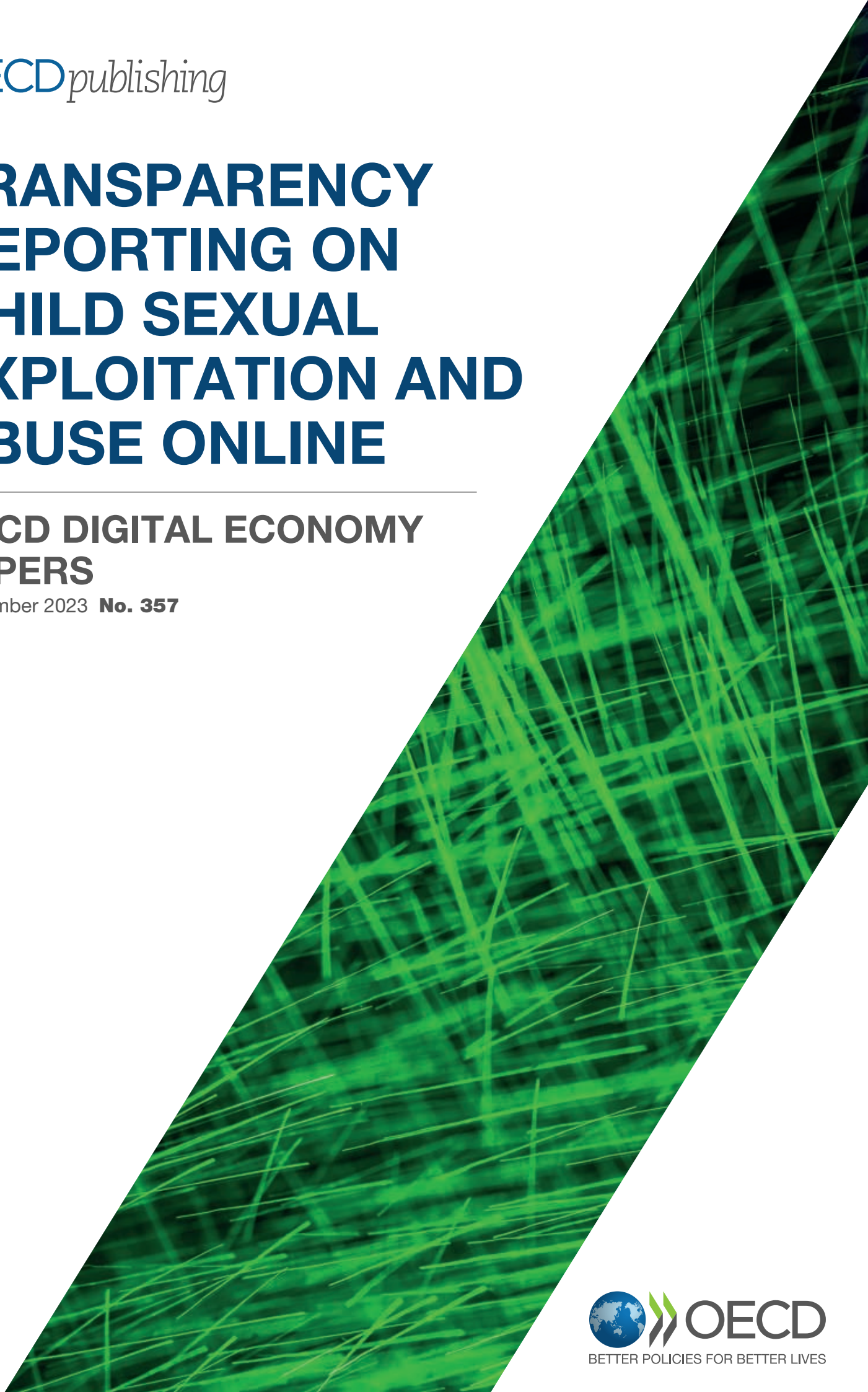
OECD publishing

TRANSPARENCY REPORTING ON CHILD SEXUAL EXPLOITATION AND ABUSE ONLINE

OECD DIGITAL ECONOMY PAPERS
September 2023 **No. 357**



OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

This report examines the policies and procedures of the world's top-50 global online content-sharing services related to child sexual exploitation and abuse (CSEA) material, providing an objective factual snapshot in time of the services' current practices, with a focus on transparency reporting. It is the first such OECD report benchmarking CSEA policies and transparency reporting practices, though it builds on three previous OECD benchmarking reports examining transparency reporting of terrorist and violent extremist content (TVEC) online. Like its TVEC counterparts, this report is intended to become part of a series.

This report was written by Dr Brian O'Neill (consultant to the OECD) under the guidance of Jeremy West, Lisa Robinson and Andras Molnar (OECD Secretariat). It incorporates oral and written feedback from OECD delegates to the Committee on Digital Economy Policy on earlier drafts, as well as feedback from the companies profiled in Annex B. This paper was approved and declassified by written procedure by the Committee on Digital Economy Policy on 28 June 2023 and prepared for publication by the OECD Secretariat.

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CDEP(2022)16/FINAL

Table of contents

TABLES

BOXES

Executive summary

This is the first OECD benchmarking report examining the policies and procedures related to child sexual exploitation and abuse (CSEA) of the world's top-50 global online content-sharing services. The report builds on three previous OECD benchmarking reports examining terrorist and violent extremist content (TVEC), applying the methodology to another form of online abuse that is widely criminalised and recognised to be a serious societal challenge. As with the TVEC reports, this CSEA report provides an objective, factual snapshot in time of current practices, providing evidence that not only facilitates better understanding of the services' relevant policies and procedures, but also of the extent and comparability of their transparency reporting.

CSEA in this context – following the approach advocated by the WeProtect Global Alliance – refers to the sexual exploitation and abuse of children that is partly or entirely facilitated by technology. Online CSEA includes the production or dissemination of child sexual abuse material online, the livestreaming of child sexual abuse, and the use of technology to make contact with potential child victims online with the intention of sexual exploitation. CSEA can take place not just online but offline, in the physical world. However, the digital environment has become an enabling environment that makes it easier for offenders to produce, store and distribute child sexual abuse material and to connect with children to engage in exploitation through digital means.

CSEA is an urgent and prominent policy challenge for which there have been several high-level calls to action. G7 Ministers have called attention to the devastating impact CSEA can have on victims and on societies and have called on digital service providers to prioritise protecting children, especially from illegal and harmful content and activity. The scale of the increase in reports of CSEA has been alarming. In 2022, the National Centre for Missing and Exploited Children (NCMEC) received more than 31.8 million reports of CSEA around the world through its CyberTipline, an increase from 21.7 million reports in 2020. Each report is an instance of apparent CSEA comprising one or more unique pieces of content. While the increase is in part attributable to better detection methods, international agencies including INTERPOL have expressed concern that CSEA continues to expand in scale and severity, potentially overwhelming the ability of law enforcement to effectively respond.

This report presents a baseline study of the policies, procedures and practices that the top-50 global online content-sharing services deploy in relation to CSEA on their platforms and services. The research is based on an analysis of the publicly available policies and other governance materials, including transparency reports (TRs), issued by the services. Following the methodology established for the benchmarking of TVEC policies, this report includes profiles for each service that summarise whether and how CSEA is defined in the relevant terms of service (ToS) or other standards, the service's policies on detection and removal of CSEA, its content moderation methods, the availability of notification and appeals procedures, the issuance of TRs and evidence of the extent to which the service has been exploited for CSEA purposes.

The key findings of this benchmarking study are:

Few services have detailed CSEA policies

- All services have some form of prohibition in their ToS that can be interpreted as covering CSEA. However, only 10 of the top 50 have a detailed policy specifically on CSEA that includes relevant explanations and examples, enabling an understanding of what CSEA-related content and activity are prohibited and how the platform addresses them.
- Twenty-five of the 50 services publish policies that prohibit CSEA either in general terms or by inclusion within a wider category of either Child Safety or Endangerment of Minors. Fourteen of these 25 services have an explicit prohibition of CSEA in their policies but do not provide a detailed explanation of what this means or how the policy is operationalised on the service.
- Just under a third, or 15 of the 50 services reviewed, address CSEA within their policy through broader prohibitions such as general prohibitions against posting illegal content. In these cases, there is no reference to CSEA or child endangerment of any kind.
- There are far fewer references in the policies to child sexual exploitation or grooming behaviour. This tends to be limited to a small number of services that have developed dedicated policies on CSEA.
- Very few services provide a definition of "child" or refer to accepted international legal definitions.

Transparency reporting is uneven and inconsistent

- Twenty of the top 50 services issue TRs on content and/or behaviour related to CSEA. In most instances, these have been introduced in the last two to three years as part of expanded transparency reporting on policy enforcement. Thirty out of the top-50 services do not issue any TR in respect of CSEA.
- Of the services that do issue TRs, there is wide variation in the nature of the information and data fields that are reported. Most services report data on the overall number of items of content actioned for violating CSEA policies. Many also report on the number of accounts identified and actioned for CSEA violations. However, the way in which these metrics are calculated is rarely disclosed. These factors contribute to a lack of clarity and consistency across TRs, which limits comparability and prevents a sector-wide perspective.
- Fifteen of the 20 TRs provide data on the proactive rate of detection. Just three of these give detail on how the CSEA violations are further classified or categorised, instead relying on aggregate statistics in terms of specific services, features or offences.

Limited information on content moderation

- Twenty-nine of the 50 top online content-sharing services state that they deploy a combination of staff, automated tools and community user reporting to detect CSEA content on their platforms. A further 21 services provide limited or no information at all about their approach to monitoring compliance on their platforms.
- Concerning CSEA in particular, just 16 of the top-50 services provide detailed information about their detection methods. Thirteen of the top-50 services provide less detailed information either in their policy and governance documents or in the respective TRs.
- Twenty-eight services state they have policies and procedures to notify users of enforcement decisions and appeal processes. It should be noted, however, that notifications or appeals processes are not always allowed for serious violations such as CSEA.

The report also includes an overview of international efforts to foster multistakeholder co-operation and key trends in international and domestic legislation on CSEA, including a brief discussion of emerging legislation and policy currently under consideration.

Introduction

This report presents the first OECD benchmarking study of transparency reporting among the top global online content-sharing services in relation to child sexual exploitation and abuse (CSEA). It forms part of an OECD series mapping the transparency reporting practices of online content-sharing services, which initially focussed on terrorist and violent extremist content (TVEC) (OECD, 2020[1]) (OECD, 2021[2]) (OECD, 2022[3]). These reports are an objective, factual snapshot of current practices, providing evidence that not only facilitates a better understanding of the services' relevant policies and procedures but also of the extent to which they engage in thorough and comparable transparency reporting. The present report aims to achieve a similar outcome, contributing an evidence base regarding CSEA to better understand industry-wide efforts to combat this serious and growing challenge.

CSEA is a major global threat that is complex, evolving and growing in scale, bringing about untold harm for child victims of abuse and undermining confidence in the digital environment. While digital technologies have brought unprecedented benefits for societies everywhere, they have also created conditions for technology-facilitated abuse, making it easier for offenders to contact potential child victims and to share images of that abuse online on a global scale (WeProtect Global Alliance, n.d.[4]). CSEA continues to evolve, and this report does not set about defining it. CSEA encompasses a variety of forms of abuse, including the production, possession, and distribution of child sexual abuse content through technology services and platforms, as well as the intentional sexual exploitation of children through technology-facilitated means (ECPAT International, 2016[5]).

The terminology used in this report is consistent with the recommended usage put forward by an interagency working group convened in 2016, which produced the Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse (the "Luxembourg Guidelines") (ECPAT International, 2016[5]). For instance, CSEA is used in preference to CSAM or "child sexual abuse material". CSAM refers only to content and not to behaviours associated with child exploitation, such as grooming or child trafficking, which are among the harms facilitated via online platforms. The Luxembourg Guidelines also make a point of discouraging the use of the term "child pornography" as an inadequate description of the phenomenon and to avoid confusion with the more general term "pornography", which refers to largely legal, adult-oriented content. However, for historical reasons, "child pornography" has a continuing usage in many legal statutes as well as in some policies included among the top global online content-sharing platforms. For this reason alone, a reference to "child pornography" is retained where that is the usage contained in the original. More generally, the terms CSEA and, specifically, online CSEA encapsulate the primary subject matter addressed in this report. Further explanations of the terminology used are provided in Annex C. Definitions.

High-level calls for action on CSEA

At a political level, it has been widely recognised that CSEA is an urgent and prominent policy challenge. Several calls to action have been issued in this regard. At a global summit in June 2022, the European Union, African Union and 17 governments from around the world joined forces with the WeProtect Global

Alliance to establish a new Global Taskforce on Child Sexual Abuse Online. The international task force aims to develop a coordinated response to CSEA and to secure engagement at national, regional and global levels (WeProtect Global Alliance, 2022[6]).

In 2020, following an action agreed at their 2019 Ministerial Meeting, the security ministers of Australia, Canada, New Zealand, the United Kingdom and the United States developed Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse. These principles, which are directed at industry, seek to provide a common and consistent framework to help combat the proliferation of CSEA online. Since then, G7 Interior Ministers have added their support (G7, 2021[7]) while a total of 16 companies have also endorsed the principles (UK Government, 2022[8]). Among other things, the principles recommend that companies regularly publish or share meaningful data and insights on their efforts to combat CSEA, noting that "*regular and transparent reporting will improve available data about the production, distribution, blocking and removal of child sexual exploitation and abuse*" (Five Country Ministerial, 2020[9]).

In 2019, G7 Digital Ministers noted the growing prevalence of CSEA and called for more to be done to enhance accountability and transparency while protecting and promoting human rights (G7, 2019[10]). Combatting CSEA was directly addressed at G7 level in 2021 when Interior and Security Ministers issued Principles for Tackling Online Violence Against Women and Girls and an Action Plan to Combat Child Sexual Exploitation and Abuse. Both recognised the importance of transparency reporting and the acute need for action on these issues (G7, 2021[7]).

In April 2021, G7 Digital Ministers adopted the Internet Safety Principles (G7, 2021[11]), which noted that "*online content that is illegal, and content that is harmful, can have a major impact on people, especially women and children, and on our societies*". These principles stress the importance of corporate responsibility and note that companies should be transparent about the presence of known illegal and harmful activity on their services and the decisions and measures that they take to improve safety; and that they should be accountable for decisions taken to counter illegal and harmful content in line with their terms and conditions.

As part of the G7 Internet Safety principles, companies were called on to "*prioritise the protection of children on their services and provide safety measures to ensure children are protected from both illegal and harmful content and activity*" (G7, 2021[11]) in line with the OECD Recommendation on Children in the Digital Environment (OECD, 2021[12]). Similarly, in 2021, the G20 adopted High Level Principles for Children Protection and Empowerment in the Digital Environment (G20, 2021[13]) which were drawn from the Recommendation. In a June 2022 communiqué, G7 Ministers affirmed their commitment to combat CSEA, pledging to "*step up (…) our efforts to combat child sexual abuse and exploitation globally, both online and offline*" (G7, 2022[14]). At their meeting in November 2022, Interior and Security Ministers addressed child sexual exploitation and abuse. Building on previous statements with respect to industry playing its part, Ministers highlighted the issue of livestreaming of child sexual abuse. It was noted that the special challenge lies in the real-time and commercial components of this crime, and that this can only be successfully tackled through multi-disciplinary co-operation (G7, 2022[15]).

The OECD Recommendation on Children in the Digital Environment was adopted by the OECD Council at Ministerial Level in May 2021 (OECD, 2021[12]). It aims to respond holistically to the needs of children in the digital environment, seeking to assist governments and other actors in implementing coherent policies and procedures that can address the delicate trade-off between enabling opportunities and protecting children from harm. Among other recommendations, it calls on governments to ensure that their legal frameworks promote responsible business conduct and provides guidance on the actions of digital service providers[1] through the accompanying "Guidelines for Digital Service Providers". The Guidelines similarly emphasise the importance of transparency and accountability (OECD, 2021[16]).

The scale and growth of CSEA online

CSEA is accelerating in scale, severity and complexity. The increase in incidence of CSEA coincided with the onset of the COVID-19 pandemic, during which digital technologies became ever more central to communities (EUROPOL, 2020[17]). In 2022, the National Centre for Missing and Exploited Children (NCMEC), which operates the United States' CyberTipline and liaises with law enforcement globally, received more than 31.8 million reports, up from 29.3 million in 2021 and 21.7 million in 2020 (NCMEC, 2023[18]). Between 2019 and 2020 a 100% increase in NCMEC reports was recorded (NCMEC, 2022[19]) (WeProtect Global Alliance, 2021[20]). In 2022, the Internet Watch Foundation (IWF) (a United Kingdom-based child protection organisation that uses technology to find and remove CSEA online), investigated a total of 375,230 reports suspected to contain child sexual abuse imagery and experienced a 20% increase in reports since 2020 (IWF, 2023[21]). In 2021 alone, the IWF observed that it had detected more CSEA than in its first 19 years of operation (IWF, 2021[22]). The Police Foundation, the United Kingdom's policing think tank, has stated the volume of online child sexual abuse offences is now so great that it had "*simply overwhelmed the ability of law enforcement agencies, internationally, to respond*" (The Police Foundation, 2022[23]). Likewise, INTERPOL has stated that CSEA is consistently rising and that 2021 was the worst year on record (INTERPOL, 2022[24]).

According to NCMEC data, most reports originate from online content-sharing services (NCMEC, 2022[19]) who predominantly use hash-based tools – a form of digital fingerprinting[2] – to detect CSEA and which automate (or partially automate) most reports (WeProtect Global Alliance, 2021[25]). Of the approximately 31.8 million reports that NCMEC received in 2022, 99% were submitted by "electronic service providers" (ESPs), and some 250,000 came from the public (NCMEC, 2023[18]). In 2022, ESPs submitted 49.4 million images to the CyberTipline of which 18.8 million (38%) were unique. Of the 37.7 million videos reported by ESPs, 8.3 million (22%) were unique. Over the course of 2022, 236 companies submitted CyberTipline reports and just 5 ESPs[3] accounted for more than 90% of the reports. NCMEC has observed trends such as the increasing use of video and livestreaming in CSEA offending (NCMEC, 2022[19]) as well as an increased reported incidence of online enticement of children for sexual acts. Online enticement or solicitation increased by 80% from 44,155 reports in 2021 to 80,524 reports in 2022 (NCMEC, 2023[18]).

CSEA is inherently global in nature, and while a report may be made in one country, it is likely that the offender and/or the victim are in one or more other country. For instance, NCMEC notes that reports to their tip line can be traced to nearly every country in the world, with 89.9% of their reports in 2022 resolved to locations outside the United States (NCMEC, 2023[18]).

A significant proportion of CSEA reports are generated by the resharing of known or previously detected imagery. When CSEA is reshared, it serves to re-victimise the abused child and to further exacerbate psychological damage even after the perpetrator may have been caught and punished (WeProtect Global Alliance, 2021[20]). Resharing of images is a clear concern. However, at the same time, significant amounts of new CSEA are being produced. For instance, in 2021, INHOPE (an international network of CSEA hotlines) reported that 82% of content exchanged between their hotlines was previously unknown, itself a 34% increase in the amount of unknown CSEA reported in 2020 (INHOPE, 2021[26]).

CSEA victims are predominantly girls, though boys are also affected, and in some cases, images or videos contain more than one child victim (WeProtect Global Alliance, 2021[20]). Most are between 3 and 13 years of age, but children 2 years of age and under are also among the victims. The Internet Watch Foundation (IWF) reports that when they see images depicting babies and toddlers, they are more likely to fall into the worst category of abuse.[4] Reports involving older children are increasing, mirroring an increase in reports of child self-generated sexual material. For example, between 2019 and 2020, the IWF saw a 77% increase in reports of self-generated material (IWF, 2021[22]). Sexual extortion, where predators demand sexual favours, money, or other benefits from a child under the threat of sharing their self-generated content (OECD, 2021[27]), has also increased (WeProtect Global Alliance, 2021[20]) (INHOPE, 2021[26]).

Outline of the report

Against that background, this benchmarking study of Transparency Reporting on Child Sexual Exploitation and Abuse Material Online by the Global Top-50 Content-Sharing Services aims to provide a robust evidence base on how leading digital service providers publicly report on their efforts to combat this threat. Following this overview of the challenge that CSEA poses in the digital environment, Section 1 explains the approach taken for the research design, data collection, analysis and compilation of the benchmarking profiles that provide the main source of evidence. Section 2 presents the main findings of the benchmarking research, highlighting key trends in policy definition, enforcement practices, content moderation strategies and transparency reporting. This is followed by an overview of the principal responses to CSEA as evidenced by recent international and inter-governmental policy initiatives (Section 3) and by relevant laws and regulations addressing CSEA (Section 4), including some that are still under consideration. Annexes to the benchmarking report contain the individual profiles of the top-50 online content-sharing services with a focus on their policies and transparency reporting practices. A summary of the definitions of key terms used in the report is also provided.

1 Scope, methodology and research design

This report examines the policies, procedures and practices that the top-50 global online content-sharing services deploy in relation to CSEA on their platforms and services. The report mirrors the approach developed and adopted for the OECD benchmarking of policies and practices regarding TVEC on the same top-50 services. This section describes the current benchmarking report's scope, methodology and research design.

Previous OECD benchmarking studies regarding TVEC examined not just the largest and most prominent online content-sharing services but the entire top-50. This has led to the creation of a powerful evidence base which, over the course of three successive benchmarking exercises (OECD, 2020[1]) (OECD, 2021[2]) (OECD, 2022[3]), has revealed trends, noteworthy practices and gaps in the upper echelon of the global technology sector's response to TVEC as an extremely harmful and illegal form of content. The starting point for the current study is to replicate this methodology for CSEA and to begin building an equivalent evidence base with a focus on publicly available policies on CSEA and transparency reporting about it.

As outlined in the first benchmarking report on current approaches to TVEC (OECD, 2020[1]), identifying the top-50 global online content-sharing services presents a range of methodological challenges. Sharing or storing content online can take very different forms. Accordingly, a wide diversity of services exist which enable the uploading, sharing, and transferring of digital content across multiple types of services. Relevant services also include communication and messaging services which facilitate audio, voice and video online communications. Following the approach adopted by the TVEC benchmarking reports, services may be divided into three broad categories:

1. Social media, video sharing services and online communications services;
2. Cloud-based file sharing services; and
3. An "Other" category that includes popular digital services for content management and online reference.

For the purposes of identifying the most popular social media platforms, video-sharing platforms and communications services, the metric of monthly active users (MAU) was selected as the most suitable measure. MAU is widely used in industry to measure online engagement and the reach of a platform and, therefore, an appropriate basis on which to rank the most used services. However, MAU is not as relevant to other types of services. Accordingly, market share was chosen to select the most prominent cloud-based file-sharing services. Finally, two additional services, the WordPress content management system and the Wikipedia reference site, were also included in the original top-50 ranking even though their popularity could not be directly determined relative to other services. However, as the TVEC benchmarking report sets out, their inclusion is warranted, given their indicative market share and monthly page views (OECD, 2020[1]).

The list of the top-50 global online content-sharing services is given in Annex A. Global top-50 most popular online content-sharing services. It was compiled based on the metrics mentioned above and is the same list that appears in the most recent TVEC benchmarking report (OECD, 2022[3]). This list includes some

notable changes from the original top-50 ranking of online services presented in the first benchmarking report (OECD, 2020[1]) and, for instance, now includes video conferencing services such as Zoom and Microsoft Teams, which rose to prominence over the course of the global COVID-19 pandemic.

The research design for the CSEA benchmarking report featured the following four main steps:

- **Step 1**: A standardised template was developed to collate the relevant data on each service, including references to their publicly available terms of service (ToS), community guidelines and other policies which describe their approaches to CSEA.
- **Step 2**: The template was used to complete profiles on the services' policies and transparency reporting practices.
- **Step 3**: Each service was contacted to solicit their feedback on the accuracy of the information in their profile and make any necessary adjustments and corrections.
- **Step 4**: Commonalities, trends and divergences among the services' policies and procedures on CSEA were identified and included in the draft report.

The research design focuses on collecting and reporting on the publicly available policies and practices of online content-sharing services in respect of the following: i) how CSEA is defined in the relevant terms of service and policies of providers; ii) policies and procedures for the detection and removal (and/or other action) of CSEA; iii) content moderation strategies; iv) notification, enforcement and appeals procedures; v) the issuance of transparency reports (TRs) concerning CSEA including their content, methodology and frequency; and vi) evidence of the extent to which the service has been used to disseminate, store or produce CSEA or to solicit children for purposes of sexual exploitation. This data is compiled using the standardised template in Step 1 and is presented in the current report in Annex B. Profiles of the top-50 services (Step 2). During Step 3, each service was contacted for feedback on the profile data collected. Twenty-six of the top-50 services, or just over half, responded and, where appropriate, revisions to the profiles were made.[5] An analysis of the commonalities, trends and divergences among the services' policies and procedures (Step 4) was then undertaken and included in the finalised study.

2 Commonalities, developments and trends in the services' approach to CSEA

This section includes findings on the main trends, commonalities and divergences among the global top-50 online content-sharing services in how they address CSEA. The data for this analysis is derived from profiles compiled for each service (presented in Annex B. Profiles of the top-50 services). Each profile presents data on the following:

- how individual services define CSEA in their Terms of Service (ToS), community guidelines or standards (Question 1);
- how services communicate their ToS, community guidelines or standards (Question 2);
- their policies on enforcing compliance with those ToS (Question 3);
- consequences for users arising from policy violations (Question 4);
- the extent to which there are notifications and appeal procedures with respect to removals, sanctions, or other actions taken for violative content (Questions 4.1 and 4.2);
- the issuance of a transparency report by the service (Question 5);
- methods deployed by services to detect CSEA (Question 6); and
- evidence of the service being used to disseminate, store or produce CSEA (Question 7).

The transparency reporting practices of online content-sharing services are the central focus of the analysis in this report. In Question 5 of the standardised profile template, the transparency reporting is examined with an analysis of the kinds of data and information made available in services' reports (if any) as well as the methodologies used to calculate or estimate the data. TRs, supported by independent reports by agencies such as NCMEC, or other third-party media reports, are then cited to confirm if CSEA has been found on the service in question.

As revealed in either the provider's own TRs and/or its reporting to public agencies, there is evidence in 35 out of the 50 top online content-sharing services that the service has been used to host or share CSEA. In nine cases, third party media reports identified that CSEA had been found on the platforms concerned.[6] In the case of six services, no information could be identified regarding the incidence of CSEA on the platform due to the lack of transparency reporting, independent reports to public agencies or other third party media reports.[7] Additionally, in some instances, providers do not disclose the extent of CSEA on individual services; instead, they provide only aggregate data.[8]

How services define CSEA

Whether services specifically state that they prohibit the use of their service for disseminating, storing or producing CSEA and, if so, whether they define CSEA in their governance documents such as their ToS,

acceptable use policies, community guidelines or equivalent, are key indicators of services' overall approach to CSEA. A clear and precise definition of CSEA is needed to articulate what is expressly forbidden on the service, either in terms of content or conduct. CSEA, as outlined further in Section 4 of this report, is internationally recognised as a serious offence, outlawed widely in international law and in the domestic legislation of most countries. CSEA is also a complex and multifaceted problem comprising different forms of abuse and exploitation facilitated by digital technologies. Therefore, it matters whether services define CSEA with specificity, and whether they provide sufficient detail to clarify the scope of the terms applicable to the service. Giving examples where relevant is one way to enhance clarity about what is permitted and not permitted on the platform. Many services also choose to have a dedicated CSEA policy to which their ToS or equivalent refers.

All 50 services were found to have some form of prohibition in place within their ToS that can be interpreted as covering CSEA. A summary of the findings is presented in Table 2-1.

Table 2-1: How services define CSEA in their ToS

Group 1	Group 2	Group 3	Group 4
Services that define CSEA and related concepts with sufficient detail to understand the scope of such terms, providing examples where appropriate **N=10**	***Services that explicitly ban the use of their technologies to post or distribute CSEA, using (but not explaining in detail) relevant terms related to child sexual exploitation and abuse*** **N=14**	***Services that include CSEA within a more general category of Child Safety / Endangerment of Minors*** **N=11**	***Services that use broad and/or general descriptions of prohibited conduct, which can be interpreted as encompassing CSEA*** **N=15**
Facebook (Meta Platforms, Inc.) Facebook Messenger (Meta Platforms, Inc.) Google Drive (Alphabet, Inc.) Instagram (Meta Platforms, Inc.) Quora (Quora, Inc.) TikTok (ByteDance Technology Co.) Twitch (Amazon.com, Inc.) Twitter (Twitter, Inc.)[9] YouTube (Alphabet, Inc.) Zoom (Zoom Video Communications, Inc.)	Ask.fm (IAC [InterActiveCorp]) DeviantArt (DeviantArt, Inc.) Dropbox (Dropbox, Inc.) LinkedIn (Microsoft, Inc.) Medium (A Medium Corporation.) Pinterest (Pinterest, Inc.) Reddit (Reddit, Inc.) Snapchat (Snap, Inc.) Tumblr (Automattic, Inc.) Viber (Rakuten, Inc.) Vimeo (Vimeo, Inc.) WhatsApp (Meta Platforms, Inc.) Wikipedia (Wikimedia Foundation Inc.) WordPress.com (Automattic, Inc.)	Discord (Discord, Inc.) Flickr (SmugMug, Inc.) KaKao Talk (Daum Kakao Corporation) Likee (BIGO Technology PTE. LTD.) Microsoft OneDrive (Microsoft, Inc.) Microsoft Teams (Microsoft, Inc.) Picsart (Picsart, Inc.) Skype (Microsoft, Inc.) VK (Mail.Ru Group) Weixin/WeChat (Tencent Holdings Ltd.) Xigua Video (ByteDance Technology Co.)	iMessage/FaceTime (Apple, Inc) Baidu Tieba (Baidu, Inc.) Douban (Information Technology Company, Inc.) Huoshan (ByteDance Technology Co.) IMO (PageBites, Inc.) iQIYI (Baidu, Inc.) Kuaishou (Beijing Kuaishou Technology Co., Ltd) LINE (Line Corporation) Odnoklassniki (Mail.Ru Group) QQ (Tencent Holdings Ltd.) QZone (Tencent Holdings Ltd.) Smule (Smule, Inc.) Telegram (Telegram Messenger LLP) Weibo (Sina Corp.) Youku Tudou (Alibaba Group Holding Limited)

Source: Annex B. Profiles of the top-50 services

For the purpose of analysis, services' approaches to defining CSEA in their ToS were divided into four groups:

- Ten of the 50 services define or describe CSEA in detail in their ToS, with relevant explanations and examples for the purpose of governing prohibited content and activity on their platforms (Group 1).
- A further 14 services have an explicit prohibition on CSEA, though with less detailed explanations of what CSEA means in this context or how the service operationalises its policy (Group 2). Nonetheless, CSEA is specifically identified as a category of misuse and prohibited content/conduct on the service.
- Eleven services include CSEA under more general categories, such as child endangerment or child safety, in which CSEA is identified and covered more broadly (Group 3). However, the relevant policy statements are framed without specific reference to threats of CSEA or potential misuse of their platforms for online exploitation and abuse.
- Finally, just under one third, or 15 out of the top-50 services, describe prohibited content in broad or general ways (such as by requiring that users not post anything that is illegal) without any specific reference to CSEA or child endangerment (Group 4).

Each of Groups 1, 2 and 3 – representing 35 of the top-50 services – may be said to have a policy in place that expressly prohibits CSEA on their platforms. However, there remains significant variation among these

services in how this is documented in practice. Services within Group 1 do so in a clear and defined way, many having a dedicated CSEA sub-policy that elaborates in greater detail the policy and the approach of the service in combatting CSEA. Services in Group 2 do so with less detail and with less explanation. Services in Group 3 explicitly prohibit CSEA but do so in the context of more general policy statements, for example, in relation to child safety or endangerment of minors. Services in Group 4 do not have an explicit prohibition of CSEA and rather rely on a more general formulation within their ToS that refers to a prohibition of any content or conduct that may be in violation of local laws or regulations.

In addition to the differing ways in which services define CSEA, there are also differences between services in how they categorise prohibited content or conduct. The posting, storing or distribution of content depicting various forms of child sexual abuse is the type of prohibited activity most frequently referenced in ToS. However, only a limited number of services refer to child exploitation, or illegal behaviour such as solicitation or attempts to interact with or 'groom' a child, for sexual purposes. Furthermore, only a few services define what is meant by a 'child' in the context of their policy on CSEA or refer to definitions such as that in the OECD Recommendation (OECD, 2021[12]), or in Article 1 of the UN Convention on the Rights of the Child (UNCRC) where a child is defined as any person under the age of 18 (UN General Assembly, 1989[28]).

Transparency reporting practices

Twenty of the top-50 online content-sharing services issue TRs on content and/or behaviour related to CSEA. Many such TRs have been introduced in the last two to three years as part of an expanded approach to transparency reporting that includes platforms' enforcement actions in respect of their Community Guidelines. Thirty of the top-50 online content-sharing services do not issue TRs in respect of CSEA. Table 2-2 gives an overview of the services that issue TRs alongside those for whom no TR on CSEA enforcement could be found.

Table 2-2: Services that issue transparency reports on CSEA

Issues a TR Yes = 20	Issues a TR No = 30
Ask.fm (IAC [InterActiveCorp]) Discord (Discord, Inc.) Dropbox (Dropbox, Inc.) Facebook (Meta Platforms, Inc.) Facebook Messenger (Meta Platforms, Inc.) Google Drive (Alphabet, Inc.) Instagram (Meta Platforms, Inc.) LINE (Line Corporation) LinkedIn (Microsoft, Inc.) Microsoft OneDrive (Microsoft, Inc.) Microsoft Teams (Microsoft, Inc.) Pinterest (Pinterest, Inc.) Reddit (Reddit, Inc.) Skype (Microsoft, Inc.) Snapchat (Snap, Inc.) Tik Tok (ByteDance Technology Co.) Twitch (Amazon.com, Inc.) Twitter (Twitter, Inc.) YouTube (Alphabet, Inc.) Zoom (Zoom Video Communications, Inc.)	Baidu Tieba (Baidu, Inc.) DeviantArt (DeviantArt, Inc.) Douban (Information Technology Company, Inc.) Flickr (SmugMug, Inc.) Huoshan (ByteDance Technology Co.) iMessage/FaceTime (Apple, Inc) IMO (PageBites, Inc.) iQIYI (Baidu, Inc.) KaKao Talk (Daum Kakao Corporation) Kuaishou (Beijing Kuaishou Technology Co., Ltd) Likee (BIGO Technology PTE. LTD.) Medium (A Medium Corporation.) Odnoklassniki (Mail.Ru Group) Picsart (Picsart, Inc.) QQ (Tencent Holdings Ltd.) Quora (Quora, Inc.) QZone (Tencent Holdings Ltd.) Smule (Smule, Inc.) Telegram (Telegram Messenger LLP) Tumblr (Automattic, Inc.) Viber (Rakuten, Inc.) Vimeo (Vimeo, Inc.) VK (Mail.Ru Group) Weibo (Sina Corp.) Weixin/WeChat (Tencent Holdings Ltd.)

	WhatsApp (Meta Platforms, Inc.) Wikipedia (Wikimedia Foundation Inc.) WordPress.com (Automattic, Inc.) Xigua Video (ByteDance Technology Co.) Youku Tudou (Alibaba Group Holding Limited)

Source: Annex B. Profiles of the top-50 services

Among the 20 services that do issue TRs, there is wide variation in the nature of the information and data fields that are reported. An overview of the main metrics included within individual TRs is given in Table 2-3. Key findings may be summarised as follows:

- Most services report data on the overall number of items of content actioned for violating the platform's policies on CSEA. Many also report on the number of accounts identified and actioned for CSEA violations. Actioning content or accounts in this context refers to enforcement actions such as suspension, removal and/or reporting to the relevant authorities for a suspected CSEA violation. Several services (8 out of the 20) also provide data on the number of appeals submitted or accounts/content reinstated following appeal.
- Fifteen of the 20 services that issue TRs provide data on the volume of CSEA that is proactively detected, i.e., the number of items of content or accounts detected directly by the platform suspected to be in violation prior to anyone reporting it. Some services also report the number of views a violative piece of content has received prior to detection or removal. Proactive detection may involve using automatic detection tools, human staff, or a combination of both to pre-screen or monitor content on the platform.
- Just 3 of the 20 services give details on how CSEA violations are further classified or categorised, e.g., with respect to content types or forms of violative behaviour. Only 4 of the 20 services provide data confirming the number of reports made to external authorities following the identification and detection of CSEA material.

Table 2-3: Metrics included with transparency reports

	CSEA identified and actioned	*User accounts identified and actioned*	*Number of reports to relevant external authorities*	*Proactive detection rate*	*Reasons/ sub-categories of CSEA*	*Content appealed*	*Content actioned by country*	*Other relevant metrics*
Ask.fm (IAC [InterActiveCorp])		●		●	●		●	
Discord (Discord, Inc.)	●	●	●	●		●		Breakdown by content service User reports actioned
Dropbox (Dropbox, Inc.)	●	●						
Facebook (Meta Platforms, Inc.)	●			●		●		Content restored with appeal
Facebook Messenger (Meta Platforms, Inc.)	●			●		●		Content restored with appeal
Google Drive (Alphabet, Inc.)	●	●	●					CSAM hashes to NCMEC
Instagram (Meta Platforms, Inc.)	●			●		●		Content restored with appeal

	CSEA identified and actioned	*User accounts identified and actioned*	*Number of reports to relevant external authorities*	*Proactive detection rate*	*Reasons/ sub-categories of CSEA*	*Content appealed*	*Content actioned by country*	*Other relevant metrics*
LINE (Line Corporation)	●			●				Staff review Breakdown by content service
LinkedIn (Microsoft, Inc.)	●							
Microsoft OneDrive (Microsoft, Inc.)	●	●		●		●		*not broken down by product
Microsoft Teams (Microsoft, Inc.)	●	●		●		●		*not broken down by product
Pinterest (Pinterest, Inc.)	●	●		●		●		
Reddit (Reddit, Inc.)	●	●		●				Violations by content type User vs. admin reports Automated vs. manual reports
Skype (Microsoft, Inc.)	●	●		●		●		*not broken by product
Snapchat (Snap, Inc.)	●	●		●				Median turnaround time
TikTok (ByteDance Technology Co.)	●	●		●	●		●	
Twitch (Amazon.com, Inc.)	●	●	●	●				Proactive vs manual detection User reports
Twitter (Twitter, Inc.)	●	●						Number of views pre-removal
YouTube (Alphabet, Inc.)	●	●	●	●	●		●	
Zoom (Zoom Video Communications, Inc.)	●	●					●	

Source: Annex B. Profiles of the top-50 services

Transparency reports also vary widely in the ways in which they present data or make it available for further analysis or comparison with data from previous TRs. Some noteworthy practices include specifying where trends can be identified, parameters can be adjusted for further analysis, and archived data from previous reports can be consulted or downloaded for further analysis. For larger platforms, the ability to examine data in relation to content or CSEA violations by country or region is especially relevant. Four services make that data available.

Several services have incorporated additional information fields relevant to their platforms that provide valuable insights in understanding challenges encountered by that service and how it is being managed. For example, data on where violations occur within a given service – e.g. in chatrooms or direct communications, in comments fields, content storage areas, etc. – is vital to understanding the behaviours and areas of a service that may be most vulnerable to misuse. Where a TR presents data for a suite of product offerings, it is especially important to understand to which service the data refers. While aggregated

data is useful for the purposes of understanding the overall scale and scope of the problems addressed, it is the detail that will be of most value in building a culture of trust and accountability. Other useful practices identified in some services' TRs include the median turnaround times for processing reports, the ability to compare proactive versus manual detection rates, and the effectiveness of user or community reporting in identifying CSEA violations.

Detecting and actioning CSEA

The policies and procedures that services have in place to minimise their use in ways that violates their ToS, community guidelines or standards were also reviewed. Forty-one of the top-50 online content-sharing services state that they deploy a combination of staff, automated tools and community user reporting to detect infringing content on their platforms.

Table 2-4 gives an overview of the information provided by services in respect of their methods to detect CSEA.

Table 2-4: Services' provision of information on methods to detect CSEA

Content moderation approaches explained with good detail N=16	Content moderation approaches explained in broader / less detailed terms N=13	Vague statements on content moderation N=12	No information on content moderation available N=9
Discord (Discord, Inc.) Facebook (Meta Platforms, Inc.) Facebook Messenger (Meta Platforms, Inc.) Google Drive (Alphabet, Inc.) Instagram (Meta Platforms, Inc.) LinkedIn (Microsoft, Inc.) Microsoft OneDrive (Microsoft, Inc.) Microsoft Teams (Microsoft, Inc.) Reddit (Reddit, Inc.) Skype (Microsoft, Inc.) Snapchat (Snap, Inc.) Tik Tok (ByteDance Technology Co.) Twitch (Amazon.com, Inc.) Twitter (Twitter, Inc.) YouTube (Alphabet, Inc.) Zoom (Zoom Video Communications, Inc.)	Ask.fm (IAC [InterActiveCorp]) DeviantArt (DeviantArt, Inc.) Dropbox (Dropbox, Inc.) Flickr (SmugMug, Inc.) IMO (PageBites, Inc.) LINE (Line Corporation) Pinterest (Pinterest, Inc.) Tumblr (Automattic, Inc.) Vimeo (Vimeo, Inc.) VK (Mail.Ru Group) WhatsApp (Meta Platforms,Inc.) Wikipedia (Wikimedia Foundation Inc.) WordPress.com (Automattic, Inc.)	KaKao Talk (Daum Kakao Corporation) Kuaishou (Beijing Kuaishou Technology Co., Ltd) Medium (A Medium Corporation.) Quora (Quora, Inc.) QQ (Tencent Holdings Ltd.) QZone (Tencent Holdings Ltd.) Odnoklassniki (Mail.Ru Group) Picsart (Picsart, Inc.) Smule (Smule, Inc.) Viber (Rakuten, Inc.) Weibo (Sina Corp.) Weixin/WeChat (Tencent Holdings Ltd.)	Baidu Tieba (Baidu, Inc.) Douban (Information Technology Company, Inc.) Huoshan (ByteDance Technology Co.) iMessage/FaceTime (Apple, Inc) iQIYI (Baidu, Inc.) Likee (BIGO Technology PTE. LTD.) Telegram (Telegram Messenger LLP) Xigua Video (ByteDance Technology Co.) Youku Tudou (Alibaba Group Holding Limited)

Source: Annex B. Profiles of the top-50 services

Sixteen of the top-50 online content-sharing services give detailed information about the methods used to detect CSEA on their platforms. This information is typically incorporated in a distinct section of the community guidelines or separate policy dedicated to combatting CSEA. For these services, TRs are also used to communicate information on the services' content moderation efforts in this area. The information provided gives greater insight into the overall scale of moderation deployed by the service, the nature of the automated tools used, innovations in approaches to moderation, trends and successes in detecting CSEA and areas where the service in question continues to focus its efforts.

Thirteen of the top-50 services provide less detailed information either in their policy and governance documents or in the respective TRs. While this group of services demonstrates implementation of accepted industry practices in detecting and removing CSEA material, there is very little information about how this is achieved in practice or what level of resources are available to the service in this work.

Twenty of the top-50 online content-sharing services give no information or provide only vague statements regarding the methods used to detect or take action on CSEA on their platforms. In the case of 12 of the services, this includes such general statements as reserving the right to monitor for infringements of their policies without any obligation to do so and without giving any further information. Nine of the services give no information at all about their approach to detecting CSEA or other illegal content.

