as machine learning classifiers or other tools to detect and remove never-before-hashed CSAM imagery.
- Prioritizing responding to reports of CSAM, including by deploying automated tools and human review.

- Ensuring you have the internal policies and processes in place to review and appropriately action reports of CSAM. Policies and procedures may address:
 - o Taking action to manage the content (e.g., by removing or "quarantining" it) and to impose consequences on the user (e.g., terminating their account).
 - o Compliance with legal reporting requirements.
 - o The preservation of evidence.
 - o Staff training and measures to support staff resilience, safety and well-being.



Principles 3 & 4: Target online grooming and preparatory behavior

Principle 3: Companies seek to identify and combat preparatory child sexual exploitation and abuse activity (such as online grooming for child sexual abuse), take appropriate action under their terms of service, and report to appropriate authorities.

Principle 4: Companies seek to identify and combat advertising, recruiting, soliciting, or procuring a child for sexual exploitation or abuse, or organizing to do so, take appropriate action under their terms of service, and report to appropriate authorities.

You may wish to consider:

- Ensuring your terms of service or other content policies explicitly prohibit grooming and other potentially illegal or inappropriate contact and conduct, particularly with minors.
- Where appropriate, ensuring contractual, and other terms prohibit the use of advertising for CSAM-related or other unlawful activities. Commercial and/or monetized content may be subject to different measures.
- Conducting public awareness-raising activities including offering digital citizenship and e-safety resources for children.
- Deploying automated tools to help detect

potential online grooming by analyzing behavior and metadata and interactive chat to detect possible signals of risky behavior.

- Prioritizing response to reports of possible grooming, trafficking, or other CSAM-related behavior, including through human review, moderation, and/or reporting to the authorities, where appropriate.
- Providing parental control and other tools to help limit unwanted contact.



Principle 5: Target livestreaming

Principle 5: Companies seek to identify and combat the use of livestreaming services for the purpose of child sexual exploitation and abuse, take appropriate action under their terms of service, and report to appropriate authorities.

You may wish to consider:

- Ensuring the prohibitions in your terms of service or other content policies have been drafted to include livestreamed CSEA.
- Providing a mechanism for user reporting, including of imminent physical harm, within your livestreaming service.
- Establishing internal policies and procedures to:
 - halt, interrupt, or disable a livestream
 - preserve evidence after becoming aware of CSEA, and
 - to report potential imminent or presumed ongoing harm to the appropriate authorities.
- Investing resources in abuse protections specific to livestreaming.



Principle 6: Search

Principle 6: Companies seek to prevent search results from surfacing child sexual exploitation and abuse, and seek to prevent automatic suggestions for such activity and material.

You may wish to consider:

- Providing a mechanism for reporting URLs that are making accessible or disseminating presumed CSAM.
- Where CSAM is confirmed, ensuring it is removed from the search index and reported to the appropriate authorities.
- Deploying machine learning systems, whether developed in-house or made available by another company or non-governmental organization, that can help detect potential CSAM content.
- Developing your search algorithms to seek to prevent images, videos, or websites containing CSAM from appearing in search results or search suggestions.
- “Turning off” auto-complete or suggestion functionality for suspected CSAM-seeking queries.
- Where available and appropriate, establishing partnerships to implement deterrence programs (e.g., in response to suspected CSAM-seeking search queries, generating information on interventions for those at risk of offending, such as links to support services, or generating warning messages).



Principle 7: A specialized approach for children

Principle 7: Companies seek to adopt enhanced safety measures with the aim of protecting

children, in particular from peers or adults seeking to engage in harmful sexual activity with children; such measures may include considering whether users are children.

You may wish to consider:

- Integrating child safety considerations into your company's policies, procedures, and processes. This might include :
 - o Having a point person or team to manage child safety issues across the company.
 - o Taking a safety-by-design approach to developing new products or features.
 - o Partnering with governments, educators, and others on awareness-raising and capacity-building projects aimed at young people.
- Indicating relevant age requirements for your platform or service.
- Making available awareness-raising and educational resources to enable children and parents to make informed decisions about their use of a product or service.
- Putting in place appropriate default security, privacy, and safety settings based on user age.

- Promoting national support services where survivors and victims can seek help and guidance.
- Working with expert organizations that support survivors of child sexual abuse to inform potential reporting pathways.
- Developing policies to address content that, while may not rise to CSAM, is sexually exploitative of minors or is part of a known series linked to CSAM.



Principles 9, 10 & 11: Collaborate & respond to the evolving threat

Principle 9: Companies seek to take an informed global approach to combating online child sexual exploitation and abuse and to take into account the evolving threat landscape as part of their design and development processes.

