Trust

Voluntary Framework for Industry Transparency

TECH COALITION

TRUST
Voluntary Framework for Industry Transparency

Table of contents

Introduction

Trust: Voluntary Framework for Industry Transparency (the Framework) has been developed by the Tech Coalition to provide principles-based guidance to tech companies seeking to build trust around their efforts to address online child sexual exploitation and abuse (CSEA) risks on their services.

This is a voluntary framework, drawing on the experience of Tech Coalition members, multi-stakeholder conversations, and extant practices in transparency reporting in relation to online harms.

In joining the Tech Coalition, member companies have demonstrated their commitment to combating CSEA, and to their accountability for those efforts.

Purpose of the Framework

The Framework aims to:

- Encourage companies to provide online CSEA transparency reporting;
- Support the development and improvement of transparency reports by providing a variety of options; and
- Increase consistency across reporting, to better enable information-sharing and accountability.

Transparency is an essential component of industry efforts to combat online CSEA. It drives accountability and plays a critical role in building trust with users, regulators, and the general public. The importance of transparency reporting in helping advance this fight is recognized in the *Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse:* Principle 11 of that document reflects the need for companies to regularly publish data or insights on their efforts.

This Framework is intended for use by trust and safety or similar personnel as a companion in their effort to develop transparency reporting for online CSEA. It provides suggested categories for transparency reporting, as well as a list of possible topics and items for inclusion, depending on the various defining factors of any given service.

Transparency reporting in this context refers to reports that explain a company's approach to addressing online CSEA, which should highlight the company's policies, explain its processes, and document the outcomes of its efforts. There is a need for specific information and data related to CSEA in company transparency reporting that is distinct from metrics on other illegal or harmful content.

The Tech Coalition recognizes that this voluntary Framework will sit alongside emerging and existing regulatory reporting requirements in many jurisdictions, and anticipates it will sit alongside important efforts by governments and other stakeholders to enhance their own transparency. While acknowledging the critical role that law enforcement agencies play in addressing this challenge, the Framework does not address the process of reporting CSEA to law enforcement or other legal or operational matters related to CSEA identification or investigations.

Why provide transparency reporting?

Company transparency reporting on online CSEA serves three main purposes:

1. Explains a company's policies and actions to address the risk of harm resulting from online CSEA;
2. Demonstrates accountability for a company and develops trust by showing how it has implemented those policies and actions, and driven any improvements; and
3. Shares knowledge and meaningful data to support the global fight against CSEA.

Transparency reports have their limits. While they can supplement and support research, they are not intended to provide confidential data sets for deep research into specific services. Transparency reporting is also not an appropriate channel for individual case transparency; transparency reports are comprised of aggregated and anonymized data. Transparency reports reflect the unique nature and services of each company's platform, and thus cannot always be directly compared with one another. Additionally, while important, company transparency reporting alone cannot be viewed as a representation of, or prevalence estimate for, the overall scope and scale of CSEA, including within any particular geography – this is a complex problem with online and offline dimensions.

Audiences for transparency reporting

While a transparency report may be reviewed for a wide range of purposes, the primary audiences include:

- The general public and users (including parents) – who are able to better understand the policies and actions of the services they or their children use;
- Victims and survivors – who can see what is being done to prevent and limit potential harm;
- Government representatives, ranging from policy-makers to law enforcement – who are able to see what actions companies are taking to address a societal harm;
- The child safety community, including child safety non-government organizations and other civil society groups – who are able to review actions being taken and identify effective techniques, as well as opportunities to better fight online CSEA; and
- Other companies – who can learn from others, share best practices, and help normalize dialogue about this challenge.

While transparency reports are not generally directed to children, children are at the center of why companies should provide transparency reports. Companies should separately provide appropriate resources to help children, their families and/or caregivers understand and mitigate the risks of CSEA on that service.

Scope

This section outlines common terms that will be used throughout the Framework. These scoping definitions are informed by international best practice, including the ECPAT 2016 Luxembourg Guidelines,[1] but have been kept deliberately broad. Given that there are no consistent global definitions, this approach is intended to encourage companies to take an inclusive approach.

It is important to note that the definitions here are only intended to help demonstrate the scope of the Framework – the Framework does not seek to provide authoritative definitions nor guidance on relevant legal thresholds. Each company will determine the content that is acceptable or violative on its services according to its policies and definitions, as well as legal obligations, and shape the company's transparency reporting accordingly.