Notification, enforcement and appeals processes

Over half, or 28 of the top-50 services, have policies and procedures regarding notifications to users regarding enforcement actions for suspected platform policy violations (see Table 2-5). These services also offer appeal processes whereby users may appeal an enforcement action, such as content removal, service suspension or some other ban when they believe an error has been made or that an enforcement action is not warranted. Fourteen of the top-50 services, or nearly 40%, do not specify if a notification mechanism or appeals procedure is available on their service. In four cases, it was not clear if the service issued notifications. Four services with notifications procedures did not specify to users if it offered an appeals procedure.

Many services indicated that where instances of CSEA or other serious violations of platform policies were concerned, immediate suspension and/or removal of content without notification would apply. Most such services also indicate that cases of detected CSEA are reported to the relevant competent authorities and/or to law enforcement.

Table 2-5: List of services that provide notification and appeals mechanisms

Services that have mechanisms for notifying users and appeals processes in place.	*Services that have mechanisms for notifying users but do not specify an appeals process.*	*Services have appeals processes but do not specify notifications to users*	*Services that specify neither appeals nor notifications.*
N= 28	**N=4**	**N=4**	**N=14**
DeviantArt (DeviantArt, Inc.) Discord (Discord, Inc.) Douban (Information Technology Company, Inc.) Facebook (Meta Platforms, Inc.) Facebook Messenger (Meta Platforms, Inc.) Flickr (SmugMug, Inc.) Google Drive (Alphabet, Inc.) Instagram (Meta Platforms, Inc.) KaKao Talk (Daum Kakao Corporation) Likee (BIGO Technology PTE. LTD.) LinkedIn (Microsoft, Inc.) Medium (A Medium Corporation.) Microsoft OneDrive (Microsoft, Inc.) Microsoft Teams (Microsoft, Inc.) Pinterest (Pinterest, Inc.) Quora (Quora, Inc.) Reddit (Reddit, Inc.) Skype (Microsoft, Inc.) Snapchat (Snap, Inc.) Tik Tok (ByteDance Technology Co.) Tumblr (Automattic, Inc.) Twitch (Amazon.com, Inc.) Twitter (Twitter, Inc.) Vimeo (Vimeo, Inc.) WhatsApp (Meta Platforms, Inc.) WordPress.com (Automattic, Inc.)	Dropbox (Dropbox, Inc.) Weibo (Sina Corp.) Picsart (Picsart, Inc.) VK (Mail.Ru Group)	Ask.fm (IAC [InterActiveCorp]) Kuaishou (Beijing Kuaishou Technology Co., Ltd) Viber (Rakuten, Inc.) Wikipedia (Wikimedia Foundation Inc.)	Baidu Tieba (Baidu, Inc.) Huoshan (ByteDance Technology Co.) iMessage/FaceTime (Apple, Inc) IMO (PageBites, Inc.) iQIYI (Baidu, Inc.) LINE (Line Corporation) Odnoklassniki (Mail.Ru Group) QQ (Tencent Holdings Ltd.) QZone (Tencent Holdings Ltd.) Smule (Smule, Inc.) Telegram (Telegram Messenger LLP) Weixin/WeChat (Tencent Holdings Ltd.) Xigua Video (ByteDance Technology Co.) Youku Tudou (Alibaba Group Holding Limited)

YouTube (Alphabet, Inc.) Zoom (Zoom Video Communications, Inc.)			

Source: Annex B. Profiles of the top-50 services

Disclosure by Chinese platforms

Twelve of the top-50 services are Chinese-based holdings.[10] Only one of these services, TikTok (ByteDance Technology C.), issues a TR. The remaining 11 provide only limited information about their approach to content moderation or processes for enforcing their ToS and policies. TikTok issues a detailed Community Guidelines Enforcement Report and presents detailed information about its content moderation enforcement towards CSEA. Few other Chinese platforms give details of their policies regarding CSEA, defining neither prohibited content nor activity in any detail. Enforcement actions are also presented in generalised and vague terms.

There is also limited publicly available information regarding the nature or prevalence of CSEA on Chinese platforms. The Chinese regulatory framework refers to obligations on providers through a self-regulatory process to uphold professional ethics and to achieve a "healthy and civilised environment" with reference to the protection of minors in their use of online platforms. A Chinese government directive states that it expects *"Chinese people to enjoy a healthy and colorful cultural life in cyberspace, show higher ethical standards, and behave well in the sphere, particularly underage Internet users"* (Xinhua, 2021[29]). Some of the secondary literature and media reports refer to cases of Chinese platforms being held to account for CSEA, but in the absence of any transparency reporting by platforms themselves or official statistics at governmental level, this is hard to verify (Bursztein, 2019[30]).

More generally, the People's Republic of China (hereafter "China") has adopted a unilateral approach of enforcing Internet industry self-regulation through its command and control style policy enforcements, requiring the Internet industry to strictly monitor the Internet space with heavy penalties for companies not following the legislation.[11] The Internet Society of China—ISC, which acts as the Chinese Internet industry association, pledges through this self-regulatory process to uphold professional ethics within the Internet industry (Mubarak, 2020[31]). Article 9 of the self-regulation pledge states that if the Internet Service Provider (ISP) discovers information that is inconsistent with the law on its network, it will voluntarily remove it. Article 11 of this self-regulation pledge requires the Internet industry to make its best efforts to take effective measures to create a healthy and civilised environment for Internet usage and to assist the users, especially children, to use the Internet healthily (China Services Info, 2018[32]).

3 International initiatives to combat CSEA

CSEA is recognised as one of the most serious challenges that the digital environment poses to children's online safety. The apparent ease with which digital services can be misused to propagate extremely harmful material such as child sexual abuse and the way platforms can be exploited by abusers to gain access to children have given rise to significant concern among governments, law enforcement, the Internet industry and child safeguarding organisations. Because of CSEA's complex and multifaceted nature, a number of national and international multistakeholder responses have developed over the years. This section outlines some of the main global initiatives to combat CSEA.

Tackling CSEA at the international level

International initiatives designed to foster increased co-operation to tackle CSEA online have been promoted by key stakeholders at inter-governmental, industry and civil society levels since the commercial Internet was in its youth. The International Conference on Combating Child Pornography on the Internet, held in Vienna in 1999, called for a policy of zero tolerance for CSEA on the Internet and the worldwide criminalisation of possession of CSAM. The Vienna statement also stressed the need to strengthen law enforcement and for closer co-operation and partnership between governments and the Internet industry (United Nations, 2000[33]).

A further early response to the threat of CSEA was the establishment of Internet hotlines or cyber tip lines for the reporting of illegal CSEA content. This response was designed in part to support and augment law enforcement efforts to address this new form of cybercrime (Carr, 2021[34]). Hotlines sought to strengthen policing of the Internet by fostering greater co-operation between law enforcement, industry, civil society and the public. In the United States, the National Center for Missing & Exploited Children (NCMEC) was first established in 1984 as a private, non-profit organisation to serve as a clearinghouse of information about missing and exploited children. In 1998, NCMEC established its CyberTipline in response to the growing problem of child sexual exploitation online. The CyberTipline continues to act as an online mechanism for members of the public and what it terms "electronic service providers" (ESPs) to report incidents of suspected child sexual exploitation. NCMEC's Exploited Child Unit (ECU) reviews all reports and forwards them to local as well as international law enforcement agencies as appropriate (NCMEC, n.d.[35]). In Canada, Cybertip.ca was established along similar lines, initially as a pilot in 2002, and from 2005 launched as Canada's national hotline, supported by a multisectoral group of industry, government, non-governmental, and law enforcement stakeholders working to address the problem of online child sexual exploitation (Cybertip.ca, n.d.[36]).

In the European Union, hotlines alongside awareness raising and industry self-regulation were key pillars of the first Action Plan for a Safer Internet (European Commission, 1999[37]). In 1999, the non-governmental organisation, Childnet International, established the International Hotline Providers in Europe Forum, providing a space for hotlines to meet and exchange information. With support from the European

Commission's Daphne programme, the initiative laid the foundation for the establishment of the INHOPE Association in 1999, now representing a network of 50 hotlines around the world (INHOPE, 2021[26]).

Currently, co-operation for the notification, reporting, processing of removal requests, investigation and prosecution for CSEA offences is led and coordinated at an international level by a diverse group of organisations, including NCMEC, the INHOPE network of 50 global hotlines, the Internet Watch Foundation (IWF),[12] INTERPOL and EUROPOL, and the International Centre for Missing and Exploited Children (ICMEC).[13] Many of these initiatives are sponsored at inter-governmental and governmental level and involve stakeholders from industry and civil society.

Among the various organisations working to combat online CSEA, one of the largest is the WeProtect Global Alliance, which brings together experts from government, the private sector, civil society and international organisations. The WeProtect Global Alliance has its origins in the 2016 merger between the Global Alliance Against Child Sexual Abuse Online launched in 2012 by the European Commission and U.S. Department of Justice, and the WeProtect project established by the United Kingdom in 2014 (WeProtect Global Alliance, n.d.[38]). Relaunched in 2020 as an independent organisation, the Alliance has developed the Model National Response (WeProtect Global Alliance, 2016[39]) and the Global Strategic Response (WeProtect Global Alliance, 2019[40]) to help identify and guide actions to combat CSEA, including technology solutions to scan, detect and remove child sexual abuse material and stop grooming attempts.

The Global Partnership and Fund to End Violence Against Children is an initiative launched in July 2016 by the UN Secretary-General (End Violence Against Children, n.d.[41]). The Partnership is the only global entity focused solely on Sustainable Development Goal (SDG) 16.2: ending abuse, exploitation, trafficking and all forms of violence against and torture of children by 2030. The End Violence Partnership is a platform for collective, evidence-based advocacy and action. It comprises an alliance of more than 700 organisations including governments, UN agencies, research institutions, international NGOs, foundations, local CSOs, private sector groups and faith networks. The partnership undertakes activities to raise awareness, influence policy, mobilise new resources, promote evidence-based solutions, and support multistakeholder efforts to end all forms of violence, abuse and neglect of children. The End Violence Fund is a flexible funding vehicle that identifies new and emerging challenges to SDG 16.2 and invests in innovative initiatives that have the potential to replicate and scale. In collaboration with the industry alliance the Tech Coalition, the Partnership has created the Tech Coalition Safe Online Research Fund with the aim of developing and supporting new technology solutions to tackle online CSEA.

The involvement of the private sector is also central to efforts to tackle and address CSEA. Industry contributes to and co-organises hotlines in many countries to facilitate the rapid removal of illegal content that may be hosted unwittingly by service providers. In many jurisdictions, industry has cooperated with law enforcement to facilitate, where legally permitted, the voluntary blocking of URLs for sites containing known CSAM material (see section 4). Cross-sectoral industry groups have also cooperated to develop new technologies to detect and remove CSEA. One such organisation, the Tech Coalition, is an alliance of global technology companies who collaborate to combat CSEA online (Tech Coalition, n.d.[42]). Its work includes fostering innovation and adoption of technologies to address CSEA, collective stakeholder action, encouraging accountability and consistency in transparency reporting, information and knowledge exchange, and investment in research.

In 2022, the Tech Coalition launched its voluntary framework for industry transparency (Tech Coalition, 2022[43]). The Trust Voluntary Framework for Industry Transparency sets out principles-based guidance to technology companies with the objective of building greater public awareness and confidence in how industry platforms address CSEA on their services. The non-binding framework recommends transparency as an essential component of the overall industry effort to combat CSEA online. It contains a recommended structure for transparency reporting, organised around the three core components of: a) *policies and practices*, in which companies should describe their approach to CSEA and provide detail on what is

prohibited on their services, b) *processes and systems*, in which companies are asked to provide information on how their policies are operationalised and enforced, and c) *outcomes*, in which companies provide some quantitative information. This framework emphasises flexibility over standardisation and comparability. Three metrics are highlighted as essential in all TRs: i) the volume of CSEA identified and actioned; ii) the number of violative user accounts identified and actioned; and iii) the number of reports submitted to relevant external authorities. All other metrics included in the framework are described as optional.

Other relevant industry-led initiatives that contribute guidance and standards on tackling CSEA include the GSMA Mobile Alliance Against Child Sexual Abuse Content (GSMA, 2016[44]), which operates globally on behalf of the mobile telecommunications sector, and the ICT Coalition for Children Online (ICT Coalition, n.d.[45]), operating in the European Union. Both organisations have developed voluntary codes of practice to which members adhere in addressing CSEA that may occur on their services.

Given the complex and multifaceted nature of child sexual abuse and exploitation, no single type of intervention can successfully address the problem. For this reason, a host of non-legislative measures to combat CSEA operate in many regions to enhance the legislative process and to ensure more effective coordination between stakeholders. Significant international multistakeholder initiatives include diverse global programmes such as *Disrupting Harm* supported by ECPAT International, INTERPOL and the UNICEF Office of Research Innocenti (End Violence Against Children, n.d.[46]); the Child Dignity Alliance (Child Dignity Alliance, n.d.[47]); and the Alliance to better protect minors online (European Commission, n.d.[48]), the European Union's initiative to increase co-operation with the private sector. In this context, non-legislative measures adopted at national level are also relevant. For example, Canada's National Strategy for the Protection of Children from Sexual Exploitation on the Internet (National Strategy), first created in 2004, is led by Public Safety Canada. The programme implements diverse measures with additional funding awarded in 2020. The Government of Canada also supports the Canadian Centre for Child Protection (C3P) which operates Cybertip.ca hotline to report suspected online sexual exploitation of children, and delivers a range of public awareness, educational, support and referral services.

The OECD is well placed to add to this body of work and to contribute to international initiatives combatting CSEA online. The OECD has already taken a leading role on policy responses to empower and protect children online through its Recommendation on Children in the Digital Environment (OECD, 2021[12]), as well as on the responsibility of business in this regard through the Guidelines for Digital Service Providers (OECD, 2021[16]), which accompany the Recommendation. Following on the Recommendation, the OECD is undertaking analytical work on digital safety by design for children. This work aims to facilitate a shared understanding of what a safety-by-design approach involves, and to help develop and implement policy responses that keep children safer online – including protecting them from CSEA.

This present report not only builds on this work but adds to the OECD's existing expertise and evidence base regarding the transparency reporting practices of online content-sharing practices. In the context of TVEC, following a multidisciplinary consultation process, the OECD developed a Voluntary Transparency Reporting Framework (VTRF) (OECD, 2022[49]). The VTRF aims to improve transparency reporting practices, to support international co-operation and information sharing, and ultimately to reduce the volume and reach of TVEC online while promoting the protection of human rights. At the same time, the OECD series of reports mapping the transparency reporting practices of online content-sharing services in relation to TVEC (OECD, 2022[3]) (OECD, 2021[2]) (OECD, 2020[1]) has contributed to a better understanding of industry-wide efforts to address TVEC. By undertaking this same mapping exercise with regard to CSEA, the OECD likewise aims to contribute important evidence to the fight against CSEA online.

Technology solutions for tackling CSEA

An area that has received increased attention is technology solutions for detecting and removing CSEA from online platforms. The use of technology in content moderation systems offers benefits in part due to its ability to work efficiently at scale while reducing the burden on human moderators having to review content at the extreme end of the harmful spectrum. Although the effectiveness and accuracy of such solutions are improving all the time (Lee, 2020[50]), concerns remain about the ability of automated processes to detect new, previously undetected CSEA reliably (Ngo et al., 2022[51]). Human moderation remains an integral part of the review process, particularly with regard to making decisions about removing content and reporting it to law enforcement.

Technologies to detect child abuse material based on digital fingerprinting or hash matching are the longest established and most widely deployed forms of automated content moderation. Hash matching technologies are used to tag, remove and prevent re-upload of known images and videos of known child sexual abuse material. PhotoDNA, the most widely known such technology, was developed by Microsoft in partnership with Dartmouth College in 2009. PhotoDNA creates a unique digital signature (known as a "hash") of an image which is then compared against signatures (hashes) of other photos to find copies of the same image. When matched with a database containing hashes of previously identified illegal images, PhotoDNA can help detect, disrupt and report the distribution of CSEA. The largest database of hashes is maintained by NCMEC and comprises approximately 1.5 million unique fingerprints of content that has been positively determined as CSEA. A PhotoDNA hash is not reversible and cannot be used to recreate an image (Microsoft, n.d.[52]). PhotoDNA is made available for free by Microsoft both on its Azure platform as well as through NCMEC and is used by over 150 organisations globally. PhotoDNA has been in use for over 10 years and is known to have a high degree of accuracy in the detection, disruption, and reporting of millions of child exploitation images (European Commission, 2022, p. 278[53]).

A similar approach called CSAI (Child Sexual Abuse Imagery) Match has been developed by YouTube to hash match video content online (Google, n.d.[54]). CSAI Match also deploys hash matching or digital fingerprinting of online video streams. When integrated into content moderation systems and connected through an API to Google's database of known CSEA content, the system can be used to detect potentially violative video content. A match using the system will identify which portion of a video matches known CSAI, prioritising it for review by a staff moderator and providing a standardised categorisation of the type of content that was matched.

An example of a widescale deployment of hash matching technology to support the detection of CSEA is Project Arachnid, hosted by the Canadian Centre for Child Protection (C3P). Project Arachnid is an automated web crawler that detects and processes tens of thousands of images per second and sends content removal notices to online service providers to remove child sexual abuse material globally (C3P, n.d.[55]). The initiative uses hashing technology to assist in matching a particular image or video against a database of known CSEA. Hashing technology can either be exact (one image is exactly the same as another), or it can be a close match such as a resized image, for example. Close matches are obtained by using "perceptual hashing technology" or Microsoft PhotoDNA software. Project Arachnid also makes an API available for companies to assist content administrators or hosting providers to proactively compare incoming or existing media on their service against Project Arachnid's list of digital fingerprints.

Identifying previously unknown child exploitative content or detecting suspicious behaviour online poses greater challenges for technology-based content moderation. The use of artificial intelligence and machine learning continues to evolve in this area and typically uses algorithmic-based classifiers and forms of pattern recognition to identify potentially violative content. By their nature, these technologies are not as accurate and need to be trained on large datasets to improve their effectiveness (Ngo et al., 2022[51]). Thorn's Safer tool (Thorn, n.d.[56]), Google's Content Safety API (Google, n.d.[57]), and Meta's AI technology (Meta, 2018[58]) are examples of technologies that use or incorporate classifiers and AI technology to detect previously unknown CSEA content. When used in conjunction with tools that detect known and previously

'hashed' CSEA where there is a likelihood of finding new patterns or matches for violative material, their effectiveness can be improved. Thorn's Safer tool, for example, uses a modular approach and is offered to online content platforms as an integrated solution including hash matching technology that can be extended by using a machine learning based classifier to detect new and previously unreported CSEA material (Thorn, n.d.[56]). This approach is also applied to the detection of grooming, or the solicitation of children for sexual purposes. Here, AI technologies are similarly deployed to assist and complement the human moderator review carried out by platforms, for example, to detect suspicious patterns in text-based communication and online conversations.

A significant challenge facing technology-based content moderation systems (as well as human moderators) is the increasing availability of end-to-end encrypted online communications services, which render existing automated technologies for pre-screening or reviewing content on the server side difficult or impossible. Encryption has been widely deployed across a range of communications services, including interpersonal communications, live video streaming, private messaging and online storage. While encryption provides essential security for personal data and data transactions, when used for illegal purposes, including to exchange or disseminate CSEA content, criminals or CSEA perpetrators can use encryption to mask their identities and transactions. Encryption technology is used to safeguard data "at rest", i.e. data stored on various devices and on digital storage services, as well as data "in motion", i.e. data transmitted or transferred from one device to another, and which may be secured using end-to-end encryption. Some stakeholders contend that existing hash matching, AI tools and other automated content moderation technologies can work in end-to-end encrypted environments if they are deployed on the client side instead of the server sider (Levy and Robinson, 2022[59]) (Safety Tech Challenge Fund, n.d.[60]). However, others assert that the "tools currently used by industry to reliably detect known child sexual abuse materials do not work in E2EE electronic communications" (European Commission, 2022, p. 285[53]).

According to Europol, the increasing use by criminals of mainstream platforms with strong encryption poses a serious and ongoing challenge for law enforcement agencies (Europol, 2021[61]). Meta's announcement in 2021 of its intention to integrate end-to-end encryption into all its messaging platforms (Meta, 2021[62]) caused alarm among child safety organisations across the world, as 95% of all reports to NCMEC in 2021 came from Facebook alone. With the integration of encryption across all of its communications services, the majority of such detection activity would no longer be possible. By way of an example, the detection of CSEA on WhatsApp, an encrypted communications service, relies on user reports, triangulation of unencrypted metadata, such as profile names and account information, as well as behavioural analysis (WhatsApp, n.d.[63]). As a result, the number of reports submitted by WhatsApp to NCMEC in 2022 was just over 1 million reports. This compares to the 21 million reports submitted by Facebook, including Facebook Messenger, which at this point is not encrypted by default (NCMEC, 2023[18]).[14]

Preserving access to encrypted communications for the purposes of law enforcement has been the subject of several legislative initiatives, none of which to date has succeeded in establishing a new standard or protocol. In the United States, the Lawful Access to Encrypted Data Act was introduced in the Senate in 2020 (U.S. Congress, 2020[64]) and, if it had been approved, would have mandated technical access to specified communications across a large section of the industry including manufacturers, online content providers and cloud hosting services. The Bill would also have supported further research into the creation of secure products and services that provide lawful access to encrypted communications (Europol & Eurojust, 2021, p. 35[65]). In the European Union, the proposal for a Regulation laying down rules to prevent and combat child sexual abuse (European Union, 2021[66]) would require online content providers to detect and remove illegal CSEA content with the stipulation that the requirements apply without prejudice to end-to-end encrypted communications. The United Kingdom's Online Safety Bill (UK Parliament, 2022[67]) also includes consideration of technologies to detect CSEA and is further discussed in the next section.

4 Existing and emerging laws and regulations on CSEA online

As indicated in Section 3, international initiatives have sought to secure a consistent approach to legislation and criminalisation of sexual offences perpetrated against children through the use of digital technologies. Lawmakers are incorporating enhanced measures in civil laws to halt the spread of CSEA through digital platforms and services. Laws to criminalise CSEA predate the Internet, though the increased prevalence and complexity through the proliferation of digital technologies has created the need for new legal frameworks. Ensuring there is an effective legal response to address online CSEA is a central pillar of the Model National Response (WeProtect Global Alliance, 2016[68]) and the End Violence Against Children Partnership (End Violence Against Children, n.d.[41]). While laws to prohibit and criminalise what is traditionally referred to as "child pornography" exist in most jurisdictions, key aspects of CSEA, particularly in its digital form, are still not systematically outlawed in many jurisdictions (ICMEC, 2018[69]). The latest edition of the periodic review of legislation undertaken by ICMEC, the Model Legislation & Global Review, points out that just 21 of 196 countries surveyed had fully comprehensive laws in place to address all aspects of child sexual abuse and exploitation (ICMEC, 2018[69]). This section highlights the main features of legislative and regulatory development as articulated within key international and regional legal instruments addressing CSEA. These are then further explored within the key trends evidenced within domestic laws (including those under consideration as well as those already enacted) as illustrated by examples from OECD member countries.

International and regional law

The UNCRC (UN General Assembly, 1989[28]) and the Optional Protocol on the Sale of Children, Child Prostitution and Child Pornography (OPSC) (UN General Assembly, 2001[70]) provide the main reference framework for many national laws in this area. Article 34 of the UNCRC requires States Parties "*to protect the child from all forms of sexual exploitation and abuse" and to take all appropriate measures to prevent:*

> *"(a) The inducement or coercion of a child to engage in any unlawful sexual activity;*
>
> *(b) The exploitative use of children in prostitution or other unlawful sexual practices;*
>
> *(c) The exploitative use of children in pornographic performances and materials.*"
>
> (UN General Assembly, 1989[28]).

The Optional Protocol, at Article 2(c), defines child pornography in a more detailed way as "*any representation, by whatever means, of a child engaged in real or simulated explicit sexual activities or any representation of the sexual parts of a child for primarily sexual purposes*" (UN General Assembly, 2001[70]). Article 3 (1) requires States Parties to criminalise acts and activities including "*producing, distributing, disseminating, importing, exporting, offering, selling or possessing...child pornography*", whether committed domestically or transnationally, on an individual or organised basis. Article 10(1) addresses the need for international co-operation in the "*prevention, detection, investigation, prosecution, and*

punishment of those responsible for acts involving the sale of children, child prostitution, child pornography and child sex tourism" (UN General Assembly, 2001[70]).

The Convention on Cybercrime, also known as the Budapest Convention, (Council of Europe, 2001[71]) is another important legal instrument in tackling CSEA. The Convention was signed in 2001 and now has 65 member states, predominantly from Europe but also from Asia, North and South America, Australia, and the Pacific. It entered into force in July 2004. The Budapest Convention is significant as one of the first international legal instruments to address cybercrime and the trans-border dimension of digital technology-enabled crime. Article 9 of the Budapest Convention deals with offences related to "child pornography" which involve a minor or any person under 18 years of age. Article 9(2) defines "Child pornography" as "*any material that visually depicts:*

- *a minor engaged in sexually explicit conduct;*
- *a person appearing to be a minor engaged in sexually explicit conduct;*
- *realistic images representing a minor engaged in sexually explicit conduct*" (Council of Europe, 2001[71]).

Article 9(1) requires States parties to make it a criminal offence to "*produce child pornography for the purpose of its distribution through a computer system; offer or make available child pornography through a computer system; distribute or transmit child pornography through a computer system; procure child pornography through a computer system for oneself or for another person; and possess child pornography in a computer system or on a computer-data storage medium*" (Council of Europe, 2001[71]).

The Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (known as the "Lanzarote Convention") (Council of Europe, 2007[72]) further advances the legal definition of CSEA-related crimes and directs States to prevent such offences and to prosecute perpetrators and protect child victims. Article 20(2) of the Lanzarote Convention defines "child pornography" as "*any material that visually depicts a child engaged in real or simulated sexually explicit conduct or any depiction of a child's sexual organs for primarily sexual purposes*". Furthermore, Article 23 identifies the sexual solicitation of children for sexual purposes through information and communication technologies (also known as "grooming") as an offence and requires parties to take necessary measures to criminalise the conduct. Overall, the Convention calls on parties to criminalise the production, offering, making available, distribution or transmission of CSEA material. It also requires criminalisation of the possession or knowingly obtaining access through information and communication technologies to CSEA content. These proscriptions are taken to comprehensively address all forms of technology-facilitated access to CSEA including through online, web access, live streaming, downloading or any other form of electronic display, thus targeting not just its supply or distribution but any form of consumption or possession through online means. Notably, the Convention calls on States to criminalise the possession of CSEA and not just its production as a means of disrupting the market for online CSEA.[15] The Convention also addresses the issue of corporate responsibility (Article 26) and the general principles and measures for international co-operation (Article 38).

The UNCRC and the OPSC were both adopted at a time when the impact of digital technologies on the perpetration of sexual offences against children was less pronounced. They remain, however, the most comprehensive and universal of legal instruments safeguarding children's rights. The UNCRC has been ratified by 196 countries while, as of 2023, 178 countries, including all OECD member countries, have ratified the OPSC (OHCHR, 2023[73]). Legal instruments from the Council of Europe are more recent in origin and have been influential in shaping legislative development. All 46 Member States of the Council of Europe have signed and ratified the Lanzarote Convention. The Budapest Convention which is more international in nature has been ratified by most OECD member countries.[16] The OECD is also an observer organisation to the Cybercrime Convention Committee.

To ensure that legislative measures keep pace with developments in digital technology, the UN Committee on the Rights of the Child has issued guidelines for implementation of the UNCRC and the OPSC in the light of new challenges and evolving online sexual abuses (UN Committee on the Rights of the Child, 2019[74]). In addition, General Comment 25 on children's rights in relation to the digital environment contains a number of recommendations which refer to obligations of the private sector. Relevant recommendations call on States Parties to:

- ensure that businesses meet their responsibilities to respect children's rights and remedy abuse (p.6, para 35)
- monitor compliance of businesses in preventing their services contributing to violation or abuse of children's rights (p.7, para 36)
- require the business sector to undertake child rights due diligence, including the use of child rights impact assessments and to disclose them to the public (p.7, para 38)
- require business enterprises to implement regulatory frameworks, industry codes and terms of services that adhere to the highest standards of ethics, privacy and safety (p.7, para 39)
- consider appropriate measures to enable the detection and reporting of child sexual exploitation and abuse or child sexual abuse material in the case of encrypted networks, noting that such measures must be strictly limited according to the principles of legality, necessity and proportionality (p.12, para 70).

The Council of Europe has similarly moved to draft guidance regarding the application of children's rights frameworks to the digital environment and to further elaborate on the obligations that should be considered in relation to digital service providers. Its Recommendation to respect, protect and fulfil the rights of the child in the digital environment (Council of Europe, 2018[75]) states that children have the right to be protected from all forms of violence, exploitation and abuse in the digital environment. The Recommendation notes that contact risks of sexual exploitation and abuse and solicitation for sexual purposes are of particular concern. It is recommended therefore that member States should "*require business enterprises to take reasonable, proportionate and effective measures to ensure that their networks or online services are not misused for criminal or other unlawful purposes in ways which may harm children, for example in relation to the production, distribution, provision of access to, advertising of or storage of child sexual abuse material or other forms of online child abuse*" (p.10, para 64). States are further urged to require that business enterprises apply hash lists with a view to ensuring their networks are not being misused to store or distribute child sexual abuse images.

A notable legal standard at the regional level is *Directive 2011/93/EU of the European Union on combating the sexual abuse and sexual exploitation of children and child pornography* (see further details in Box 1) (European Union, 2011[76]). The Directive entered into force in 2011 and builds on the Council of Europe's Lanzarote Convention. It prescribes in greater detail how Member States should address the role of the Internet and information and communication technologies (ICT) in the distribution of child sexual exploitation and abuse. Member States are required to take necessary measures to prevent the use of the Internet for child sexual abuse, exploitation, or the dissemination of child pornography.

Box 1. Directive 2011/93/EU of the European Union on combating the sexual abuse and sexual exploitation of children and child pornography

Directive 2011/93 of the EU contains a number of articles that are particularly relevant to the creation of domestic legislation concerning online CSEA. For example:

- Article 2 (c) defines "child pornography" as

(i) *any material that visually depicts a child engaged in real or simulated sexually explicit conduct;*

(ii) *any depiction of the sexual organs of a child for primarily sexual purposes;*

(iii) *any material that visually depicts any person appearing to be a child engaged in real or simulated sexually explicit conduct or any depiction of the sexual organs of any person appearing to be a child, for primarily sexual purposes; or*

(iv) *realistic images of a child engaged in sexually explicit conduct or realistic images of the sexual organs of a child, for primarily sexual purposes.*

- Article 2 (e) defines "pornographic performance" as "*a live exhibition aimed at an audience, including by means of information and communication technology of a (i) a child engaged in real or simulated sexually explicit conduct; or (ii) the sexual organs of a child for primarily sexual purposes*".
- Article 4 outlines offences concerning sexual exploitation requiring Member States to take the necessary measures to criminalise the following conduct:
  - causing or recruiting a child to participate in pornographic performances;
  - profiting from or otherwise exploiting a child for such purposes;
  - coercing or forcing a child to participate in pornographic performances;
  - threatening a child for such purposes; or
  - knowingly attending pornographic performances involving the participation of a child.
- Article 5 sets out the possession, distribution and production of CSEA by means of information and communication technology as distinct offences punishable by minimum terms of imprisonment.
- Article 6 creates the specific offence of the solicitation of a child by means information and communication technology whereby an adult seeks to meet with a child for sexual purposes or for the purpose of producing CSEA.
- Article 25 deals with measures against websites containing or disseminating CSEA. Member States are required to ensure the prompt removal of web pages containing or disseminating child pornography hosted in their territory and to endeavour to obtain the removal of such pages hosted outside of their territory. Member States may also take measures to block access to web pages containing or disseminating child pornography within their territory. Where blocking is introduced, transparent procedures and adequate safeguards are required to ensure that the restriction is limited to what is necessary and proportionate, and that users are informed of the reason for the restriction.

Regarding new laws under consideration, in December 2019, the UN General Assembly adopted resolution 74/247, through which an open-ended ad hoc intergovernmental committee of experts, representative of all regions, was established to elaborate a comprehensive international convention on countering the use of information and communications technologies for criminal purposes (UN General Assembly, 2022[77]). A consolidated negotiating document on the future convention was introduced in November 2022 and is currently subject to negotiations among UN Member States with a view to reach agreement on its final wording in 2024. The draft chapter on criminalisation includes, *inter alia*, provisions on tackling cybercrimes related to child sexual abuse or to grooming or procuring of a child for sexual purposes.

In May 2022, the European Commission proposed a new EU Regulation laying down rules to prevent and combat child sexual abuse (see further details in Box 2) (European Commission, 2022[78]). The proposal is designed to complement the Digital Services Act (European Commission, 2022[79])[17] but with a specific focus on combatting online CSEA. The new rules propose the creation of a new independent EU Centre

on Child Sexual Abuse, clear obligations for service providers to detect, report, remove and block access to online child sexual abuse material, as well as specific prevention, prosecution, and protection responsibilities for national authorities. The proposed CSA Regulation would also replace the current temporary derogation from the e-Privacy Directive 2002/58/EC, which allows providers of interpersonal communications services to scan communications and process personal and other data for the purpose of combatting online child sexual abuse (European Union, 2021[66]).

Box 2. Proposal for a Regulation on preventing and combatting the sexual abuse and sexual exploitation of children (the 'CSA Regulation')

The proposed legislation introduces new rules to help EU countries to detect and report child sexual abuse online; prevent child sexual abuse; and support victims.

A summary of the proposed rules includes:

- *Mandatory risk assessment and risk mitigation measures* are proposed for hosting or interpersonal communication service providers (Article 3). Providers are required to implement appropriate mitigation measures for the risks identified (Article 4) and to report on the outcome of both the risk assessment and the mitigation measures to the Coordinating Authorities designated by the Member States (Article 5). App stores are required to assess the risk that their services may be misused to disseminate child sexual abuse material or for the solicitation or grooming of children (Article 6).
- *Detection orders* may be issued to a provider to detect certain types of online child sexual abuse on the relevant service where a coordinating authority becomes aware of a significant risk of its being misused or the purpose of online child sexual abuse (Articles 7 and 8).
- Providers will be obliged to report any instance of potential online child sexual abuse on their services to the EU Centre (Article 12) with specific technical requirements that the relevant report must fulfil (Article 13).
- National coordinating authorities can *issue removal orders* if illegal child sexual abuse material is not swiftly removed by providers (Article 14). Internet access providers will also be required to disable access to images and videos that cannot be removed, for example, where they are hosted outside the EU in non-cooperative jurisdictions (Articles 16 and 17).
- The proposal also lays out an exemption from liability for child sexual abuse offenses for providers of relevant information society services who comply with the Regulation (Article 19).

Under the proposed Regulation, the Commission will set up a European centre to prevent and fight child sexual abuse and support victims (the EU Centre). Its aim is to provide oversight, transparency and accountability. The proposed EU Centre will coordinate actions to fight against child sexual abuse, from detection and reporting, to prevention and assistance to victims. It will support law enforcement to act on reports and will work closely with similar centres in the United States, Canada and Australia. It will also provide companies with indicators to find and report online child sexual abuse.

Source: (European Commission, 2022[78])

Domestic legislation and online CSEA offences

National legislation to address offences related to the production, distribution, and possession of CSEA have evolved in line with international law as well as the changing threat environment. ICMEC's Model Legislation and Global Review observes a growing international consensus on definitions of CSEA with

118 countries deemed to have sufficient legislation to address CSEA-related offences (ICMEC, 2018, p. 5[69]). However, considerable variation remains across jurisdictions with inconsistencies or gaps in the definitions of CSEA, including various forms of technology-facilitated CSEA such as live-streaming or self-generated CSEA imagery as well as extra-territorial CSEA offences. In addition, national legislation regarding the liability of online content-sharing services and obligations on platforms to report and/or remove CSEA remains fragmented. The following provides a brief overview of selected examples of domestic legislation in OECD member countries to illustrate key trends.

Offences related to child sexual abuse and the possession of child sex abuse material are typically covered in domestic criminal legislation and reflect the international consensus that the abuse of children by adults for sexual purposes is a heinous crime, meriting the strongest of responses. The technological recording and depiction of such abuse and its online distribution is recognised to be a further compounding of a deeply serious offence which is also a re-victimisation of the abused subject. Accordingly, many countries have sought to update the relevant offences in their penal code to include reference to specific offences concerning the production, distribution or possession of material depicting real or simulated sexual activity involving minors. The solicitation of a child for sexual purposes through technological means is also typically included as a specific offence under criminal law. Some relevant examples are expanded on below.

In Australia, laws relating to CSEA exist both at the federal and the state level. Australia's federal Criminal Code Act 1995 outlines specific offences relating to CSEA in the context of communications services (Australian Government, 1995[80]). Offences include the use of the Internet or telecommunications services by a person transmitting, accessing and soliciting sexual abuse material of a child under 18 years. Solicitation, such as an adult procuring and grooming a child for sexual activity, as well as sexual activity which occurs remotely using technology such as through live webcam streaming or peer-to-peer networks and by causing a child to engage in sexual activity with another person are specific offences. Similarly, telecommunications-based child exploitation offences also cover the range of activities a person can engage in when using the Internet, mobile phones and other applications to procure or access child sexual abuse material. Such activities include viewing, copying, downloading, sending, exchanging, soliciting and making material available to others (CDPP, n.d.[81]).

In Canada, laws to protect children from CSEA exist both at the federal and at the provincial and territorial levels. Canada's Criminal Code has a comprehensive range of prohibitions against all forms of sexual abuse and exploitation of children. Section 163(1) of the Criminal Code (Government of Canada, 1985[82]) defines "child pornography" as:

- *Any visual representation of explicit sexual activity with a person who is, or who is depicted as being under the age of 18;*
- *Any written material, visual representation or audio recording that advocates or counsels sexual activity with a person under the age of 18;*
- *Any audio recording that has as its dominant characteristic the description, presentation or representation, for a sexual purpose, of sexual activity with a person under the age of 18 years that would be an offence under the Act.*

The making, distribution, possession and accessing of "child pornography" are each made indictable[18] offences. Canada's Criminal Code also creates specific offences relating to trafficking of a person under the age of 18 years for sexual or other purposes. The 2015 Tougher Penalties for Child Predators Act S.C. (2015, c. 23) amended the Criminal Code to strengthen penalties for CSEA perpetrators.

Under French law, CSEA and CSAM are classified as material that is prejudicial to human dignity for which strict laws founded on multilateral human rights treaties are enacted. The Civil Code defines minors as individuals of either sex under the age of 18 years (Article 488). In 1998, the French Parliament passed Law 98-468 (République Française, 1998[83]) to bring together all relevant provisions relating to sexual offences against children in traditional media, on the Internet and through computer-generated child sexual

abuse material under the child protection provisions of the French Penal Code (Eko, 2016[84]). French law distinguishes several categories of sexual offences against children and provides for offences of rape, sexual assault, indecent assault and corruption of minors. CSEA is specifically addressed in Articles 227-23 – 227-27 of the Criminal Code (République Française, 2021[85]). For example, Section 227-23, (Endangerment of Minors), specifically outlaws "*the taking, recording or transmitting the picture or representation of a minor with a view to circulating it, where that image or representation has a pornographic character"*. Strict penalties apply to the possession and distribution of pornographic material involving minors, including pornographic images of a person whose physical appearance is that of a minor. The offence is aggravated when electronic communications are involved. Section 227-24-1 addresses soliciting sexual relations with a minor and increases penalties where a communications network is used.

Germany's legal provisions to deal with CSEA are contained within the Criminal Code (StGB), the Youth Protection Act (JuSchG) and the Interstate Treaty on the Protection of Minors in the Media (JMStV). In November 2008, the German Criminal Code was updated to implement the framework decision of the Council of the European Union on combating sexual exploitation of children and child pornography, which had been in effect since 2004. Amendments in 2008, in 2015 and in 2021 were introduced to strengthen and harmonise legislation on child sexual abuse offences, child trafficking, and to raise the age of protection from sexual abuse and exploitation to 18 years. Section 184 of the Criminal Code prohibits making pornographic material available to minors under the age of 18. The Criminal Code (Sections 184b and 184c) distinguishes between child sexual abuse material involving children (under 14 years of age) and juveniles (between 14 and 18). The possession of either child or juvenile sexual abuse material is considered a punishable offence if it involves an actual, or in the case of child sexual abuse material, a realistic event (§ 184 b-c StGB). Section 184 also prohibits the use of pornographic or sexually explicit material in schools. An exception is made if sexually explicit content is produced by consenting adolescents and is intended solely for personal use, thus seeking to avoid criminalisation of children under laws designed to protect them.[19]

In Korea, the Act on the Protection of Children and Youth against Sexual Abuse (Republic of Korea, 2020[86]) is the primary legislation relating to CSEA and implements key provisions of the OPSC which Korea ratified in 2004. Under the Act, children or youth refers to anyone under the age of 19 years.[20] Child or youth sexual exploitation material is defined as the "*depiction of children or youth, or persons or representations that can be obviously perceived as children or youths, doing any act such as engaging in any other sexual act, in the form of a film, video, game software, or picture, image, etc. displayed on computers or other communications media*" (Article 2(5)). The Act prohibits and punishes the production, import and export, as well as the sale, loan, distribution or provision of child abuse or sexual exploitation material for commercial purposes (Article 11 (1) and (2)). The possession of child or youth sexual exploitation material is also punishable (Article 11-5), where possession may be taken to mean viewing, accessing, or downloading CSEA content. The Act was further amended in 2021 to include crimes of online grooming. Article 15-2 creates an offence to converse with a child or youth for the purpose of sexual exploitation. Under the revised law, people repeatedly attempting to lure children by sending online sexual messages, or trying to exploit them sexually, face severe penalties.

New Zealand does not have a distinct CSEA offence. Instead, child sexual abuse material is included as a subset of publications considered "objectionable" under the Films, Videos, and Publications Classification Act 1993 (the Classification Act) (New Zealand Department of Internal Affairs, 1993[87]). Under the Classification Act, one of the reasons a publication may be deemed objectionable is if it promotes or supports, or tends to promote or support, "*the exploitation of children, or young persons, or both, for sexual purposes"* (section 3(2)(a)). The possession, making, supply and distribution of such objectionable material is subject to severe penalties. The Films, Videos, and Publications Classification Amendment Act 2015 introduced several significant changes to the enforcement provisions of the Classification Act. Possession of objectionable content is punishable by imprisonment of up to 10 years while distributing objectionable content, such as making the content available for other people to access online, carries a maximum term

of imprisonment of up to 14 years. The New Zealand Crimes Act 1961, Section 98AA (New Zealand, Ministry of Justice, 2021[88]), was also updated in 2015 to criminalise domestic trafficking and some aspects of grooming.

The Republic of Türkiye's (hereafter "Türkiye") Criminal Code (Law No. 5237) (Government of Türkiye, 2005[89]) criminalises the production, distribution, possession, and access of CSEA material, and establishes penalties for offenders (Article 226(3)). Under the Criminal Code it is an offence to acts as an intermediary for the purpose of broadcasting such material, or to show obscene sexual material (including CSEA) to a child (Article 226(5)). Public awareness measures complement the legal framework, and Türkiye seeks to encourage public participation in reporting illegal online content. For example, Türkiye's Internet Hotline serves as a platform for individuals to report instances of CSEA material and other illegal content found online (Türkiye Information Technologies and Communications Authority, 2022[90]).