Principle 10: Companies support opportunities to share relevant expertise, helpful practices, data, and tools where appropriate and feasible.

Principle 11: Companies seek to regularly publish or share meaningful data and insights on their efforts to combat child sexual exploitation and abuse.

You may wish to consider:

- Joining collaborative industry associations or multi-stakeholder bodies dedicated to combating CSEA.
- Sharing information and expertise with other companies and non-governmental organizations.
- Developing ways to feed updated threat information into relevant internal processes, including design, development, and moderation decisions.
- Familiarizing yourself with evidence-based research on CSEA risks and child-protection systems.
- Releasing periodic transparency reports with metrics on actions taken regarding



Principle 8: Victim/survivor considerations

Principle 8: Companies seek to take appropriate action, including providing reporting options, on material that may not be illegal on its face, but with appropriate context and confirmation may be connected with child sexual exploitation and abuse.

You may wish to consider:

- Providing guidance for users on reporting inappropriate or concerning content, contact, or conduct.
- Prioritizing possible CSEA-related content for human review.

CSAM and CSEA.

- For members of the Technology Coalition, familiarizing yourself with the Transparency Report guide.

Examples, useful links, and other resources

Terms of service and other content policies

Facebook's Community Standards on child nudity and sexual exploitation

Roblox Community Rules: Consequences and Violations

Community standards for Xbox

YouTube policies: Child safety on YouTube

Roblox Community Rules: Child endangerment

Snapchat: Community Guidelines

Twitter Rules and policies: Child sexual exploitation policy

Microsoft Services Agreement

Trust and safety centers and other guidance

Google Safety Center

Microsoft: Online Safety

Snapchat Safety Center

Twitter Help Center

Facebook Help Center

Microsoft: Stay alert to online grooming

Staying Safe on Snapchat

Xbox Family Settings App

Control your Twitter experience

Google: Family Link

Roblox: Safety Features: Chat, Privacy & Filtering

Reporting advice and mechanisms

Google For Families Help: Report inappropriate content or behavior toward children

Snapchat Support: Report abuse on Snapchat

Facebook Help Center: How to Report Things on Facebook

Twitter: Report a child sexual exploitation issue

Google: Reporting child endangerment and

Protecting Children

Microsoft: Report Abuse in OneDrive

Instagram: Report something: Exploitation: Human trafficking

Report a concern to Bing

YouTube Help: Report inappropriate content

INHOPE's hotline referral site (international)

The National Center for Missing and Exploited Children's (NCMEC) CyberTipline (U.S.)

The Child Helpline International Network (international reporting)

Canadian Centre for Child Protection (Canada's tipline)

Transparency reports

Twitter: Transparency

Google Transparency Report / YouTube Community Guidelines Enforcement Report

Facebook Transparency Report

Microsoft: Digital Safety Content Report

Snapchat: Transparency Report

Technical tools

PhotoDNA, a robust hash-matching technology for CSAM detection.

Safer, a third-party CSAM detection platform, offered by Thorn.

CSAI Match, YouTube's tool to detect known hashes in video content.

Google's Content Safety API, which can help prioritize potentially illegal content for human review.

Facebook's video-matching technology, *PDQ-TMKF*.

The Internet Watch Foundation offers a range of services, including URL lists and hashes.

The anti-grooming starter kit, available from Thorn

Civil society and non-government organizations

The WePROTECT Global Alliance

Crimes Against Children Conference

Global Partnership to End Violence Against Children

Child Dignity Alliance

ECPAT International

Thorn

The Lucy Faithfull Foundation

The Marie Collins Foundation

Child Helpline International

World Childhood Foundation

eSafety Outreach and Education

Google: *Be Internet Awesome*

Microsoft: *Digital Civility*

Roblox: *Digital Wellbeing*

Facebook: *Get Digital, Digital Literacy Library*

The Alannah and Madeline Foundation's *eSmart Digital Licence* and *Media Literacy Lab*

NCMEC's *NetSmartz*

NSPCC's *resources* for parents on how to talk to your child about online safety

Canadian Center for Child Protection's *Don't Get Sextorted* resource

Thorn's *NoFiltr* initiative

Young and eSafe *curriculum* from the Australian eSafety Commissioner

Netsafe New Zealand's *resources* for young people

Other resources

The Technology Coalition resources for members include a Starter Kit, model policies, and mentoring

Facebook's Cross-Industry Child Safety Hackathon

The Australian eSafety Commissioner's *Safety by Design* principles

ITU Child Online Protection: *Guidelines for Industry*

Supporting the Voluntary Principles: Next steps

WePROTECT Global Alliance is keen to highlight endorsement of the Voluntary Principles and associated good practice to drive greater international tech collaboration – please contact info@weprotectga.org or further information and support