- **Child sexual exploitation and abuse (CSEA):** Any online content or conduct that depicts, instructs in or promotes/encourages child sexual exploitation and abuse. We have used this as an overarching term in the Framework to include:
  - **CSAM (child sexual abuse material)/CSEAI (child sexual exploitation and abuse imagery):** Any visual depiction of sexually explicit images of, or conduct of a child.[2] These two acronyms are sometimes used interchangeably.
  - **Grooming behavior:** Actions taken by a user of an electronic service to solicit or encourage a child to generate CSEAI, participate in sexual activity, or meet in real life for the purpose of sexual exploitation. May include coercive or threatening behavior.
- **Content** may manifest in a range of formats, including: imagery, video, livestreaming, audio, and/or text.
- **Conduct** refers to the actions taken by users.

1 https://ecpat.org/luxembourg-guidelines/
2 This Framework does not define "child" as interpretation will vary depending on legal context and company policies.

A Principles-Based Approach

The following principles provide a general basis for considering how to approach transparency reporting. Given the diversity of online services available, a principles-based approach provides companies with the ability to tailor an individual strategy while still achieving the overall goal of providing greater accountability and information about their efforts to address CSEA. Each service may face unique risks, depending on its purpose and features, user base, business model, and a range of other factors. Effective practices for one service will not necessarily suit another, and highly prescriptive approaches to trust and safety practices may be too narrow, or have unintended consequences.

This Framework aims to recognize the diverse service landscape and provide flexibility for different companies to adapt their transparency reporting practices, while also defining a common framework. This commonality is important, because without some consistency in measures and approach, it becomes challenging to maximize the value of transparency reporting and to get a complete picture of efforts across companies and begin to understand prevalence and other key trend data.

1. Reporting should support trust and accountability
Transparency builds trust and demonstrates a willingness to be held accountable for decisions and actions. Transparency reports should therefore be designed to build trust with users, suppliers, employees, investors and government authorities by demonstrating that the company has appropriate policies and procedures, and is applying them consistently and fairly. The data-gathering process should also provide insights that allow a company to continuously improve their policies and practices and systems and processes. And, as outlined earlier, transparency reporting on CSEA supports efforts to address an important, whole-of-society issue.

2. Reporting should reflect the unique nature of each company's service(s)
There is a diverse range of digital products and services that may be impacted by CSEA, ranging from social media, to communications, gaming, productivity, cloud storage, infrastructure provision, to newer concepts like the "metaverse," and beyond. Each service differs with regard to how risks manifest on its platform, the specific steps it can take to address the risk, and the results of those protective measures. Each service also has its own unique technology infrastructure. Thus, each service's transparency reporting efforts should be proportionate and tailored to its specific business case, risk profile, practices, and technology.

3. Reporting will depend on service maturity
Every company is at a different stage of maturity (broadly defined) and capacity. Maturity may also differ among differing services offered by a single company. Transparency reporting follows the development and implementation of risk mitigation, content policy, tools and process. It requires building out data collection, reporting and analysis capacity to produce the relevant information. New and smaller companies or services may prioritize building their risk mitigation capacity (including child-safe design and other elements of safety by design) and should be given adequate time to develop the capacity to produce and publish transparency reports. As companies and services mature, they should equally seek to mature and grow their transparency reporting capability in a proportionate manner – the Framework is designed to help advance reporting. Companies may also seek to go beyond the Framework as their business and approach matures.

4. Reporting should be regular and evolve over time
Companies should aim to provide reporting on a regular cadence: whether annually, biannually, or quarterly. More regular reports provide an opportunity to share more up-to-date information, so companies should aim to report at least annually. Companies should also seek to provide comparative information and metrics across reports to enable readers to track trends and to draw comparisons over time. Having said that, transparency reporting is a dynamic and iterative process. Each company begins transparency reporting with the resources at hand and builds from there, gaining insights from both internal and external feedback. Companies should seek views and feedback from a range of stakeholders.

5. Reporting should not compromise privacy or safety
Companies should strive to be as open as feasible, while not compromising other important interests, including privacy and safety. Transparency reporting should not in any way infringe on individual privacy. Great care should be taken to ensure that any data provided can not be tied to specific individuals. Transparency reporting should also be carefully calibrated to ensure that the information released does not compromise a company's safety efforts and inadvertently enable bad actors to subvert safety measures. When data is appropriately aggregated and anonymized, the provision of transparency reporting provides important visibility into the way a company is protecting the rights of its users, including the rights of children.

Recommended Report Categories

The following transparency report framework seeks to help companies organize reporting in a way that will allow different audiences to understand how companies combat CSEA on their service(s). To that end, we recommend a reporting structure that starts with **policies and practices**, allowing a company to describe their approach to CSEA and what they prohibit on their services, followed by a descriptive summary of the **processes and systems** the company uses in combating CSEA, and ends with numerical reporting on the **outcomes** that flow from the company's overall approach.

The following sections offer different elements in each of these categories that a company may choose to report on, in line with the principles outlined above. A company may seek to refine its reporting as its approach to transparency matures but not every element or suggested data point will be relevant or even possible for every service, even the most mature services.