In the United Kingdom, offences related to CSEA are covered under the Protection of Children Act 1978 (PCA 1978) at section 1 (UK Government, 1978[91]), and the Criminal Justice Act 1988 (CJA 1988) at section 160 (UK Government, 1988[92]).[21] These Acts create offences in respect of indecent images or pseudo-images of a child. A pseudo-photograph means an image, whether made by computer graphics or in any other way, which appears to be a photograph. A child is any person aged under 18 years of age (s7(6) of PCA 1978). Under Section 1 of PCA 1978, it is an offence to *"make any indecent photographs or pseudo-photographs of a child; distribute or show such indecent photographs or pseudo-photographs; have in a person's possession such indecent photographs or pseudo-photographs with a view to their being distributed or shown by themselves or others".* Section 160 of CJA 1988 covers the offence of possession of an indecent image of a child. There is no requirement that the defendant have any motive in relation to making or distributing the image – all that is required is that the defendant had the image in their possession. The Coroners and Justice Act, 2009 made the possession of non-photographic sexualised depictions of children, e.g., computer-generated imagery or virtual pornography, a criminal offence (section 62). Livestreaming of child sexual abuse is also covered under existing legislation whereby a person who has merely viewed an image or video of livestreamed CSEA is deemed to have committed an offence and that causing a livestream to be displayed on a device is considered to be "production" of child abuse imagery.[22] The Sexual Offences Act (UK Government, 2003[93]) deals with offences of solicitation of a child and makes it illegal for an adult to intentionally communicate with a child under 16 for a sexual purpose. The offence is still committed whether or not the child communicates with the adult (s15a).

In the United States, images of child sexual abuse or exploitation are illegal under federal as well as state law and are not protected under First Amendment rights[23] (United States, Department of Justice, n.d.[94]). Section 2256 of Title 18, United States Code, defines "child pornography" as any visual depiction of sexually explicit conduct involving a minor (someone under 18 years of age) (U.S. Congress, 1982[95]). Visual depictions include photographs, videos, digital or computer-generated images indistinguishable from an actual minor, and images created, adapted, or modified, but appear to depict an identifiable, actual minor. Section 2251 criminalises solicitation of a child or any attempt to persuade, induce, entice, or coerce a minor to engage in sexually explicit conduct for purposes of producing visual depictions of that conduct. Notably, the legal definition of sexually explicit conduct does not require that an image depict a child engaging in sexual activity. A picture of a naked child may constitute illegal child sexual exploitation if it is sufficiently sexually suggestive. Additionally, the age of consent for sexual activity in a given state is irrelevant; any depiction of a minor under 18 years of age engaging in sexually explicit conduct is illegal.

Extra-territorial offences

The borderless nature of the Internet in the context of CSEA creates additional challenges for legislators in establishing effective measures to deal with crimes committed in other jurisdictions. Perpetrators may

seek to exploit gaps in the legislative framework in some jurisdictions to perpetrate the exploitation of children through what is referred to as the sexual exploitation of children in travel and tourism (SECTT). The growth of international travel and tourism in tandem with mobile technologies, peer-to-peer networking and the relative anonymity afforded by many digital services, have brought about increased opportunities for sexual exploitation of children in locations where child protection services and law enforcement support may be less well resourced. ECPAT International has called for legal reforms in all countries to clearly define and prohibit child sexual exploitation in travel and tourism and for increased international co-operation to tackle extra-territorial offences (ECPAT International, 2016[96]).

In Australia, Division 272 of the Criminal Code (Australian Government, 1995[80]) targets child sex offences committed by Australians against children who are living overseas. These include possession of child abuse material, engaging in sexual activity with children overseas including where the technology is used to enable the offence to occur remotely. This can be in circumstances where the offender travels overseas to engage in sexual activity or where the offences occur over the Internet through sexual acts online or involve the transmission of material from the place where children are procured and groomed.

Section 227-27-1 of France's Penal Code likewise outlaws trafficking of children for sexual and other exploitative purposes. Where sexual abuse or exploitation is committed abroad against a minor by a French national or a person habitually resident in France, French law applies (République Française, 1998[83]).

Provisions are also contained in the German Criminal Code (StGB) to prosecute child sex tourism (German Federal Ministry of Justice, 2021[97]). German extraterritorial jurisdiction applies to acts committed abroad such as the criminal offences of trafficking and the distribution of pornography, regardless of the legality of those acts in the country where they took place.

The New Zealand Crimes Act 1961, Section 98AA (New Zealand, Ministry of Justice, 2021[88]), covers child sexual exploitation and trafficking in other jurisdictions. A person who sells, buys, barters, rents, hires or in any other way enters into an arrangement involving a person under 18 years for the purpose of commercial sexual exploitation can be sentenced up to 14 years in jail. Under Section 131B, a person can be sentenced to seven years for meeting a child under the age of 18 following sexual grooming.

Industry responsibility, mandatory reporting and administrative powers of regulators

Balancing the need for inclusion within countries' criminal law of various offences of CSEA is a corresponding provision within civil law to address the responsibility of industry providers to identify and eliminate CSEA as well as mandatory reporting obligations for ISPs and other industry actors to report suspected CSEA to law enforcement or other mandated agency (ICMEC, 2018[69]). Such provisions are more recent in origin and are currently evolving in many jurisdictions, moving away from self-regulatory arrangements regarding industry reporting obligations.

Australia's Online Safety Act 2021 (Government of Australia, 2021[98]) updated its existing provisions related to industry obligations through a range of expanded statutory schemes, functions and powers delegated to the eSafety Commissioner. These include a new adult cyber abuse scheme and updated schemes on cyber bullying, image-based abuse and illegal and restricted online content, including child sexual exploitation and abuse material. Under the latter scheme, the eSafety Commissioner can investigate content complaints from the public and take action to facilitate the removal of child sexual exploitation material. While this is primarily done through eSafety's membership in the INHOPE global network, the Commissioner also has powers to issue removal notices to services where this material is hosted, regardless of their location. In certain circumstances, the eSafety Commissioner may also issue link deletion and app removal notices.

The Online Safety Act 2021 also provides for industry associations across eight sections of the online industry (such as providers of social media, messaging, search engine and app distribution services) to develop codes with relevant measures to regulate illegal and restricted content for participants of their industry section. The codes are submitted to the Commissioner for consideration; if they meet statutory requirements and provide appropriate community safeguards, they are registered and become mandatory and enforceable. If a code does not meet the threshold, the Commissioner is able to issue an industry standard which dictates the relevant mandatory and enforceable rules for that section.[24]

The Act also provides the eSafety Commissioner with powers to require online service providers to report on how they are meeting a list of Basic Online Safety Expectations. This includes reporting on the reasonable steps a service provider is taking to proactively minimise material or activity that is unlawful or harmful, including child sexual exploitation material, and ensuring users can use a service in a safe manner. The obligation to respond to a reporting requirement is enforceable and backed by civil penalties and other mechanisms. The eSafety Commissioner can publish summaries of the information received through the notices, with the goal of improving industry's transparency and accountability. The first transparency notices were issued in August 2022, and a report published in December 2022 (eSafety Commissioner, 2022[99]).[25]

In Canada, An Act respecting the mandatory reporting of Internet child pornography by persons who provide an Internet service (Government of Canada, 2011[100]) requires ISPs to report tips they receive regarding any website that may possibly contain child sexual exploitation or abuse material to the Canadian Centre for Child Protection (through Cybertip.ca) and to notify law enforcement if they believe that a CSEA-related offence has been committed using their Internet service. Canada's provinces also implement equivalent reporting legislation based on the definition of "child pornography" in the Criminal Code to make it mandatory for any individual to report instances of suspected CSEA to a designated regulated organisation, agency or person such as the national hotline, Cybertip.ca.

In France, the Law on Confidence in Digital Economy (LECN) (République Française, 2004[101]) outlines the liability and role of ISPs. Act No. 2007-297 of 5 March 2007 extended the application of sexual offences against children to Internet communications, which places obligations on ISPs through a system of self-regulation (République Française, 2007[102]). This includes relevant obligations regarding legal notices, mandatory online information, and obligations and liability related to content and hosting providers. ISPs, while having an obligation to support the fight against the dissemination of criminal offences relating to CSEA content and the abuse of minors, are not liable for information transmitted or hosted unless they have actual knowledge of illegal activity or do not act expeditiously to remove or disable access when made aware of such content (Article 6-I-7 of Law No 2004-575). Public authorities also have powers under the law to remove or block of content determined to be illegal. However, proposed legislation which would require social media platforms to remove CSEA as well as terrorist content within one hour of flagging were later struck down by France's Constitutional Council in 2020 (Thurman, Nalmpatian and Obster, 2022[103]).

The protection of minors from harmful media is regulated in Germany through the Youth Protection Act (JuSchG) (German Federal Ministry of Justice, 2002[104]). The Act was last updated in May 2021 to take account of evolving risks in the online world such as cyber-grooming and cyber-bullying. It is also intended to minimise the risk of children under 14 years and adolescents, 14 to 18 years, becoming victims of online sexual violence. Under the Act, online services are obliged to undertake appropriate precautionary measures to protect the personal integrity of children and adolescents. These include the provision of secure default settings for limiting the risk of specific sites being used by minors, low-threshold reporting and help systems, using general terms and conditions that are suitable for children and adolescents or implementing age verification systems. The Act also updates age classifications for video games and films and extends the regulations to streaming platforms and online gaming platforms. Implementation is overseen by The Federal Agency for the Protection of Children and Young People in the Media (Bundeszentrale für Kinder- und Jugendmedienschutz).

Germany's Network Enforcement Act (NetzDG) (German Federal Ministry of Justice, 2017[105]) which came into effect in 2017 imposed large fines for social media platforms for non-compliance with existing legal obligations to remove content that is "clearly illegal" within 24 hours after receiving a user complaint. NetzDG is applicable only to social media networks that have 2 million or more registered users in Germany. The Act was amended in 2021 and requires platforms to report specific offences – including the possession, acquisition or distribution of child pornography material – directly to the Federal Criminal Police Office. The updated NetzDG aims to enhance the user-friendliness of the reporting channels for complaints about unlawful content and introduces an appeals procedure for measures taken by the social network provider. Under the EU Audiovisual Media Services Directive, video-sharing platform services are included in the scope of the NetzDG.

In Korea, the Act on Promotion of Information and Communication Network Utilization and Information Protection (Republic of Korea, 2020[106]), and the Telecommunication Business Act (Republic of Korea, 2020[107])define "*child and adolescent sexual exploitation materials*" as "*illegally filmed materials or the like"* ("illegally filmed materials") and impose various obligations on online service providers to prevent the circulation of such materials. Article 44-9 of the Act on Promotion of Information and Communication Network Utilization and Information Protection requires online service providers to designate a person responsible for preventing the circulation of Illegal filmed materials.[26] Online service providers are required to take necessary measures to prevent circulation, such as deleting illegal filmed materials and blocking access. Article 64-5 of the same Act requires the disclosure of transparency reports on the disposal of illegally filmed materials, every year. The presidential Decree in Article 22-5(2) of the Telecommunications Business Act obliges online service providers to take technical and managerial measures, as follows: a) provision of reporting functions for information suspected to be illegally filmed materials, b) restriction of the transmission of search results, if the user's search query corresponds to illegally filmed materials, c) technical filtering if the information that users intend to publish corresponds to illegally filmed materials, and d) advance warning to users that circulation prevention measures such as deletion and access blocking are being taken and that they may be punished for circulating illegally filmed materials pursuant to the relevant laws.

An amendment to New Zealand's Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Act 2021 (2021/43) (New Zealand, Department of Internal Affairs, 2021[108]) was passed by Parliament to facilitate the urgent prevention and mitigation of harms caused by objectionable publications. This was one of a number of measures introduced in the aftermath of the Christchurch mosque shootings on 15 March 2019. The Act makes live-streaming of objectionable content a criminal offence and confers additional authority on the Chief Censor to issue removal notices for offending material. The legislation provides that the "safe harbour" provisions in the Harmful Digital Communications Act 2016 (HDC Act) do not apply in respect of online content providers' liability under this Act. It also facilitates the setting up of future government-backed mechanisms for blocking or filtering objectionable online content. Review and appeal processes set out in regulations would apply to decisions relating to blocking websites, online applications or their equivalent. Decisions relating to the blocking of websites may also be challenged through judicial review. New Zealand's Department of Internal Affairs operates a Digital Child Exploitation Filtering system whereby ISPs can voluntarily block known child sexual abuse content. The Department of Internal Affairs also provides a Content Complaint reporting mechanism whereby members of the public can report online content that promotes or supports various categories of objectionable content, including the sexual exploitation of children.

Türkiye's Law on the Regulation of Internet Publications and Combating Crimes Committed through Such Publications (Law No. 5651) (Government of Türkiye, 2015[109]) regulates the responsibilities and liabilities of content providers, hosting providers, access providers and public use providers (Articles 1, 2). The law sets out provisions aimed at combatting the spread of illegal content on the Internet, including CSEA. It includes measures for blocking access to and removing such content and establishes penalty should

hosting or access providers fail to comply with a blocking or removal request (Articles 8, 9). The law requires technical measures to combat the dissemination of CSEA material in collaboration with ISPs and technology companies. One initiative in place to support the implementation of the law is Türkiye's "Safe Internet Service" is a filtering system that works in conjunction with ISPs to block access to websites containing illegal content, including CSEA material (Türkiye Information Technologies and Communications Authority, 2018[110]).

In the United Kingdom, the Coroners and Justice Act 2009 (Section 68 and schedule 13) addresses the obligations of information society services providers (United Kingdom, 2009[111]). This law implements provisions of the European Union's e-Commerce Directive (European Union, 2000[112]), limiting the liability of ISPs from prosecution for the distribution of illegal content where the service provider had no actual knowledge when the information was provided that it contained offending material or, on obtaining actual knowledge that the information contained offending material, the service provider expeditiously removed the information or disabled access to it.

Regarding legislation under consideration, the United Kingdom's Online Safety Bill (UK Parliament, 2022[67]) was published in draft form in May 2021. It aims to improve online safety by placing new duties on online platforms to protect their users, requiring them to take action against both illegal content and content that is harmful to children. All services in scope of the legislation are required to prevent individuals from encountering priority illegal content. This includes offences relating to CSEA, including grooming, livestreaming and the sharing of child sexual abuse material. Additionally, providers who publish or place pornographic content on their services are required under the proposed legislation to prevent children from accessing that content. All companies would be required to have clear and accessible ways for users, including children, to report harmful content or challenge wrongful takedown. The designated regulator, Ofcom, will have the power to require platforms, where necessary and proportionate, to use accredited technology, or use best endeavours to develop or source technology, to identify CSEA material on public or private parts of a service. In-scope services will also be required to report all detected CSEA content to the United Kingdom's National Crime Agency, unless it is already being reported to a similar body outside the United Kingdom. Contributing further to this debate, the Report of the Independent Inquiry into Child Sexual Abuse recommended that the UK government makes it mandatory for all regulated providers of search services and user-to-user services to pre-screen for known child sexual abuse material (IICSA, 2022, p. 339[113]) and to require services to implement more stringent age verification measures (IICSA, 2022, p. 323[113]).

United States Federal law obliges interactive computer service (ICS) providers to report apparent violations of the statutes that involve child pornography to the CyberTipline operated by NCMEC (U.S. Congress, 1982[95]). NCMEC is required by federal law to make these provider reports available to law enforcement agencies. NCMEC in turn receives legal protection from any claims arising from the performance of its CyberTipline responsibilities and other actions, with certain exceptions. Currently, nothing in U.S. federal law requires providers to monitor their services or content for CSAM in the first instance. Under the law, although providers must report CSAM to NCMEC, which must then make the reports available to law enforcement, providers are not obligated to "affirmatively search, screen, or scan for" these violations. Nevertheless, many providers opt to voluntarily detect, remove, and report CSAM on their platforms.

Conclusion

Child sexual abuse and exploitation has been a persistent and challenging menace overshadowing the many benefits the wider digital environment affords. While CSEA predates the Internet and has long posed a threat to children's safety, technology-facilitated abuse has exacerbated the problem, making it easier to produce, store and distribute child sexual abuse material while also creating new opportunities for child sexual exploitation, such as through livestreaming. Despite many political calls to action, the proliferation of online CSEA is a societal challenge that is growing in scale and complexity. Policy frameworks such as the Model National Response (WeProtect Global Alliance, 2016[39]) argue that an effective response to online child sexual exploitation requires diverse cross sector inputs and multidisciplinary and multistakeholder co-operation. The focus of this report is on the part played by technology providers and online content-sharing services in mitigating risks of their platforms being misused and exploited for online CSEA. Specifically, this report has sought to establish a benchmarking baseline of transparency reporting practices of the top-50 online content-sharing services in how they define and respond to CSEA and make public data on the effectiveness of their efforts in tackling this deeply serious issue.

The report has found that there is a fragmented response by online content-sharing services to this complex and evolving problem. While no provider wishes to see illegal content or conduct on their service, the extent to which there is a public-facing policy setting out the principles, the specific definitions, and the way such policies are to be enforced varies considerably among the top global platforms. Many of the largest providers have developed robust policies in this regard and, as detailed in other sections of the report, have played an important role in developing technology solutions, supporting cross sector initiatives and building wider industry consensus in how to address the problem. But despite these good practices, the report finds transparency reporting is still very undeveloped and uneven across the sector. Sixty percent of the top global platforms do not issue a transparency report of any kind on how they tackle CSEA. Among those that do, data is presented in inconsistent and uneven ways. Information on content moderation strategies both within in TRs and in other governance documentation made available by platforms also suggests uneven and inconsistent practices and makes a wider assessment of the impact of service providers' efforts very difficult.

As detailed in Section 4 of this report, legislation and policy to tackle online CSEA continues to evolve. A greater emphasis is now placed on the obligations of service providers to demonstrate how they manage the risks of their services being exploited for purposes of CSEA. Transparency reporting will play a vital role in this process and will help to underpin collective action on the part of industry in sharing expertise, knowledge and meaningful data on measurable progress in combatting online CSEA. The purpose of this benchmarking study is to make available a snapshot of the current situation and to provide a baseline account how the top-50 global online content-sharing services combat online CSEA. Building on the OECD's existing expertise on children in the digital environment and on the transparency reporting practices of online content-sharing services, the objective going forward is to provide a foundation for monitoring progress in how industry addresses this challenge.

Annex A. Global top-50 most popular online content-sharing services

Rank	Name of service (parent company)	Monthly active users (*a*) or unique visitors (*b*) (millions)	Type of service	Issues CSEA transparency reports	Provided feedback / comments on its profile
1	Facebook (Meta Platforms, Inc.)	2,853*(a)* (as of July 2021) (Datareportal, 2021[114])	Social networking platform	Y	Y
2	YouTube (Alphabet, Inc.)	2,291*(a)* (as of July 2021) (Datareportal, 2021[114])	Video streaming platform	Y	Y
3	Zoom (Zoom Video Communications, Inc.)	2,100*(b)* (as of June 2021) (Statista, 2021[115])	Video chat and voice calls app	Y	Y
4	WhatsApp (Meta Platforms, Inc.)	2,000*(a)* (as of July 2021) (Datareportal, 2021[114])	Messaging app	N	Y
5	iMessage/FaceTime (Apple, Inc)	1,650*(a)* (as of January 2021) (Kastrenakes, 2021[116]))	Messaging and video chat apps	N	N
6	Instagram (Meta Platforms, Inc.)	1,386*(a)* (as of July 2021) (Datareportal, 2021[114])	Social networking platform	Y	Y
7	Facebook Messenger (Meta Platforms, Inc.)	1,300*(a)* (as of July 2021) (Datareportal, 2021[114])	Messaging app	Y (included in Facebook's)	Y
8	Weixin/WeChat (Tencent Holdings Ltd.)	1,242*(a)* (as of July 2021) (Datareportal, 2021[114])	Social networking/content-sharing/messaging platform	N	N
9	Viber (Rakuten, Inc.)	820*(a)* (as of January 2021)	Messaging app	N	Y

		(99 Firms, 2021[117])			
10	Tik Tok (ByteDance Technology Co.)	732*(a)* (as of July 2021) (Datareportal, 2021[114])	Short video app	Y	N
11	QQ (Tencent Holdings Ltd.)	606*(a)* (as of July 2021) (Datareportal, 2021[114]) as of July 2021	Instant messaging and web portal site	N	N
12	Youku Tudou (Alibaba Group Holding Limited)	600*(a)* (as of January 2021) (V-click Technology, 2021[118])	Video streaming platform (user-generated and syndicated content)	N	N
13	Telegram (Telegram Messenger LLP)	550*(a)* (as of July 2021) (Datareportal, 2021[114])	Messaging app	N	N
14	QZone (Tencent Holdings Ltd.)	548*(a)* (as of January 2021) (Warner, 2021[119])	Social networking platform	N	N
15	Weibo (Sina Corp.)	530*(a)* (as of July 2021) (Datareportal, 2021[114])	Social networking platform	N	N
16	Snapchat (Snap, Inc.)	514*(a)* (as of July 2021) (Datareportal, 2021[114])	Social networking platform	Y	Y
17	Kuaishou (Beijing Kuaishou Technology Co., Ltd)	481*(a)* (as of July 2021) (Datareportal, 2021[114])	Short video app	N	N
18	iQIYI (Baidu, Inc.)	480*(a)* (as of January 2020 (Statista, 2021[120])	Video streaming platform (user-generated and syndicated content)	N	N
19	Pinterest (Pinterest, Inc.)	478*(a)* (as of July 2021) (Datareportal, 2021[114])	Social networking platform	Y	N
20	Reddit (Reddit, Inc.)	430*(a)* (as of July 2021) (Datareportal, 2021[114])	Social news aggregation, web content ranking and discussion website	Y	N
21	Twitter (Twitter, Inc.)	397*(a)* (as of July 2021) (Datareportal, 2021[114])	Short messages-focused social networking platform	Y	Y
22	Tumblr (Automattic, Inc.)	327*(a)* (as of January 2021) (Finances Online, 2021[121])	Microblogging and social networking platform	N	Y
23	LinkedIn (Microsoft, Inc.)	310*(a)* (as of January 2021) (99 Firms, 2021[122])	Jobs-focused social networking platform	Y	Y
24	Douban (Information	300*(a)* (as of	Social networking	N	N

	Technology Company, Inc.)	July 2021) (Marketing to China, 2021[123])	platform		
25	Baidu Tieba (Baidu, Inc.)	300*(a)* (as of January 2021) (Marketing to China, 2021[124])	Online communications platform	N	N
26	Quora (Quora, Inc.)	300*(a)* (as of July 2021) (Datareportal, 2021[114])	Question-and-answer website	N	N
27	Teams (Microsoft, Inc.)	250*(a)* (as of July 2021) (Techcircle, 2021[125])	Online collaboration platform	Y	Y
28	IMO (PageBites, Inc.)	212*(a)* (as of January 2020) (Smith, 2021[126])	Video chat and voice calls app	N	N
29	Ask.fm (IAC [InterActiveCorp])	180*(b)* (as of January 2021) (Ask.fm, 2021[127])	Social networking platform	Y	Y
30	Vimeo (Vimeo, Inc.)	170*(a)* (August 2020) (Startup Talky, 2020[128])	Video streaming app	N	Y
31	Medium (A Medium Corporation.)	170*(b)* (as of January 2021) (Willens, 2021[129])	Online publishing platform	N	N
32	LINE (Line Corporation)	167*(a)* (as of July 2020) (Line Corporation, 2020[130])	Messaging app	Y	N
33	Picsart (Picsart, Inc.)	150*(a)* (as of February 2020) (Sensor Tower, 2020[131])	Photo and video app	N	Y
34	Discord (Discord, Inc.)	150*(a)* (as of January 2021) (Discord, 2021[132])	Chat platform	Y	Y
35	Twitch (Amazon.com, Inc.)	140*(a)* (as of January 2021) (Dean, 2021[133])	Livestreaming platform	Y	Y
36	Likee (BIGO Technology PTE. LTD.)	115*(a)* (as of January 2021) (JOYY Inc., 2021[134])	Streaming platform	N	Y
37	Skype (Microsoft, Inc.)	100*(a)* (as of March 2020) (Lardinois, 2020[135])	Video chat and voice calls app	Y	Y
38	VK (Mail.Ru Group)	97*(a)* (as of January 2021) (Mail.ru, 2021[136])	Social networking platform	N	Y
39	Xigua Video (ByteDance Technology Co.)	85*(a)* (as of May 2021) (Statista, 2021[137])	Short video streaming app	N	N

40	Odnoklassniki (Mail.Ru Group)	71*(b)* (as of January 2021) (Mail.ru, 2021[136])	Social networking platform	N	N
41	Flickr (SmugMug, Inc.)	60*(b)* (as of January 2021) (Flickr, 2021[138])	Image and video hosting service	N	Y
42	Huoshan (ByteDance Technology Co.)	59*(a)* (as of May 2021) (Statista, 2021[137])	Short video streaming app	N	N
43	KaKao Talk (Daum Kakao Corporation)	52*(a)* (as of December) (Statista, 2021[139])	Messaging app	N	N
44	Smule (Smule, Inc.)	50*(a)* (as of September 2020) (Audiens, 2020[140])	User-generated music-video sharing platform	N	N
45	Deviantart (DeviantArt, Inc.)	45*(b)* (as of January 2021) (DeviantArt, 2021[141])	Online artwork, videography and photography platform	N	N

Monthly active user (MAU) data are unavailable for certain other online content-sharing services, yet the metrics that are available suggest that they should be included in the top-50 list. The table therefore continues below with five more services, but without ranks because metrics other than MAU indicate their significance, so a proper comparison with the services above was not possible. In any event, for purposes of this report, the overall composition of the group of 50 is more important than the individual rankings.

Name of service (parent company)	**Indicative Global Market Share**	**Type of market/service**	**Transparency report on CSEA**	**Provided feedback / comments on its profile**
Google Drive (Alphabet, Inc.)	35.69% (as of January 2021) (Datanyze, 2021[142])	Cloud-based file sharing	Y	Y
Dropbox (Dropbox, Inc.)	20.42% (as January 2021) (Datanyze, 2021[142])	Cloud-based file sharing	Y	Y
Microsoft OneDrive (Microsoft, Inc.)	13.66% (as of January 2021) (Datanyze, 2021[142])	Cloud-based file sharing	Y	Y

Name of service (parent company)	**Indicative Global Market Share or monthly average unique devices (millions)**	**Type of market/service**	**Transparency report on CSEA**	**Provided feedback / comments on its profile**
Wordpress.com (Automattic, Inc.)	62% (as of January 2021) (Envisage Digital, 2021[143])	Content management system	N	Y
Wikipedia (Wikimedia Foundation)	2,000 (as of September 2021) (Wikimedia, 2021[144])	Online encyclopaedia	N	Y

Annex B. Profiles of the top-50 services

1. Facebook (Meta Platforms, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	A definition of CSEA is provided in Facebook's Community Standards under the policy on "Child Sexual Exploitation, Abuse and Nudity". Facebook defines a violation as any content or activity that sexually exploits or endangers children. The policy expressly prohibits the posting of content or activity "*that threatens, depicts, praises, supports, provides instructions for, makes statements of intent, admits participation in or shares links of the sexual exploitation of children (real or non-real minors, toddlers or babies)*" (Meta, n.d.[145]) The policy outlines a list of violative content, including but not limited to: • Explicit sexual content of any kind that depicts sexual intercourse. • Depictions of children with sexual elements. • Content of children in a sexual fetish context. • Content that supports, promotes, advocates or encourages participation in pedophilia. • Content that identifies or mocks alleged victims of child sexual exploitation. Solicitation is defined as content that solicits child sexual abuse material (CSAM), or other nude or sexualised imagery of children, and/or seeks real world sexual encounters with children. Inappropriate interactions with children come within the scope of solicitation under this policy, which may include exposing children to sexual material or engaging in implicitly sexual conversations with children in private messaging. Sextortion is encompassed in the policy, prohibiting threats to exploit minors through exposing intimate imagery or information with the intent of coercing money, favours or intimate imagery. Threatening to or stating an intent to share private sexual conversations or

	intimate imagery, or actually sharing such content violates exploitative Meta's Child Sexual Exploitation, Abuse and Nudity policy (Meta, n.d.[145]). Sexualisation of children more generally is taken to include activity to solicit imagery of children that is sexually exploitative such as depictions that shows children in a sexualised context or groups or pages targeted towards sexualising children. A prohibition on posting content depicting child nudity applies more generally and is included within the policy given the potential risk of such content being exploited. In cases where a professional news agency posts imagery depicting child nudity, for example, in the context of famine, genocide, war crimes, or crimes against humanity, a warning label may be applied. Finally, Facebook's ToS set out at a high level the obligations and responsibilities of users to not share or do anything that is unlawful, misleading, discriminatory or fraudulent, or that breaches the platforms Terms, Community Standards or other policies.
2. How are the ToS or Community Guidelines/Standards communicated?	Facebook Terms of Service are available at: Facebook Terms and Policies The policy on Child Sexual Exploitation, Abuse and Nudity is available at: https://transparency.fb.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Facebook enforces its policies and Community Standards through a combined process using technology, human review teams and user reporting to prevent misuse of its platform. Facebook describes its approach to content enforcement as a three-part strategy to "*remove, reduce and inform*" in respect of content or conduct that may be in breach of its policies. According to the company, Facebook implements this strategy "*in order to detect, review and take action on millions of pieces of content every day on its platforms*" (Meta, n.d.[146]). The key elements of the approach include: • *Removal*: content removal is enforced for any content that goes against Facebook's Community Standards, including pages or communities that are found to repeatedly violate its policies. Where a piece of content is removed, a strike is applied against the account holder who posted the content. Repeated violations despite warnings may lead to account restrictions, disabling or suspension. • *Reducing*: limiting the distribution of problematic content is a further element of the strategy. This refers to content that may not violate the Community Standards but may still be problematic. Accordingly, Facebook will avoid recommending content that may be low-quality, objectionable, sensitive or inappropriate for younger viewers

	to limit its visibilty. Warning screens may be applied for potentially sensitive content. • *Informing*: this involves providing context on content that may be sensitive or misleading – even if it doesn't explicitly violate Facebook's Community Standards. A warning screen may, for example, be applied over potentially sensitive content.

4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	As stated in Facebook's ToS, a range of enforcement actions may be applied which include the removal of content, blocking access to certain features, disabling an account, or contacting law enforcement. Relevant actions include: • *Removing violating content*: if content is found to be in breach of the Community Guidelines, it will be removed and users will be informed as to the reason for its removal. A strike system is used to count violations. Depending on the nature of the violation and the number of strikes a user has accumulated, an account may be restricted or disabled. Facebook will also remove pages or groups that violate its Community Guidelines. • *Restricting accounts*: following a series of strikes or repeated warnings, accounts may be restricted from creating content, such as posting, commenting, using Facebook Live or creating a Page. Restrictions may be just for one day for an initial violation or for longer for repeated offences. However, accounts found to violate CSEA policies will be automatically disabled and the process of counting strikes will not apply. • *Disabling accounts*: where an account holder continues to post violative content despite warnings, Facebook will disable the account. In addition, an account may be permanently disabled, depending on the severity and frequency of the violations. In the case of posting child sexual exploitation content, the account will be immediately disabled. Accounts of dangerous individuals, convicted sex offenders or accounts created to get around its restrictions are also immediately disabled. When the platform is made aware of apparent CSEA, it is also reported to NCMEC as required under U.S. law. In addition, according to Meta, it has expanded its work to detect and remove users or groups, including Facebook profiles, Pages, groups and Instagram accounts, that violate its policies, and has also made it easier to report violating content. For example, when Meta identifies potentially suspicious adults on Facebook and Instagram, it works to prevent them from discovering and connecting with accounts of minors (Meta, 2021[147]). This intervention can take a number of different forms including: • Not recommending young people's accounts to them by removing them from "People You May Know." • Not allowing them to comment on young people's posts, and not allowing them to see young people's comments on other people's posts. • Not providing an option to friend or follow a young person's account after searching for their username.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes. According to Facebook, if content is found to go against its Community Standards, it will be removed and the account holder notified. The account may also be restricted or disabled and the user

	notified depending on the nature of the violation and the history of previous violations.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, removals or other content enforcement decisions may be appealed. In the first instance, users whose content has been removed following a content moderation decision may appeal by submitting a request for a review. According to Facebook, the decision will be reviewed again by its moderation team, normally within 24 hours. If the complainant is still not satisfied with the outcome of the review, they may appeal the decision to the Oversight Board established by Meta. However, Facebook states that the board only selects a certain number of appeals to review and not all decisions are eligible for this appeal.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, Meta produces a quarterly Community Standards Enforcement Report for both Facebook and Instagram. Reports are published in the Meta Transparency Center (Meta, n.d.[148]). Enforcement decisions are summarised under 11 main categories, including the category of "Child Endangerment: Nudity and Physical Abuse and Sexual Exploitation". As this combines a number of separate policy violations, the sub-heading of "Child Endangerment: Sexual Exploitation and Child Nudity and Sexual Exploitation" contains the relevant data on Facebook's handling of CSEA.
5.1 What information/fields of data are included in the TRs?	Data is reported under the following headings: • Content Actioned • Content Appealed • Content Restored with appeal • Proactive rate • Content Restored without appeal
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	• *Content actioned* is defined as the total number of content items that Facebook takes action on for child nudity and physical abuse; child sexual exploitation; and child nudity and sexual exploitation. Content actioned refers to the discrete number of pieces of content where an action has been taken for violating its Community Guidelines. The metric represents the overall scale of enforcement activity. Where a piece of content has been found to violate multiple standards, an enforcement action is attributed to one primary violation, typically involving the most severe standard. How this is ranked is not disclosed by Meta. • *Proactive Rate* shows the percentage of all content or accounts acted on that were found and flagged before users reported them to the platform. This metric is used by Facebook as an indicator of how effectively it detects violations. • *Appealed Content* refers to the number of pieces of content (such as posts, photos, videos or comments) that people appeal after an action has been for going against Facebook

	policies. • *Restored Content* refers to the total number of content items restored either from a successful appeal decision or without appeal (for example when an error in the original moderation decision was discovered).
5.3 Frequency/timing with which TRs are issued	The Community Standards Enforcement Report is released on a quarterly basis.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Facebook uses a combination of user reporting, proactive detection techologies and human staff review to detect and combat child sexual exploitation and abuse material on its platform. Among the detection technologies used are photo and video-matching technologies to help detect, remove, and report the sharing of images and videos that exploit children. These technologies create a unique digital signature of an image (or "hash") which is then compared against a database containing signatures (hashes) of previously identified illegal images to find copies of the same image. These technologies are used across Facebook public surfaces, as well as on non-encrypted information available on its private messaging services, including profile pictures and user reports. Hash matching techologies are also used to detect links to other Internet sites shared on its apps sharing known CSEA. In addition to photo- and video-matching technology to detect known images, Facebook uses artificial intelligence and machine learning to proactively detect child nudity and previously unknown and new child-exploitative content (Meta, 2018[58]) on public and non-encrypted surfaces within its apps and services. According to Meta, technology is employed to detect and prevent possible child exploitation such as inappropriate interactions between adults and children on its private messaging services (Meta, 2021[147]). Techniques developed by Meta use a combination of information from its public platforms and critical traffic data to detect if an adult is attempting to contact children. This is used to inform a decision on possible intervention such as to restrict all or some of the functions a user has available to them. Meta reports that in 2022, more than 40,000 people worked across safety and security worldwide at the company (Meta, 2022[149]). This includes specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigations who review potentially violating content and report apparent child sexual exploitation to NCMEC.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. As reported in the Facebook Community Standards Enforcement Report, in Q2, 2022, 20.4 million pieces of content were actioned for "child endangerment: sexual exploitation" (Meta, n.d.[148]). This represented an increase on the 16.5 million pieces of content actioned in Q2, 2021. According to Facebook, this was due

	to an increase in enforcement on viral content and old, violating content detected by its media-matching. The proactive rate, according to the report for this period, was 99.1% where content was detected and removed before anyone saw it. NCMEC also reports that in 2022, Facebook notified 21,165,208 cases of CSEA to the Centre (NCMEC, 2023[18]). This compares to 22,118,952 reports submitted in 2021, the largest number of reports ever reported by an electonic service provider to NCMEC's CyberTipline reporting service (NCMEC, 2022[150]).

2. YouTube (Alphabet, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	YouTube's Community Guidelines outlines its approach to CSEA which contains a general prohibition on content that features the sexualisation of children. CSEA is defined as "*sexually explicit content featuring minors and content that sexually exploits minors*". This forms part of a wider prohibition of explicit or pornographic content of any kind on the platform. Any content found to contain child sexual abuse imagery is reported to NCMEC. YouTube's Child Safety Policy states that the platform has a zero-tolerance policy for predatory behaviour on the platform (YouTube, n.d.[151]). This includes, for example, videos that feature minors involved in provocative, sexual, or sexually suggestive activities, challenges and dares, such as kissing or groping. YouTube policies on CSAM and CSEA form part of Google's wider strategy towards combatting child sexual abuse and exploitation on its services. The strategy is documented in the Protecting Children Google microsite (Google, n.d.[57])
2. How are the ToS or Community Guidelines/Standards communicated?	YouTube Terms of Service are available at: Terms of Service (youtube.com) YouTube Community Guidelines are available at: https://www.youtube.com/intl/ALL_ie/howyoutubeworks/policies/community-guidelines/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	YouTube uses a combination of automated detection technology, community moderation and human review to prevent its services being exploited for child sexual abuse. The YouTube ToS state that while the user remains legally responsible for content they submit to the service, the platform may use automated systems that analyse the content to help detect infringement and abuse, such as spam, malware, and illegal content (described further in Section 6 below).
4. What are the service's policies and procedures for enforcing its ToS or Community	A strikes policy applies for violations of YouTube's policies. For a first offence, a user will receive a warning with no penalty to the channel. After one warning, a Community Guidelines strike will be added to

Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	the channel and the account will have temporary restrictions applied including not being allowed to upload videos, live streams, or stories for a 1-week period. Channels that receive three strikes within a 90-day period will be terminated. Channels that are dedicated to violating YouTube policies or that have a single case of severe abuse of the platform, will bypass the strikes system and be terminated.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, if content is removed for violation of YouTube Community Guidelines, a notice is sent to the account holder. For a first time violation, a creator or account holder will receive a warning with no penalty to the channel. After one warning, users are notified and a Community Guidelines strike will be applied. Additional restrictions as outlined above may be placed on the channel.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Enforcement decisions such as strikes, content removal and account terminations can be appealed. An appeal is initiated in the user's dashboard in the YouTube app. Appeals may be lodged up to 30 days after the warning or strike was first issued and each strike may only be appealed once. An appeal request will result in one of the following outcomes: • If the content is found to have followed the Community Guidelines, it will be reinstated and the strike removed from the channel. Where an appeal is granted, the next offence will be a warning. • If the content is found to have followed the Community Guidelines, but isn't appropriate for all audiences, an age-restriction will be applied. If it's a video, it won't be visible to users who are signed out, are under 18 years of age, or have Restricted Mode turned on. If it's a custom thumbnail, it will be removed. • If the content is found to have been in violation of the Community Guidelines, the strike will stay and the video will remain down from the site. There is no additional penalty for appeals that are rejected.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, YouTube issues a Transparency Report which is published on the Google Transparency Report site (Google, n.d.[152]) In 2021, Google launched a dedicated Transparency Report on efforts to combat online child sexual abuse material across Google services including YouTube. The report also provides data on the number of accounts disabled for violations of its child safety policies, including CSAM violations across its services (Google, n.d.[153]).
5.1 What information/fields of data are included in the TRs?	The following fields of information are included in the YouTube Transparency Report: • channels removed, by number • channels removed, by removal reason • videos removed, by number • videos removed, by source of first detection • videos removed, by views • videos removed, by removal reason

	• videos removed, by country/region • comments removed, by number • comments removed, by source of first detection • comments removed, by removal reason Data for each may be searched to include or exclude automated flagging. The Google CSAM Transparency report is published twice yearly rather than quarterly and reports on a subset of categories. The following fields of data are included: • cyberTipline reports to NCMEC • total pieces of content reported to NCMEC • accounts disabled for CSAM violations • URLs de-indexed for CSAM from Google Search • CSAM hashes contributed to the NCMEC database
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	• *Removed channel by number:* A YouTube channel is terminated if it accrues three Community Guidelines strikes in 90 days, has a single case of severe abuse (such as predatory behavior), or is determined to be wholly dedicated to violating our guidelines (as is often the case with spam accounts). When a channel is terminated, all of its videos are removed. • *Channels removed, by removal reason:* Data shows the volume of channels removed by YouTube, by the reason a channel was removed. The majority of channel terminations are a result of accounts being dedicated to spam or adult sexual content. Child safety accounted for just 1.4% in the reporting period Q1 2022. • *Removed videos by number:* shows the number of videos removed by YouTube for violating its Community Guidelines per quarter. Data can be reported either including or excluding automated flagging. • *Videos removed, by source of first detection:* This shows the volume of videos removed by YouTube, by source of first detection (automated flagging or human detection). Flags from human detection can come from a user or a member of YouTube's Trusted Flagger program. • *Videos removed, by views:* This shows the percentage of video removals that occurred before they received any views versus those that occurred after receiving some views. • *Videos removed, by removal reason:* This shows the volume of videos removed by YouTube, by the reason a video was removed. These removal reasons correspond to YouTube's

	Community Guidelines. The category of Child Safety therefore is broader than CSAM violations and may include other forms of child endangerment such as dangerous acts involving children, cyberbullying and harassment. The Google CSAM report allows for a more detailed review of data regarding CSAM on the platform. Its key reporting categories are determined as follows: • *CyberTipline reports to NCMEC:* When CSAM is identified on the platform, a "CyberTipline" report is sent to NCMEC. This may include information identifying the user, the minor victim, and/or other helpful contextual facts. It may be the case that more than one report is sent on a particular user or piece of content — for example, in cases where content is identified from multiple sources. • *Total pieces of content reported to NCMEC:* A single report may contain one or more pieces of content depending on the circumstances. This content could include, for example, images, videos, URL links, and/or text soliciting CSAM. A single piece of content may be identified in more than one account or on more than one occasion, so this metric may include pieces of content reported more than once. • *Accounts disabled for CSAM violations:* If CSAM is identified in a user's account, a CyberTipline report to NCMEC and account may be disabled. Users are notified of the account termination and are given the opportunity to appeal. • *URLs de-indexed for CSAM from Google Search:* This metric represents the number of URLs removed from the Search index. Google does not control over the content on third-party web pages. When CSAM is identified on third-party web pages, it is de-indexed and the URL removed from Search results. The content from the third-party page remains online. • *CSAM hashes contributed to the NCMEC database:* when a new item of CSAM is identified, a hash of the content is created and added to Google's internal repository. This is also shared with NCMEC. This metric represents the cumulative number of hashes Google has contributed to this effort.
5.3 Frequency/timing with which TRs are issued	Google Transparency reports are published on a quarterly basis. The Google CSAM report is published at six monthly intervals.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the	A combination of community reporting, technology detection and human staff review is used to detect CSEA. Automated technology detection is used to identify and remove potentially violative content, and to remove re-uploads of content

service use to detect CSEA?	previously reviewed and determined to be in breach of YouTube policies. Technologies include machine learning classifiers and hash-matching technology, which creates a "hash", or unique digital fingerprint, for an image or a video so it can be compared with hashes of known CSAM. When CSAM is found, it is reported to NCMEC. Content flagged by the YouTube user community is further assessed by trained human staff reviewers who may remove content that violates platform policies. Content that is not in violation but may be appropriate for all audiences can also be age-restricted. Reviewers' inputs are then used to train and improve the accuracy of AI systems at a much larger scale. YouTube also operates a Trusted Flagger program in conjunction with government agencies and non-governmental organisations (NGOs). The programme includes a dedicated reporting channel that government agencies and NGOs can use to contact YouTube directly. Trusted Flagger reports are prioritised for review because of their high degree of accuracy though are not subject to any differential policy treatment.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes, as reported in YouTube's Transparency Report, in Q2 2022, approximately 1.38m videos were removed for reasons of child safety, representing just over 30% of all removals. This is a composite category however and includes violations other than CSAM (Google, n.d.[152]). According to NCMEC, a total of 2,174,548 reports were submitted by Google in 2022 (including all Google products and YouTube) for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 875,783 reports submitted in 2021 (NCMEC, 2022[150]).

3. Zoom (Zoom Video Communications, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Zoom defines child exploitation material in its Acceptable Use Guidelines as "*any content that depicts or promotes sexual abuse or activity involving children"*. Such content is not tolerated on the platform and accounts that violate the policy may be permanently suspended. The Sensitive Content Policy adds further detail and prohibits the following examples of child sexual exploitation material: • visual depictions of a child engaging in sexually explicit or sexually suggestive acts; • illustrated, computer-generated or other forms of realistic depictions of a human child in a sexually explicit context, or engaging in sexually explicit acts;

	• sexualised commentaries about or directed at a known or unknown minor; • links to third-party sites that host child sexual exploitation material; • sharing fantasies about or promoting engagement in child sexual exploitation; • expressing a desire to obtain materials that feature child sexual exploitation; • 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; • sending sexually explicit media to a child; • engaging or trying to engage a child in a sexually explicit conversation; • trying to obtain sexually explicit media from a child or trying to engage a child in sexual activity through blackmail or other incentives; and • identifying alleged victims of childhood sexual exploitation by name or image. Furthermore, Zoom's Terms of Service (ToS) prohibits use of the platform for any activity that is illegal, fraudulent, false, or misleading, or that uses the services to communicate any message or material that is harassing, libelous, threatening, obscene, or indecent. Zoom states that its overall policy is to provide users with an experience that is open and diverse, and free from harmful or malicious activity. In line with this approach, it sets out a zero-tolerance policy around child sexual abuse and exploitation material and underlines its commitment to continue to implement new strategies to combat it.
2. How are the ToS or Community Guidelines/Standards communicated?	The Zoom Terms of Service are available at: Zoom Terms of Service | Zoom Zoom's Acceptable Use Guidelines are available at: https://explore.zoom.us/en/acceptable-use-guidelines/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Zoom's general approach to keeping its platform safe and free of child sexual abuse is to rely on reports and public-facing information to discover potential violations while also deploying automated content moderation technologies to detect potential violations of its policies including the dissemination of CSAM (Zoom, n.d.[154]). Zoom's developed its Acceptable Use Guidelines (formerly known as "Community Standards") first in October 2020 following the rapid growth in its customer base with the onset of the global COVID-19 pandemic. The Guidelines describe the types of content and behavior that are prohibited on the platform and encourage users to report any violations using a dedicated report form.
4. What are the service's policies and procedures for enforcing its ToS or Community	Under its ToS, Zoom may investigate any complaints and violations that come to its attention and may take any (or no) action that it believes is appropriate, including, but not limited to issuing warnings,

Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	removing the content or terminating accounts and/or user profiles. When potentially violative activity is detected, actions that may be taken include: • *Event(s) Suspended:* whereby Zoom will end or prevent a particular event from taking place. • *OnZoom/Zoom Events Host(s) Suspended:* Zoom may block one or more hosts of OnZoom or Zoom Events. • *Strike Issued:* The user will receive a strike for a violation. Strikes expire after 180 days and do not affect the user's ability to use the platform unless they accumulate. Depending on the reason for the strike, either one or two additional strikes within the same 180 day period will result in a suspension against the user. • *User(s) Suspended:* The user may be deactivated and/or blocked. They are prohibited from using Zoom unless they successfully appeal the decision.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, where a breach of its Acceptable Use Guidelines is found to have taken place, users are notified by email of the action taken against the account and the reason for the enforcement decision.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may submit a request for a review using the appeal request form. Details are provided when action is taken against an account due to a violation of the Terms of Service or Acceptable Use Guidelines. However, appeals for certain issue types, such as those involving CSEA or references to terrorism and violent extremism, will not be granted.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, Zoom publishes a report on its Acceptable Use Guidelines Enforcement which it initiated in October 2020 (Zoom, n.d.[155]).
5.1 What information/fields of data are included in the TRs?	Data is presented for each of the categories of violation as defined in the Zoom Acceptable Use Guidelines. Fields of data are as follows: • Reporter Country • Issue Type (Cateogory of Violation) • Report Month • Number of reports actioned • Action type
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Individual items within the Acceptable Use Guidelines Enforcement Report are defined as follows: • *Dismissed:* No action was taken. • *Duplicate*: Two or more reports about the same issue from the same reporter. • *Event(s) Suspended:* Zoom ended or prevented a particular event from taking place. • *OnZoom/Zoom Events Host(s) Suspended:* Zoom blocked one or more hosts of OnZoom or Zoom Events. • *Strike Issued:* The user received a strike. Strikes expire after 180 days and do not affect the user's ability to use the platform unless they accumulate. Depending on the reason for the strike, either one or two additional strikes within the same 180-day period will

	result in a suspension against the user. • *User(s) Suspended:* The user was deactivated and/or blocked. They are prohibited from using Zoom unless they successfully appeal the decision.
5.3 Frequency/timing with which TRs are issued	The report on Acceptable Use Guidelines Enforcement is a cumulative report which is updated monthly since the publication of the first report in July 2021.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Zoom employs user community reporting, technoloogy and human staff review in monitoring for violations of its policies, including CSEA. According to Zoom, content moderation is staffed by a Trust and Safety team of staff reviewers supported by automated content moderation technologies to review reports of alleged violations of its policies. Zoom reports that it prioritises response to CSEA-related reports and has a dedicated NCMEC API that allows the platform to report these instances directly from its dashboard to NCMEC. Zoom's system of content moderation is organised in 4 discrete tiers (Zoom, n.d.[156]). Most reports go first to Tier I analysts. Tier I reviews reports flagged or submitted by people or automated tools for various categories of violations, including child sexual abuse material (CSAM), spam, violent extremist groups, and hateful conduct, among others. Tier I decisions that cannot be resolved quickly are escalated to Tier II or III, which also handle some subject matters in the first instance. Tier IV involves an Appeals Panel which reviews decisions about reports escalated from Tier III and deals with the most controversial or hard-to-classify matters. According to Zoom, it responds to user reports of potential violations by investigating and taking action as quickly as possible (Zoom, n.d.[154]). Where violations relate to child sexual exploitation material, Zoom makes a report to NCMEC or other entities as required by law. Where warranted, it may also make a report to law enforcement. Cases from unknown countries are the result of testing and developing hash-matching capabilities to scan certain customer content data, such as files uploaded to persistent chat, Zoom Room backgrounds, and profile pictures. These technologies work by comparing a hash value assigned to a piece of known CSAM against a hash of a file uploaded into the Zoom system. If the known CSAM hash matches the file upload, a report is generated for Zoom's Trust and Safety team to conduct a human review. From February 2022 – January 2023, 99.75% of these reports were dismissed after human review as false positives. Zoom reports that it is currently working with a range of stakeholders to improve the true positive rate of its CSAM detection capabilities.
7. Has this service been used to disseminate, store, or produce	Yes, as reported in its Acceptable Use Guidelines Enforcement Report, for the 12-month period February 2022 to January 2023,

CSEA, or to solicit children for sexual purposes?	11,263 reports of child sexual exploitation were processed (10,676 were from an unknown country and 99.75% of these cases were dismissed*). Of cases from known countries (560 cases), 31.6% were dismissed. In 50.53% or 283 cases from known countries, users were suspended from the service. A strike was issued in 17.14% or 96 cases from known countries. According to NCMEC, a total of 136 reports were submitted by Zoom in 2022 for the online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 548 reports submitted in 2021 (NCMEC, 2022[150]).

4. WhatsApp (Meta Platforms, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	The WhatsApp Terms of Service do not provide a specific definition of CSEA. However, under "Acceptable Use of Our Services", users are expressly prohibited from engaging in any activity that may be "*illegal, obscene, defamatory, threatening, intimidating. … or encourage conduct that would be illegal or otherwise inappropriate, such as endangering or exploiting children*". WhatsApp states that it has zero tolerance for child sexual exploitation and abuse, and that it bans users when it becomes aware they are sharing content that exploits or endangers children (WhatsApp, n.d.[63]).
2. How are the ToS or Community Guidelines/Standards communicated?	The WhatsApp Terms of Service are available at: Terms of Service (whatsapp.com)
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	WhatsApp ToS state that it works to prohibit misuse of its services including harmful conduct towards others or breaching its terms. Where it learns of people or activity violating its policies, WhatsApp states that it will take appropriate action, including by removing such people or activity or contacting law enforcement. As an end-to-end encrypted messaging service, the platform does not and cannot routinely scan messages for violative content. Instead, WhatsApp states that it seeks to prevent abuse or misuse happening in the first place using a combination of proactive detection and design features to mitigate risks of its services being misused (see Section 6 below).
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	According to WhatsApp ToS, action may be taken against a user's account for any violation of its terms or policies. Enforcement actions range from disabling, suspending or terminating an account. Where an account is suspended or terminated, the user is prohibited from creating another account without the company's permission.

4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, where an account is subject to an enforcement action, the user will receive the following message: "*This account is not allowed to use WhatsApp. Accounts are banned for violations of the Terms of Service, for example if it involves spam, scams or if it puts WhatsApp users' safety at risk*".
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, if a user believes their account was suspended in error, they can submit a request for a review and add details to support their case. Users are then contacted directly following a review of the decision.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No. To date, WhatsApp has only issued a Transparency Report for the Indian market to comply with the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 (PRS Legislative Research, 2021[157])
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	WhatsApp uses a combination of human staff review and automated content moderation to combat child sexual exploitation and to detect and prevent the distribution of child sexual abuse material. This applies only to unencrypted information and metadata that may point to suspicious content. Detection methods include the use of advanced automated technology, including photo- and video-matching technology, to proactively scan unencrypted information such as profile and group photos and user reports for known child exploitation imagery. WhatsApp also deploys its own technology to detect new, unknown CSEA material within this unencrypted information including machine learning classifiers to both scan text surfaces, such as user profiles and group descriptions, and evaluate group information and behavior for suspected sharing of CSEA material. In addition to proactive detection, WhatsApp encourages users to report problematic content to the platform. Safety by design techniques are employed to mitigate risks. For example, a user cannot search for people they do not know on WhatsApp. A phone number is needed to connect with a new contact. When a user first receives a message from someone outside of their address book, they are asked if they wish to block or report them. The number of chats to which a user can forward a message to at once to help is also limited to help prevent the spread of harmful viral content. WhatsApp also states that it works with app store providers to

	prevent the proliferation of apps that contain CSEA or that attempt to connect people interested in sharing such content via group invite links. WhatsApp also restricts the listing of invite links by popular search engines.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to WhatsApp, on average 300,000 accounts are banned each month for sharing child exploitation imagery (WhatsApp, n.d.[63]). In 2022, WhasApp made 1,017,555 reports of CSEA to the NCMEC CyberTipline (NCMEC, 2023[18]). This compares to 1,372,696 reports submitted in 2021 (NCMEC, 2022[150]).

5. iMessage/FaceTime (Apple, Inc)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No specific definition of CSEA is given for iMessage. However, the Apple Media Services Terms and Conditions which apply to use of all Apple Internet services, prohibit the posting or submission of any objectionable, offensive, unlawful, deceptive, inaccurate, or harmful content. It is also prohibited for users to request personal information from a minor, or to plan or engage in any illegal, fraudulent, or manipulative activity.
2. How are the ToS or Community Guidelines/Standards communicated?	The Apple Media Services Terms and Conditions (ToS) are published online on the Apple website and are available at: Legal – Apple Media Services – Apple Additional legal resources and contact information are available at: https://www.apple.com/ie/legal/more-resources/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	No details are provided of Apple's approach to monitoring for potential breaches of its policies. However, as an end-to-end encrypted instant messaging service, Apple does not have access to message content or to photos, videos, or other file attachments which are also encrypted. A number of additional safety features focused on child protection have also been developed by Apple (see Section 6 below).
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	The ToS state that Apple may, at its sole discretion, suspend, disable and/or terminate the accounts of users who have been identified as repeatedly engaging in infringing activities or for other related reasons. The Apple Media Services Terms and Conditions do not specify any particular procedures in this regard. Reference is made under its Submission Guidelines that Apple may monitor and decide to remove or edit any submitted material.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No notification procedures are specified.
4.2. Are there processes under which users can appeal content	No appeal processes are specified.

removals, account suspensions or other enforcement decisions?	
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No, Apple's Transparency Report only gives details of government information requests for customer data (Apple, n.d.[158]). It does not give any information about content or activity related to CSEA or about its enforcement of its Terms and Conditions.
5.1 What information/fields of data are included in the TRs?	No relevant information is given.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable as no relevant information is given.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	In August 2021, Apple announced new measures to combat the spread of CSEA via its devices and services, and specifically uploading photos depicting CSEA to its iCloud service. The measures involved the planned deployment of NeuralHash, which is client-side software meant to allow hashing technologies like photoDNA to operate within an E2EE system. NeuralHash would enable images on a user's device to be compared against a database of known CSEA images provided by NCMEC. If an on-device match was found Apple would be notified. Apple would then manually review the report to confirm the match, disable the user's account, and send a report to NCMEC (Apple, 2021[159]). A range of privacy specialists criticised the plan, which was withdrawn (Simon, 2022[160]). See also: (Patel, 2021[161]) Also at this time, Apple introduced a communication safety feature in iMessage to enhance child safety (Apple, n.d.[162]). When activated, the Message app issues a notification if a child account (up to age 18) in a Family Sharing plan attempts to receive or send photos that contain nudity. This is an opt-in feature must be activated by the parent/guardian account holder and is currently available in Australia, Canada, New Zealand, the United Kingdom, and the United States.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. In 2022, Apple Inc reported 234 cases of online exploitation of children, including child sexual abuse material to NCMEC (NCMEC, 2023[18]). This compares to 160 reports submitted in 2021 (NCMEC, 2022[150]).

6. Instagram (Meta Platforms, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Facebook and Instagram share content policies. Content that is considered violating on Facebook is also considered violating on Instagram. The shared Child Sexual Exploitation policy sets out that Meta, as the parent company, does not allow content or activity that sexually exploits or endangers children. Child sexual exploitation is

	defined as "*content or activity that threatens, depicts, praises, supports, provides instructions for, makes statements of intent, admits participation in or shares links of the sexual exploitation of children (real or non-real minors, toddlers or babies)"* (Meta, n.d.[145]). See Section 1 of the Facebook profile for details of the applicable policy. Furthermore, Instagram's Terms of Use prohibit the use of the platform for anything unlawful, misleading, or fraudulent or for any illegal or unauthorised purpose. Instagram's Community Guidelines highlight that the platform has zero tolerance when it comes to sharing sexual content involving minors or threatening to post intimate images of others.
2. How are the ToS or Community Guidelines/Standards communicated?	Instragram's Terms of Use are available at: Terms of Use | Instagram Help Center Instragram's Community Guidelines are available at: https://help.instagram.com/477434105621119/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Instagram's approach is to employ a combination of user reporting, human content moderation and artificial intelligence technologies to review content that may violate its ToS or Community Guidelines. According to the Meta Transparency Centre, when content that contains nudity or physical abuse or content that sexually exploits children on Facebook or Instagram is detected, it is removed, regardless of the context or the person's motivation for sharing it. Instragram may also disable the account of the person who shared it, unless it appears the intent was not malicious (for example, to spread awareness of child exploitation) Apparent child exploitation is also reported to NCMEC (Meta, n.d.[145]). See also Section 3 of the Facebook profile for details of the applicable enforcement policy. Instagram offers a feature called Account Status which allows user to check if they have posted content in violation of the Community Guidelines, and if such posts may lead to their account being taken down (Instagram, n.d.[163]). For professional accounts Account Status allows them to see if they've recently or repeatedly posted content or have something on their profile (such as profile photo or bio) that violates the Recommendations Guidelines. Recommendations Guidelines help Instagram decide which public accounts' content may be eligible to be recommended in places such as Explore, Reels and Feed Recommendations to people who don't already follow them.
4. What are the service's policies and procedures for enforcing its ToS or Community	Instagram states that it may remove any content or information that violates the Terms of Use, its Community Guidelines or other lawful requirements. Enforcement actions include suspension, withdrawal

Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	of all or part of the service, including immediately terminating or disabling access if continued use poses a threat to the community or if it creates a risk of legal exposure for the company. The same detection technologies to identify suspicious accounts and prevent interactions with accounts of minors, as referenced in Section 4 of the Facebook profile, apply equally on Instagram.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, notification is provided in cases where content is removed. For serious violations, accounts may be disabled without warning (Instagram, n.d.[164]).
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. If a user believes their content should not have been taken down or their account was disabled in error, they may appeal the decision by following appeal procedures outlined in the Help Centre (Instagram, n.d.[165]) Appeals can be submitted by tapping the “Ask for a review” in the notification that is given for removal of content. Users can also request a review from Account Status or Support Inbox. Once a request has been submitted the content will be reviewed again by Instagram. Once a decision is made, the user receives a push notification. If the content was removed in errror, the user is informed and their content will be reposted. If the decision is confirmed and a user still does not agree, they may be able to appeal to the Oversight Board established by Meta though it is noted the board only selects a certain number of eligible appeals.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, Meta produces a quarterly Community Standards Enforcement Report for both Facebook and Instagram. Reports are published in the Meta Transparency Center (Meta, n.d.[148]). As detailed in Section 5 of the Facebook profile, enforcement decisions are summarised under 11 main categories, including the category of “Child Endangerment: Nudity and Physical Abuse and Sexual Exploitation”. As this combines a number of separate policy violations, the sub-heading of “Child Endangerment: Sexual Exploitation and Child Nudity and Sexual Exploitation” contains the relevant data on Facebook’s handling of CSEA.

5.1 What information/fields of data are included in the TRs?	See Section 5.1 of the Facebook Profile.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	See Section 5.2 of the Facebook Profile.
5.3 Frequency/timing with which TRs are issued	The Community Standards Enforcement Report is released on a quarterly basis.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Instagram uses a combination of proactive detection techologies and human review to detect and combat child sexual exploitation and abuse material on its platform. The Community Guidelines encourage users to report anything they see that may violate its guidelines. According to the Help Centre, artificial intelligence (AI) technology is central to Instagram's content review process and can detect and remove content that goes against its Community Guidelines before anyone reports it (Instagram, n.d.[166]). Other times, AI technology sends content to human review teams for analysis and decision. See Section 6 of the Facebook profile for details of the methods used to detect CSEA.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes, according to its TR, in the period October to December 2022, Meta took action against 9.7m pieces of content on Instagram for child sexual exploitation (Meta, n.d.[148]).Over 99% of content was proactively found and actioned by Meta with less than 1% reported by users. According to NCMEC, in 2022, Instagram reported 5,007,902 cases of CSEA to its CyberTipline (NCMEC, 2023[18]). This compares to 3,393,654 reports submitted in 2021 (NCMEC, 2022[150]).

7. Facebook Messenger (Meta Platforms, Inc)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	See also Profile #1 – Facebook. Messenger does not have its own Terms of Service or Community Guidelines. As part of the Meta suite of products, Messenger shares its Terms of Service with Facebook. It offers messaging, voice and video calling services within Meta products and accordingly combined Meta policies apply. Meta's Child Sexual Exploitation, Abuse and Nudity policy (Meta, n.d.[145]) is the applicable policy, the provisions of which are summarised in the Facebook profile.
2. How are the ToS or Community Guidelines/Standards communicated?	The "Legal and policies" tab within Messenger account settings links to Facebook Terms of Service and Community Standards and are available at: https://www.facebook.com/policies_center/

3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	See Section 3 of the Facebook profile.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Enforcement actions are as outlined in the Facebook ToS. See Section 4 of the Facebook profile.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	See Section 4.1 of the Facebook profile.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	See Section 4.2 of the Facebook profile.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	See Section 5 of the Facebook profile. Note there is no breakdown specifically for Messenger.
5.1 What information/fields of data are included in the TRs?	See Section 5.1 of the Facebook profile.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	See Section 5.2 of the Facebook profile.
5.3 Frequency/timing with which TRs are issued	See Section 5.3 of the Facebook profile.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to Meta, Messenger uses artificial intelligence to identify unusual behavioral patterns to detect phishing, scamming, and other harmful activities and removes these from the platform. It also provides a range of safeguards for minors such as limiting who can message them and restricting how they can be found in search. Messenger uses machine learning to detect and disable accounts which engage in inappropriate interactions with children (Messenger, n.d.[167]). Messenger also relies on user reporting to flag any concerns or violations of Facebook Community Standards. The Help Centre reminds users that "*images and videos of children being physically abused or sexually exploited are against Messenger policies*" and encourages users to contact law enforcement immediately if someone sends them an image of a child being physically or sexually abused (Messenger, n.d.[168]). Users are also urged to report

	potentially violative content using the Messenger reporting tools. End-to-end encryption is offered as a feature within Messenger, building on the approach developed at WhatsApp (which has been end-to-end encrypted by default since 2016), limiting the ability of the platform to detect CSEA or other illegal content. In its White Paper, "Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging" (Meta, 2022[169]), Meta argues that even without accessing/scanning the contents of users' private messages (unless reported), the platform can with the use of AI technologies identify suspicious behavior and disrupt potential harm before it happens. Meta also refers to its upstream detection methods, which include disrupting entire networks of bad actors before they can use messaging to cause harm in the first place. In other words, Meta contends that the platform will detect harmful behavioral patterns using non-content signals, content on non-encrypted surfaces like Facebook and Instagram, and user reports of messaging content to identify and respond to potential abuse. (See also Section 6 of the WhatsApp profile). Additionally, according to Meta, it offers prevention tools and controls that operate across its platforms regarding who can message all users, including minors, as well as features like blocking and deleting friend requests. For minors in particular, Meta reports that it seeks to raise awareness of these controls through in-app education, like Safety Notices (Meta, 2021[147]). Meta also indicates that it provides easier reporting mechanisms for minors. Should a minor block or delete a friend request, they are subsequently asked if they want to make a report. These reporting tools are also described as easier to find. Meta also reports that using the option "involves a child" when reporting harm, helps to prioritise the report for review and action. (See also Section 6 of the Facebook profile.)
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. See Section 7 of the Facebook profile. As reported in the Facebook Community Standards Enforcement Report, in Q2, 2022, 20.4 million pieces of content were actioned for "child endangerment: sexual exploitation" (Meta, n.d.[148]). Reports associated with Messenger are included in the overall totals. A separate breakdown for Messenger-specific reports is not available.

8. Weixin/WeChat (Tencent Holdings Ltd.)

1. How is online child sexual exploitation and abuse (CSEA)	Weixin/WeChat does not provide a specific definition of CSEA. However, WeChat's Acceptable Use Policy (AUP) states that it

defined in the Terms of Service (ToS) or Community Guidelines/Standards?	"*prohibits any content or behaviour relating to, depicting, promoting or encouraging participation in, or soliciting any of the following: Child nudity and exploitation – including any content where adults are soliciting minors (and vice versa)"* (Section 3 – Personal Safety). The AUP also prohibits depicting, promoting or encouraging participation in, or soliciting any content or behaviour related to sexual exploitation more generally including acts or photos involving non-consenting adults, paid sexual services, and other types of pornography (whether its public distribution was consented to or otherwise). This is also covered under the general prohibition of any violent, criminal, illegal, or inappropriate content or activities on WeChat. The AUP also contains a general category of Objectionable Content (Section 5) defined as "*any content or behaviour that is reasonably likely to cause upset and/or distress, either to the subject and/or to the public*" and includes nudity and sexual activity, and sexual solicitation – e.g., sharing of pornography or explicit offers of/requests for sexual services.
2. How are the ToS or Community Guidelines/Standards communicated?	The WeChat Acceptable Use Policy is available at: https://www.wechat.com/en/acceptable_use_policy.html The Terms of Service are available at: https://www.wechat.com/en/service_terms.html
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	WeChat uses a combination of user reporting as well as automated technologies to detect, block and remove harmful content. WeChat encourages users to report any unsafe or illegal content or behaviour they may come across on the platform. Users can report potentially violative content or block unwanted communications using a reporting mechanism provided in app. Where content is reported to WeChat, according to the AUP it will be investigated and the outcome communicated via the WeChat Team Official Account in the user's profile.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	According to WeChat's AUP, the platform may at its discretion carry out any of the following actions in response to a breach of its policies: • Issue a warning regarding the offending behaviour. • Hide or remove the content relating to a suspected breach. • Display a warning notice to recipients of the relevant content. • Restrict the user from accessing account functions or suspend or terminate the user's account. • Notify and cooperate with appropriate governmental and/or law enforcement authorities in the relevant jurisdiction where a suspected offence has been committed.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Enforcement policies refer to issuing a warning. However, no further details of notification procedures are provided.

4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No information is given about appeal processes.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to WeChat, the platform deploys automated processes to detect and prevent harmful content that breaches its policies or other applicable laws. This includes preventing the uploading of content and monitoring of content on the platform. The policy states that the platform may "*refuse or remove any harmful content available on or transmitted through WeChat in breach of this Policy*." WeChat also provides a user reporting mechanism where potentially violative content or abuse may be reported. Reports are investigated by WeChat staff though no further details of the review process are provided.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. While there is little information publicly available about the prevalence or extent of CSEA on the WeChat platform BBC news in 2021 reported that a number of Chinese online platforms including Weixin were fined in July 2021 by the Cyberspace Administration of China (CAC) for prominent online problems that endanger the physical and mental health of minors including "engaging in child porn" (BBC, 2021[170]).

9. Viber (Rakuten, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Viber does not provide a definition of CSEA. However, its Acceptable Use Policy (AUP) states that: "*We prohibit pornographic content, nudity in certain non-sexual contexts may be permitted, such as breastfeeding, we will report child sexual content to the authorities"*. Prohibited sexual content includes content that seeks to exploit or harm children by exposing them to inappropriate content or content that may impair their physical, mental or moral development. Grooming behaviour such as asking children for personally identifiable details is strictly prohibited. Additionally, content that includes or glorifies the sexual abuse and sexual exploitation of children, including child pornography is said to constitute a serious

	violation of the fundamental rights of children, especially child victims, and is strictly prohibited. Viber states that it will remove, and report to the applicable authorities, any content that exploits children immediately once it becomes aware of it. Viber's Terms of Service state that users may only use its services for lawful purposes and in accordance with applicable law. Users are prohibited from storing, distributing, or transmitting any unlawful material through their use of Viber.
2. How are the ToS or Community Guidelines/Standards communicated?	Viber's Acceptable Use Policy is published on its website at: https://www.viber.com/en/terms/viber-public-content-policy/ and is available under Terms & Policies in the Viber app. Viber's Terms of Service are available at: https://www.viber.com/en/terms/viber-terms-use/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Viber employs a combination of user reporting and automated technologies to moderate content on its platform. User reporting is especially relevant to "admins" and "superadmins" who have the responsibility of moderating channels (or communities) on Viber. Admins are also responsible for accepting or banning users. The AUP states that the platform reserves the right to use artificial intelligence and machine learning tools as well as human staff reviwers teams to pre-moderate and review content published for any potential illegal content or content that violates the AUP and Viber Terms.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Viber's enforcement actions may include any of the following: • the platform may remove any content that breaches its ToS or policies, • it may terminate or limit the visibility of an offender's account • it may notify law enforcement. Viber states it can remove reported content, at its sole discretion, if it finds it to be in breach of the AUP, Viber Terms or applicable law.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Viber states that it make best efforts to notify the parties of an enforcement decision but does not commit to doing so.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	An appeals procedure is outlined in Viber's AUP. According to Viber, in the event that action is taken against any particular user such as having their content removed or services withdrawn, they may appeal the decision by contacting the platform through a designated "Contact Us" form. The person "aking the appeal is asked to state why they feel the decision was incorrect. If the appeal is granted, the user is notified and the content reinstated, or account is reactivated. Any strikes or restrictions applied will be removed.
5. Does the service issue transparency reports (TRs)	No.

specifically on content and/or behaviour related to CSEA?	
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Very little information is available about content moderation approaches on Viber. The AUP refers to the use of artificial intelligence and machine learning tools as well as human teams to pre-moderate and review content published for any potential illegal content or content that violates its policies. According to Viber, the platform also relies on user reports and competent authorities' reports to detect CSEA. In May 2022, Viber signed the EU Code of Conduct against illegal hate speech online at which it confirmed that it had dedicated "resources to train and support moderation teams to assess and remove any hateful or illicit content within 24 hours of being reported so that all users can feel good about using and enjoying content on Viber." (Rakuten Viber, 2022[171]).
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. A study undertaken for the Australian Institute of Criminology found evidence of the use of Viber for CSEA live streaming (Napier, Teunissen and Boxall, 2021[172]). A U.S. Attorney's Office press release also refers to findings for the use of Viber to download child sexual abuse material (U.S. Attorney's Office, 2022[173])

10. Tik Tok (ByteDance Technology Co.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	TikTok defines child sexual abuse material (CSAM) as "*any visual depiction of sexually explicit nudity or conduct of a minor, whether captured by predatory adults, peers, or self-generated by minors*" (Community Guidelines, 'Sexual Exploitation of Minors). More generally, sexual exploitation of a minor is taken to mean "*any abuse of a position of power or trust for sexual purposes, including profiting financially, socially, sexually, or politically from the exploitation of a minor*". For the purposes of the policy, a minor is defined as any person under the age of 18. TikTok's Community Guidelines expressly prohibit the posting, streaming or sharing of any of the following: • Content that shares, reshares, offers to trade or sell, or directs users off platform to obtain or distribute CSAM; • Content that engages with minors in a sexualised way, or otherwise sexualizes a minor (e.g., via product features like duets);

	• Content that depicts, solicits, glorifies, or encourages child abuse imagery including nudity, sexualised minors, or sexual activity with minors; • Content that depicts, promotes, normalises, or glorifies pedophilia or the sexual assault of a minor; • Content that re-victimized or capitalises on minor victims of abuse by third party reshares or reenactments of assault or confessions. Grooming behaviour is highlighted as a specific form of sexual exploitation of a minor and is defined as "*activity in which an adult builds an emotional relationship with a minor in order to gain the minor's trust for the purposes of future or ongoing sexual contact, sexual abuse, trafficking, or other exploitation*". In this context, the following behaviours are prohibited on the platform: • Grooming advances; • Content that depicts, promotes, normalizes, or glorifies grooming behaviors; • Content that solicits real-world contact between a minor and an adult or between minors with a significant age difference; • Content that displays or offers nudity to minors; • Content that solicits minors to connect with an adult on another platform, website, or other digital space; • Any solicitation of nude imagery or sexual contact, through blackmail or other means of coercion. The Community Guidelines further prohibit content that depicts nudity or sexual activity involving minors. This include digtally created or manipulated content. The following are given as examples of prohibited content: • Content that depicts or implies minor sexual activities including penetrative and non-penetrative sex, oral sex, or intimate kissing; • Content that depicts sexual arousal or sexual stimulation involving a minor; • Content that depicts a sexual fetish involving a minor; • Content that depicts exposed genitals, buttocks, the pubic region, or female nipples of a minor; • Content that contains sexually explicit language depicting or describing a minor; • Content depicting a minor that contains sexually explicit song lyrics; • Content with sexually explicit dancing of a minor, including twerking, breast shaking, pelvic thrusting, or fondling the groin or breasts of oneself or another; • Content depicting a minor undressing; • Content depicting a minor in minimal clothing that is not

	situationally relevant to the location; • Sexualised comments, emojis, text, or other graphics used to veil or imply nudity or sexual activity of a minor. TikTok Terms of Service reiterate that Community Guidelines apply to all users and all content on the platform. Service users undertake not post to anything illegal, engage with minors in an exploitative or inappropriate way, or to post, live stream or otherwise distribute any content on the platform that is is obscene, pornographic, hateful or inflammatory or which promotes sexually explicit material (e.g. by linking to adult or pornographic websites).
2. How are the ToS or Community Guidelines/Standards communicated?	The TikTok Community Guidelines are available at: https://www.tiktok.com/community-guidelines?lang=en#31 The TikTok Terms of Service are available at: https://www.tiktok.com/legal/terms-of-service-eea?lang=en
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	TikTok's policies as expressed in the Community Guidelines aim to set norms and a common code of conduct that provides for "*a safe and welcoming space for everyone*". TikTok prioritises proactive enforcement of its policies using a mix of technology and human moderation with the aim of removing harmful material before people report potentially violative content to the platform. TikTok also encourages its community members to use the tools provided to report any content or account they believe violates its Community Guidelines.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	TikTok reserves the right to investigate any suspected breach of its terms or Community Guidelines during which it may remove some or all of the potentially violative content. The platform states that it may also suspend access to some or all of its features in accordance with the seriousness of the suspected breach. Mutiple violations will result in the termination of an offender's account.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, according to the Community Guidelines, individuals are notified of any decisions taken and that they can appeal them if they believe no violation has occurred.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. Enforcement decisions such as content removal, account suspensions or termination may be appealed using the appeal mechanism provided on the platform. This will trigger a further review of the decision as to whether a violation has occurred.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, TikTok publishes Transparency Reports regarding enforcement of its Community Guidelines, law enforcement requests, governement requests for content removals and intellectual property requests.
5.1 What information/fields of data are included in the TRs?	The following fields of data are included: • Total videos removed/total videos published by quarter • Total videos removed/restored, by type and quarter • Total video removal, by policy • Ads policy enforcement

	• Spam account activity • Fake engagement • Total account removal, by quarter and reason • Proactive removal volume and rates, by country • Proactive removal rate, by quarter/policy • Video removals and rates, by sub-policy According to the 2022 report for Q1, 41.7% of the total number of videos removed were for reasons of minor safety. Of these, three quarters (74.6%) were for reasons of nudity and sexual activity involving minors; 1.7% were for grooming behaviour and 1.9% for sexual exploitation of minors.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	For the purposes of TikTok transparency reporting, video removals are reported by the policy violation concerned. A video may violate multiple policies and each violation is reflected. Only videos that have been reviewed by moderators are included in the sub-policy dashboard. The cateogory of Minor safety comprises the following sub-policy categories: • Grooming behaviour • Sexual exploitation of minors • Physical and psychological harm of minors • Harmful activities of minors • Nudity and sexual activity involving minors The “nudity and sexual activity involving minors” sub-policy prohibits a broad range of content, including “minors in minimal clothing” and “sexually explicit dancing”; these two categories represent the majority of content removed under that sub-policy. This accounts for the largest proportion of content removals under Minor Safety (74.6%). Child Sexual Abuse Material (CSAM) or sexual exploitation of minors is reported separately and makes up a smaller proportion of removals (1.9% in Q1 2022).
5.3 Frequency/timing with which TRs are issued	Community Guidelines Enforcement Reports have been published on a twice yearly basis since July 2019 and on a quarterly basis since January 2021. All reports are available in the TikTok Transparency Centre (TikTok, n.d.[174])
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	A combination of automated detection technologies, community reporting and human review is used to detect CSEA violations on the platform. According to TikTok, it prioritises automated detection and removal of violative content before anyone reports it. The platform uses AI and machine learning technologies to detect known CSEA content based on specific criteria involving a range of different automated systems including the NCMEC hash list and the IWF database. TikTok uses PhotoDNA, which works to proactively identify illegal and harmful content. The decision engine also ranks potentially

	violating content to help moderation teams review the most urgent content first. The platform also filters red-flag language and shares information with NCMEC about situations that may indicate grooming behavior as defined in its policies as well as reflecting industry norms. TikTok has regional Trust & Safety hubs in California, Dublin, and Singapore which oversee the development and implementation of moderation policies across its services and localized as appropriate within each market.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to TikTok's transparency reports (see 5.1 and 5.2 above), CSAM has been detected on TikTok services. According to NCMEC, a total of 288,125 were submitted by TikTok in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 154,618 reports submitted in 2021 (NCMEC, 2022[150]).

11. QQ (Tencent Holdings Ltd.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA Is provided. However, the QQ Acceptable Use Policy (AUP) prohibits content or activity "*that seeks to harm or exploit any person (whether adult or child) in any way, including through bullying, harassment or threats of violence*". A more general prohibition applies to "*content that is pornographic, suggestive, violent, or otherwise adult in nature"* as well as content that violates any law or regulation or conduct that engages in illegal activity. Tencent's Service Agreement states that through their use of the service, users agree not to:""*to publish, deliver, transmit or store any content that contravenes national law, or threatens the national security, reunification of the nation, social stability, or anything that is inappropriate, insulting, defamatory, obscene, violent and against the national laws, regulations and policies*".
2. How are the ToS or Community Guidelines/Standards communicated?	The Tencent Acceptable Use Policy is available at: https://www.tencent.com/en-us/acceptable-use-policy.html The Tencent Service Agreement is available at: https://www.tencent.com/en-us/service-agreement.html
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Tencent employs a combination of user reporting and automated tools to monitor any violation of its policies. Users are encouraged to report using the available mechanisms any content that is unsafe, malicious or violates its terms. Tencent uses both automated and manual review to review and verify reports (Tencent, n.d.[175]).
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when	According to Tencent, where a piece of content is found to have violated its policies, its dissemination will be limited on QQ. Accounts found to have violated Tencent policies will be issued with a warning, and may have their services restricted or withdrawn as appropriate.

violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Reference is made to warnings that may be issued for policy violations. However, no information is given in the AUP or elsewhere in the Help Centre.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeal processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to Tencent, a combination of human and automated review is used to monitor content on the service. Users can report violations through the reporting function. The reporting function allows users to submit reports about suspected malicious content on a QQ or qZone account or group account. According to the platform, the QQ Security Center will verify and deal with all such reports in a timely manner to protect the rights and interests of its users. QQ also uses tools to proactively discover policy violations. However, there is very little public information available on this.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. While there is little information publicly available about the prevalence or extent of CSEA on QQ, BBC news in 2021 reported that a number of Chinese online platforms including QQ were fined in July 2021 by the Cyberspace Administration of China (CAC) for prominent online problems that endanger the physical and mental health of minors including “engaging in child porn” (BBC, 2021[170]).

12. Youku Tudou (Alibaba Group Holding Limited)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Youku Tudou does not specifically define CSEA. However, the platform’s User Service Agreement lists among the prohibited activities inducing minors to engage in illegal and criminal activities including pornography. Endangering social morality or disseminating content that is prohibited by relevant laws, administrative regulations and State regulations are also prohibited.
2. How are the ToS or Community	The User Agreement / Terms of Service are published at: https://terms.alicdn.com/legal-

Guidelines/Standards communicated?	agreement/terms/suit_bu1_unification/suit_bu1_unification20200514208_14749.html
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Youku Tudou broadly states that the platform has the right to manage any information uploaded, published or transmitted. For any violations of its policies or applicable laws, it will suspend transmission and implement measures to stop the spread of viral information and report to the relevant authorities.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	As stated in the User Agreement, Youku Todou has the right to manage any information uploaded or posted to its platform and to block or remove any content found to be in violation of its policies or applicable laws. Enforcement options include content removal, preventing the spread of information, keeping relevant records, and reporting offences to the relevant authorities. If a user or account holder is found to be in breach of its policies, Youku Tudou may: • Restrict participation in activities, • Suspend the provision of some or all services, etc., • Deduct damages. If offending behavior is deemed to constitute a fundamental breach of contract, the platform may close the account and terminate the provision of services to the user. If the behavior on the platform is deemed to have violated relevant laws and regulations, it may be reported and usage records and other information submitted to the relevant authorities in accordance with the law.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No notifications are specified.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeals processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff)	No information is given about Youku Tudou's content moderation or review processes to detect CSEA.

reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. While there is little information publicly avaiable about the prevalence of CSEA on Youku Tudou, media reports refer to actions taken by Chin''s National Office Against Pornography and Illegal Publications against online services including Youku Tudou for child abuse content (Shumin, 2018[176])

13. Telegram (Telegram Messenger LLP)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided. However, Telegram ToS state that the posting of illegal pornographic content on publicly viewable Telegram channels, bots, etc. is prohibited. However, these ToS do not appear refer to secret chats or end-to-end encrypted messages. No other acceptable use policy or community guidelines are provided.
2. How are the ToS or Community Guidelines/Standards communicated?	Telegram Terms of Service are available at: https://telegram.org/tos The Telegram Privacy Policy which forms part of the Terms of Service are available at: https://telegram.org/privacy#1-2-terms-of-service
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Telegram facilitates user reporting of abuse on any publicly available content on Telegram such as sticker sets, channels, and bots. Where a user comes across any content on Telegram that they think may be illegal, they are encouraged to report it to the platform using an in-app reporting mechanism or by email to abuse@telegram.org. There is also a dedicated Stop Child Abuse channel on Telegram (stopCA@telegram.org) where users can report suspect CSAM or CSEA. Telegram states that it engages in proactive detection of content prohibited by its publicly available ToS. However, no information is provided of its policies in this regard. Conversations or so-called 'secret chats' in Telegram are end-to-end encrypted. The encryption is device-specific and is not part of the Telegram cloud. Message, photos, videos and files can also be ordered to self-destruct or be deleted from both sides of a conversation. Accordingly, Telegram has no access to the content.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	No details are given in the ToS regarding enforcement policies or procedures. The Telegram FAQ states that it processes legitimate requests to remove illegal public content (e.g., sticker sets, bots, and channels) within the app, including the removal of porn bots. When the platform receives a complaint regarding the legality of public content, it performs the necessary legal checks and removes the content when deemed appropriate.

	However, Telegram's key feature is its secrecy and for content not covered by end-to-end encryption, Telegram uses a distributed infrastructure in multiple data centers around the globe that are controlled by different legal entities spread across different jurisdictions. As a result, several court orders from different jurisdictions are required to compel Telegram to release any data. According to Telegram, so far this has not happened.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No notifications are specified.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeal processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Telegram does not issue a Transparency Report as such regarding its handling of CSEA. However, its Stop Child Abuse channel (topCA@telegram.org) publishes daily updates on banned CA-related content. No further detail or breakdown of statistics is available.
5.1 What information/fields of data are included in the TRs?	Total number of groups and channels related to child abuse that have have been banned.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No information given.
5.3 Frequency/timing with which TRs are issued	A daily update is provided on the StopCA channel.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	No information available.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. The Telegram Stop Child Abuse channel publishes daily updates on banned CA-related content. For example, a total of 26880 groups and channels related to child abuse were banned during the month of July 2022 (Telegram, n.d.[177]). In 2018, the Telegram app was removed from the Apple App Store due to dissemination of illegal content, specifically child pornography. Fixes were put in place (not specified) to prevent a recurrence and the app was restored to the App Store (Hall, 2018[178]). According to a BBC investigation in 2019, images of child sexual abuse were found to be openly traded on encrypted conversations in Telegram. The investigation found that paedophiles were using Telegram to give people access to abuse material, and that links to Telegram groups were buried in the public comments section of YouTube videos (BBC, 2019[179]).