A. Policies and Practices

This category refers to the provision of qualitative or descriptive information about the company's overall approach to CSEA. Below are some examples of the types of things that may be relevant, depending on the company or service(s).

Any information provided on any of the suggested topics in this section should be carefully calibrated to provide clear information without exposing any detail or other information that could be exploited by offenders.

In language that is easy to understand, companies should consider providing information on, or link to public information on, topics such as the following:

- Description of the company or service's policies with respect to CSEA, including what constitutes violative content and conduct. This may include referring to terms of service, community guidelines, or similar content standards.
- Where applicable: explanation of any updates or changes to relevant policies that have taken place over time (for example, highlight updates that were implemented since the last reporting period).
- Description of the potential consequences of breaching those policies for users.
- Description of any policies or other relevant information that are available specifically for children or young people.
- Description of the policy and process for appeals related to CSEA.
- Simple definitions of any technical terms used in the transparency report.
- Description of when reports are made, and to which reporting or referral agency (e.g., the National Center for Missing and Exploited Children).
- Description of the company's processes for responding to law enforcement requests.
- Description of the company's membership in relevant industry organizations, such as the Tech Coalition, and other commitments such as support of the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse.
- Description of any relevant child safety partnerships (for instance, for awareness-raising, or with support services).
- Description of any other relevant cross-industry or other collaboration (e.g., use of industry or other hash databases or URL lists).

Companies may also wish to add any other relevant information that helps to contextualize their approach to addressing CSEA. This may include pointing to data from the National Center for Missing and Exploited Children, or similar.

B. Processes and Systems

This category is an opportunity to provide qualitative information on how policies are operationalized and enforced. This means providing information on the current moderation systems and processes by providing descriptions and information on some or all of the following areas, as applicable.

Any information provided on any of the suggested topics in this section should be carefully calibrated to provide clear information without exposing any detail or other information that could be exploited by offenders.

1. CSEA Prevention and Detection

- A description of steps the company or service takes to prevent its exploitation for the generation, storage, or dissemination of CSEA.
- A description of the technology and other measures used to detect apparent CSEA.
- Where applicable: a description of processes and systems to facilitate user reporting and other flagging of potential CSEA. This may include information about trusted flagger programs.
- A description of any new safety tools or measures that have been developed or deployed since the last reporting period.
- Information on any risk assessment or other evaluation processes that the company undertakes to understand the unique risks related to CSEA on its service(s). It may be appropriate to comment on the frequency of such assessments and the methodology used, depending on the maturity of any such program.

2.Mitigation

- A description of the kinds of actions taken to respond to threats, incidents and bad actors, as well as any additional measures taken to reduce identified CSEA risks. This may include an overview of the company's approach to safety by design.

3. Moderation and Enforcement

- A description of the company or service's processes for actioning CSEA, once detected.
- A description of the moderation and enforcement options that may be taken in response to a breach of the company or service's terms of service or other relevant policies.
- A description of the company or service's processes for dealing with illegal content and/or content where there is an imminent risk of harm.

C. Outcomes

The final recommended reporting category is the provision of data that helps illustrate the impact of the company's policies and practices, and systems and processes. As outlined earlier, the contents of a report will depend on the maturity of the business, and the relevance of different items to the service(s) provided. A company should choose metrics to help build trust and drive accountability based on their unique approach and service(s). It may be helpful to explain in the report why there is a focus on some outcome data, or why some metrics may not be appropriate for the service(s).

Outcome data should include numerical data and made available in a downloadable format, where possible. Where possible, it also should be supported by graphs or charts as well as narrative comment to help build understanding for a range of audiences, with a focus on accessibility. It may be appropriate to accompany any trends, spikes, or significant differences with commentary or explanation.

This section of the Framework provides a set of recommended **Starting Metrics** on which all companies should provide reporting, as well as suggestions for additional sub-categories of data that companies may choose to provide.

Individual reports will therefore be company-specific but will provide some comparability.

Starting Metrics

Regardless of diversity across industry and of the different processes employed, all companies should provide data on the following items in their transparency report:

- **Action:** CSEA identified and actioned.[3]
- **Action:** User accounts identified and actioned.
- **Reporting:** Number of reports to relevant external authorities.

Reporting will differ, depending on where the company is headquartered and/or operating.

3 Actions taken for 'Actioned' items should be described clearly. e.g, User warned; user suspended temporarily; User terminated, etc.

Additional Outcome Metrics

The following provides a menu of suggested **Additional Outcome Metrics**, broken down by suggested category. Not all of these metrics will be relevant for each company and where examples are given, these are examples only.