14. qZone (Tencent Holdings Ltd.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided. Qzone shares Terms of Service with QQ Instant Messaging (see Profile #11 for QQ) as part of the Tencent Holdings suite of services. The QQ International Service Agreement, or Terms of Service, prohibit use of the software or any of the services offered by Tencent for any illegal purpose or in any form inconsistent with its policies. As stated in its Acceptable Use Policy (AUP), content or activity "*that seeks to harm or exploit any person (whether adult or child) in any way, including through bullying, harassment or threats of violence*" is expressly prohibited. A more general prohibition applies to "*content that is pornographic, suggestive, violent, or otherwise adult in nature"* as well as content that violates any law or regulation or conduct that engages in illegal activity.
2. How are the ToS or Community Guidelines/Standards communicated?	The Tencent Acceptable Use Policy is available at: https://www.tencent.com/en-us/acceptable-use-policy.html The Tencent Service Agreement is available at: https://www.tencent.com/en-us/service-agreement.html
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Tencent implements a combination of user reporting and automated tools to monitor any violation of its policies. The Tencent Help Centre encourages users to use the reporting function to flag any content that is unsafe, malicious or violates its terms. Tencent uses both automated and manual review to review and verify reports.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	According to Tencent, where a piece of content is found to have violated its policies, its dissemination will be limited on qZone Accounts found to have violated Tencent policies will be issued with a warning, and may have services restricted or withdrawn as appropriate.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Reference is made to warnings that may be issued for policy violations. However, no information is given in the AUP or elsewhere in the Help Centre.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeal processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.

5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to Tencent, a combination of human and automated review is used to monitor content on the service. Users can report violations through the reporting function. The reporting function allows use to submit reports about suspected malicious content on a QQ or qZone account or group account. According to qZone, the QQ Security Center will verify and deal with all reports in a timely manner to protect the rights and interests of its users. qZone also uses tools to proactively discover policy violations. However, no public information is available on this.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. While there is little information publicly available about the prevalence or extent of CSEA on qZone, BBC news in 2021 reported that a number of Chinese online platforms including qZone were fined in July 2021 by the Cyberspace Administration of China (CAC) for prominent online problems that endanger the physical and mental health of minors including “engaging in child porn” (BBC, 2021[170]). Other media reports also refer to actions taken by Chin''s National Office Against Pornography and Illegal Publications against online services including Youku Tudou for child abuse content (Shumin, 2018[176])

15. Weibo (Sina Corp.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided. CSEA is covered in a general way under the Weibo Service Use Agreement (or ToS) which prohibits the uploading or sharing of any content that is“"*false, impersonating, harassing, defamatory, offensive, abusive, intimidating, racially discriminatory, defamatory, revealing of privacy, pornographic, obscene, malicious plagiarism, violence, gore, suicide, self-harm or any other illegal information* “ (4.10.4). Under its rules of use, service users must abide by relevant laws and regulations and undertake not to share any content that “*contains illegal and bad information*” (4.1). Users are also encouraged to report any content that infringes Weibo’s ToS.

2. How are the ToS or Community Guidelines/Standards communicated?	The Weibo Service Use Agreement is available at: http://weibo.com/signup/v5/protocol
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	The ToS states that Weibo has the right to review, supervise and process users' behaviors and information on Weibo services, including but not limited to user information (account information, personal information, etc.), published content (location, text, pictures, audio, videos, trademarks, patents, publications, etc.), user behavior (building relationships, information, comments, private messages, participating in topics, participating in activities, marketing information release, reporting complaints, etc.) (4.12).
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	In its ToS, Weibo reserves the right to take any of the following enforcement actions in response to a violation of its policies: • Change, delete or block relevant content; • Warning offending accounts and banning accounts; • Freeze user account funds to make up for losses caused by users to the Weibo operator, its affiliates, and others; • Change, restrict or prohibit some or all of the functions of the offending account; • Suspend, restrict or terminate the use''s right to use the Weibo service, cancel the user account, etc.; • Report to relevant regulatory authorities or state authorities; • Other measures deemed reasonable by the Weibo operator.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	The ToS refer to Weibo's right to take any enforcement action without notice though it is also stated that users will be notified where possible after an action has been taken.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeal processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Weibo employs a combination of platform content moderation and user reporting. A reporting mechanism is available where users can report any content or behaviour suspected to be illegal or which infringe the platform's policies. The ToS state that Weibo personnel will verify and deal with reports as soon as possible. f The ToS also refer to the platform's right to review all content published on the

	platform. However, no further detail is provided about the processes involved.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. While there is little information publicly available about the prevalence or extent of CSEA on Weibo, BBC news in 2021 reported that a number of Chinese online platforms including Weibo were fined in July 2021 by the Cyberspace Administration of China (CAC) for prominent online problems that endanger the physical and mental health of minors including "engaging in child porn" (BBC, 2021[170]).

16. Snapchat (Snap, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Snapchat's Community Guidelines prohibits "*"any activity that involves sexual exploitation or abuse of a minor, including sharing child sexual exploitation or abuse imagery, grooming, or sexual extortion (sextortion)."* Snap's Transparency Report defines CSEA as "as content that contains sexual images of a minor and all forms of child sexual abuse material (CSAM), as well as grooming or enticement of a minor for any sexual purpose" (Snap Inc., n.d.[180]). The Community Guidelines also expressly prohibit acts of solicitation such as asking a minor to send sexually explicit content. Snapchat further prohibits accounts that promote or distribute pornographic content. All cases of child sexual exploitation are, according to the platform, reported to the authorities. In the Terms of Service, users agree not to violate any applicable law or regulation in connection with their access to or use of Snapchat. Snapchat's Transparency Report states that the sexual exploitation of any member of its community, especially minors, is illegal, unacceptable, and prohibited by its Community Guidelines. Preventing, detecting, and eradicating Child Sexual Abuse Material (CSAM) on the platform it declares is a top priority for Snapchat, and for which it is continuously developing its capabilities to address CSAM and other types of child sexually exploitative content.
2. How are the ToS or Community Guidelines/Standards communicated?	Snapchat Community Guidelines are available at: https://values.snap.com/privacy/transparency/community-guidelines Snapchat Terms of Service are available at: https://www.snap.com/terms
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	According to the platform, Snapchat applies safety by design principles to help keep its community safe and to prevent misuse. Public content on Snapchat, such as its Discover content platform and its Spotlight entertainment platform, is curated or pre-moderated to ensure it complies with the platform's guidelines before it can reach a larger audience. AI and machine learn'ng tools are used to proactively detect illegal content and activity. Search results are

	blocked for certain keywords for illegal content. To mitigate risks of potential grooming of minors for sexual purposes, there are restrictions in place to prevent under-18s showing up in search results or as a friend suggestion to someone else unless they have multiple friends in common. Under-18s also need to be friends with (i.e., affirmatively accept) another user before they can communicate directly.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Any violation of Snapchat's Community Guidelines may result in a warning to the user, removal of the offending content, termination or limiting the visibility of the offending account and/or notification to law enforcement. Snap Inc. also reserves the right to remove users whom it has reason to believe pose a danger to others, on or off of Snapchat. If a user's account is terminated for violations of Snapchat's policies, the offender is prohibited from using Snapchat again.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	While no notification details are specified publicly, according to Snapchat if reporting users have provided an email address associated with their Snapchat account, they will receive an email notifying them that Snap has taken action against the content or account that they reported. In the case of users that have been the subject of a report, they will see an in-app warning when Snap deletes or locks their content or account, and will be logged out when their account is locked. An example of the text they see is as follows: “"You're receiving this warning because we have removed your content for violating our Community Guidelines prohibiting Sexual Content. Additional violations of this provision of our Community Guidelines will lead to your account being locked or deleted. You must fully read the Community Guidelines before proceeding”"
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. Users are able to appeal their account suspensions at https://help.snapchat.com/hc/en-gb/requests/new?co=true&ticket_form_id=7058755437844
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Snapchat has issued Transparency Reports since 2015 on legal requests for account information. In 2020, Transparency Reports were expanded to give details of violations of its Terms of Service or Community Guidelines (Snap Inc., 2022[181]).
5.1 What information/fields of data are included in the TRs?	Transparency Reports present data on the following: • Total content and accounts reported • Reason or category of violation • Volume of content or accounts actioned for violation • % of Total Content Enforced • Unique Accounts Enforced • Median Turnaround Time (minutes)
5.2 Methodologies for determining/calculating/estimating the information/data included	• Total content and accounts reported includes all pieces of content and accounts processed by Snapchat's Trust & Safety whether detected proactively or as a result of being

in the TRs	reported by users. • Reason refers to the category of violation as defined by Snapchat's Community Guidelines. • Content enforced refers to all accounts and content where enforcement action was taken including removing the offending content or terminating the account in question.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are issued on a six monthly basis.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	As stated in Snapchat's Transparency Report, its Trust and Safety teams use active technology detection tools, such as PhotoDNA robust hash-matching and Google's Child Sexual Abuse Imagery (CSAI) Match to identify known CSEA images and videos and report them to NCMEC. Reporting mechanisms in app, through a website form and via direct messaging allow users to report any content or behaviour they suspect to be in breach of its policies. All reports are reviewed by a member of Snapchat's Trust & Safety Team. If the content is found to violate Snapchat's Community Guidelines, the user may be warned, the content may be removed or the account suspended/terminiated. If necessary, law enforcement will also be contacted.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. In the period January 1 to June 30, 2022, a total of 201,527 unique accounts were deleted for CSEA, 746,051 pieces of content were actioned, and 285,470 submissions were made to NCMEC. 94% of the total CSEA violations were detected proactively (Snap Inc., 2022[181]). According to NCMEC, a total 551,086 reports were submitted by Snapchat in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 512,522 reports submitted in 2021 (NCMEC, 2022[150]).

17. Kuaishou (Beijing Kuaishou Technology Co., Ltd)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided. However, the Kuaishou Terms of Service (the Kuaishou Software License and Service Agreement) prohibits the uploading, downloading, sending or transmission of any content in violation of national laws. This includes "*spreading eroticism or obscenity*" which is expressly prohibited as is engaging in any potentially illegal transactions or activities.
2. How are the ToS or Community Guidelines/ Standards communicated?	The Kuaishou Terms of Service are available at: https://www.kuaishou.com/about/policy
3. What are the service's policies and procedures for preventing its use in a manner	Kuaishou states in the ToS that it reserves the right to examine or verify any content uploaded or published on the platform and the right to deal with it in accordance with its policies and with applicable

that violates its ToS or Community Guidelines/Standards?	national laws.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	For any violations of its policies, Kuaishou has the right to take any of the following enforcement measures: restricting or prohibiting use of the platform; closing or deactivating an offender's account; retaining the offending content and delivering or reporting it to relevant authorities.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No details of notification procedures are given. According to Kuaishou, where users have questions about the reason for an account enforcement ban, they are informed they can contact Customer Service setting out their problem, and the staff will reply and deal with it in time.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. An appeals process is referred to in the Kuaishou FAQs. Where an account has been banned for violations, a user can submit an appeal for further review.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	There is limited information available about Kuaishou's approach to content moderation and detection methods for CSEA. The platform provides a user reporting mechanism where users can report suspected violations or potentially illegal content. These reports are reviewed and verified by moderators. Kuaishou also states that it has the right to review and verify all content uploaded or published by users though the precise nature or the technologies involved are not specified.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. While there is little information publicly available about the prevalence or extent of CSEA on the Kuaishou platform, BBC news in 2021 reported that a number of Chinese online platforms including Kuaishou Technology were fined in July 2021 by the Cyberspace Administration of China (CAC) for prominent online problems that endanger the physical and mental health of minors including "engaging in child porn" (BBC, 2021[170]).

18. iQIYI (Baidu, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided. However, the iQIYI Terms of Service state that users must abide by its policies and all applicable laws and refrain from activity that would harm minors in any way, or engage in activities, violations or crimes infringing upon the lawful rights and interests of others. iQIYI states that it attaches great importance to the protection of minors.
2. How are the ToS or Community Guidelines/Standards communicated?	The iQIYI Terms of Service are available at: https://www.iq.com/intl-common/international-useragreement.html?lang=en_us
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	iQIYI states that it will monitor and scrutinise content uploaded on its platform as required by applicable laws and/or regulations. Users must acknowledge that confidentiality with respect to any content, whether it is published or not, is not guaranteed.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	If found to have violated any of its policies, iQIYI has the right to suspend or permanently terminate, cancel or withdraw an offending account, and to withdraw or end the provision of further iQIYI Services to the individual in question.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No notification processes are specified.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeals process is specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL	The company states that it can monitor and scrutinise content uploaded to its platform. However, no information is given of the processes involved.

sharing database) does the service use to detect CSEA?	
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. While there is little information publicly available about the prevalence of CSEA on the iQIYI service, media reports refer to actions taken by Chin"s National Office Against Pornography and Illegal Publications against online services including iQIYI for child abuse content (Shumin, 2018[176])

19. Pinterest (Pinterest, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	CSEA is not specifically defined. However, Pinterest states that it does not tolerate child sexual exploitation (CSE) of any kind on its platform and that it enforces a strict, zero-tolerance policy for any content including imagery, video, or text that might exploit or endanger minors. Sexualisation or sexual exploitation of minors is expressly prohibited and is taken to include grooming, sexual remarks or inappropriate imagery. Where any such content is detected, it is reported to relevant authorities such as NCMEC. Pinterest furthermore states that its CSE policy goes further than its legal obligations to prohibit any content that contributes to the sexualization of minors. For example, content that suggests the sexualization of minors in the form of cartoons or animé is probhibited as is the intentional misuse of content depicting minors engaging in non-sexualized activities, like modeling clothing or participating in athletics. Pinterest Community Guidelines forbid pornographic content more generally and prohibits content including 'vivid sexual descriptions' or 'graphic depictions of sexual activity'. Its Community Guidelines and all related policies also apply to comments posted on Pins as well as content uploaded and shared on the platform.
2. How are the ToS or Community Guidelines/Standards communicated?	The Pinterest Community Guidelines are available at: https://policy.pinterest.com/en-gb/community-guidelines Pinterest Terms of Service are available at: https://policy.pinterest.com/en-gb/terms-of-service
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Pinterest states that safety and positivity are at the heart of its design approach. Its policies are designed to protect users from seeing unsafe or harmful content, including misinformation. It states that the platform has a focus on wellness and positivity and a strong stance against harmful and illegal content. (Pinterest, n.d.[182])
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when	The Pinterest ToS state that it reserves the right to terminate or suspend access to the platform, including without notice if there is a good reason, for any violation of its Community Guidelines. Enforcement actions include blocking, limiting the distribution of or

violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	deactivating content and the accounts, individuals and groups that create or spread that content, based on how much harm it poses.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, Pinterest provides a notification when an account is suspended. However, it does not always provide notification when content that goes against its community guidelines is removed.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, an appeals process is available for account suspensions but not for content removal. Details of the appeals process are given in the Help Centre.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Pinterest has published a biannual Transparency Report since 2013. In Q4 2020, this was expanded to include details of enforcement of its Community Guidelines. The July-December 2021 report provided more detailed information regarding enforcement actions against child sexual exploitation (CSE) content.
5.1 What information/fields of data are included in the TRs?	Metrics for CSEA include: • reach and actioned user reports; • numbers of images deactivated; • proactivity rate; • numbers of appeals, and • appeals upheld.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	• Statistics for actioned user reports include all reports where violating content was found. • Reach is defined as the number of people who saw the offending content before it was removed. • Content seen by 0 people refers to content that was proactively detected before being seen on the platform. For CSEA violations, all deactivations are counted even if other actions may have already been taken against the Pin, board or user. For example, if a Pin has been automatically deactivated—meaning no one on the platform can see it—for violating the Spam policy and is subsequently found to contains material that violates its CSE policy, the Pin is counted in both the Spam and CSE deactivation numbers. This is for the purpose of providing more accurate insight into violations.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are issued every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Pinterest uses a combination of automated content moderation, human staff review and user reporting to detect and action content that violates its policies. Pinterest proactively detects CSE images and videos through its own internal tools and shared industry tools such as PhotoDNA, which uses a shared industry hash database of known CSAM, and CSAI Match to identify video content. Pinterest states that it works closely with NCMEC to combat CSEA, and reports content violations in appropriate circumstances under the law. Pinterest also has a user reporting tool whereby any piece of content or Pin, comments, photos, messages or profiles may be

	reported. Reports may be submitted anonymously. All reports are reviewed by staff reviewers and action taken if a violation is found.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to the Transparency Report for July–- December 2021, 2,545 distinct images, which comprised 104,029 Pins, were deactivated for violating Pinteres''s CSE policy. The majority were proactively detected and 98% were seen by fewer than 100 users. 3,110 account appeals were received, 2,120 accounts were reinstated (Pinterest, n.d.[183]). According to NCMEC, a total of 34,310 reports were submitted by Pinterest in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 2,283 reports submitted in 2021 (NCMEC, 2022[150]).

20. Reddit (Reddit, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Reddit's approach to CSEA is outlined in its Content Policy which prohibits any sexual or suggestive content involving minors or someone who appears to be a minor. This includes "*sexual abuse imagery, child pornography, and any other content, including fantasy content (e.g. stories, "loli"/anime cartoons), that depicts, encourages or promotes pedophilia, child sexual exploitation, or otherwise sexualizes minors or someone who appears to be a minor* " (Rule 4). The Content Policy outlines rules that are platform-wide and that apply to all communities and everyone on Reddit. Reddit's Content Policy for Live Video and Audio prohibits any content that contains nudity, pornography or sexually suggestive content that may be considered "NSFW" (or "Not Safe for Work") or inappropriate for viewing in a public or formal setting such as a workplace. Broadcasts may not contain activities that are illegal. Reddit's Terms of Service (the User Agreement) stipulates all users must comply with its Terms and all applicable laws, rules, and regulations. Reddit states in its Transparency Report that it has a zero tolerance for content that puts children at risk. When it finds child sexual abuse material ("CSAM") on the platform, it is removed immediately and the offending user is permanently suspended from Reddit. Reddit also reports the relevant users/content to the NCMEC and preserves relevant user data as required by law (Reddit, 2021[184]).
2. How are the ToS or Community Guidelines/Standards communicated?	Reddit's Content Policy is available at: https://www.redditinc.com/policies/content-policy and its User Agreement at: https://www.redditinc.com/policies/user-agreement

	The Content Policy for Live Video and Audio is available at: https://www.redditinc.com/policies/broadcasting-content-policy
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Reddit has a multi-layered moderation system which is applied across its network of communities, relying to a great extent on individual communities to self-regulate and apply the shared rules of the platform. Reddit is a network of communities that are moderated by volunteer community moderators who review and apply the rules specific to each community. All communities are obliged to comply with the Content Policy which are the platform-wide rules that apply to everyone on Reddit. These rules are enforced by 'admins' who are Reddit staff reviewers and who implement platform-wide policy. According to Reddit's ToS, the platform has no obligation to screen, edit, or monitor a user's content, though it may at its sole discretion, delete or remove content at any time and for any reason, including for violating its terms or Content Policy.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Reddit's Content Policy outlines a tiered approach to enforcement which includes: • An informal warning • More formal warnings • Temporary or permanent suspension of accounts • Removal of privileges from, or adding restrictions to, accounts • Adding restrictions to Reddit communities, such as adding NSFW tags or Quarantining • Removal of content • Banning of Reddit communities Moderators within communities have a range of enforcement options that they may exercise, including modifying lists of approved submitters, taking action on posts such as marking them as spam or in breach of community rules, removing posts, banning or muting users (Reddit, n.d.[185]).
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, notifications are generally sent regarding Reddit's various enforcement actions. When community moderators take action on a piece of content, they may add an optional reason for the removal which will appear next to the post. Moderator tools include pre-defined messages that can be sent at the time a post or comment is removed. The Moderator Guidelines recommend this as good practice to help educate the submitter and cut down on queries or appeals that can take up the moderator's time. A suspension from a subreddit community is notified via a private message. If the suspension is temporary, a visual reminder of the suspension will appear on each page visited and any time a forbidden action is attempted. Permanent suspensions will be publicly communicated via the user page and accessible by other users.

	An account suspension is an action taken by Reddit's administration for security purposes or to enforce its Content Policy. If an account is suspended, the user will receive an administrator message in their Reddit inbox explaining a reason for the site-wide suspension.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, account suspensions may be appealed. If a user believes a suspension for a content violation was applied incorrectly, they can submit an appeal using an internal appeal form. All appeals are reviewed and if upheld, Reddit will reverse the suspension.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Reddit publishes an annual Transparency Report about content that was removed from Reddit, accounts that were sanctioned, and legal requests to remove content or disclose private user data. Content removals for violations of the Content Policy such as minor sexualisation are included.
5.1 What information/fields of data are included in the TRs?	The main relevant information fields include the following: • Content created on Reddit by year • Content removed from Reddit for the year in question for any reason (including spam) • Content removed by moderators • Content removed by admins (not including spam and other content manipulation) vs. content removed by mods • Content removed by admins (content manipulation & spam removals vs. Content Policy removals) • User reports for Content Violations • Posts & comments removed by Content Policy violation • Subreddits quarantined vs. removed • Subreddit removal reasons • Private messages removed by Content Policy violation • Admin account sanctions per rule • Admins: manual vs. automated action • Appeals against admin action and appeals by reason • CSAM reports–- automated vs. manual user reporting
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	As outlined in its Transparency Report, the following explanations for data fields are given: • **Content removals** are broken down by mod removals, admin removals, and volume of user reports. Removals performed by mods can be based on any reason specific to the rules of a given community, and are not necessarily an indication of content being in violation of Reddit's Content Policy. Admin removals include manual review and action, as well as with the help of automated tools and user reports. • **User reports for potential content violations** include all reports submitted including duplicate reports, those already actioned and those not deemed actionable. The percentage of actionable reports is also given. • **Violations of Content Policy** are broken down by content type (Posts & Comments, Subreddits or communities, and Private Messages) as well as by category of violation. As such, this includes total content created vs. reported/flagged

	vs. removed; the source of the reports flagged (whether by users, or flagged by Reddit automation). • **Account suspensions** includes the total number of temporary and permanent account suspensions handed out by admins. Reasons for suspension are also given. Similar data is provided for appeals of account suspensions. In the 2021 Transparency Report, details for CSAM violations included the total volume of reports detected and identified as CSAM with a percentage breakdown by User Reports, Photo DNA technology for images, and YouTube CSAI technology for videos.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are published annually at: https://www.redditinc.com/policies/transparency-report-2021-2/ Previous transparency reports since 2014 are available at: https://www.reddit.com/wiki/transparency/
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Reddit uses a combination of multi-level community moderation, as well as platform level human review and automated tools to detect violative content. As described in its Transparency Report, content moderation on Reddit happens through a layered, community-driven approach. Reddit's Content Policy serves as a set of principles-based rules that apply to all users and content on Reddit. Users then create their own communities (known as "subreddits") and establish additional rules that are tailored to a community's unique needs. Users who write and enforce these community-specific rules are volunteer moderators (known as "mods"), and they perform the majority of community moderation actions without involvement from Reddit, Inc. Reddit regards this self-moderation effort at the community level continues to be the most effect solution the scalability of moderating content online. Reddit employees (known as "admins") are responsible for the Content Policy and enforce it across Reddit with the help of mods, who apply the Content Policy to their communities in addition to their own specific rules. For the purposes of detecting CSEA, Reddit also uses its own automated tools as well as PhotoDNA hash-matching technologies for images and Google CSAI technology for known video content containing CSAM.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to the 2021 Transparency Report, Reddit removed 9,258 pieces of content identified as CSAM and made 10,059 CyberTipline reports to NCMEC. This represented an increase in NCMEC reporting from the 2,233 reports made in 2020. According to NCMEC, a total of 52,592 reports were submitted in 2022 by Reddit to the CyberTipline (NCMEC, 2023[18]). This compares to 10,059 reports submitted in 2021 (NCMEC, 2022[150]).

21. Twitter (Twitter, Inc.)[27]

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Twitter's Child Sexual Exploitation policy defines CSEA as "*any content that depicts or promotes child sexual exploitation including, but not limited to*: • *visual depictions of a child engaging in sexually explicit or sexually suggestive acts;* • *illustrated, computer-generated or other forms of realistic depictions of a human child in a sexually explicit context, or engaging in sexually explicit acts;* • *sexualized commentaries about or directed at a known or unknown minor; and* • *links to third-party sites that host child sexual exploitation material."* (Child Sexual Exploitation Policy, October 2020) The policy also prohibits activity that may constitute solicitation of a minor for sexual purposes or 'grooming' (e.g. sending sexually explicit media to a child; trying to engage a child in a sexually explicit conversation; trying to obtain sexually explicit media from a child or trying to engage a child in sexual activity through blackmail). Content or activity that normalises CSEA in any way is also expressly prohibited (e.g. sharing fantasies about or promoting engagement in child sexual exploitation; expressing a desire to obtain materials that feature child sexual exploitation; promoting or normalising sexual attraction to minors as a form of identity or sexual orientation). Content may include media, text, illustrated, or computer-generated images. Regardless of the intent, viewing, sharing, or linking to child sexual exploitation material is a violation of its policy.
2. How are the ToS or Community Guidelines/Standards communicated?	The Twitter Terms of Service are available at: https://twitter.com/en/tos The Child sexual exploitation policy is available at: https://help.twitter.com/en/rules-and-policies/sexual-exploitation-policy The Twitter Rules or Community Guidelines for the platform are available at: https://help.twitter.com/en/rules-and-policies/twitter-rules
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Twitter uses a combination of user reporting, machine learning and human review to monitor potential violations of its policies on the platform. Breaches of its policies, whether reported by users or detected through the platform's own systems, are directed to human moderators who review potential rule violations. Twitter's stated purpose is to support public conversation and to apply its rules so as "*to ensure all people can participate in the public conversation freely and safely*" (Introduction to the 'The Twitter Rules').

	Twitter's ToS state that it reserves the right to remove content that violates the User Agreement, including for unlawful conduct, or harassment, or to suspend or terminate an account for any reason including for unlawful conduct. Additionally, account owners may be asked to verify ownership in order to prevent violators operating multiple accounts for abusive purposes.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Twitter's Enforcement Policy sets out the consequences for users for violating terms. Enforcement actions may include the following: • Tweet-level enforcement which may include labelling a tweet, limiting its visibility or its removal. • Direct Message Level enforcement which may include blocking the person sending the message, removal of the content or hiding its visibility in a group conversation. • Account level enforcement which may include editing of content, placing the account in read only mode, verifying account ownership or suspension of an account. Action is taken at the account level for repeated violations or particularly egregious infringements of the Twitter Rules, such as CSEA. The consequence for violating Twitter's child sexual exploitation policy is immediate and permanent suspension. In addition, violators will be prohibited from creating any new accounts in the future. When made aware of CSEA material, including links to images of or content promoting child exploitation, Twitter removes the material from the site without further notice and makes a report to NCMEC as required under U.S. law.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes. When a tweet has been found to be a violation of its rules, the account holder is informed and instucted to remove it before they can tweet again. An email notification is sent identifying the tweet or tweets in question and which policies have been violated. The account holder is then required to remove the tweet or appeal if they believe an error has been made. In the interim, following an enforcement action the tweet is hidden from public view behind a notice stating that it is no longer available because of a violation of the Twitter Rules.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users can appeal enforcement actions such as tweet removals, or account suspensions if they believe an error has been made. Appeals are submitted using the platform interface or by filing a report. Upon appeal, if it is found that a suspension is valid, a response to the appeal is given with information on the policy that the account has violated. If the appeal is upheld, the account / content removed is restored. According to Twitter, appeals must be made from the account that has been blocked or locked.

5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, Twitter issues a Transparency Report every six months at: https://transparency.twitter.com/ Transparency Reports include sections covering information requests, removal requests, copyright notices, trademark notices, email security, Twitter Rules enforcement, platform manipulation, and State-backed information operations.
5.1 What information/fields of data are included in the TRs?	Metrics provided in the Transparency Report include: • Numbers of accounts action • Numbers of accounts suspended • Content removed metrics • Impressions of violative tweets (From Jan 2021) • Categories for accounts actioned
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Metrics related to enforcement actions within Transparency Reports include: accounts actioned, content removed, and accounts suspended. These are defined as follows: • *"Accounts actioned*" are the number of unique accounts that were suspended or had some content removed for violating the Twitter Rules, including auto-actioned content. • *"Content removed"* reflects the number of unique pieces of content (such as Tweets or an accoun''s profile image, banner, or bio) that Twitter required account owners to remove for violating the Twitter Rules, including auto-actioned content. • *"Accounts suspende""* reflects the number of unique accounts that were suspended for violating the Twitter Rules. In 2021, Twitter introduced a new metric of "impressions" to capture the number of views a Tweet received prior to being removed. For the purposes of the Transparency Report, an impression is defined as any time at least half of the area of a given Tweet is visible to a user for at least half a second (including while scrolling). This also includes views by logged-out users.
5.3 Frequency/timing with which TRs are issued	Transparency reports are published every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	A range of methods are used to monitor content on Twitter for the purposes of detecting CSEA, including user reporting as well as proactive detection through the use of AI technology and human staff review. Twitter encourages its users to report violations of the Twitter Rules. Reports submitted by users are reviewed by staff moderators who decide whether the content violates Twitter's rules. Reports may be made through in-app reporting or by using a dedicated web form. Both registered users and unregistered people can report breaches of content rules and suspected illegal content. In addition, a reporting mechanism for any aspects of safety and sensitive content on the platform is available to all through the dedicated page in the Help Center. According to Twitter, the platform has a global safety team that manages enforcement of the Twitter Rules with 24/7 coverage

	in every supported language on Twitter. Twitter also uses internal, proprietary tools to proactively detect violations of the Twitter Rules, including the posting of CSEA. Some of the same machine learning technology used to track spam, platform manipulation and other rule violations is also used to detect suspicious behaviour and potential abuse. According to Twitter, increasing use of AI and machine learning tools has contributed to signficant ehnacement of proactive detection instead of relying on reports from people on Twitter (Twitter, 2019[186]). Twitter has also extended its *#ThereIsHelp* to CSEA. When users attempt to search terms associated with CSEA, an automated prompt provides information about Twitter's zero tolerance policy and directs users to help resources and local prevention programs.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to Twitte"s Transparency Report, 453,754 unique accounts were suspended in the period January to June 2021 for violating its child sexual exploitation policies. 89% of these were detected proactively by internal proprietary tools and industry hash sharing initiatives (Twitter, 2022[187]). There was a 31% increase in the number of accounts actioned for violating CSE policies in the period July to December 2021. In January 2023, Twitter Safety reported that it had suspended approximately 404,000 accounts for creating, distributing or engaging with CSE content. This represented a 112% increase since November 2022 (Twitter, 2023[188]). According to NCMEC, a total 98,050 reports were submitted by Twitter in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 86,666 reports submitted in 2021 (NCMEC, 2022[150]).

22. Tumblr (Automattic, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Tumblr does not provide a specific definition of CSEA. However, Tumblr's Community Guidelines expressly prohibit the posting or soliciting of "*content that features the abuse of a minor, that includes suggestive or sexual content involving a minor or anyone that appears to be a minor, or that facilitates or promotes child sexual abuse*" (Community Guidelines, 'Harm to Minors'). Content in this context may include photos of real individuals, illustrations, animation, or text.
2. How are the ToS or Community Guidelines/Standards communicated?	Tumblr's Community Guidelines are available at: https://www.tumblr.com/policy/en/community The Tumblr Terms of Service are available at: https://www.tumblr.com/policy/en/terms-of-service
3. What are the service's policies and procedures for	Tumblr uses both user reporting and content moderation to prevent its platform being used for CSEA. User reporting is available across

preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	the platform where any form of content may be reported for suspected violation of the Community Guidelines. Content moderation uses a mix of machine-learning classification and human moderation by its Trust & Safety team to review and classify posts that may be in breach of its polices.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Where a user is found to have violated the Tumblr Community Guidelines, they receive a notice via email in which they are asked to explain or correct their behaviour and action may be taken against their account. Repeat violations of the Community Guidelines may result in permanent blog or account suspension. Tumblr states that it reserves the right to suspend accounts, or remove content, without notice, for any reason, but particularly to protect its services, infrastructure, users, and community.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, users receive notification via email and in a banner on the flagged content.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal content moderation decisions and account suspensions. Where content is flagged for a violation of Community Guidelines, the poster can request using the appeal mechanism for a further review. All reviews are carried out by the content moderation team. Content already reviewed by a member of the moderation teamt cannot be appealed, and this information will be published in a banner on the post.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No. Tumblr publishes transparency reports in respect of government requests for user information and for copyright- and trademark-related content removals only. It does not publish any data regarding enforcement of its Community Guidelines or content moderation decisions.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Transparency Reports for legal requests are published on a six monthly basis and are available at: https://transparency.automattic.com/tumblr/
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to Tumblr, a mix of machine-learning classification and human staff moderation is used to moderate content on the platform. Machine learning is largely used to detect potentially violative content. More nuanced contextual decisions are managed by human reviewers (Tumblr, n.d.[189]). All uploaded content is screened by content moderation software. If the software detects possible violative content in any media uploaded, it is escalated to a team of moderators for confirmation. A user reporting mechanism is also available where users can report any content they believe violates the platfor''s Community Guidelines. Content within the dashboard, blogs, tag pages, and

	search results may be reported. Pre-defined categories and context fields are used to describe the nature of the violation. All reports are reviewed by its Trust & Safety team who review the reported content and take the appropriate action. All suspected CSEA content is reported to child protection organisations and law enforcement around the world, including NCMEC as required under U.S. law.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, a total of 4,845 reports were submitted by Tumblr in 2021 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 4,511 reports submitted in 2021 (NCMEC, 2022[150]).

23. LinkedIn (Microsoft, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	LinkedIn does not give a specific definition of CSEA. However, it states that it has zero tolerance for child sexual exploitation. Under its Professional Community Policies, LinkedIn prohibits the sharing, posting or soliciting of any CSEA material. Use of the platform to "*in any way facilitate, encourage, or engage in the abuse or exploitation of children"* is prohibited. When it becomes aware of any apparent child exploitation, LinkedIn will report it to NCMEC. LinkedIn's Professional Community Policies also articulate a more general prohibtion on nudity or adult content. This includes any content including pornography that contains depictions of real or simulated sex acts, erotic literature, and other graphic depictions of sex acts performed alone or with others. As per its Terms of Service or User Agreement, users agree to abide by its Professional Community Guidelines when they sign up for the service.
2. How are the ToS or Community Guidelines/Standards communicated?	The LinkedIn Professional Community Policies are available at: https://www.linkedin.com/legal/professional-community-policies The LinkedIn User Agreement is available at: https://www.linkedin.com/legal/user-agreement#dos
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	LinkedIn uses a combination of automated safety technology, human staff content moderation to monitor and remove content violations. Users areencouraged to report any content such as conversations, posts, pages or groups, that violate its Professional Community Policy. Automated systems are used to detect spam and violative content.

4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	If an account, or content posted to that account, is found to have violated the Professional Community Policies or the User Agreement, LinkedIn may remove the content or place a restriction on the account. Depending on the severity of the violation, the account may be restricted indefinitely. The following are given as examples of conduct or activity that may result in account restriction: • An unusually large number of page views from the account. • The name used in the account profile is in violation of the User Agreement. • Inappropriate or illegal activity is detected on the account. • A history of repetitive abusive behavior on the account. • The account may have been compromised.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, according to LinkedIn, users will receive an in-app notification or email notifying them that content or activity does not comply with its policies.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, an appeals process is available. If a user believes that their account has been restricted or content removed in error, they can appeal and ask for a further review of the decision. The appeal process within the app is triggered by replying to the notification sent regarding the enforcement action. After a review is completed, the user will receive one of the following updates: • If the content doesn't go against the Professional Community Policies, it will be made available on LinkedIn. • If it is found that the post does go against its policies, only the user will be able to access the post.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, LinkedIn publishes a Transparency Report on actions taken on content that violated its Professional Community Policies and User Agreement (LinkedIn, n.d.[190]).
5.1 What information/fields of data are included in the TRs?	The only data provided is the total number of pieces of content removed and the category of violation.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No explanation is provided, though it may be assumed that content removed refers to the total individual pieces of content (conversations, posts, pages, groups) actioned for policy violation.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are published every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	LinkedIn's system of content moderation is built around three main layers which are used to filter out content that violates its policies whether in the feed or in private messages. Automated detection is described as the first layer of prevention. All content created on LinkedIn is automatically filtered for bad or violative content. According to LinkedIn, automated filters work within 300 milliseconds of creation content to prevent anyone but the author viewing the content. AI models are used to better identify and restrict similar content from being posted in the future.

	A second layer uses a combination of automated and human-led detection. AI systems may flag potentially violative content but requires further human review for confirmation. If found to be in breach of its policies, the content is removed from the platform. User reporting is regard as a third, further layer of abuse prevention. Users are encouraged to report any suspicious content on the platform which is then sent to the LinkedIn team of reviewers for further evaluation and removed if found to be in violation of the platform policies. In its 2021 Transparency Report, it is claimed that 99.6% of content violations were removed through automated processes (LinkedIn, 2022[191]).
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to the LinkedIn Transparency Report for the period January to June 2022, 1663 pieces of content were removed for violations of its policy on child exploitation (LinkedIn, n.d.[190]). This compares to 125 pieces of content removed in the previous six month. According to NCMEC, 201 reports were submitted by LinkedIn in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 110 reports submitted in 2021 (NCMEC, 2022[150]).

24. Douban (Information Technology Company, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided. Among its list of prohibited activities, Douban's Community Guiding Principles forbid use of the platform for spreading obscenity, abuse, harassment, use of threatening words to coerce others to obey or engage in unlawful behaviour. However, the Community Guidelines do not specify CSEA or offences of child pornography. Douba''s""Usage Agreement" require that all user behavior and published content should comply with its ""Community Guideline"" and other policies published by the service.
2. How are the ToS or Community Guidelines/Standards communicated?	The Douban Community Guidelines are available at: https://www.douban.com/about/guideline
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Douban's Legal Notice refers in a general way to its right to review content uploaded to its service and to delete material that is in violation of its policies. However, no further information regarding its overall approach to platform safety is provided.

4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	If an account is found to violate the Community Guidelines, or publishes content that violates the Community Guidelines, varying enforcement measures may apply. These include removal of the content, banning according the account associated with the violation, and/or termination of the the account without the opportunity for reinstatement.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, according to the Douban Help Centre, users are notified through their Douban mailbox of any enforcement decisions. Where content is posted that is suspected to contain content that violates laws and regulations or community guidelines, it is submitted for review. While being reviewed, the content is temporarily visible only to the user. If approved for posting, the content will be published automatically; if the review is not passed, the content will be automatically sent to the mailbox of to the account (the notification will not be sent if the mailbox is not set).
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal an enforcement decision as follows: • For an account ban, a user can log in on the Douban webpage and appeal according to the prompts on the page. • If banned from contributing posts, the user can appeal according to the prompts on the page when publishish content on the Douban webpage.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	No details are provided of Douban's methods for monitoring content on its service beyond reference to the review of suspected violations of its Community Guidelines.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Unknown.

25. Baidu Tieba (Baidu, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided. However, Baidu Tieba's Terms of Service prohibit any activity on its platform that spreads obscentity or pornography (2.3.7) or any content that is vulgar, obscene, or otherwise morally objectionable (2.3.11). The posting of any illegal or infringing remarks that contain obscene or pornographic material will result in the suspension or deletion of the account and reporting to the authorities (4.1).
2. How are the ToS or Community Guidelines/Standards communicated?	The Baidu Tieba Terms of Service are available at: https://gsp0.baidu.com/5aAHeD3nKhI2p27j8IqW0jdnxx1xbK/tb/eula.html
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	The platform has a reporting mechanism in place through which users can report any content that may breach its rules. Baidu Tieba reserves the right to withdraw services, suspend and delete account for violations of its policies. Baidu has also implemented special initiatives under its protection of minors policy dedicated to creating a safe and healthy online environment for minors with sustainable content services. This includes the establishment of a hotline for reporting "harmful information involving minors" to improve all links of services relating to the protection of minors. Also, under this policy Baidu has introduced a "teen network mode" with restricted functions and limits on screen time (Baidu, 2021[192]).
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Baidu Tieba's ToS state that it has the right to suspend or terminate the provision of its service to users for any reason including any of the following breaches of its policies: • Violation of laws and regulations or the provisions of its ToS; • Activity affecting user experience; • Activity creating potential safety hazards; • Violation of Baidu Tieb''s operating principles, or other management requirements of Baidu. Enforcement options including withdrawal of services and suspension or deletion of an offende''s account. In cases of serious violations, reports are made to the relevant law enforcement authorities.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No notification processes are specified.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeals processes are specifed.
5. Does the service issue transparency reports (TRs) specifically on content and/or	No.

behaviour related to CSEA?	
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Baidu deploys a combination of human and automated content moderation processes as well as user reporting to prevent its services being misused. However, no further details are available about its use of technology to detect violations of its policies or CSEA specifically.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Unknown.

26. Quora (Quora, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Quora has a dedicated policy on Sexual Exploitation and Abuse that expressly prohibits "*Sexually explicit or suggestive content (written or visual) involving children and minors"*. Examples of prohibited content include: • Sexually explicit or suggestive content involving children and minors (this can include content involving minors who are fully clothed and not engaged in overtly sexual acts) • Descriptions, whether real or fantasy, of engaging in sexual interactions with children and minors • Soliciting or exchanging sexually explicit or suggestive images involving children and minors, or Spaces dedicated to such behavior • Sending or requesting sexually explicit media to or from a child or minor • Content describing different ways to coerce children and minors into sexual interactions • Advocating for or glorifying minor–- non-minor relations • Sharing of external links to content that would violate this policy • Sharing sexually explicit content in Spaces that are directed towards minors (ages 13-17) • Grooming behavior, such as an adult attempting to engage in sexually explicit conversations with a minor on or off the platform Quora's policy on sex-related content states that while it allows

	discussion of adult sex-related topics, this is only allowed on pages restricted to adult topics and tagged with the appropriate Quora Adult topic(s). Under this policy, solicitation or content that advertises or promotes prostitution, sex trafficking, or sexual exploitation of children is not permitted (Quora, n.d.[193]). Spaces or communities that appear to function as a place for exchanging sexually explicit images of minors will be removed even if the images are being exchanged off the platform.
2. How are the ToS or Community Guidelines/Standards communicated?	Quora's Acceptable Use Policy is available at: https://www.quora.com/about/acceptable_use The Terms of Service are available at: https://www.quora.com/about/tos
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Quora uses internal content moderation systems and user reporting to prevent abuses of its policies. Content moderation is undertaken by its 'admins' or Quora staff moderation team members who monitor content and activity for violations of its policies. Quora also encourages its users to report violations using the platform's reporting tool. Quora states that it has the sole authority and final decision as to whether content or behavior violates its policies.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	According to Quora, enforcement actions may include but are not limited to written warnings, removal of content, adding warning tags to content, or the limitation or termination of a user's access to Quora. A user found to be in violation of its policies may receive an 'edit-block' or a ban for any of the following reasons: • If they vandalise content on the site that is editable by everyone, including questions and answer summaries. • If they engage in one or more actions which violate the 'Be Nice, Be Respectful' policy in questions, answers, or comments posted on the platform. • If they post a significant number of questions, answers, and/or comments that are''t helpful. • If they repeatedly violate Quora policies and/or do not change behavior after receiving a content warning. For violations involving CSEA, Quora states that it will remove it, report it to relevant authorities, and the account in question will be permanently banned.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, Quora moderation decisions are communicated to users through their profile settings. Quora also notifies affected users about requests from law enforcement or other government authorities, if legally permitted.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, if users have received a notification from the Quora Moderation team about their content being restricted or removed, they may appeal the decision by using the appeal mechanism provided on the notification from Quora.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of	Not applicable.

data are included in the TRs?	
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Quora uses both human reviewers and automated technology for moderation of content on its platform, including the detection of CSEA. However, very little information is available about the technologies in use or the scale of resources available to its moderation team. User reporting is available using the“"Repor”" tool which is available across the platform. Reports of potential policy violations are sent to Quora admins for review, who will determine whether a violation has occurred and of the content should be removed from the site.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, 2,242 reports were submitted by Quora in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 25 reports submitted in 2021 (NCMEC, 2022[150]).

27. Microsoft Teams (Microsoft, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Microsoft Teams is one of a number of Microsoft products that is covered by the Microsoft Services Agreement. This does not provide a specific definition of CSEA. However, the Microsoft Services Agreement includes a Code of Conduct that outlines what is allowed and what is prohibited when using a Microsoft account. This prohibits any activity that exploits, harms, or threatens to harm children. Microsoft states in its Digital Safety Content Report that it has a long-standing commitment to online child safety, and that it develops both tools and multistakeholder partnerships to help address this issue (Microsoft, n.d.[194]). As specified in its Code of Conduct, part of the Microsoft Services Agreeement, it prohibits “any activity that exploits, harms, or threatens to harm children” across its products and services – including but not limited to distribution of child sexual exploitation and abuse imagery (CSEAI), and grooming of children for sexual purposes.
2. How are the ToS or Community Guidelines/Standards communicated?	The Microsoft Services Agreement is available at: https://www.microsoft.com/en-us/servicesagreement[28]
3. What are the service’s policies and procedures for preventing its use in a manner	Microsoft deploys a combination of content moderation, staff moderation and user reporting to prevent misuse of its platforms and services and to prevent activity that may violate its ToS. Microsoft

that violates its ToS or Community Guidelines/Standards?	deploys tools to detect CSEA, including hash-matching technology (e.g., PhotoDNA) and other forms of proactive detection. Microsoft has made available in-product reporting for products such as OneDrive, Skype, Xbox, and Bing, whereby users can report suspected child exploitaiton or other violating content.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Microsoft states that it may disable a user's account for any suspicious activity or for violating the Microsoft Services Agreement such as by hosting photos, video or other content in violation of the Code of Conduct. Microsoft may also block delivery of a communication (like email, file sharing or instant message) or may remove or refuse to publish a user's content for any reason. When investigating alleged violations of its policies, Microsoft reserves the right to but states that it does not have the obligation to, review any content in order to resolve the issue. Microsoft also states that it removes content that contains apparent CSEAI. Microsoft also reports all apparent incidence of CSEA or grooming of children for sexual purposes to NCMEC via the CyberTipline, as required by U.S. law.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Details of notifications of content removals or account suspensions are not specified. The Microsoft Services Agreement refers in a general way to service notifications which it may send at its discretion in connection with a user's account via email or via SMS (text message), or by in-product messages.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal an account suspension by completing a form and submitting supporting information with their request for a review.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Microsoft's Digital Safety Content Report (DSCR) contains its Transparency Report covering actions that it has taken in relation to child sexual exploitation and abuse imagery (CSEAI), grooming of children for sexual purposes, terrorist and violent extremist content (TVEC), as well as non-consensual intimate imagery (NCII) (Microsoft, n.d.[194]). The report is an aggregate one for all Microsoft hosted consumer services and a breakdown by individual products - including OneDrive, Outlook, Skype and Xbox - is not available.
5.1 What information/fields of data are included in the TRs?	Microsoft includes the following information regarding enforcement for CSEA violations: • Content Actioned • Content Detected Proactively • Accounts Actioned • Accounts Reinstated
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	• *Content actioned* refers to removal of a piece of user-generated content from any of its products and services and/or block user access to a piece of user-generated content. With regard to Bing, "content actioned" may also mean filtering or de-listing a URL from the search engine

	index. • *Account actioned* refers to when Microsoft suspends or blocks access to an account, or restricts access to content within the account. • *Proactive detection* refers to Microsoft-initiated flagging of content on its products or services, whether through automated or manual review. • *Accounts reinstated* refer to actioned accounts that were fully restored including content and account access, upon appeal.
5.3 Frequency/timing with which TRs are issued	The Digital Safety Content Report is published every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to its Digital Safety Content Report, Microsoft deploys tools to detect child sexual exploitation and abuse imagery (CSEAI), including hash-matching technology such as PhotoDNA and other forms of proactive detection (Microsoft, n.d.[194]). "Proactive detection" refers to Microsoft-initiated flagging of content on its products or services, whether through automated or manual review. Microsoft developed PhotoDNA in partnership with Dartmouth College in 2009 to help find duplicates of known child sexual exploitation and abuse imagery (Microsoft, n.d.[52]). PhotoDNA is now an industry-standard technology used by organisations around the world and is deployed across Microsoft's consumer products and services.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	According to the Digital Safety Content Report for January-June 2022, Microsoft actioned 40,722 pieces of content and 10,207 consumer accounts associated with CSEAI or grooming of children for sexual purposes during this period (Microsoft, n.d.[194]). However, Microsoft does not break down information on CSEA violations by individual products and therefore the extent of prevalence of CSEA involving the use of Teams is unknown. Microsoft detected 98.7 percent of the content that was actioned, while the remainder was reported to Microsoft by users or third parties. Of the accounts actioned for CSEAI, 0.56 per cent were reinstated upon appeal. According to NCMEC, in 2022, Microsoft – Online Operations submitted 107,274 reports to its CyberTipline for child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 78,603 reports submitted by Microsoft in 2021 (NCMEC, 2022[150]).

28. IMO (PageBites, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	A specific definition of CSEA is not provided. However, IMO's Community Guidelines state that there is zero-tolerance for child sexual exploitation on the platform, which is taken to include though not limited to "*speech and acts such as spreading child nudity content, inciting underage users to commit crimes, etc.*" Where cases

	are confirmed, accounts are permanently suspended. The IMO Acceptable Use Policy (dated August 14, 2014) includes a reference which prohibits distributing defamatory, obscene, or unlawfully pornographic content. Through their use of its services, users are deemed to have accepted its policies and to have agreed to abide by its terms.
2. How are the ToS or Community Guidelines/Standards communicated?	IMO Community Guidelines are available at: https://imo.im/policies/community_guidelines.html IMO Terms of Service are available at: https://imo.im/policies/terms_of_service.html The Acceptable Use Policy is available at: https://imo.im/policies/acceptable_use_policy.html
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	IMO states that it is committed to creating a safe and friendly community at all times. While under no obligation to review content, it reserves the right to do so at any time. IMO may report any activity that it suspects violates any law or regulation to appropriate law enforcement officials, regulators, or other appropriate third parties. It also has the right to remove, screen, edit, or disable access to any content that that it considers to be in violation of its terms or otherwise harmful to the IMO service. IMO also encourages users to report any potential violations to the platform.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	The Community Guidelines refer to the removal of any harmful content detected and the temporary or permanent blocking of accounts according to the specific guidelines.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No notification procedures are specified.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeals processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.

6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to IMO's Community Guidelines, the platform deploys advanced artificial intelligence technology to prevent posting harmful content that could impact its community. Harmful content identified by its AI technology will be deleted and, for more serious cases, accounts will be blocked temporarily or deleted permanently according to the specific guidelines. It also states that it has a 24/7 global team which reviews all user reports and removes content and accounts that do not meet its guidelines. IMO encourages its users to report anything they come across that violates its community guidelines using the platfor''s report button. Harmful content can also be removed, it adds, through mutual co-operation among community members. For example, where a user sends potentially infringing content in a group chat message, users are encouraged to persuade them to stop posting or contact the group administrator for help. Users are also advised to report any potential criminal activities to local law enforcement.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Unknown.

29. Ask.fm (IAC [InterActiveCorp])

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	A specific definition of CSEA is not provided. However, Ask.fm's Terms of Use expressly prohibit the posting or sharing of any content that "*Is horrible, shocking, distressing, obscene or pornographic, contains any pictures of naked people, is sexually explicit, depicts graphic violence, shows or encourages physical, sexual or psychological exploitation of anyone but especially children, human trafficking specifically related to children, or sexual grooming, or any content which, by posting or sharing it, may be considered as an act of harassment to others*". The Community Guidelines further state that sexual and abusive content related to children is illegal and it will be reported to police. The Community Guidelines and Terms of Service were last updated in March 2022.
2. How are the ToS or Community Guidelines/Standards communicated?	The Terms of Use are available at: https://about.ask.fm/legal/en/terms.html The Community Guidelines are available at: https://about.ask.fm/community-guidelines/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	The ToS prohibit the posting or sharing of any content in breach of its ToS or Community Guidelines. Users are encouraged to report all potential violations of ASKfm policies using the reporting tools in the platform's feed, inbox, chat message or on the whole profile. ASKfm states that while it has no obligation to monitor access to or use of the service for violations of its policies, it reserves the right to do so for any purpose lding compliance with applicable llws. It also has the right to block or otherwise deal with content that it determines to be

	objectionable or in violation of its policies. Through accepting the ToS, users give their consent to ASKfm to monitor and block content that it considers to be harassing or bullying.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	The ToS state that ASKfm may suspend or terminate a user's access or a member's account for violation of its policies. It may also block users from accessing and using the service by using IP blockers or other solutions as appropriate. All reports of sexual exploitation of minors, child human trafficking, grooming and other serious illegal offenses received by its moderation team are reported to law enforcement and/or NCMEC. Profiles which are created only for prohibited activities or for distribution of prohibited content and spam are banned straight away.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	ASKfm does not commit to providing a notification for any enforcement decisions but responds to queries submitted to its Help Centre. In 2022, its support team received about 750 requests from banned users, and about 300 asked to unban their profiles. An internal violation counter is maintained and uses algorithms to detect offending profiles and ban them from using ASKfm.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal enforcement decisions including content removals and account bans. An online appeal form is available where users can submit further information and request a review.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. A Transparency Report is published on the website of ASKfm. A summary is provided of actions taken and reasons for banning accounts. Charts provide summaries of key metrics but there is no access to the actual data, limiting its use for transparency purposes.
5.1 What information/fields of data are included in the TRs?	The following fields of data are provided: • Number of account bans • Users banned by category • Ban reasons by category through use of hash list • Top banned reasons by country • Users banned by reason and by country • Violative content detected through automated technology • External requests to remove content
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No information is provided.

5.3 Frequency/timing with which TRs are issued	Transparency Reports are published annually.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	ASKfm deploys a combination of user community reporting as well human and automated content moderation. The platform uses pre-moderation tools which help define malicious content without the need for human input. Its pattern system includes words and word expressions in different languages, blacklisted websites, suspicious website links, webpages with adult content and other forms of clickbait. ASKfm has a range of partnerships with external organisations which it claims help to inform its moderation training materials and improve efficiency of its content moderation. Improvements to the moderation process introduced in 2021 included technical improvements to the moderation system interface.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. ASKfm states that 40 cases of CSAM were reported to authorities and 150 reports were submitted to NCMEC in 2022. According to NCMEC, in 2022, 26 reports were submitted by Ask.fm for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 117 reports in 2021 (NCMEC, 2022[150]).

30. Vimeo (Vimeo, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Vimeo does not provide a specific definition of CSEA. However, under its Acceptable Use Community Guidelines, content that exploits or endangers minors is prohibited. This is defined as content "*that was created through the exploitation of children or that is harmful to children"* and is taken to include: • Child sexual abuse material (CSAM) • Content that sexualises minors • Content that appeals to minors but contains adult themes • Videos that invite minors to engage in harmful or dangerous activities, whether through express invitation or example. Content featuring child nudity is not permitted and, according to Vimeo, will be removed regardless of the intention or who posts it, in order to mitigate risks of its being used for harmful purposes. The Vimeo Terms of Service similarly prohibit content or behaviour on the platform that exploits or endangers minors.
2. How are the ToS or Community Guidelines/Standards communicated?	Vimeo Terms of Service are available at: https://vimeo.com/terms#acceptable_use_policy and Vimeo Acceptable Use Community Guidelines are available at: https://vimeo.com/help/guidelines
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or	According to its Community Guidelines, Vimeo endeavors to review specific content that is flagged by its users, third parties, or by its software-based moderation systems in order to prevent the misuse of its services. It does not commit to review every piece of content

Community Guidelines/Standards?	uploaded to the platform. Vimeo states that it does not review content for all possible violations or to "pre-clear" any content before submission.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Violations of the Community Guidelines may result in suspension or removal of videos, account privileges, or an entire account. According to the platform, account removals will occur in severe cases, such as for wilful or repeated violation of its terms or for uploading extremely inappropriate content. Where an account is permanently removed, a user is banned from creating a new account. Vimeo states that if it locates any content suspected of containing CSAM, the account is immediately removed and the incident reported to NCMEC.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, Vimeo states that it endeavors to notify account holders of any enforcement decisions by emailing the registered email addresses on file. However, suspected CSEA is immediately removed without notification.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. If a user believes that a mistake has been made in moderating their account, they may submit a request to reconsider the decision. An Appeals Form is available for users to provide an explanation of why they believe a moderation decision was made in error. Vimeo undertakes to provide a response within 30 days and may seek additional information in the course of the review. If an appeal is granted, Vimeo will either restore the materials or allow them to be resubmitted. Vimeo reserves the right not to allow appeals in cases of extreme content, such as CSEA.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Vimeo applies a combination of user reporting, human review and automated content moderation on its platform. Users are encouraged to report any conduct or content that may violate the pl'tform's terms by either flagging it or by contacting its support function directly. Vimeo reviews all reports flagged by its users or detected by its automated systems but does not commit to reviewing or "pre-clearing" all content on its platform. However, Vimeo does not provide any further information about its automated processes for content moderation.

	Vimeo employs Thorn as a vendor who provides a hashset of over 5 million hashes of known CSAM. According to Vimeo, uploads from unpaid accounts (where nearly all CSAM uploads on the service occur) are scanned against this hashset. When a match is detected the following occurs: • the content is immediately removed • the account is immediately removed • all content uploaded to that account is automatically reported to NCMEC through their API. Should CSEA be missed by the by the hash detection process, and discovered by moderators the same procedure occurs. Vimeo states that it provides any necessary assistance to law enforcement.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, 368 reports were submitted by Vimeo in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 360 reports submitted in 2021 (NCMEC, 2022[150]).

31. Medium (A Medium Corporation.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No specific definition of CSEA is provided. However, content which promotes the sexual, violent, or other exploitation of minors, including the sexualisation of fictional minors is expressly prohibited in its Community Guidelines('the Medium Rules). By registering for the service and signing in to use the platform, users agree to its terms and to use it in accordance with applicable laws.
2. How are the ToS or Community Guidelines/Standards communicated?	The Medium Rules are available at: https://help.medium.com/hc/en-us/articles/213477928-Medium-Rules The Medium Terms of Service are available at: https://policy.medium.com/medium-terms-of-service-9db0094a1e0f
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Medium relies on both user reporting as well as its own platform moderation systems to prevent its misuse or violations of its policies. According to Medium, each user or participant is responsible for maintaining the platform's standards and is encouraged to report any violations they may come across. Medium states that it reserves the right to suspend accounts or remove content, without notice, for any reason, to protect its services, infrastructure, users, or community. Attempts to evade suspension by creating new accounts or posts results in termination of those accounts and posts.
4. What are the service's policies and procedures for	Violations of its rules may result in consequences such as account restrictions, limited distribution of posts, and suspension of an

enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	offender's account. Medium states that it has the sole authority and final decision as to whether content or behaviour violates its rules. Under its policies, Medium may suspend or limit the distribution of controversial or extreme content at its discretion, including potentially harmful misinformation and intentionally deceptive disinformation. For all reported content, it takes into account factors such as newsworthiness, the context and nature of the posted information, reasonable likelihood, breadth, and intensity of foreseeable social harm, and applicable laws.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, account holders are notified when Medium investigates or disables content associated with a potentially violative account, unless the account is automated or operating in bad faith, or if notifying the offender is likely to cause, maintain or exacerbate harm to someone.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. If a user believes their content or account has been restricted or disabled in error, or believe there is relevant context that the platform was not aware of in making the decision, they can appeal by emailing the platform.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No. Medium issued a Transparency Report in 2015 in relation to law enforcement requests for user information or content removal (Medium, 2015[195]). This reported zero requests. No subsequent reports have been issued.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Medium uses a combination of user reporting and platform-level content moderation. According to Medium, it evaluates flagged and reported content according to its rules and takes appropriate actions against violations of them, including warnings, suspension, and decreased distribution. Medium does not provide any further information about its content moderation procedures or monitoring algorithms.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, just two reports were submitted by Medium for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 113 reports submitted in 2021 (NCMEC, 2022[150]).

32. LINE (Line Corporation)

1. How is online child sexual exploitation and abuse (CSEA)	No definition of CSEA is provided. However, the Line Terms and Conditions of Use prohibt the posting or transmitting of any content

defined in the Terms of Service (ToS) or Community Guidelines/Standards?	that contains explicit sexual expression or that amounts to child pornography or child abuse. More generally, the service prohibits any content that violates applicable laws and regulations or that may be in violation of public order, morals or customs.
2. How are the ToS or Community Guidelines/Standards communicated?	The Terms and Conditions of Use are available at: https://terms.line.me/line_terms?lang=en
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	LINE states that it may check and confirm content submitted by users to the extent permissible under law, or when it is necessary to confirm compliance with related laws and regulations or the provisions set out in its terms. However, it further states that it is not obligated to undertake such confirmation.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	LINE uses a combination of user reporting and platform-level moderation to review unencrypted content on its platform. If LINE believes that a user has violated or may violate any of its terms or applicable laws or regulations, it reserves the right to restrict access to the service and to delete content without providing prior notice to the user.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No information is given about notification procedures. LINE's terms state that it reserves the right to restrict access to or withdraw services or remove content without prior notice to the user.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeals process is specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. LINE issues a six monthly Transparency Report in respect of content moderation.
5.1 What information/fields of data are included in the TRs?	Data is included for the following areas: • Percentage of content Suspended by Automatic Check • Percentage Suspended by Manual Check • Breakdown of types of suspended content by manual check • Content services covered by the report
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No details are specified.
5.3 Frequency/timing with which TRs are issued	Every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the	LINE employs a two-step process to monitor posts on the platform's Timeline (Current LINE VOOM), as well as its LINE LIVE, LINE Manga and other LINE services. In the first instance, content posted by users on supported LINE services is checked by its automated monitorings system to ensure that it does not contain any prohibited

service use to detect CSEA?	language, break any service rules, or violate the ToS or relevant laws. If objectionable content is found by the monitoring system, it is immediately suspended after being posted. Next, a monitoring team checks any content the monitoring system cannot classify. The monitoring team compares the content against a set of evaluation criteria and previous examples to make a decision on whether or not the content is safe. If the monitoring team determines the posted content is in violation of the terms of service or relevant laws, it is suspended. This two-step process, the company states, is designed to help ensure that any post that violates the terms of service or relevant laws are not further circulated on LINE's various community platforms (Line, 2021[196]). LINE is unable to monitor any message a user sends/receives on a regular LINE chat room unless the user sends unencrypted chat data to LINE by using the reporting tool.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. In its Transparency Report for the period January to June 2021, LINE reports that 23% of the total suspended content of approximately 15 million items was for the category of "Obscene content: Child pornography, videos of sexual content and genitalia, etc." While a precise breakdown is not available, it may be assumed that a proportion of such content included CSEA.

33. Picsart (Picsart, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Picsart does not specifically define CSEA. However, its Community Guidelines under the heading of Minor Safety prohibit any content that depicts or promotes the sexual exploitation of children, including illustrated or computer-generated content of that nature. This is taken to include: • Content depicting nudity or sexual activity involving minors. • Content depicting sexual fetishism or arousal involving minors. • Content depicting sexually suggestive language objectifying or otherwise involving minors. • Accounts created for the sole purpose of sexualising or inappropriately admiring children. The Community Guidelines further prohibit the posting of any content that contains: • links to third-party sites that host material involving the sexual exploitation of minors. • content depicting child abuse or the infliction of physical or emotional trauma or other harm on a minor. • content that promotes physical abuse, neglect, or other forms of abuse towards minors. • content that promotes or glorifies pedophelia.

	Users are also expressly forbidden to follow or like users on the platform for the purposes of making sexual remarks or engaging in grooming behavior, including building an emotional relationship with a minor for the purposes of sexual abuse, exploitation, or trafficking. Picsart states that it is deeply committed to minor safety and has zero tolerance for content or acts that jeopardise the safety of children. Any content that violates its Community Guidelines on minor safety is reported to relevant authorities or law enforcement. Picsart's Terms of Use further emphasise that users are responsible for all activity on their account and must not engage in any conduct that is illegal, abuses others or misuses the service.
2. How are the ToS or Community Guidelines/Standards communicated?	Picsart's Community Guidelines are available at: https://picsart.com/community-guidelines Picsart Terms of Use are available at: https://picsart.com/terms-of-use
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Picsart states that keeping its platform safe is a top priority and that its Trust & Safety team thoroughly investigates all reports of violations. Picsart will remove content that violates its Community Guidelines and restrict or ban accounts with severe or repeated violations. As stated in its ToS, content remains the responsibility of the person who posts it, and while its terms prohibit certain conduct and content on the service, Picsart may not monitor or control any such content posted.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	As stated in its Community Guidelines, Picsart may suspend access to certain features or terminate the account of offenders. Content that violates its Community Guidelines may be removed. In certain circumstances, it may also report an account to the relevant authorities or law enforcement. According to Picsart, determining whether there has been a violation of its Community Guidelines can be nuanced, and it reserves the right to make decisions considered appropriate for the Picsart community.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, according to Picsart, users are notified of actions taken. However they do not notify of high risk enforcement actions (including CSEA) unless required by law enforcement.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	A formal appeals process is not available. According to Picsart, users may write to help@picsart.com , however a review of any enforcement actions is not guaranteed.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included	Not applicable.

in the TRs	
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Users are encouraged to report any content or conduct that they believe may violate the platform's policies. According to the Picsart Community Guidelines, all reports of violations are investigated by its Trust & Safety team. Picsart has partnered with Thorn to help eradicate and prevent the spread of child sexual abuse material. According to Picsart, the platform employs specialised moderation teams to ensure platform safety. This team works 24/7 across multiple languages. Picsart also reports that it has dedicated Investigations and Law Enforcement Response personnel who work to resolve escalated safety or legal issues, including CSAM.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, 549 reports were submitted by Picsart in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 316 reports submitted in 2021 (NCMEC, 2022[150]).

34. Discord (Discord, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA Is provided. However, Discord's Community Guidelines contains a Youth Safety policy that prohibits users from soliciting, sharing, or making attempts to distribute content that depicts, promotes, or attempts to normalise child sexual abuse. Users are prohibited from posting content "*that in any way sexualizes children (...) (including) real as well as manipulated media, animation (such as lolicon), and any type of digital creation*". Discord also prohibits users from soliciting sexual content from or engaging in any sexual conduct ("grooming") with anyone under the age of 18. Discord's Terms of Service forbids any content or conduct that is illegal, or use of the services to do harm to oneself or others. Discord further states that it has a zero-tolerance policy for anyone who endangers children. Users who upload abuse material of minors to Discord are reported to NCMEC and removed from the platform. Grooming and endangerment cases are quickly escalated to the proper authorities (Discord, 2022[197]).
2. How are the ToS or Community Guidelines/Standards communicated?	The Discord Community Guidelines are available at: https://discord.com/guidelines The Discord Terms of Service are available at: https://discord.com/terms Users are required to agree to Discord's Terms of Service and Community Guidelines when registering to use the platform. Any changes and updates to them are communicated to users ahead of time via in-product communications and emails associated with their

	account.
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Discord implements a range of policies to uphold its Community Guidelines and prevent misuse of its services. Discord receives reports from users and trusted reporters, reports from volunteer community moderators, and also employs human staff review to monitor for violations. Discord uses technology such as PhotoDNA to proactively detect CSAM. When Discord has data suggesting that a user is engaging in illegal activity or violating their policies, they may investigate their activity on Discord, and their messages to proactively detect accomplices and determine whether violations have occurred (Discord, 2022[197]). According to the platform, Discord is not an anonymous platform. Users may be identified by reference to verified email addresses and IP addresses, and conversations are not end-to-end encrypted. The company investigates concerning behaviour when alerted and takes action as appropriate. Discord's policies also state that relevant off-platform behavior may be considered when assessing violations of specific Community Guidelines. Off-platform behaviour refers to any activity taking place outside of Discord, either in other digital spaces or in a physical community. This may include inappropriate contact with minors (or "grooming") made on other platforms or in a physical space. If it becomes aware of any such behaviour, Discord'sTrust & Safety team may launch an investigation into an account, including reviewing the user's activity and posts.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Subject to applicable law, Discord reserves the right to suspend or terminate a use''s account and/or access to some or all of its services with or without notice, at its discretion, for breaches of its policies. Discord's Trust & Safety team reviews reports by users, moderators, or trusted reports. When someone violates its guidelines, Discord may take a number of enforcement steps against them including: issuing warnings; removing content; suspending or removing the account(s) and/or server(s) responsible; and potentially reporting to law enforcement. Discord reports all cases of CSEA to NCMEC.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	The Discord Terms of Service state that the platform will give advance notice for any account termination if reasonable to do so or required by applicable law. Discord's Transparency Report explains that warnings are issued to accounts and servers for most violations, and are frequently used to correct problematic behavior that does not require immediate permanent removal from the platform. Regarding CSEA, Discord does not issue warnings but rather immediately disables and reports the account to NCMEC and removes the content.
4.2. Are there processes under	Appeals are not available in cases involving CSEA. However, in

which users can appeal content removals, account suspensions or other enforcement decisions?	general, users may appeal any enforcement decision by submitting an appeal to the Trust and Safety team using the online contact form. According to its Transparency Report, Discord takes seriously all user appeals and considers any additional context or information it may not have known at the time of the original decision. Discord reinstate accounts if it is determined that a mistake was made, or if it determines that the user (in good faith) has recognised the violation (of a lower-harm issue only) and will abide by the Community Guidelines once back on Discord (Discord, 2022[197]).
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, Discord issues a Transparency Report every three months which includes information on enforcement decisions and government information requests. Discord's Transparency Reports provides details of enforcement decisions for category violations as defined in its Community Guidelines. From 2021, Discord recategorised its safety matrix form 14 categories to 1 categories. The category of Child Safety replaces the previous category of CSAM and Exploitative Content. CSAM now appears as a subcategory of Child Safety.
5.1 What information/fields of data are included in the TRs?	The following fields of data are included in the Transparency Report: • User reports received by Category • Reports Received by Category • Action Rates of Reports Received by Category • Actions Taken (Account and Server Warnings) • Actions Taken (Accounts Disabled) • Actions Taken (Servers Removed – Proactive vs Reactive) • Appeals
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No information is provided.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are issued every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Discord deploys a combination of community/user reporting, human moderator review and automated technologies to detect CSEA. • Discord scans images uploaded to the platform using industry-standard PhotoDNA to detect matches to known child sexual abuse material. • CSEA can be reported by users, including community moderators and trusted reporters. • Auto moderation is also available within communities set up on Discord and supplements manual moderation by community moderators. Auto moderation can be used to set up keyword filters that automatically trigger moderation actions such as blocking messages containing specific keywords from being sent and logging flagged messages as alerts for review. • The platform has also established a Discord Moderator Academy which team trains and empowers moderators to

	keep their communities safe and healthy. • Discord works closely with industry groups and partner organisations to help make Discord a safe and welcoming place for all. This includes the Family Online Safety Institute and the Technology Coalition. Additionally, as a member of the EU Internet Forum, Discord collaborates with other companies and governments to exchange best practices on combatting CSEA online. According to Discord, its Trust & Safety team works with cutting-edge technology to detect and respond to abuse, both proactively and from user reports. Approoximately, 15% of Discord's employees work in Trust and Safety.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to its Transparency Report for the period April – June 2022, Discord disabled 532,498 accounts and removed 15,163 servers for Child Safety in the second quarter of 2022. Servers removed for CSAM increased to 6,640, up from 1,271, with a proactive removal rate of 95%, an increase from the 52% CSAM server proactive removal rate in the first quarter of 2022, which, according to Discord, resulted from the introduction of new tools built to identify and detect servers hosting this content (Discord, 2022[197]). According to the TR, 58,179 accounts and 57,943 reports of images or videos, many flagged through PhotoDNA, were reported to NCMEC. 290 grooming or endangerment reports were also delivered to NCMEC in 2022 (Discord, 2022[197]). According to NCMEC, Discord submitted a total of 169,800 reports in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 29,606 reports submitted in 2021 (NCMEC, 2022[150]).

35. Twitch (Amazon.com, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No specific definition of CSEA is provided. The Twitch Community Guidelines contain a dedicated Youth Safety policy which expressly prohibits any content or activity that endangers youth including content that features or promotes child sexual abuse material (CSAM), sexual misconduct or grooming of youth defined as minors under 18. Examples of prohibited content and activity include: • sexually explicit content or sexualized images of youth • sharing links to third-party sites that contain content prohibited by this policy • content that promotes, encourages, provides instruction to, or admits participation in the sexual exploitation or sexualisation of youth • content that constitutes or facilitates inappropriate

	interactions with youth, including grooming, purposefully exposing youth to sexually explicit language or sexual material, and engaging in sexual conversations • content that attempts to exploit youth by coercing money, favors or intimate imagery with threats to expose intimate imagery or information • content depicting nudity of youth • identifying alleged victims of CSAM by name or image. The Community Guidelines define minors as anyone under 18 years. It is further stated that all illegal content or activity is reported to NCMEC. The Twitch Terms of Service prohibit the use of Twitch Services for any illegal purpose, or in violation of any local, state, national, or international law or regulation.
2. How are the ToS or Community Guidelines/Standards communicated?	The Twitch Community Guidelines are available at: https://safety.twitch.tv/s/article/Community-Guidelines?language=en_US The Twitch Terms of Service are available at: https://www.twitch.tv/p/en/legal/terms-of-service/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Twitch states that it endeavours to promote safety on the platform and prevent its misuse through a combination of user community moderation of its content supported by automated processes and human staff review. When using its services, all users agree to abide by the platform's policies and all applicable laws. The Community Guidelines provide the policy basis of what is permitted on the platform and is enforced through a series of graduated sanctions. Twitch's Trust & Safety team monitors content and reviews reports by users, moderators, or trusted reporters in upholding its guidelines.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Twitch reserves the right to suspend any account at any time for any conduct that it determines to be inappropriate or harmful. Enforcement actions may include: removal of content, a strike on the account, and/or suspension of account(s). According to the platform, a number of factors are taken into consideration when reviewing reports of violations, including the intent and context, the potential harm to the community, legal obligations and other factors. If any content that contains the violation has been recorded on its service, it will be removed. Enforcement depends on the nature of the violation, and can range from a warning, a temporary suspension (1-30 days), or for the most serious offenses, an indefinite suspension from Twitch. The various enforcement options are explained as follows: • *Warnings*: A warning is a courtesy notice for some violations. Twitch may also remove content associated with the violation. Repeating a violation for which a warning has already been issued, or committing a similar violation, will result in a suspension.

	• *Temporary suspension*: a temporary suspensions may range from one to 30 days. After the suspension is complete, a user will be able to access the service once again. A record is kept of past violations, and multiple suspensions over time can lead to an indefinite suspension. During a suspension, users will not have access to watching streams, broadcasting, chatting, creating other accounts or appearing/participating in the stream of a third party channel. • *Indefinite suspension*: this is applied for the most serious offenses, including engagement with CSAM or the sexual exploitation, sexual misconduct or grooming of youth, and involves the immediate and indefinite suspension of an account with no opportunity to appeal.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes. According to Twitch, notifications are issued to account holders specifying the type of enforcement decision (a warning, content removal, or temporary/indefinite account suspension).
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal enforcement decisions on their account using the platform's appeals portal. Appeals can be requested for enforcements issued within the previous 60 days, or for the most recent enforcement if currently serving an indefinite suspension. According to Twitch, appeals are reviewed in the order they are received and no guarantee is given that enforcements will be overturned. For suspensions of 30 days or less, a user may only submit 1 appeal per enforcement. For indefinite suspensions, only 1 appeal may be submitted in a 6 month period per indefinite enforcement.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes, Twitch issues Transparency Reports every six months (Twitch, n.d.[198]).
5.1 What information/fields of data are included in the TRs?	The following fields of data are reported in the Transparency Report for the period July to December 2021: • *Moderation Coverage*: percentage of minutes watched by automod and by human moderators (including third party moderators). • *Proactive and Manual Removals of Chat Messages*: volume of chat messages removed manually and proactively. • *Channel Enforcement Actions:* timeouts, channel bans and total number of enforcement actions per channel. • *Reports made on Twitch* total number of user reports submitted by category. • *Enforcements*: total number of enforcement actions taken with further detail given by category of violation. Twitch reports on the number of NCMEC Cyber Tips submitted by the platform. The report for H2, 2021 shows an increase from 2,615 in H1 2021 to 4,006 in H2 2021 (+53% HoH). This equates to a 73% increase in tips per thousand hours watched.

5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No further information on its methodologies is provided.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are issued on a six monthly basis.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Twitch deploys a combination of community user reporting, human moderation and automated technologies to detect violations of its Community Guidelines, including CSEA. As noted in its Transparency Report, the approach to safety on the platform is tailored to the specific nature of the service which involves interaction around live, often ephemeral content. Twitch describes its approach as a layered one which, beginning with its Community Guidelines as the foundation of its safety strategy, applies progressive safety levels across the service. Service level safety comprises the following distinct functions: • *Machine Detection:* Automated technologies are used to scan content on the service and flag it for review by human moderators. Ephemeral, live-streaming content is challenging for machine detection but is, according to the company, viable and useful on Twitch. • *User Reporting:* User reporting is regarded as particularly effective on Twitch because the vast majority of the content on Twitch - video and chat - is public. Creators, mods, and viewers are encourgaed to report content that violates its policies. User reports are sent to a team of content moderation professionals to review. • *Review and Enforcement:* this is the responsibility of trained professionals who review user reports and content that is flagged by its machine detection tools. Reports are prioritized so that the most harmful behavior can be dealt with most quickly. Review time for any given report is dependent on a number of factors including the severity of the report, the availability of evidence to support the report, and the current volume of the report queue. Twitch has a dedicated Law Enforcement Response (LER) team to investigate the most egregious reports, and liaises with law enforcement as necessary.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, a total 14,508 reports were submitted by Twitch in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 6,629 reports submitted in 2021 (NCMEC, 2022[150]).

36. Likee (BIGO Technology PTE. LTD.)

1. How is online child sexual exploitation and abuse (CSEA)	Likee's Community Guidelines prohibit a range of CSEA-related content which is defined as "*any content involving child abuse, sexual*

defined in the Terms of Service (ToS) or Community Guidelines/Standards?	*content involving minors and any conduct or content that may affect the safety of minors*". As stated in its Community Guidelines, child abuse is taken to refer to physical or psychological harm caused to minors; physical abuse; the intentional infliction of physical harm on a child; and psychological abuse as in the harming of minors through the threat of physical or sexual violence, bullying, or insults. The following content is prohibited on the platform: • content describing the physical or psychological abuse of minors • content encouraging or extorting the propagation of pornography by children • sexual content involving minors • content including sexual or pornographic language involving minors Likee states that it is committed to the safety of minors. It prohibits the publication of content that puts minors at risk. It does not allow the description or transmission of content involving the abuse of minors, nude images of minors, or sexual exploitation of minors. In addition, it does not allow content depicting minors engaging in illegal activities. The following content is specifically prohibited: • content that shows the private parts of minors • content describing the sexual exploitation of minors • content describing sexual behavior involving minors More generally, Likee states that it does not allow pornographic content, including animated pornographic content for the reasons that there are many risks associated with sexualized content, such as legal consequences in certain jurisdictions as well as the fact that in some cultures, sexual content can be offensive. In the Terms of Use, users agree that they will not use or attempt to use any method, device, software or technologies to harm others or interfere with the functioning of Likee Services.
2. How are the ToS or Community Guidelines/Standards communicated?	Likee Community Guidelines are available at: https://mobile.likee.video/live/page-about/community.html. Likee Terms of Use are available at: https://likee.video/live/page-about/user-agreement.html
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	The Likee Terms of Use broadly state that users have responsibility for any user generated content (UGC) they submit, including its legality, reliability, accuracy and appropriateness. Likee states that while it sometimes reviews UGC contributed by users, it is not obligated to do so. As such, it may review, monitor, display, reject, refuse to post, store, maintain, accept or remove any UGC posted, and may delete, move, re-format, remove or refuse to post or otherwise make use of UGC without notice to users. In this context, it may address UGC that comes to its attention that may be deemed offensive, obscene, violent, harassing, threatening, abusive, illegal or otherwise objectionable or inappropriate.

	The platform also has a Parental Controls feature that filters out content on the Likee app that may not be age-appropriate for a younger audience. In addition, the feature automatically makes all content published by a child account private by default, blocks live broadcasts, publishing, and viewing of certain content, and terminates in-app communication and messaging. Furthermore, displaying users' location and commentary information is automatically set to 'off' when the parental control mode is on. Parents can enable and disable the Parental Control function using a secure password.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Likee states that it strives to protect all its users by ensuring content on the platform meets the standards set out in its Community Guidelines. Accordingly, it states that it will remove any content that violates its policies. User accounts involved in serious or repeated violations will be penalised or banned. If necessary, violations will be reported to relevant legal authorities and the platform will cooperate in investigations to ensure community safety. Users who breach Likee's Terms of Service or Community Guidelines may be sanctioned in a number of ways depending on the nature of the breach as well as other factors. Sanctions may include content removal, restrictions on certain features or account ban. Likee also states in its ToS that it reserves the right to fully cooperate with any law enforcement authorities or court order requesting or directing it to disclose the identity or other information of anyone providing any UGC on or through Likee Services.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	According to Likee, the platform provides real time in-app notifications to users to inform them of content removal, account bans and feature bans. Likee states that its policy is to notify users of requests for their information prior to disclosure where it is legally permissable to do so.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	While not specified publicly, Likee reports that users can appeal against sanctions for content removal and account bans by clicking the "submit appeal" button. For content that is made "hard to find" or where content is "not actively promoted", users are unable to appeal. Users can appeal "feature bans" and "device bans" (which arise as a result of an accumulation of content-level violations), but they can not appeal the underlying content level violations and/or account bans.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.

5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	While this information is not publicly available, Likee states that it places emphasis on its own proactive moderation to review and detect uploaded content in breach of its policies. Users may report and flag potentially harmful material using the platform reporting tools. Trusted third-party flaggers also monitor and flag harmful content.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	No information available.

37. Skype (Microsoft, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Skype is one of a number of Microsoft products that is covered by the Microsoft Services Agreement. This does not provide a specific definition of CSEA. However, the Microsoft Services Agreement includes a Code of Conduct that outlines what is allowed and what is prohibited when using a Microsoft account. This prohibits any activity that exploits, harms, or threatens to harm children. Microsoft states in its Digital Safety Content Report that it has a long-standing commitment to online child safety, and that it develops both tools and multistakeholder partnerships to help address this issue (Microsoft, n.d.[194]). As specified in its Code of Conduct, part of the Microsoft Services Agreeemnt, it prohibits “any activity that exploits, harms, or threatens to harm children” across its products and services – including but not limited to distribution of child sexual exploitation and abuse imagery (CSEAI), and grooming of children for sexual purposes.
2. How are the ToS or Community Guidelines/Standards communicated?	The Microsoft Services Agreement is available at: https://www.microsoft.com/en-us/servicesagreement
3. What are the service’s policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Microsoft deploys a combination of content moderation, staff moderation and user reporting to prevent misuse of its platforms and services and to prevent activity that may violate its ToS. Microsoft deploys tools to detect CSEA, including hash-matching technology (e.g., PhotoDNA) and other forms of proactive detection. Microsoft has made available in-product reporting for products such as OneDrive, Skype, Xbox, and Bing, whereby users can report

	suspected child exploitaiton or other violating content.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Microsoft states that it may disable a user's account for any suspicious activity or for violating the Microsoft Services Agreement such as by hosting photos, video or other content in violation of the Code of Conduct. Microsoft may also block delivery of a communication (like email, file sharing or instant message) or may remove or refuse to publish a user's content for any reason. When investigating alleged violations of its policies, Microsoft reserves the right to, but states that it does not have the obligation to, review any content in order to resolve the issue. Microsoft also states that it removes content that contains apparent CSEAI. Microsoft also reports all apparent incidence of CSEA or grooming of children for sexual purposes to NCMEC via the CyberTipline.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Details of notifications of content removals or account suspensions are not specified. The Microsoft Services Agreement refers in a general way to service notifications which it may send at its discretion in connection with a user's account Microsoft account via email or via SMS (text message), or by in-product messages.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal an account suspension by completing a form and submitting supporting information with their request for a review.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Microsoft's Digital Safety Content Report (DSCR) contains its Transparency Report covering actions that it has taken in relation to child sexual exploitation and abuse imagery (CSEAI), grooming of children for sexual purposes, terrorist and violent extremist content (TVEC), as well as non-consensual intimate imagery (NCII) (Microsoft, n.d.[194]). The report is an aggregate one for all Microsoft hosted consumer services and a breakdown by individual products - including OneDrive, Outlook, Skype and Xbox - is not available.
5.1 What information/fields of data are included in the TRs?	Microsoft includes the following information regarding enforcements for CSEA violations: • Content Actioned • Content Detected Proactively • Accounts Actioned • Accounts Reinstated
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	• *Content actioned* refers to removal of a piece of user-generated content from any of its products and services and/or block user access to a piece of user-generated content. With regard to Bing, "content actioned" may also mean filtering or de-listing a URL from the search engine index. • *Account actioned* refers to when Microsoft suspends or blocks access to an account, or restricts access to content within the account.

	• *Proactive detection* refers to Microsoft-initiated flagging of content on its products or services, whether through automated or manual review. • *Accounts reinstated* refer to actioned accounts that were fully restored including content and account access, upon appeal.
5.3 Frequency/timing with which TRs are issued	The Digital Safety Content Report is published every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to its Digital Safety Content Report, Microsoft deploys tools to detect child sexual exploitation and abuse imagery (CSEAI), including hash-matching technology such as PhotoDNA and other forms of proactive detection (Microsoft, n.d.[194]). "Proactive detection" refers to Microsoft-initiated flagging of content on its products or services, whether through automated or manual review. Microsoft developed PhotoDNA in partnership with Dartmouth College in 2009 to help find duplicates of known child sexual exploitation and abuse imagery (Microsoft, n.d.[52]). PhotoDNA is now an industry-standard technology used by organisations around the world and is deployed across Microsoft's consumer products and services.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	According to the Digital Safety Content Report for January-June 2022, Microsoft actioned 40,722 pieces of content and 10,207 consumer accounts associated with CSEAI or grooming of children for sexual purposes during this period (Microsoft, n.d.[194]). However, Microsoft does not break down information on CSEA violations by individual products and therefore the extent of prevalence of CSEA involving the use of Skype is unknown. Microsoft detected 98.7 percent of the content that was actioned, while the remainder was reported to Microsoft by users or third parties. Of the accounts actioned for CSEAI, 0.56percent were reinstated upon appeal. According to NCMEC, in 2022, Microsoft – Online Operations submitted 107,274 reports to its CyberTipline for child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 78,603 reports submitted by Microsoft in 2021 (NCMEC, 2022[150]).