A company may choose to use different terminology (e.g., distribution versus publish versus dissemination) depending on the functionality of its service(s). Companies should provide definitions or explanations of what they are reporting.

1. Discovery

Data points on discovery may help illustrate the relative effectiveness of different methods in flagging and/or detecting CSEA on the service(s). Items to consider providing data on:

- Total volume, broken down by flagging method. Examples may include:
  - User reporting
  - Trusted flaggers
  - Government or law enforcement reporting
  - Proactive tools or other technology
  - Review by moderation teams

2. Action

Data points on content actioned should help provide insights on the volume and nature of confirmed CSEA on the service(s), as well as visibility of the actions taken in response. The list below provides different suggested ways to break out the data, noting that care should be taken to avoid any risk of double-counting. Items to consider providing data on include:

- Volume or percentage of CSEA actioned, broken down by action, as applicable to the service(s):
  - Content removed
  - Content disabled
  - Sites deindexed
  - Other actions, as relevant to the service
- Volume or percentage actioned, broken down by content format, as applicable to the service(s):
  - Images
  - Videos
  - Chats
  - Livestreams
  - Other content formats
- Volume or percentage actioned, broken down by policy violation, as applicable. Examples may include:
  - CSEAI
  - Grooming
- Timeframes to action.
- Number of accounts actioned, broken down by type of action. Examples may include:
  - Warning
  - Temporary suspension
  - Permanent closure
  - Other, as applicable

3. Reporting

Data points on reporting should provide clarity on the authorities to which the company is reporting and the nature and volume of that reporting.

- Provide data on the types and number of statutory reports made and to which relevant authorities (e.g., to the National Center for Missing and Exploited Children)

4. Law enforcement requests

Data points on law enforcement requests should help provide visibility on which countries are seeking information from companies. Items to consider providing data on include:

- The total number of law enforcement requests for information, by country, and the percentage of requests where information was provided.
- Number of valid vs invalid requests.
- Number of requests by type. Examples may include:
  - Subpoena
  - Search warrant
- Number of requests by information disclosed. Examples may include:
  - Content
  - Non-content metadata

5. Appeals

Data points related to appeals can provide insights into potential error rates and other challenges. Items to consider providing data on include:

- Total number of user appeals
- Appeal success rate.
- Appeals consequences, including the total number or percentage of accounts or content reinstated.

6. Other insights

Data points on other topics can help provide greater insight into the nature and scale of the harm and how a particular service may be exploited. The data points below are more challenging to gather and may not be appropriate for all services. Items to consider providing data on include:

- CSEA insights:
  - Dissemination of CSEA (e.g., shares, posts etc before content was actioned). Examples may include:
    - Number of instances of CSEA dissemination
    - Number of unique and/or duplicate images disseminated
    - Number of duplicate images disseminated
    - Time to action CSEA
    - Number of unique views or shares
  - Prevention of CSEA. Examples may include:
    - Volume of potential offenders clicking through to support or deterrent messaging
    - Volume of CSEA blocked at upload
- Consider options to develop estimates on prevalence of CSEA, grooming or other violating conduct and/or conduct.
- Data on a geography or regional basis, where possible (this may include in relation to content and/or accounts actioned).
- Other supplemental information that contributes to, or further illuminates, aspects of the transparency report. This might include relevant research findings or other reporting, such as trend data.

Development and Review of the Framework

Members of the Tech Coalition developed this Framework and it was approved by the Tech Coalition Board in May 2022. It incorporates member companies' experiences and includes input received from civil society, academics, governments, and other companies during a consultation process facilitated by the WeProtect Global Alliance. The Tech Coalition acknowledges all the stakeholders who took the time to engage and to provide feedback.

The Tech Coalition will continue to receive feedback on the Framework, including to hear the experiences of companies developing their first CSEA transparency reports. It will review the Framework two years after its launch, and at regular intervals thereafter, to ensure it is keeping pace with technological and other developments.

Example Reporting and Resources

Click on the logos below to see Tech Coalition industry member reports and resources

Adobe amazon CLOUDFLARE DISCORD

Dropbox Google MEGA Meta

Microsoft Pinterest Snap Inc. TikTok

verizon yahoo! yubo

TECH COALITION

About Tech Coalition

The Tech Coalition facilitates the global tech industry's fight against the online sexual abuse and exploitation of children. We are an alliance of technology companies of varying sizes and sectors that work together to drive critical advances in technology and adoption of best practices for keeping children safe online. The Tech Coalition convenes and aligns the global tech industry, pooling their knowledge and expertise, to help all our members better prevent, detect, report, and remove online child sexual abuse content. This coalition represents a powerful core of expertise that is moving the tech industry towards a digital world where children are free to play, learn, and explore without fear of harm.

To learn more visit www.technologycoalition.org