38. VK (Mail.Ru Group)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	VK does not supply a specific definition of CSEA. However, its Terms of Service prohibit the sharing of any content that infringes on the rights of minors or "*that is vulgar or obscene, contains pornographic images and texts or sexual scenes with the participation of minors*" VK Top, 6.3.4). The VK Safety Guidelines and Platform Standards also expressly prohibit the solicitation or sexual exploitation of children or adults and child pornography.

2. How are the ToS or Community Guidelines/Standards communicated?	The VK Terms of Service are available at: https://vk.com/terms The VK Safety Guidelines and Platform Standards are available at: https://vk.com/safety
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	VK employs what it calls a hybrid method of moderation which it says can respond quickly to reports from users, social organisations, and government regulatory agencies, in addition to its proactive internal monitoring. VK encourages users to report any content or conduct that breaches its policies using the platform's reporting mechanism. According to VK, its team reviews every report and removes content that violates the VK Terms of Service or applicable laws. It also blocks communities and profiles used by scammers to spread offending content. According to the platform, its response time is never more than an hour and is usually a matter of minutes.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	VK immediately removes any content found to be in violation of its policies or platform standards. VK reserves the right but is not obliged to monitor content on the platform. The ToS state that it may delete or remove any content or users without notice at its own discretion for any reason which it believes breaches its terms or threatens the security of other users or third parties. This includes the right to right to remove a user's personal page and/or suspend, limit or terminate the user's access to any of the VK site services.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes. While it is not specified publicly, when a Community page or a personal profile is blocked, according to VK, users are shown a notification with the reasons for blocking.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, according to VK, users can contact the support service, and based on the information provided by the user, the decision may be revised.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	VK does not issue Transparency Reports. However, statistics on enforcement decisions are included within the Safety Guidelines under the various categories of content outlined in its Safety Guidelines (VK, n.d.[199]).
5.1 What information/fields of data are included in the TRs?	Data included for various categories of content outlined in the Safety Guidelines as follows: • number of pieces of content blocked • number of profiles blocked • number of communities blocked
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No information is given.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the	VK's hybrid content moderation approach uses a combination of internal monitoring, user reports and automated detection technologies. Its monitoring system uses a neural network to automatically look for and block dangerous content. VK states that all hashtags related to harmful or illegal topics such as CSEA

service use to detect CSEA?	automatically appear in its system as soon as they are posted, enabling a quick response and removal of violative content. Any content that is connected with child exploitation, including pornographic materials and information regarding child trafficking or prostitution is quickly deleted, according to the platform.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. As reported by VK in Section 5 above, according to VK, in 2021, more than 1.3 million pieces of content, as well as 490,000 unique profiles and 10,000 communities, were blocked for distributing child exploitation or child sexual abuse material (VK, n.d.[199]).

39. Xigua Video (ByteDance Technology Co.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Xigua Video does not provide a specific definition of CSEA. However, the User Service Agreement prohibits the uploading or sharing of any content that violates the lawful rights and interests of minors or harming the physical and mental health of minors. More generally, users are required to abide by all applicable laws and agree not to disseminate any content that is unlawful or in violation of the service's policies as outlined in the User Service Agreement. Terms of Use for Minors require parental consent for users under the age of 18. Minors must also abide by the "National Youth Network Civilization Convention".
2. How are the ToS or Community Guidelines/Standards communicated?	The Xigua Video User Service Agreement is available at: https://www.ixigua.com/user_agreement/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	The platform has a user complaint and reporting feature and invites users to participate in maintaining safety on the platform. Users are encouraged to report to the company any violations of laws and regulations, illegal communication activities, illegal and harmful information, etc. The company states that it will accept and handle all complaints and reports in a timely manner.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	The platform reserves the right to take action against any detected violation of its policies and may without notice refuse to publish, immediately stop transmission of information, delete content or follow-up, issue a short-term prohibition of content or follow-up, and limit certain account features or access to services. The platform also reserves the right to keep relevant records of suspected violations of laws and regulations and suspected illegal and criminal acts, and to report violations to the relevant competent authorities in accordance with local or national laws. The company reserves the right not to restore deleted content.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No details of notification procedures are specified.
4.2. Are there processes under which users can appeal content	No appeals processes are specified.

removals, account suspensions or other enforcement decisions?	
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	No information is provided about Xigua Video content moderation policies. Reference is made in the User Service Agreement to the prompt review of all submitted user reports of violations of its policies. The company reserves the right to review any account registration or content uploaded though no further details are given. While details of its content moderation resources and technology used are not available publicly, according to reports from former employees, ByteDance employs about 20,000 content moderators to monitor content in China and deploys a range tools and algorithms to monitor, delete or alter content. The most popular livestream rooms are also reportedly closely monitored (Lu, 2021[200]).
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Unknown.

40. Odnoklassniki (Mail.Ru Group)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Odnoklassniki, the social network known as OK, does not provide a specific definition of CSEA. However, its Terms of Use prohibit the sharing of any content containing pornographic images of minors (7.4.9), information that infringes the rights of minors (7.4.13), or any information that is prohibited by legislation in force (7.4.14). In registering for the service, users agree to abide by the terms, comply with all relevant regulations and all applicable laws.
2. How are the ToS or Community Guidelines/Standards communicated?	The Terms of Use are available at: https://ok.ru/regulations
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or	According to the ToS, Odnoklassniki may delete, without notice, any content that violates its policies and/or may violate the laws of the Russian Federation, or may infringe the rights of other users or third parties, cause them harm or potential harm, or threaten their safety

Community Guidelines/Standards?	(6.4.2). As a social network, the company states that it does not perform and has no technical capability to perform pre-moderation of information and content posted by users and is not responsible for its content (8.6). The service has a reporting mechanism where users can submit reports or complaints regarding any content or conduct that they believe may violate the platform's policies or may be illegal.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	According to the ToS, Odnoklassniki has the right to suspend, restrict, or terminate a user's access to all or any sections and/or elements of the social network, the personal page, communities, groups on the platform. It may also restrict access to or delete communities and groups created by a user, at any time without justification, with or without prior notice, in accordance with applicable laws (6.4.4). The Customer Support Team in responding to reports or complaints may block accounts for any of the following reasons: • collecting personal information about other users; • knowingly providing false or fictitious information about yourself; • numerous and systematic insults towards other users using strong and obscene language; • inciting social, racial, national, or religious hatred; • posting pornographic content, as well as links to Internet sites with such content; • sending spam; • registration of multiple accounts by the same person; • posting information about topics and activities related to occultism on the site; • accessing other users' profiles without authorisation; • any other violation as specified in the Service Agreement.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	The platform states that it may, but is not obliged to, issue warnings or notices to users regarding non-compliance with its terms of use.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeals processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.

5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Odnoklassniki uses a combination of user reporting and internal content moderation on the platform. Few details are available about its content moderation processes. The Help section provides details of its report mechanism which can be submitted within the app or through a contact form to report to the Customer Support Team.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Unknown.

41. Flickr (SmugMug, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Flickr does not provide a specific definition of CSEA. However, Flickr's Community Guidelines underlines its zero tolerance policy towards harmful content involving minors. Flickr states that "*In protecting children from criminal predators, particularly in crimes involving child sexual abuse, [Flickr] will aggressively report and cooperate with law enforcement with the goal of prosecuting to the full extent of the law. This includes, but is not limited to, images, video, comments, faves, and other communications"*. Moreover, Flickr's Terms of Service prohibit the sharing of any content that is obscene, pornographic, indecent, lewd, or sexually suggestive. Any user content that that would constitute, encourage or provide instructions for a criminal offense is also expressly prohibited.
2. How are the ToS or Community Guidelines/Standards communicated?	The Flickr Terms and Conditions of Use are available at: https://www.flickr.com/help/terms The Flickr Community Guidelines are available at: https://www.flickr.com/help/guidelines
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Flickr uses a combination of automated moderation technology, human staff review and user reporting as part of its overall content moderation strategy. Flickr encourages its members to report any potential violating content to its Trust & Safety team
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	The Flickr ToS states that it may terminate a subscription or a user's access to Flickr services, at any time for any reason at its sole discretion. If users violate the Terms of Use, Flickr may require remedy of any violation and may take any other actions it deems appropriate. As stated in the ToS, in the absence of a legal requirement to do so, Flickr may refrain from notifying a user of Flickr's disclosures to governmental authorities where such notification may jeopardize a law enforcement investigation. Flickr may engage service providers for assistance with carrying out any obligation or exercising any right

	under the Flickr ToS.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes, enforcement decisions such as content removal or re-labelling as a result of auto-moderation or human staff review are notified to the user.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may request a review following any moderation decisions imposed by the platform. If an account has been reviewed by Flickr staff and marked as unsafe, the user is required to amend the photo content in question or to re-label it according to Flickr's guidelines. The user may then seek a further review by Flickr staff by completing an online review request form.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Flickr uses a combination of automated moderation technology, human staff review and user reporting as part of its overall content moderation strategy. Flickr applies auto-moderation to all newly uploaded content for potential violations of its policies and to ensure the appropriate labelling is applied. Users are also required to moderate their own content in conjunction with its moderation guidelines. The Moderation Bot detects explicit content from new uploads and automatically updates mis-moderated content to the correct moderation levels according to Flickr's established policies. When the system detects mis-labelled content in a user's account, the user receives a private notification that lets them know about the mismatch and directs them to the photo in question. Users are encouraged to report any content they come across which may be in violation of Flickr's policies. All reports are reviewed by the Trust and Safety Team who monitor newly uploaded as well as existing content on the platform. According to the company, while some of its activities to combat CSEA are very visible, much of the important work happens behind the scenes. As one of the largest online photography communities, Flickr states that it has a responsibility to ensure the safety of the most vulnerable, especially children. Flickr also partners with a range of organisations such as NCMEC, Thorn, WeProtect and the Technology Coalition in developing best practices for combating

	CSAM—and to share information with other companies.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes, according to NCMEC, a total 802 reports were submitted by Flickr in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 1,169` reports submitted in 2021 (NCMEC, 2022[150]).

42. Huoshan (ByteDance Technology Co.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	No definition of CSEA is provided Through the Huoshan User Service Agreement, users agree to abide by all applicable laws, including the lawful rights and interests of citizens, social public order, and morality. The platform's policies prohibit any content or activity that: • disseminates or spreads violence, obscenity, pornography, gambling, murder, terror or abetting crime • is insulting or defaming others and infringing on their legitimate rights and interests • is intimidating or threatening others with violence, • Contains content that is horrific, violent and bloody, highly dangerous, or harmful to the performer's own or others' physical or mental health. Terms of Use for Minors require parental consent for users under the age of 18. Minors must also abide by the Minor users must abide by the "National Youth Network Civilization Convention".
2. How are the ToS or Community Guidelines/Standards communicated?	The Huoshan User Service Agreement is available at https://www.huoshanzhibo.com/agreement/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	The User Service Agreement states that Huoshan may review any application for account registration, and has the right to terminate services to users at any time for violations of its policies. Houshan states that it strives to strengthen its information security management capabilities, improve the self-regulation of posting information, interactive exchanges and comments, and fulfill its social responsibilities to comply with national laws and regulations and respect the legitimate rights and interests of citizens.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	In cases of violations of its policies, Huoshan reserves the right to independently judge and take measures such as issuing warnings, refusal to post, immediate suspension of transmission of information, deletion of content or comments, short-term ban on posting content or comments, restriction of some or all functions of the account, termination of the provision of services, or permanent closure of the account, etc. For suspected violations of laws and regulations, or illegal and criminal acts, Huoshan states that it will keep the relevant records and report to the relevant competent authorities, and to report and

	cooperate with the relevant authorities. The company also has the right not to restore the deleted content.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	No notification procedures are specified.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeals processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Huoshan encourages its users to report to the platform any illegal and unlawful acts, illegal dissemination activities, illegal and harmful information or content in accordance with the company's public complaint and reporting system. Huoshan undertakes to handle all complaints and reports in a timely manner in order to maintain the safety of its platform. While details of its content moderation resources and technology used are not available publicly, according to reports from former employees, ByteDance employs about 20,000 content moderators to monitor content in China and deploys a range tools and algorithms to monitor, delete or alter content. The most popular livestream rooms are also reportedly closely monitored (Lu, 2021[200]).
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Unknown.

43. KaKao Talk (Daum Kakao Corporation)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Kakao does not provide a specific definition of CSEA. However, its Terms of Service prohibit the uploading or sharing of any pornographic or obscene content, or content that may violate any applicable laws or policies outlined by the company. The Kakao Operation Policy further elaborates on the platform's zero-tolerance policy for sex offences against children or juveniles. Kakao Talk also has a specific "Child/Adolescent Sexual Protection Policy". Under

	these policies, it is forbidden to: • produce, provide, advertise and/or introduce child/juvenile exploitation material • wittingly possess or use child/juvenile exploitation material • help lure children or juveniles into being involved in the production of sexual exploitation • provide lewd content or sexual exploitation material to children or juveniles • prostitute children or juveniles • conspire or describe sex offences against children or juveniles • groom children or juveniles • sexually objectify children or juveniles • any other attempts to encourage sex offenses against children or juveniles Violations of its policy result in immediate termination of the account and reporting of the offence to law enforcement.
2. How are the ToS or Community Guidelines/Standards communicated?	The Kakao Terms of Service are available at: https://www.kakao.com/en/terms Operation Policy is available at: https://www.kakao.com/policy/oppolicy?lang=en
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	The company's ToS state that it takes preventive action against misuse of its service and adopts security measures for such aspects as incoming/outgoing emails to protect users from spam (i.e., phishing, virus infringement, personal information theft, and other illegal and speculative junk mails, etc.). Additional spam operation policies and functions are provided if recommended by related organisations or deemed necessary for user protection (Article 6). Under its Youth Protection Policy, Kakao takes steps to prevent young people from being exposed to harmful information. The platform deploys a "Harmful to Teenagers" filter and has various measures to prevent harmful information from spreading. It takes steps to extensively control the range of prohibited words, and illegal or harmful content to teenagers. It also seeks to systematically manage hazardous services requiring adult authentication by limiting the scope of use.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Violations of the platform's policies or relevant laws and regulations may result in restriction of the user's Kakao Account and use of its services. The service may be restricted temporarily or permanently, depending on the number of violations accumulated. For more serious offences or any explicitly unlawful activities, immediate and permanent suspension is implemented, regardless of the accumulated number of violations.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes. Where sanctions or enforcement decisions are imposed, KaKao states that it will notify the user through in-service notifications and e-mail as quickly as possible, except in cases where immediate

	action is required to protect other users. In the case of severe violations such as for CSEA, the most strict restrictions are imposed on the use of the relevant account and service without notification.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may raise an appeal via the KaKao Customer Center if they are dissatisfied with an enforcement decision. Kakao states it will reply as to whether to accept the appeal after reviewing it.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No. Kakao's Transparency Report (Kakao, n.d.[201]) is solely for the purpose of providing statistics on governments' requests for user data and it does not provide any details of its content moderartion policy or CSEA enforcement policy.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Kakao uses a combination of technology and human staff review to moderate content on its platform. It encourages users to report the occurrence of any violations of its policies, and specifically sex offenses against children or juveniles or any such possible situations to its 24/7 Report Center. A reporting mechanism is provided within each of services. The company undertakes to receive all such reports and take the necessary measures immediately and to preserve the anonymity of the reporter.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to media reports, CSEA has been found on a number of chat rooms, including Kakao in Korea (Jang, 2022[202])

44. Smule (Smule, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Smule does not provide a specific definition of CSEA. However, its ToS prohibit the posting of any objectionable content which it describes as content that: • Bullies, harasses, threatens, intimidates, abuses, or demeans. • Promotes bigotry, discrimination, hatred, intolerance, or racism. • Is pornographic, obscene, or vulgar. • Is hateful, offensive, or shocking. • Incites violence. • Is fraudulent, false, deceptive, misleading, or defamatory. • May jeopardise another user's security or right to privacy. Moreover, the ToS expressly prohibits use of the service for any or

	all unlawful purposes, though CSEA is not specifically identified among such uses.
2. How are the ToS or Community Guidelines/Standards communicated?	The Smule Terms of Service are available at: https://www.smule.com/en/termsofservice The Smule Community Guidelines are available at: https://www.smule.com/en/s/communityguidelines
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Smule states that it does not pre-screen any user content, but reserves the right to do so and to remove or delete any content that it considers to violate its terms or applicable law or that it considers to be objectionable content. It also reviews content in response to complaints from other users. Smule also reserves the right but not the obligation to take remedial action in connection with any objectionable content as defined in its Community Guidelines.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	For any violations of its policies, Smule states that it may disallow, cancel, remove, or reassign certain usernames and permalinks and suspend or terminate an offender's account. Smule states that it may also monitor or review the service for violations of its terms and for compliance with its policies. It may refuse, restrict access to or the availability of any user content or services for the purposes of protecting its members. In the case of serious violations, Smule may suspend an offender's account immediately without notification. It may also report to law enforcement authorities and/or take legal action against anyone who violates its terms.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Notifications are issued at the service's discretion. Formal notification procedures are not specified.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	No appeal processes are specified.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Details of Smule's approach to content moderation are not provided. The company states that it reserves the right to review any material flagged by Smule members and may remove it if deemed inappropriate or unsafe for the community, or if it otherwise violates its guidelines or the ToS. Content will be removed, and the user may be banned, if Smule determines content in violation of its policies is being posted. Users are encouraged to moderate their own behavior

	and to report others who do not adhere to its guidelines. A reporting feature is provided on the platform. Users may also contact the Customer Service team directly.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, two reports were submitted by Smule in 2021 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2022[150]). No reports from Smule were recorded in the 2022 CyberTipline Report (NCMEC, 2023[18]).

45. DeviantArt (DeviantArt, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	DeviantArt defines CSEA as any material that depicts a minor in an explicit, lewd, arousing, or sexually suggestive manner. DeviantArt expressly prohibits the posting of child sexually exploitative material or content that endangers children. Such exploitative content is prohibited, whether those children are real or fictional. DeviantArt policies also prohibit the manipulation of media with indecent intent, as for example through digital editing software or through the use of artificial intelligence or machine learning. Examples may include manipulation to present a person in a state of nudity, depict them engaged in an act of sexual conduct, or otherwise place them into a situation intended to serve as sexually arousing stimuli without either their knowledge or consent. Its staff remove any content in which a person has had their image manipulated for what is considered to be an indecent purpose. Staff may also remove family photographs which include nude images of children even if shared with good intentions, because of the potential for abuse by others and to help avoid the possibility of other people reusing or misappropriating the images for the purpose of exploitation. The Terms of Service further prohibit the posting of content "*that may harm minors in any way, including, but not limited to, uploading, posting, or otherwise transmitting content that violates child pornography laws, child sexual exploitation laws or laws prohibiting the depiction of minors engaged in sexual conduct, or submitting any personally identifiable information about any child under the age of 13".* Furthermore, its ToS prohibit the posting of pornographic, obscene, offensive, blasphemous, unlawful, threatening, menacing, abusive, harmful, an invasion of privacy or publicity rights, defamatory, libelous, vulgar, illegal or otherwise objectionable.
2. How are the ToS or Community Guidelines/Standards communicated?	The DeviantArt Terms of Service are available at: https://www.deviantart.com/about/policy/service/ The Etiquette Policy or Community Guidelines are available at: https://www.deviantart.com/about/policy/etiquette/ General policies and explanatory articles are also given in the Help Centre: https://www.deviantartsupport.com/en/policies/general-policies

3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	DeviantArt states that it has no ability to control content uploaded to its service and does not have any obligation to monitor such content for any purpose. Users agree through the ToS to abide by the platform's policies
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	In its ToS, DeviantArt states that it may at any time, and without notice, suspend or terminate any part of the service, or refuse to fulfill any order, or terminate membership and delete any content stored on the platform at its sole discretion for failure to comply with its terms or applicable law. Accounts found to be in violation of its policies or which engage in abusive or disruptive community activity, can be subjected to an account suspension. During a suspension the user's account will no longer be publicly visible and users will not be able to post content or interact with the community in general. Suspensions form part of the permanent record associated with an offending profile. If the user is subject to further disciplinary action, previously recorded suspension(s) will be factored in. This may lead to a longer suspension or, in the case of repeat offenders, result in any new suspension being escalated to an account termination.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Notifications are issued at the service's discretion. According to the Help Centre, users will receive notification of an enforcement action, which may include a private message or reason concerning why the action was taken, and a timer will be added to the user's profile page.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal a suspension or account termination by contacting DeviantArt Customer Support directly, and each appeal is evaluated,. If an appeal is granted, the previous account termination remains on the record associated with the active profile. Any problematic behavior which prompts administrative action on a user's account after an appeal can result in an immediate indefinite suspension, and the user will not be able to rejoin the platform.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No.
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Not applicable.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff)	DeviantArt relies on a combination of user reporting and human staff review to moderate content on its platform. In July 2020, it announced a more proactive approach to remove violative content

reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	before it is reported, while still allowing for user reporting and staff review. In particular, DeviantArt reports that technology for detecting violations was introduced so that the moderation team could work through reports more rapidly and efficiently. The company stated that while content removal decisions are still made by the moderation team rather than automatically by artificial intelligence,the new technology, would serve to help the moderation team take a more proactive approach to reviewing deviations. Users may also participate as an administrator or member of a "Group" or community based on common interests. Administrators assist in the moderation of the group activities in accordance with the platform's policies. Users may also report abuse or abusive behaviour directly to the Customer Support Team. When DeviantArt becomes aware of apparent child exploitation, it is reported it to the NCMEC, in compliance with US law.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, 11 reports were submitted by Deviantart in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 6 reports submitted in 2021 (NCMEC, 2022[150]).

46. Google Drive (Alphabet, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Google Drive does not provide a specific definition of CSEA. However, the relevant policies which include Google Terms of Service, Google Drive Additional Terms of Service and additional Program Policies applying to Drive and other Google services, specify obligations on users to abide by all applicable policies and laws in the uploading and sharing of content. Google Program Policies prohibit the creation, uploading or distribution of content that exploits or abuses children. This includes all child sexual abuse materials. More broadly, Google prohibits the use of its products to endanger children. This includes but is not limited to, predatory behaviour towards children such as: • *'Child grooming'* (for example, befriending a child online to facilitate, either online or offline, sexual contact and/or exchanging sexual imagery with that child); • *'Sextortion'* (for example, threatening or blackmailing a child by using real or alleged access to a child's intimate images); • *Sexualisation of a minor* (for example, imagery that depicts, encourages or promotes the sexual abuse of children or the portrayal of children in a manner that could result in the sexual exploitation of children); and • *Trafficking of a child* (for example, advertising or solicitation of a child for commercial sexual exploitation).

2. How are the ToS or Community Guidelines/Standards communicated?	The Google Terms of Service are available at: https://policies.google.com/terms Google Drive Additional Terms of Service are available at: https://www.google.com/drive/terms-of-service/ Google Programme Policies (Abuse programme policies and enforcement) are available at: https://support.google.com/docs/answer/148505#zippy=%2Cchild-sexual-abuse-and-exploitation
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Google's Program Policies which apply to Drive, Docs, Sheets, Slides, Forms and Sites address the prevention of abuse on its platforms and services. The relevant policies outline the various forms of prohibited content and activities when using Google products. The ToS for Google Drive, for instance, state that Google may review content to determine whether it is illegal or violates its Program Policies, and may remove or refuse to display content that it reasonably believes violates Google policies or applicable law. At the same time, the ToS states this does not necessarily mean that it does review all such content, and users are reminded that they should not assume so. Google's Terms of Service prohibit using any of Google's platforms or services to store or share CSEA. Across Google, the company states, its teams work to identify, remove, and report this content, using a combination of automated detection tools and specially-trained reviewers. Google also receive reports from third parties and its users, which complement this ongoing work.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Relevant enforcement actions are outlined in Google's Program Policies. These state that Google may review content for violations of its policies and take actions such as access restriction of content, removal of content, and limitation or termination of a user's access to Google products. Enforcement actions that Google may take on violative material may include: • Removing the file from the account. • Restrict sharing of a file. • Limiting who can view the file. • Disabling access to one or more Google products. • Deleting the Google Account. • Reporting illegal materials to the appropriate law-enforcement authorities.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Yes. Where a file has been flagged for a violation in Google Drive, the owner will see a flag next to the filename and they won't be able to share it. The file will no longer be publicly accessible, even to people who have the link. In instances where an account is disabled due to a policy violation, a disable notification will appear when the user attempts to log into a Google product. In some cases, the account owner will get an email or text message from Google to tell them that their account is disabled.

4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. Owners of files that have been found to be in breach of Google policies may request a review of the violation. The request for a review can be accessed from the file location or - in cases where an account has been disabled - they can complete an appeal form.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Google publishes a combined Transparency Report for Google products and services. However, this does not contain a specific breakdown for enforcement actions related to Google Drive. Transparency Reports containing data on government requests for user data as well as relevant actions related to privacy, security, access to information and content. A summary report on Google's efforts to combat online child sexual abuse material is included within the Transparency Report.
5.1 What information/fields of data are included in the TRs?	The following information on Google's efforts to combat CSEA are included: • CyberTipline reports to NCMEC • Total pieces of content reported to NCMEC • Accounts disabled for CSAM violations • URLs reported and de-indexed for CSAM from Google Search • CSAM hashes contributed to the NCMEC database
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	The following explanations for individual fields of data are detailed in the Google Transparency Report: *Reports to NCMEC:* A single report may contain one or more pieces of content depending on the circumstances. This content could include, for example, images, videos, URL links, and/or text soliciting CSAM. A single piece of content may be identified in more than one account or on more than one occasion, so this metric may include pieces of content reported more than once. *Accounts disabled for CSAM:* When CSAM is identified in a user's Google account, Google sends a CyberTipline report to NCMEC and may disable the account. Users are notified of the account termination and are given the opportunity to appeal. *URLs reported and de-indexed for CSAM from Google Search:* This metric represents the number of URLs reported and removed from the Search index. *CSAM hashes contributed to the NCMEC database:* When Google identifies new CSAM it may create a hash of the content and add that to its internal repository. Hashing technology allows Google to find previously identified CSAM. It also shares hash values with NCMEC so that other providers can access these hashes as well.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are issued approximately every 6 months
6. What methods (for example, monitoring algorithms, user flaggers, human (staff)	Google states that it invests heavily in fighting child sexual exploitation online and uses technology to deter, detect, and remove CSEA from its platforms. This includes automated detection and

reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	human review, in addition to relying on reports submitted by its users and third parties such as NGOs, to detect, remove, and report CSAM on its platforms. Google deploys hash matching, including YouTube's CSAI Match, to detect known CSAM. It also deploys machine learning classifiers to discover never-before-seen CSAM, which is then confirmed by its specialist review teams. Using these classifiers, Google created the Content Safety API, which it provides to other platforms to help them prioritise abuse content for human review. Users are also encouraged to report any content on a Google product that may exploit a child using the 'Report abuse' mechanism. Google states that it will remove such content and take appropriate action, which may include reporting to NCMEC, limiting access to product features, and disabling accounts. Users are also asked to contact the police immediately if they believe that a child is in danger of or has been subject to abuse, exploitation or trafficking.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes, according to its Transparency Report, Google made 334,215 reports to NCMEC in the period July - December 2021. This total is for Google products excluding YouTube which is reported separately. While a breakdown is not available specifically for Google Drive, it can be reasonably assumed that the aggregate includes Drive given its key role as file storage for any Google account. According to NCMEC, a total of 2,174,548 reports were submitted by Google in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement. This figure includes data for YouTube (NCMEC, 2023[18]). This compares to 875,783 reports submitted in 2021 (NCMEC, 2022[150]).

47. Dropbox (Dropbox, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Dropbox does not provide a specific definition of CSEA. Its ToS prohibit any content or activity not permitted by applicable laws or regulations and requires users to adhere to its Acceptable Use Policy (AUP). The AUP includes the following specific prohibition forbidding users to "publish, share, or store materials that constitute child sexually exploitative material (including material which may not be illegal child sexual abuse material but which nonetheless sexually exploits or promotes the sexual exploitation of minors), unlawful pornography, or are otherwise indecent."
2. How are the ToS or Community Guidelines/Standards communicated?	The Terms of Service are available at: https://www.dropbox.com/terms The Acceptable Use Policy is available at: https://www.dropbox.com/acceptable_use
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or	Dropbox states that it reserves the right to take appropriate action in response to violations of its Terms and its Acceptable Use Policy, which could include removing or disabling access to content, suspending a user's access to the Services, or terminating an

Community Guidelines/Standards?	account. At the same time, Dropbox states that it is not responsible for the content people store and share via its services. ’
4. What are the service’s policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Dropbox states that it reserves the right to take appropriate action in response to violations of its Acceptable Use Policy, which could include removing or disabling access to content, suspending a user’s access to the Services, or terminating an account.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	In cases of breaches of its ToS, Dropbox provides reasonable advance notice via the email address associated with the user’s account and gives the user an opportunity to export his or her content. If after such notice the user fails to take the steps Dropbox requires, Dropbox will terminate or suspend the user’s access to Dropbox’s services. Dropbox does not provide advance notice when a user is in material breach of the ToS, when doing so would cause Dropbox legal liability or compromise its ability to provide its services to other users, or when Dropbox is prohibited from doing so by law.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Users may appeal adverse enforcement decisions by contacting Dropbox’s Support Team. More information can be found on Dropbox’s Help Center: https://help.dropbox.com/account-access/disabled-accounts
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Dropbox has published a biannual report since 2012 on government information requests and some of its enforcement decisions including those made for child sexual abuse material.
5.1 What information/fields of data are included in the TRs?	• Number of accounts actioned • Number of individual pieces of content actioned. Content or accounts actioned refers to the disabling of access for violation of the ToS.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	No details provided.
5.3 Frequency/timing with which TRs are issued	Transparency reports are issued every 6 months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	Dropbox states that it will swiftly disable any account found with content containing CSEA. Dropbox uses a variety of tools, including industry-standard automated detection technology, and human review to find potentially violating content and action it as appropriate. It also encourages its users to report inappropriate content they come across through its reporting tool or by completing

	a report form. When Dropbox becomes aware of instances of apparent CSEA, it disables the account and makes a report to NCMEC, in accordance with US law. From July through December 2021, Dropbox submitted 24,115 CyberTip reports to NCMEC and actioned 22,799 distinct accounts and 425,847 individual pieces of content for violating its policies against child sexual abuse and exploitation material.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	According to the Dropbox Transparency Report, from January through June 2022, Dropbox submitted 26,730 CyberTip reports to NCMEC and actioned 25,336 distinct accounts and 382,261 individual pieces of content for violating its policies against child sexual abuse and exploitation material. The NCMEC annual report for 2022 confirms a total of 45,992 reports were submitted by Dropbox for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 48,371 reports submitted in 2021 (NCMEC, 2022[150]).

48. Microsoft OneDrive (Microsoft, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	OneDrive is one of a number of Microsoft products that is covered by the Microsoft Services Agreement. This does not provide a specific definition of CSEA. However, the Microsoft Services Agreement includes a Code of Conduct that outlines what is allowed and what is prohibited when using a Microsoft account. This prohibits any activity that exploits, harms, or threatens to harm children. Microsoft states in its Digital Safety Content Report that it has a long-standing commitment to online child safety, and that it develops both tools and multistakeholder partnerships to help address this issue (Microsoft, n.d.[194]). As specified in its Code of Conduct, part of the Microsoft Services Agreeement, it prohibits "any activity that exploits, harms, or threatens to harm children" across its products and services – including but not limited to distribution of child sexual exploitation and abuse imagery (CSEAI), and grooming of children for sexual purposes.
2. How are the ToS or Community Guidelines/Standards communicated?	The Microsoft Services Agreement is available at: https://www.microsoft.com/en-us/servicesagreement
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Microsoft deploys a combination of content moderation, staff moderation and user reporting to prevent misuse of its platforms and services and to prevent activity that may violate its ToS. Microsoft deploys tools to detect CSEA, including hash-matching technology (e.g., PhotoDNA) and other forms of proactive detection. Microsoft has made available in-product reporting for products such as OneDrive, Skype, Xbox, and Bing, whereby users can report

	suspected child exploitaiton or other violating content.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	Microsoft states that it may disable a user's account for any suspicious activity or for violating the Microsoft Services Agreement such as by hosting photos, video or other content in violation of the Code of Conduct. Microsoft may also block delivery of a communication (like email, file sharing or instant message) or may remove or refuse to publish a user's content for any reason. When investigating alleged violations of its policies, Microsoft reserves the right to but states that it does not have the obligation to, review any content in order to resolve the issue. Microsoft also states that it removes content that contains apparent CSEAI. Microsoft also reports all apparent incidence of CSEA or grooming of children for sexual purposes to the NCMEC via the CyberTipline, as required by U.S. law.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Details of notifications of content removals or account suspensions are not specified. The Microsoft Services Agreement refers in a general way to service notifications which it may send at its discretion in connection with a user's account Microsoft account via email or via SMS (text message), or by in-product messages.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, users may appeal an account suspension by completing a form and submitting supporting information with their request for a review.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	Yes. Microsoft's Digital Safety Content Report (DSCR) contains its Transparency Report covering actions that it has taken in relation to child sexual exploitation and abuse imagery (CSEAI), grooming of children for sexual purposes, terrorist and violent extremist content (TVEC), as well as non-consensual intimate imagery (NCII) (Microsoft, n.d.[194]). The report is an aggregate one for all Microsoft hosted consumer services and a breakdown by individual products - including OneDrive, Outlook, Skype and Xbox - is not available.
5.1 What information/fields of data are included in the TRs?	Microsoft includes the following information regarding enforcements for CSEA violations: • Content Actioned • Content Detected Proactively • Accounts Actioned • Accounts Reinstated
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	• *Content actioned* refers to removal of a piece of user-generated content from any of its products and services and/or block user access to a piece of user-generated content. With regard to Bing, "content actioned" may also mean filtering or de-listing a URL from the search engine index. • *Account actioned* refers to when Microsoft suspends or blocks access to an account, or restricts access to content within the account.

	• *Proactive detection* refers to Microsoft-initiated flagging of content on its products or services, whether through automated or manual review. • *Accounts reinstated* refer to actioned accounts that were fully restored including content and account access, upon appeal.
5.3 Frequency/timing with which TRs are issued	The Digital Safety Content Report is published every six months.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	According to its Digital Safety Content Report, Microsoft deploys tools to detect child sexual exploitation and abuse imagery (CSEAI), including hash-matching technology such as PhotoDNA and other forms of proactive detection (Microsoft, n.d.[194]). "Proactive detection" refers to Microsoft-initiated flagging of content on its products or services, whether through automated or manual review. Microsoft developed PhotoDNA in partnership with Dartmouth College in 2009 to help find duplicates of known child sexual exploitation and abuse imagery (Microsoft, n.d.[52]). PhotoDNA is now an industry-standard technology used by organisations around the world and is deployed across Microsoft's consumer products and services.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	According to the Digital Safety Content Report for January-June 2022, Microsoft actioned 40,722 pieces of content and 10,207 consumer accounts associated with CSEAI or grooming of children for sexual purposes during this period. However, Microsoft does not break down information on CSEA violations by individual products and therefore the extent of prevalence of CSEA involving the use of Microsoft OneDrive is unknown. Microsoft detected 98.7 percent of the content that was actioned, while the remainder was reported to Microsoft by users or third parties. Of the accounts actioned for CSEAI, 0.56 per cent were reinstated upon appeal. According to NCMEC, in 2022, Microsoft – Online Operations submitted 107,274 reports to its CyberTipline for child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 78,603 reports submitted by Microsoft in 2021 (NCMEC, 2022[150]).

49. WordPress.com (Automattic, Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	A specific definition of CSEA is not provided. However, WordPress.com User Guidelines prohibit the publishing of any illegal content or conduct via its services. Under its Mature Content policy, the posting of images of child sexual abuse material or content that promotes pedophilia, such as sites with galleries of images of children where the images, content surrounding the images, or the intent of the blog is sexually suggestive, are expressly forbidden.

	More generally, mature content on WordPress.com, including text, images and videos that contain nudity, offensive language, and mature subject material is permitted. However, websites that contain such content must be marked as Mature in its system. Pornography, defined as visual depictions of sexually explicit acts, are prohibited.
2. How are the ToS or Community Guidelines/Standards communicated?	The WordPress.com Terms of Service are available at: https://en-gb.wordpress.com/tos/ The User Guidelines are available at: https://wordpress.com/support/user-guidelines/
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or Community Guidelines/Standards?	Websites found to be hosting material in violation of its policies, such as its Mature Content policy, will, according to the company, be suspended. Indicative proscribed content is not intended to be exhaustive and further interpretation is down to its sole discretion. In cases of child pornography, WordPress.com states that it will report all incidences to NCMEC and will fully cooperate with law enforcement. The WordPress.com ToS state that it does not review, and can't review, all of the content (like text, photo, video, audio, code, computer software, items for sale, and other materials) posted to or made available through its services by users or other websites that link to, or are linked from, its services. As such, it states that it is not responsible for any use or effects of content or third-party websites.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	If a site or any of its content is found to be in violation of its policies, WordPress.com states that it will remove the content, disable certain features on the account, and/or suspend the site entirely. WordPress.com states it acts on all information received in order to investigate potential breaches and enforces its policies on a daily basis.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	Depending on the scenario, the company will email the account holder or add a warning notification in their dashboard. The notification will contain a link that the account owner may use to contact its support team regarding the issue. Alternatively, a contact support form or email may be used.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes, enforcement decisions may be appealed. Account holders may contact the support team via the link on their dashboard or through a contact support form. All submissions are reviewed by a staff member who will reply with a decision as soon as possible.
5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No. Automattic, the parent company of WordPress.com and Tumblr, publishes transparency reports in respect of government requests for user information and for copyright- and trademark-related content removals. It does not publish any data regarding enforcement of its Community Guidelines or content moderation decisions (Automattic, n.d.[203]).
5.1 What information/fields of data are included in the TRs?	Not applicable.

5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are published on a six monthly basis.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	WordPress.com predominantly relies on its users to report any content or activity that may violate its policies. It uses automated technologies to detect potentially violative content. However, not all content created using its services is hosted by WordPress.com and may be hosted on third party services. Details of its trust and safety resources or the number of human staff reviewers are not publicly available. In a press interview with Automattic's CEO, it was reported that a global team of 400 staff moderate content on Tumblr and WordPress.com (Patel, 2022[204]).
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, 190 reports were submitted by WordPress.com/Automattic for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 310 reports submitted in 2021 (NCMEC, 2022[150]).

50. Wikipedia (Wikimedia Foundation Inc.)

1. How is online child sexual exploitation and abuse (CSEA) defined in the Terms of Service (ToS) or Community Guidelines/Standards?	Wikipedia does not provide a specific definition of CSEA. However, the Wikimedia's Terms of Use (ToS) which govern content on Wikipedia, prohibit the posting of child pornography or any other content that violates applicable law concerning child pornography (Section 4. Refraining from Certain Activities). Moreover, the ToS prohibit the posting or trafficking in obscene material that is unlawful or using the services in a manner that is inconsistent with applicable law. Wikipedia's Child Protection Policy which governs the the behavior and actions of adult editors on the platform with regards to children, states that Wikipedia does not tolerate inappropriate adult–child relationships in any form. According to the policy, editors who attempt to use Wikipedia to pursue or facilitate inappropriate adult–child relationships, who advocate inappropriate adult–child relationships on- or off-wiki (e.g. by expressing the view that inappropriate relationships are not harmful to children), or who identify themselves as pedophiles, will be blocked and banned indefinitely.
2. How are the ToS or Community Guidelines/Standards communicated?	The Wikimedia ToS are available at: https://meta.wikimedia.org/wiki/Terms_of_use The Wikipedia Child Protection Policy is available at: https://en.wikipedia.org/wiki/Wikipedia:Child_protection
3. What are the service's policies and procedures for preventing its use in a manner that violates its ToS or	Wikipedia largely relies on its community - the network of users who contribute to its various sites or Projects - to uphold and enforce its policies. The community both creates and enforces policies for the specific Project editions (such as the different language editions for

Community Guidelines/Standards?	the Wikipedia Project or the Wikimedia Commons multi-lingual edition). According to the Wikimedia ToS, contributors, editors, or authors, are all required to follow the policies that govern each of the independent Project editions, the largest of which is Wikipedia. Wikimedia states that it does not take an editorial role and that Wikipedia is collaboratively edited with all of the content provided by users. As such, Wikimedia states that it does not generally monitor or edit the content of the Project websites, and does not take any responsibility for the content.
4. What are the service's policies and procedures for enforcing its ToS or Community Guidelines/Standards when violative content or behaviour is detected? Please be sure to mention all sanctions and consequences that may apply.	According to the Wikimedia ToS, the community has the primary role in creating and enforcing policies applicable to the Wikipedia platform. Wikimedia states that it rarely intervenes in community decisions about policy and its enforcement, except in unusual cases where it is called upon to address especially problematic behavior. In such cases, it reserves the right to: • Investigate whether use of the service has been in violation of its policies or other applicable law • Detect, prevent, or otherwise address fraud, security, or technical issues or respond to user support requests; • Refuse, disable, or restrict editing access of any user who violates its ToS; • For actions violating its policies, including repeat copyright infringement, ban a user from editing or contributing or block a user's account or access; • Take legal action against users who violate its policies (including reports to law enforcement authorities); and Where any individual has had their account or access blocked under these provisions, they are prohibited from creating or using another account or seeking access to services without explicit permission. Especially problematic users who have had accounts or access blocked on multiple Project editions may be subject to a ban from all of the Project editions, in accordance with the Global Ban Policy.
4.1. Are users notified of content removals account suspensions or other enforcement decisions?	The ToS refer to warnings that may be issued as part of overall enforcement policy. However, no details of notification procedures are given.
4.2. Are there processes under which users can appeal content removals, account suspensions or other enforcement decisions?	Yes. The Wikimedia ToS encourages users to seek resolution for any issues or disagreements through the dispute resolution procedures or mechanisms provided by the Projects or and the Wikimedia Foundation. The Dispute Resolution Policy provides for engagement and discussion with the relevant editor when content or conduct enforcement decisions are made. Where this does not succeed in resolving the matter, discussion can be extended to other parties to produce a consensus using, for example, Wikipedia's Dispute resolution noticeboard. Serious matters, including those involving legal concerns, are referred to the Arbitration Committee. Concerns in relation to suspected CSEA or breaches of the Wikipedia Child Protection Policy are referred directly to Wikimedia.

5. Does the service issue transparency reports (TRs) specifically on content and/or behaviour related to CSEA?	No. Wikimedia's Transparency Report contains information only on government requests to alter or remove content from the projects, and to provide non-public information about users. It does not contain information about content removed for CSEA or violation of child protection policies (Wikimedia Foundation, n.d.[205]).
5.1 What information/fields of data are included in the TRs?	Not applicable.
5.2 Methodologies for determining/calculating/estimating the information/data included in the TRs	Not applicable.
5.3 Frequency/timing with which TRs are issued	Transparency Reports are published twice a year.
6. What methods (for example, monitoring algorithms, user flaggers, human (staff) reviewers, hash-sharing/URL sharing database) does the service use to detect CSEA?	The Wikipedia community has the primary role in creating and enforcing policies applying to the site. The Wikimedia Foundation rarely intervenes in community decisions about policy and its enforcement except in case of problematic or dangerous behavior. Users are encouraged to report breaches of the ToS and specifically the Child Protection Policy to the Wikimedia Foundation. The ToS specify that editors attempting to pursue or facilitate inappropriate adult–child relationships, or otherwise breaching trust and safety, should be reported by email (to legal-reports@wikimedia.org), as should reports of images that raise concerns.
7. Has this service been used to disseminate, store, or produce CSEA, or to solicit children for sexual purposes?	Yes. According to NCMEC, 29 reports were submitted by the Wikimedia Foundation in 2022 for online exploitation of children, including child sexual abuse material, child sex trafficking and online enticement (NCMEC, 2023[18]). This compares to 8 reports submitted in 2021 (NCMEC, 2022[150]).

Annex C. Definitions

For the purposes of this report, the following definitions are provided for terms used throughout:

Content: Any type of digital information carried on an online content-sharing service that may serve as a medium for CSEA, such as comments, pictures, videos, files, posts, links, chatroom chats, blogs or messages.

Content-Sharing Service: Any online service that enables the transfer, transmission and dissemination of content, in whatever form, whether one-to-one, one-to-few or one-to-many and irrespective of whether the content is public-facing, semi-private or private. All of the services profiled in this report are Online Content-Sharing Services.

Child sexual abuse material (CSAM): The term "child sexual abuse material" refers to the depiction or reproduction of children in a sexualised context or with reference to sexual activities involving children. Its definition follows on extant legal definitions and is increasingly used in preference to the term "child pornography" to emphasise that such representations constitute a form of child sexual abuse (ECPAT International, 2016[5]).

Child sexual abuse and exploitation (CSEA): this formulation as used throughout this report includes the reference to CSAM above as well as to all forms of child sexual exploitation such as enticing/manipulating/ threatening a child into performing sexual acts and the soliciting and/or grooming potential child victims online with a view to exploiting them sexually. CSEA and online CSEA are used interchangeably to denote that sexual exploitation and sexual abuse of children that takes place through the Internet, or with some connection to the online environment.

Online Platform: A digital service that facilitates interactions between two or more distinct but interdependent sets of users (whether firms or individuals) who interact through the service via the Internet.

Social Media (or Social Networking) Service: Any online service that allows individuals to build a public or semi-public profile of themselves, upload and access Content shared by other users, interact and establish connections with other users, and express their views and interests.

References

99 Firms (2021), *LinkedIn Statistics*, https://99firms.com/blog/linkedin-statistics/. [122]

99 Firms (2021), *Viber Statistics*, https://99firms.com/blog/viber-statistics/. [117]

Apple (2021), *Expanded Protections for Children: Frequently Asked Questions*, https://www.apple.com/child-safety/pdf/Expanded_Protections_for_Children_Frequently_Asked_Questions.pdf. [159]

Apple (n.d.), *Expanded Protections for Children: Communication safety in Messages*, https://www.apple.com/child-safety/. [162]

Apple (n.d.), *Transparency Report*, https://www.apple.com/legal/transparency/. [158]

Ask.fm (2021), *ASKfm Safety Guide for Schools & Educators*, https://safety.ask.fm/ask-fm-safety-guide-for-schools-educators/. [127]

Audiens (2020), *Game developer, Smule, increase their high value VIP customers by 21%*, https://audiens.com/smule-increase-their-high-value-vip-customers-by-21/. [140]

Australian Government (1995), *Criminal Code Act 1995*, https://www.legislation.gov.au/Details/C2019C00043. [80]

Auttomatic (n.d.), *Government Takedown Demands*, https://transparency.automattic.com/wordpress-dot-com/government-takedown-demands/. [203]

Baidu (2021), *Environmental, Social and Governance Report*, https://esg.baidu.com/resource/1ecd4d7e-c148-6e91-8dd5-556f11239ccc/Baidu%202021%20Environmental,%20Social%20and%20Governance%20Report.pdf. [192]

BBC (2021), *China: Taobao, Weibo fined for illegal child content*, https://www.bbc.com/news/business-57911207. [170]

BBC (2019), *Child abuse images being traded via secure apps*, https://www.bbc.com/news/technology-47279256. [179]

Bulger, M. (2017), "Where policy and practice collide: Comparing United States, South African and European Union approaches to protecting children online", *New Media and Society*, Vol. 19/5, pp. 750-764. [207]

Bursztein, E. (2019), *Rethinking the Detection of Child Sexual Abuse Imagery on the Internet*, [30]

https://doi.org/10.1145/3308558.3313482.

C3P (n.d.), *Project Arachnid*, https://www.projectarachnid.ca/en/. [55]

CDPP (n.d.), *Child Exploitation*, https://www.cdpp.gov.au/crimes-we-prosecute/child-exploitation. [81]

Child Dignity Alliance (n.d.), *Promoting Child Dignity Summit 2019*, https://www.childdignity.com/. [47]

China Services Info (2018), *Public Pledge of Self-Regulation and Professional Ethics for China Internet Industry*, https://govt.chinadaily.com.cn/s/201812/26/WS5c23261f498eb4f01ff253d2/public-pledge-of-self-regulation-and-professional-ethics-for-china-internet-industry.html. [32]

Cornish, P. (ed.) (2021), *Online Child Safety*, Oxford University Press. [34]

Council of Europe (2018), *Recommendation CM/Rec(2018)7 of the Committee of Ministers to member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment*, https://rm.coe.int/guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-th/16808d881a. [75]

Council of Europe (2007), *Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse*, https://rm.coe.int/1680084822. [72]

Council of Europe (2001), *Convention on Cybercrime (ETS No. 185)*, https://rm.coe.int/1680081561. [71]

Cybertip.ca (n.d.), *Cybertip.ca's History*, https://www.cybertip.ca/en/about/history/. [36]

Datanyze (2021), *File Sharing Software Market Share*, https://www.datanyze.com/market-share/file-sharing--198. [142]

Datareportal (2021), *Global Social Media Stats*, https://datareportal.com/social-media-users. [114]

Dean, B. (2021), *Twitch Usage and Growth Statistics: How Many People Use Twitch in 2021?*, https://backlinko.com/twitch-users#monthly-active-users. [133]

DeviantArt (2021), *About DeviantArt*, https://www.deviantart.com/about/. [141]

Discord (2022), *Discord Transparency Report: April - June 2022*, https://discord.com/blog/discord-transparency-report-q2-2022. [197]

Discord (2021), *An Update on Our Business*, https://discord.com/blog/an-update-on-our-business. [132]

ECPAT International (2016), *Terminology Guidelines for the Protection of Children from Sexual Exploitation and Sexual Abuse*, https://ecpat.org/luxembourg-guidelines/. [5]

ECPAT International (2016), *The Global Study Report on Sexual Exploitation of Children in Travel and Tourism*, http://globalstudysectt.org/. [96]

Eko, L. (2016), *Regulation of Online Pedopornography (Child Pornography) in the United States and France*, Palgrave Macmillan US. [84]

End Violence Against Children (n.d.), *Disrupting Harm*, https://www.end-violence.org/disrupting-harm. [46]

End Violence Against Children (n.d.), *The Global Partnership and Fund to End Violence Against Children*, https://www.end-violence.org/. [41]

Envisage Digital (2021), *WordPress Market Share in 2021*, https://www.envisagedigital.co.uk/wordpress-market-share/. [143]

eSafety Commissioner (2022), *Basic Online Safety Expectations Summary of industry responses to the first mandatory transparency notices*, https://www.esafety.gov.au/industry/basic-online-safety-expectations/responses-to-transparency-notices. [99]

European Commission (2022), *Impact Assessment Report accompanying the document "Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022SC0209&qid=1652786640660. [53]

European Commission (2022), *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL laying down rules to prevent and combat child sexual abuse COM/2022/209 final*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A209%3AFIN&qid=1652451192472. [78]

European Commission (2022), *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)*, https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32022R2065. [79]

European Commission (1999), *Action plan for a Safer Internet 1999-2004*, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV%3Al24190. [37]

European Commission (n.d.), *Alliance to better protect minors online*, https://digital-strategy.ec.europa.eu/en/policies/protect-minors-online#. [48]

European Union (2021), *Regulation (EU) 2021/1232 of the European Parliament and of the Council of 14 July 2021 on a temporary derogation from certain provisions of Directive 2002/58/EC as regards the use of technologies by providers of number-independent interpersonal communica*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021R1232. [66]

European Union (2011), *Directive 2011/92/EU on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing*, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0093. [76]

European Union (2000), *Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce')*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000L0031. [112]

EUROPOL (2020), *Exploiting Isolation: Offenders and victims of online child sexual abuse during the COVID-19 pandemic*, https://www.europol.europa.eu/cms/sites/default/files/documents/europol_covid_report-cse_jun2020v.3_0.pdf. [17]

Europol (2021), *IOCTA 2021: internet organised crime threat assessment 2021*, https://data.europa.eu/doi/10.2813/113799. [61]

Europol & Eurojust (2021), *Third Report of the Observatory Function on Encryption*, https://www.europol.europa.eu/publications-events/publications/third-report-of-observatory-function-encryption#:~:text=category%20provides%20for%20tools%20to,use%20of%20the%20provisions%20mentioned. [65]

Finances Online (2021), *Number of Tumblr Blogs in 2021/2022: User Demographics, Growth, and Revenue*, https://financesonline.com/number-of-tumblr-blogs/. [121]

Five Country Ministerial (2020), *Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse*, https://www.weprotect.org/wp-content/uploads/11-Voluntary-principles-detailed.pdf. [9]

Flickr (2021), *Work at Flickr*, https://www.flickr.com/jobs/. [138]

G20 (2021), *G20 High Level Principles for Children Protection and Empowerment in the Digital Environment*, https://assets.innovazione.gov.it/1628084642-declaration-of-g20-digital-ministers-2021final.pdf. [13]

G7 (2022), *G7 Interior and Security Ministers' Statement*, https://www.bmi.bund.de/SharedDocs/downloads/EN/news/g7-iasm-statement.html. [15]

G7 (2022), *G7 Leaders' Communiqué*, https://id.ambafrance.org/G7-Leaders-Communique. [14]

G7 (2021), *G7 Interior and Security Ministers' Meeting, September 2021*, https://www.gov.uk/government/publications/g7-interior-and-security-ministers-meeting-september-2021/annex-2-protecting-against-online-exploitation-violence-and-abuse-accessible-version. [7]

G7 (2021), *G7 Internet Safety Principles*, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986161/Annex_3__Internet_Safety_Principles.pdf. [11]

G7 (2019), *G7 Digital Ministers: Building Digital Trust Together*, https://www.economie.gouv.fr/files/files/2019/G7/G7Num/Chairs_summary_version_finale_ENG.pdf. [10]

German Federal Ministry of Justice (2021), *Criminal Code in the version published on 13 November 1998 (Federal Law Gazette I, p. 3322), as last amended by Article 2 of the Act of 22 November 2021 (Federal Law Gazette I, p. 4906)*, https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html#p0084. [97]

German Federal Ministry of Justice (2017), *Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)*, https://www.bmj.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf;jsessionid=9A77598EC70A85C6A5DBCF84D8F1A355.1_cid334?__blob=publicationFile&v=2. [105]

German Federal Ministry of Justice (2002), *Youth Protection Act (JuSchG)*, https://www.gesetze-im-internet.de/juschg/__1.html. [104]

Google (n.d.), *Developing and sharing tools to fight child sexual abuse*, https://protectingchildren.google/#tools-to-fight-csam. [54]

Google (n.d.), *Fighting child sexual abuse online*, https://protectingchildren.google/. [57]

Google (n.d.), *Google's efforts to combat online child sexual abuse material*, https://transparencyreport.google.com/child-sexual-abuse-material/reporting. [153]

Google (n.d.), *YouTube Community Guidelines enforcement*, https://transparencyreport.google.com/youtube-policy/removals?hl=en. [152]

Government of Australia (2021), *Online Safety Act 2021*, https://www.legislation.gov.au/Details/C2021A00076. [98]

Government of Canada (2011), *An Act respecting the mandatory reporting of Internet child pornography by persons who provide an Internet service (S.C. 2011, c. 4)*, https://laws-lois.justice.gc.ca/eng/acts/I-20.7/. [100]

Government of Canada (1985), *Criminal Code (R.S.C., 1985, c. C-46)*, https://laws-lois.justice.gc.ca/eng/acts/c-46/. [82]

Government of Türkiye (2015), *Law on the Regulation of Internet Publications and Combating Crimes Committed through Such Publications*, https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2016)026-e. [109]

Government of Türkiye (2005), *Criminal Code of Türkiye*, https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2016)011-e. [89]

GSMA (2016), *Notice and Takedown: Company policies and practices to remove online child sexual abuse material*, https://www.gsma.com/publicpolicy/resources/notice-takedown-company-policies-practices-remove-online-child-sexual-abuse-material. [44]

Hall, Z. (2018), *Apple abruptly pulled Telegram last week when it learned app was serving child pornography*, https://9to5mac.com/2018/02/05/apple-telegram-illegal-content/. [178]

ICMEC (2018), *Child Sexual Abuse Material: Model Legislation & Global Review, 9th Edition*, https://cdn.icmec.org/wp-content/uploads/2018/12/CSAM-Model-Law-9th-Ed-FINAL-12-3-18-1.pdf. [69]

ICT Coalition (n.d.), *ICT Coalition for Children Online*, https://www.ictcoalition.eu/. [45]

IICSA (2022), *The Report of the Independent Inquiry into Child Sexual Abuse*, https://www.iicsa.org.uk/document/report-independent-inquiry-child-sexual-abuse-october-2022-0. [113]

INHOPE (2021), "Annual Report 2021", https://inhope.org/media/pages/articles/annual-reports/8fd77f3014-1652348841/inhope-annual-report-2021.pdf. [26]

Instagram (n.d.), *Account Status on Instagram*, https://help.instagram.com/539126347315373. [163]

Instagram (n.d.), *How does Instagram use artificial intelligence to moderate content?*, https://help.instagram.com/423837189385631. [166]

Instagram (n.d.), *I don't think Instagram should have taken down my post*, https://help.instagram.com/280908123309761/?helpref=uf_share. [165]

Instagram (n.d.), *What can I do if my Instagram account has been disabled?*, https://help.instagram.com/366993040048856/. [164]

INTERPOL (2022), *INTERPOL Secretary General: Online child sexual abuse at record levels*, [24]

https://www.interpol.int/en/News-and-Events/News/2022/INTERPOL-Secretary-General-Online-child-sexual-abuse-at-record-levels.

IWF (2023), *IWF Annual Report 2022 #BehindTheScreens*, https://www.iwf.org.uk/about-us/who-we-are/annual-report-2022/. [21]

IWF (2021), *The Annual Report 2021*, Internet Watch Foundation, https://www.iwf.org.uk/about-us/who-we-are/annual-report-2021/. [22]

Jang, H. (2022), *Tech Companies Sit on Sidelines While Korean Children Are Drawn Into Digital Sex Trafficking*, https://techpolicy.press/tech-companies-sit-on-sidelines-while-korean-children-are-drawn-into-digital-sex-trafficking/. [202]

JOYY Inc. (2021), *JOYY Reports First Quarter 2021 Unaudited Financial Results*, https://ir.joyy.sg/node/9156/pdf. [134]

Kakao (n.d.), *Kakao Transparency Report*, https://privacy.kakao.com/transparency/report?lang=en. [201]

Kastrenakes, J. (2021), *Apple says there are now over 1 billion active iPhones*, https://www.theverge.com/2021/1/27/22253162/iphone-users-total-number-billion-apple-tim-cook-q1-2021. [116]

Lardinois, F. (2020), *Microsoft Teams is coming to consumers — but Skype is here to stay*, https://techcrunch.com/2020/03/30/microsoft-teams-is-coming-to-consumers-but-skype-is-here-to-stay/. [135]

Lee, H. (2020), "Detecting child sexual abuse material: A comprehensive survey", *Forensic Science International: Digital Investigation*, Vol. 34, pp. https://doi.org/10.1016/j.fsidi.2020.301022. [50]

Levy, I. and C. Robinson (2022), *Thoughts on Child Safety on Commodity Platforms*, https://arxiv.org/pdf/2207.09506.pdf. [59]

Line (2021), *LINE Content Moderation Report*, https://linecorp.com/en/security/moderation/2021h1. [196]

Line Corporation (2020), *LINE Q3 2020 Earnings Results*, https://d.line-scdn.net/stf/linecorp/en/ir/all/FY20Q3_earning_releases_EN.pdf. [130]

LinkedIn (2022), *How our content abuse defense systems work to keep members safe*, https://engineering.linkedin.com/blog/2022/how-our-content-abuse-defense-systems-work-to-keep-members-safe. [191]

LinkedIn (n.d.), *Community Report*, https://about.linkedin.com/transparency/community-report. [190]

Lu, S. (2021), *I helped build ByteDance's vast censorship machine*, https://www.protocol.com/china/i-built-bytedance-censorship-machine. [200]

Mail.ru (2021), *Social Networks*, https://corp.mail.ru/en/company/social/. [136]

Marketing to China (2021), *Guide to Douban Marketing*, https://marketingtochina.com/guide-to-douban-marketing/. [123]

Marketing to China (2021), *Top 10 Chinese Social Media for Marketing (updated 2021)*, https://www.marketingtochina.com/top-10-social-media-in-china-for-marketing/. [124]

Medium (2015), *Law Enforcement Demands for User Information or Content Removal (2014)*, https://medium.com/transparency-report/government-requests-for-information-or-content-removal-9b23349b0e73. [195]

Messenger (n.d.), *Open up as you are – privately, safely, confidently*, https://www.messenger.com/privacy. [167]

Messenger (n.d.), *Privacy and safety in Messenger*, https://www.facebook.com/help/messenger-app/1064701417063145. [168]

Meta (2022), *Meta Launches New Content Moderation Tool as It Takes Chair of Counter-Terrorism NGO*, https://about.fb.com/news/2022/12/meta-launches-new-content-moderation-tool/. [149]

Meta (2022), *Meta's Approach to Safer Private Messaging on Messenger and Instagram Direct Messaging*, https://messengernews.fb.com/wp-content/uploads/2021/12/Metas-approach-to-safer-private-messaging-on-MSGR-and-IG-DMs-4.pdf. [169]

Meta (2021), *A Privacy-Focused Vision for Social Networking*, https://www.facebook.com/notes/2420600258234172/. [62]

Meta (2021), *Preventing Child Exploitation on Our Apps*, https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps/. [147]

Meta (2018), *New Technology to Fight Child Exploitation*, https://about.fb.com/news/2018/10/fighting-child-exploitation/. [58]

Meta (n.d.), *Child Sexual Exploitation, Abuse and Nudity*, https://transparency.fb.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/. [145]

Meta (n.d.), *How Meta enforces its policies*, https://transparency.fb.com/enforcement/. [146]

Meta (n.d.), *Transparency Reports*, https://transparency.fb.com/data/. [148]

Microsoft (n.d.), *Digital Safety Content Report*, https://www.microsoft.com/en-us/corporate-responsibility/digital-safety-content-report. [194]

Microsoft (n.d.), *PhotoDNA*, https://www.microsoft.com/en-us/photodna. [52]

Mubarak, A. (2020), "A Study on Internet Industry Self-regulation in China and Its Implications for Child Protection in Cyberspace", *The International Journal of Community and Social Development*, Vol. 2/3, pp. 297–309. [31]

Nair, A. (2019), *The Regulation of Internet Pornography Issues and Challenges*, Routledge. [206]

Napier, S., C. Teunissen and H. Boxall (2021), *Live streaming of child sexual abuse: An analysis of offender chat logs*, https://www.aic.gov.au/sites/default/files/2021-10/ti639_live_streaming_of_child_sexual_abuse.pdf. [172]

NCMEC (2023), *CyberTipline 2022 Report*, https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata. [18]

NCMEC (2022), *2021 CyberTipline Reports by Electronic Service Providers*, [150]

https://www.missingkids.org/content/dam/missingkids/pdfs/2021-reports-by-esp.pdf.

NCMEC (2022), *National Centre for Missing and Exploited Children: Our 2021 Impact*, https://www.missingkids.org/ourwork/impact. [19]

NCMEC (n.d.), *What is the CyberTipline?*, https://www.missingkids.org/blog/2022/what-is-cybertipline. [35]

New Zealand Department of Internal Affairs (1993), *Films, Videos, and Publications Classification Act 1993*, https://www.legislation.govt.nz/act/public/1993/0094/55.0/DLM312895.html. [87]

New Zealand, Department of Internal Affairs (2021), *Films, Videos, and Publications Classification (Urgent Interim Classification of Publications and Prevention of Online Harm) Amendment Act 2021*, https://www.legislation.govt.nz/act/public/2021/0043/latest/LMS294551.html. [108]

New Zealand, Ministry of Justice (2021), *Crimes Act 1961*, https://www.legislation.govt.nz/act/public/1961/0043/137.0/DLM328025.html. [88]

Ngo, V. et al. (2022), *Investigation, Detection and Prevention of Online Child Sexual Abuse Materials: A Comprehensive Survey*. [51]

OECD (2022), "Transparency reporting on terrorist and violent extremist content online 2022", *OECD Digital Economy Papers*, Vol. No. 334/https://doi.org/10.1787/a1621fc3-en. [3]

OECD (2022), *Voluntary Transparency Reporting Framework pilot*, https://www.oecd-vtrf-pilot.org/. [49]

OECD (2021), *Children in the digital environment: Revised typology of risks*, OECD Publishing, https://www.oecd-ilibrary.org/science-and-technology/children-in-the-digital-environment_9b8f222e-en. [27]

OECD (2021), *OECD Guidelines for Digital Service Providers*, https://legalinstruments.oecd.org/api/download/?uri=/private/temp/cecf7c1a-2590-4aaf-98d7-2a5f74290b92.pdf&name=Guidelines%20for%20Digital%20Service%20Providers.pdf. [16]

OECD (2021), *Recommendation of the Council on Children in the Digital Environment*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389. [12]

OECD (2021), *Transparency reporting on terrorist and violent extremist content online : An update on the global top 50 content sharing services*, OECD Publishing, Paris, https://doi.org/10.1787/8af4ab29-en. [2]

OECD (2020), *Current approaches to terrorist and violent extremist content among the global top 50 online content-sharing services*, OECD Publishing, https://doi.org/10.1787/68058b95-en. [1]

OHCHR (2023), *Status of Ratification*, https://indicators.ohchr.org/. [73]

Patel, N. (2022), *How WordPress and Tumblr are keeping the internet weird*, https://www.theverge.com/2022/3/15/22977857/wordpress-tumblr-simplenote-internet-automattic-matt-mullenweg-interview. [204]

Patel, N. (2021), *Here's why Apple's new child safety features are so controversial*, https://www.theverge.com/22617554/apple-csam-child-safety-features-jen-king-riana-pfefferkorn-interview-decoder. [161]

Pinterest (n.d.), *Review the Pinner Promise*, https://help.pinterest.com/en/article/review-the-pinner- [182]

promise.

Pinterest (n.d.), *Transparency Report*, https://policy.pinterest.com/en/transparency-report. [183]

PRS Legislative Research (2021), *The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*, https://prsindia.org/billtrack/the-information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021. [157]

Quora (n.d.), *What is Quora's policy on content that advocates certain harmful actions?*, https://help.quora.com/hc/en-us/articles/360015476372-What-is-Quora-s-policy-on-content-that-advocates-certain-harmful-actions-]. [193]

Rakuten Viber (2022), *Viber Signs EU Code of Conduct: At Viber, There's No Place for Hate*, https://www.viber.com/en/blog/2022-05-11/viber-eu-code-of-conduct/. [171]

Reddit (2021), *Transparency Report 2021*, https://www.redditinc.com/policies/transparency-report-2021-2/. [184]

Reddit (n.d.), *User Management - moderators and permissions*, https://mods.reddithelp.com/hc/en-us/articles/360009381491-User-Management-moderators-and-permissions. [185]

Republic of Korea (2020), *Act on the Protection of Children and Youth against Sexual Abuse (Act No. 9765)*, https://elaw.klri.re.kr/eng_service/lawView.do?hseq=56569&lang=ENG. [86]

Republic of Korea (2020), *Telecommunications Business Act*, https://elaw.klri.re.kr/eng_service/lawView.do?hseq=60897&lang=ENG. [107]

Republic of Korea (2020), *The Act on Promotion of Information and Communications Network Utilization and Data Protection (the "Network Act")*, https://elaw.klri.re.kr/eng_service/lawView.do?hseq=60899&lang=ENG. [106]

Republic of Korea (2009), *Act on the Protection of Children and Youth from Sexual Abuse (Act No. 9765)*, https://www.moleg.go.kr/index.es?sid=a3. [208]

République Française (2021), *Code pénal*, https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000043409170. [85]

République Française (2007), *Loi n° 2007-297 du 5 mars 2007 relative à la prévention de la délinquance (1)*, https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000615568. [102]

République Française (2004), *Loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique*, https://www.legifrance.gouv.fr/loda/id/LEGITEXT000005789847. [101]

République Française (1998), *Loi n° 98-468 du 17 juin 1998 relative à la prévention et à la répression des infractions sexuelles ainsi qu'à la protection des mineurs*, https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000556901/. [83]

Safety Tech Challenge Fund (n.d.), *Safety Tech Challenge Fund: Meet the Teams*, https://express.adobe.com/page/vxBuQnoqvGhYE/. [60]

Sensor Tower (2020), *How PicsArt has Thrived in the Competitive Photo & Video Category*, https://sensortower.com/blog/picsart-interview. [131]

Shumin, L. (2018), *China's Major Online Video Platforms Clamp Down on Disturbing Children's Videos*, https://www.yicaiglobal.com/news/china-major-online-video-platforms-clamp-down-on- [176]

disturbing-children-videos.

Simon, M. (2022), *Apple and CSAM Scanning: The latest news*, https://www.macworld.com/article/352875/ios-15-csam-scanning-icloud-photos-messages-siri-search-faq.html. [160]

Smith, C. (2021), *IMO Statistics, User Counts and Facts (2021)*, https://expandedramblings.com/index.php/imo-facts-and-statistics/. [126]

Snap Inc. (2022), *Transparency Report January 1, 2022 – June 30, 2022*, https://www.snap.com/en-US/privacy/transparency. [181]

Snap Inc. (n.d.), *Transparency Report Glossary*, https://values.snap.com/privacy/transparency/glossary. [180]

Startup Talky (2020), *YouTube vs Vimeo: A Detailed Comparison*, https://startuptalky.com/youtube-vs-vimeo/. [128]

Statista (2021), *Average number of monthly active users (MAU) of Chinese video app iQiyi from 2016 to 2020*, https://www.statista.com/statistics/1106091/china-online-video-platform-iqiyi-mobile-app-monthly-active-user-number/. [120]

Statista (2021), *Number of global monthly active Kakaotalk users from 1st quarter 2013 to 4th quarter 2020*, https://www.statista.com/statistics/278846/kakaotalk-monthly-active-users-mau/. [139]

Statista (2021), *Number of monthly active users of popular short video apps in China as of May 2021*, https://www.statista.com/statistics/910633/china-monthly-active-users-across-leading-short-video-apps/. [137]

Statista (2021), *Total global visitor traffic to Zoom.us 2021*, https://www.statista.com/statistics/1259905/zoom-website-traffic/. [115]

Tech Coalition (2022), *Trust: Voluntary Framework for Industry Transparency*, https://www.technologycoalition.org/knowledge-hub/trust-voluntary-framework-for-industry-transparency. [43]

Tech Coalition (n.d.), *Who We Are*, https://www.technologycoalition.org/about. [42]

Techcircle (2021), *Microsoft Teams reaches 250 million global MAU milestone*, https://www.techcircle.in/2021/07/29/microsoft-teams-reaches-250-million-global-mau-milestone. [125]

Telegram (n.d.), *Stop Child Abuse*, https://t.me/s/stopCA. [177]

Tencent (n.d.), *User Reporting Function*, https://kf.qq.com/faq/161227viue2M1612276jAFNB.html. [175]

The Police Foundation (2022), *Turning the Tide Against Online Child Sexual Abuse*, https://www.police-foundation.org.uk/2017/wp-content/uploads/2022/07/turning_the_tide_FINAL-.pdf. [23]

Thorn (n.d.), *Safer: How it Works*, https://safer.io/about/. [56]

Thurman, N., A. Nalmpatian and F. Obster (2022), "Lessons from France on the regulation of Internet pornography: How displacement effects, circumvention, and legislative scope may limit [103]

the efficacy of Article 23", *Policy & Internet*, Vol. 14/3, pp. 690-710.

TikTok (n.d.), *TikTok Transparency Center*, https://www.tiktok.com/transparency/en-us/. [174]

Tumblr (n.d.), *Content moderation on Tumblr*, https://help.tumblr.com/hc/en-us/articles/360011799473-Content-moderation-on-Tumblr. [189]

Türkiye Information Technologies and Communications Authority (2022), *ihbar web*, https://www.ihbarweb.org.tr/. [90]

Türkiye Information Technologies and Communications Authority (2018), *Güvenli internet merkezi*, https://www.gim.org.tr/. [110]

Twitch (n.d.), *Transparency Report*, https://safety.twitch.tv/s/article/H2-2021-Transparency-Report?language=en_US. [198]

Twitter (2023), *CSE Account Suspensions*, https://twitter.com/TwitterSafety/status/1620908367412707328?s=20&t=6FPHrY_iLZue3lVr2Nvh9g. [188]

Twitter (2022), *Sharing our latest transparency update, marking decade long commitment*, https://blog.twitter.com/en_us/topics/company/2022/ttr-20. [187]

Twitter (2019), *A healthier Twitter: Progress and more to do*, https://blog.twitter.com/en_us/topics/company/2019/health-update. [186]

U.S. Attorney's Office (2022), *St. Charles County Man Admits Downloading and Sharing Child Pornography*, https://www.justice.gov/usao-edmo/pr/st-charles-county-man-admits-downloading-and-sharing-child-pornography. [173]

U.S. Congress (2020), *S.4051 - Lawful Access to Encrypted Data Act*, https://www.congress.gov/bill/116th-congress/senate-bill/4051/text. [64]

U.S. Congress (1982), *United States Code: Sexual Exploitation of Children, 18 U.S.C. §§ 2251-2253 (1982)*, https://www.loc.gov/item/uscode1982-007018110/. [95]

UK Government (2022), *Five Country Ministerial Statement on 2-year anniversary of Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse (accessible version)*, https://www.gov.uk/government/publications/voluntary-principles-to-counter-online-child-sexual-exploitation-and-abuse/five-country-ministerial-statement-on-2-year-anniversary-of-voluntary-principles-to-counter-online-child-sexual-exploitation-and-abuse-ac. [8]

UK Government (2003), *Sexual Offences Act 2003*, https://www.legislation.gov.uk/ukpga/2003/42/contents. [93]

UK Government (1988), *Criminal Justice Act 1988*, https://www.legislation.gov.uk/ukpga/1988/33/contents. [92]

UK Government (1978), *Protection of Children Act 1978*, https://www.legislation.gov.uk/ukpga/1978/37. [91]

UK Parliament (2022), *Online Safety Bill*, https://bills.parliament.uk/bills/3137. [67]

UN Committee on the Rights of the Child (2019), *Guidelines regarding the implementation of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child* [74]

prostitution and child pornography, https://www.ohchr.org/en/documents/legal-standards-and-guidelines/crcc156-guidelines-regarding-implementation-optional.

UN General Assembly (2022), *Third intersessional consultation of the Ad Hoc Committee*, https://www.unodc.org/unodc/en/cybercrime/ad_hoc_committee/intersessional-consultations/3rd-intersessional-consultation.html. [77]

UN General Assembly (2001), *Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography*, https://www.ohchr.org/en/instruments-mechanisms/instruments/optional-protocol-convention-rights-child-sale-children-child. [70]

UN General Assembly (1989), *Convention on the Rights of the Child*, http://www2.ohchr.org/english/law/crc.htm. [28]

United Kingdom (2009), *Coroners and Justice Act 2009*, https://www.legislation.gov.uk/ukpga/2009/25/section/68. [111]

United Nations (2000), *Vienna commitment against child pornography on the internet : conclusions and recommendations of the international conference "Combating Child Pornography on the Internet", 29 September to 1 October 1999*, https://digitallibrary.un.org/record/432766?ln=en. [33]

United States, Department of Justice (n.d.), *Citizen's Guide to U.S. Federal Law on Child Pornography*, https://www.justice.gov/criminal-ceos/citizens-guide-us-federal-law-child-pornography. [94]

V-click Technology (2021), *Stats summary about China social media in 2020 when covid-19 this the world*, https://www.v-click.co.th/china-social-media-in-2020-covid/. [118]

VK (n.d.), *Social Accountability*, https://vk.com/safety. [199]

Warner, A. (2021), *Which Social Media Platform Has the Most Users?*, https://www.websiteplanet.com/blog/social-media-platform-users/. [119]

WeProtect Global Alliance (2022), *Global Taskforce on Child Sexual Abuse Online*, https://www.weprotect.org/library/global-taskforce-on-child-sexual-abuse-online/. [6]

WeProtect Global Alliance (2021), *Findings from WeProtect Global Alliance / Tech Coalition Survey of Technology Companies*, https://www.weprotect.org/wp-content/uploads/Survey-of-technology-companies-2021.pdf. [25]

WeProtect Global Alliance (2021), *Global Threat Assessment 2021*, https://www.weprotect.org/global-threat-assessment-21/. [20]

WeProtect Global Alliance (2019), *Global Strategic Response*, https://www.weprotect.org/library/global-strategic-response/. [40]

WeProtect Global Alliance (2016), *Model National Response*, https://www.weprotect.org/model-national-response/. [39]

WeProtect Global Alliance (2016), *Preventing and Tackling Child Sexual Exploitation and Abuse (CSEA): A Model National Response*, https://www.weprotect.org/model-national-response/. [68]

WeProtect Global Alliance (n.d.), *Defining child sexual exploitation and abuse online*, https://www.weprotect.org/issue/. [4]

WeProtect Global Alliance (n.d.), *Timeline*, https://www.weprotect.org/who-we-are/timeline/. [38]

WhatsApp (n.d.), *How WhatsApp Helps Fight Child Exploitation*, https://faq.whatsapp.com/154956905959033/?helpref=uf_share. [63]

Wikimedia (2021), *Unique devices*, https://www.envisagedigital.co.uk/wordpress-market-share/. [144]

Wikimedia Foundation (n.d.), *Transparency Report*, https://wikimediafoundation.org/about/transparency/2020-2/. [205]

Willens, M. (2021), *Four years into a subscription strategy, Medium still doesn't spend money to acquire subscribers*, https://digiday.com/media/four-years-into-a-subscription-strategy-medium-still-doesnt-spend-money-to-acquire-subscribers/. [129]

Xinhua (2021), *China issues guideline on developing civilized cyberspace*, http://english.www.gov.cn/policies/latestreleases/202109/14/content_WS614097a4c6d0df57f98e028b.html. [29]

YouTube (n.d.), *Child safety policy*, https://support.google.com/youtube/answer/2801999. [151]

Zoom (n.d.), *Community Standards Enforcement*, https://explore.zoom.us/en/trust/community-standards-enforcement/. [155]

Zoom (n.d.), *Expression, Safety and Process at Zoom*, https://explore.zoom.us/en/expression-safety-and-process-at-zoom/. [154]

Zoom (n.d.), *Our Tier Review System*, https://explore.zoom.us/en/content-moderation-process/. [156]

Notes

[1] The Recommendation defines Digital Service Providers as "any natural or legal person that provides products and services, electronically and at a distance" (OECD, 2021[12]).

[2] See Section 3, para 53, "Technology solutions for tackling CSEA".

[3] Facebook, Instagram, Google, WhatsApp, and Omegle.

[4] IWF categorises child sexual abuse images and videos based on UK law according to the levels in the Sentencing Council's Sexual Offences Definitive Guidelines. Category A content, the worst category of abuse consists of "*Images involving penetrative sexual activity; images involving sexual activity with an animal or sadism*". (IWF, 2021[16]).

[5] Responses were received from: Ask.fm (IAC [InterActiveCorp]), Discord (Discord, Inc.), Dropbox (Dropbox, Inc.), Facebook (Meta Platforms, Inc.), Facebook Messenger (Meta Platforms, Inc.), Flickr (SmugMug, Inc.), Google Drive (Alphabet, Inc.), Instagram (Meta Platforms, Inc.), Likee (BIGO Technology PTE. LTD.), LinkedIn (Microsoft, Inc.), Microsoft OneDrive (Microsoft, Inc.), Microsoft Teams (Microsoft, Inc.), PicsArt (PicsArt, Inc.), Skype (Microsoft, Inc.), Snapchat (Snap, Inc.), Tumblr (Automattic, Inc.), Twitch (Amazon.com, Inc.), Twitter (Twitter, Inc.), Viber (Rakuten, Inc.), Vimeo (Vimeo, Inc.), VK (Mail.Ru Group), WhatsApp (Meta Platforms, Inc.), Wikipedia (Wikimedia Foundation), Wordpress (Automattic, Inc.), YouTube (Alphabet, Inc.), Zoom (Zoom Video Communications, Inc.).

[6] The nine platforms where evidence from third party media reports is cited regarding the dissemination of CSEA are: iQIYI (Baidu, Inc.), KaKao Talk (Daum Kakao Corporation), Kuaishou (Beijing Kuaishou Technology Co., Ltd), QQ (Tencent Holdings Ltd.), QZone (Tencent Holdings Ltd.), Viber (Rakuten, Inc.), Weibo (Sina Corp.), Weixin/WeChat (Tencent Holdings Ltd.), Youku Tudou (Alibaba Group Holding Limited).

[7] The six services where it is unknown if the service has been used to disseminate CSEA are: Baidu Tieba (Baidu, Inc.), Douban (Information Technology Company, Inc.), Huoshan (ByteDance Technology Co.), IMO (PageBites, Inc.), Likee (BIGO Technology PTE. LTD.), Odnoklassniki (Mail.Ru Group) and Xigua Video (ByteDance Technology Co.).

[8] For example, Microsoft's Digital Safety Content Report is an aggregate one for all Microsoft hosted consumer services and a breakdown by individual products - including OneDrive, Outlook, Skype and Xbox - is not available. According to the Digital Safety Content Report for July-December 2021, Microsoft actioned 36,918 pieces of content and 11,805 consumer accounts associated with CSEA during this period.

[9] In July 2023 Twitter (operated by Twitter Inc) was rebranded as X (operated by X Corp). At the time of drafting this report and reviewing the company's policies and procedures it was operating under Twitter Inc. As such, this report refers to Twitter Inc, and the analysis is relevant to the company's policies and procedures as of late 2022 / early 2023 prior to its rebranding.

[10] The 12 Chinese owned services are: Baidu Tieba (Baidu, Inc.), Douban (Information Technology Company, Inc.), Huoshan (ByteDance Technology Co.), iQIYI (Baidu, Inc.), Kuaishou (Beijing Kuaishou Technology Co., Ltd), QQ (Tencent Holdings Ltd.), QZone (Tencent Holdings Ltd.), Tik Tok (ByteDance

Technology Co.), Weibo (Sina Corp.), Weixin/WeChat (Tencent Holdings Ltd.), Xigua Video (ByteDance Technology Co.), Youku Tudou (Alibaba Group Holding Limited).

[11] For example, BBC news reported in July 2021 that the Cyberspace Administration of China (CAC), the government oversight body had fined leading Chinese digital providers such as QQ, Alibaba and Weibo for hosting illegal child abuse content. The platforms were given a deadline to "rectify" and "clean up" all illegal content on their services (BBC, 2021[170])- Rdd AAB Mdv r +=€hina: Taobao, Weibo fined for illegal child content ╟+10 I t ₭+1/ 10- @ư` ₭ akd ` s9https://www.bbc.com/news/business-57911207.

[12] The IWF is a United Kingdom based child protection organisation that uses technology to find and remove CSEA online.

[13] ICMEC is an international NGO that develops resources for governments, law enforcement, NGOs, and families on prevention of CSEA. ICMEC's programmes include the Model Legislation and Global Review now in its 9th edition (ICMEC, 2018[69]).

[14] See (European Commission, 2022, p. 284 and ff[53]) for a discussion of the implications of end-to-end encryption for detecting and combatting child sexual abuse.

[15] See (Nair, 2019[206])for a discussion of this point. The United Kingdom was one of the first countries to make mere possession of child pornography a criminal offence under section 160(1) of the Criminal Justice Act 1988, in addition to its production or distribution. The offence of possession has become even more significant in the digital age given the relative ease of access to CSEA (Nair, 2019, p. 58 and following[206]).

[16] All but three of the 38 OECD member countries are parties to the Budapest Convention. Three countries (Korea, Mexico and New Zealand) have yet to ratify the treaty. In 2021, New Zealand indicated its intention to do so. See New Zealand government press release "New Zealand to join the Council of Europe Convention on Cybercrime", 18 February 2021. Available at: https://www.beehive.govt.nz/release/new-zealand-join-council-europe-convention-cybercrime.

[17] In December 2020, the European Commission proposed two legislative initiatives to upgrade rules governing digital services in the EU: the Digital Services Act (DSA) and the Digital Markets Act (DMA). The EU Digital Services Act was signed by the Presidents of the European Parliament and Council on 19 October 2022.The DSA will be directly applicable across the EU and will apply 15 months or from 1 January 2024, whichever comes later, after entry into force.

[18] An indictable offence is the most serious category of offences in Commonwealth countries. They are tried in superior courts, and generally punishable by imprisonment.

[19] See (Bulger, 2017[207]) for a discussion of this issue of the anomalies that have arisen in cases of laws dealing with so-called "sexting".

[20] Per the Act on the Protection of Children and Youth against Sexual Abuse: "Provided, That persons for whom the first day of January of the year in which they reach 19 years of age has arrived shall be excluded" (Republic of Korea, 2009[208]).

[21] United Kingdom laws with relevance to England, Wales, Scotland and Northern Ireland.

[22] See R v Smith and Jayson (2003)1 Cr.App.R.13. Available at: https://vlex.co.uk/vid/r-v-smith-graham-793762225.

[23] The First Amendment to the U.S. Constitution states that *"Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances".* First Amendment rights thereby seeks to protect freedom of speech, the press, assembly, and the right to petition the Government for a redress of grievances.

[24] Information on the Industry Codes is available on the eSafety Commissioner's website at: Industry codes | eSafety Commissioner.

[25] Reports of responses to transparency notices are published on the eSafety Commissioner's website at: https://www.esafety.gov.au/industry/basic-online-safety-expectations/responses-to-transparency-notices.

[26] Including "Child or youth sexual exploitation materials" defined in subparagraph 5 of Article 2 of the Act on the Protection of Children and Youth against Sex Offenses.

[27] In July 2023 Twitter (operated by Twitter Inc) was rebranded as X (operated by X Corp). At the time of drafting this report and reviewing the company's policies and procedures it was operating under Twitter Inc. As such, this report refers to Twitter Inc, and the analysis is relevant to the company's policies and procedures as of late 2022 / early 2023 prior to its rebranding.

[28] Note that the Microsoft Services Agreement applies only to consumer use of Teams, and not to enterprise use. In an enterprise context, Microsoft acts as a data processor and the enterprise customer controls all customer content, including end user content – any rights for the service provider to access and/or process the organisational customer's content are defined in (and constrained by) the legal agreement.