VIOLENCE AGAINST WOMEN AND CHILDREN HELPDESK

December 2024

# Lessons from the Online Child Safety Sector

Jo Feather and Valeria Miglio

**Query:** What lessons can be learnt from multi-stakeholder, cross-sectoral engagement, specifically with the technology sector, to promote Child Online Safety that can be applied to global efforts to end Technology-Facilitated Gender Based Violence, including but not limited to non-consensual intimate image abuse? Please identify best practice examples and initiatives or learning from failure in the following areas: (a) technology tools and innovations (b) data, and evidence generation;  (; (c) advocacy and policy / guideline development.

INTERNATIONAL RESCUE COMMITTEE
Raising Voices
Social Development Direct
Breakthrough
care

## Summary

There is strong consensus in the literature and among key informants for this rapid query that an effective response to online safety necessitates collaboration across the entire ecosystem. This includes the technology industry, governments, law enforcement, users, parents, children and the broader community, including civil society.  This rapid query highlights a number of lessons learnt from multi-stakeholder collaboration across the technology sector on online child safety to draw out their potential relevant to designing and implementing interventions to address technology-facilitated gender-based violence (TFGBV).

- **Consistency of terminology** remains a key challenge in the online child safety space and there is also not yet consistency in terminology and definitions used for typologies of TFGBV. Although a number of coordinated initiatives are underway to identify and agree categorisation and definitions.[1] It will be important to ensure that this work recognises TFGBV as a continuum of gender-based violence (GBV), that forms of TFGBV are wide and varied, and that many of the characteristics of TFGBV are shared with other forms of GBV. Whilst TFGBV also has distinct characteristics related to the digital nature of abuse, including the scale, speed and impact with which violence can happen. Efforts to understand and measure, and indeed address, TFGBV must be situated within the contexts of both GBV and digital exclusion, both of which are underpinned by structural gender inequality.[2]
- **Consensus around the need to address child online safety** has been an important motivator for the progress made to date in this sector, which has been reinforced by legislation and regulation which has driven this work. Supporting a similar degree of consensus for all forms of TFGBV would support a more coordinated response to all types of online harms.  Whilst those working in this area are largely aligned, the wider technology, and legislative worlds do not always cohere with these perspectives. It is worth highlighting that coordinated enforcement mechanisms have been easier to develop for CSAM, perhaps due to the stronger legal and international consensus on severity and criminality of these activities.
- **Aligning with other global digital agendas*.*** Due to limited resources for this work it is imperative that stakeholders consider entry points and synergies to connect efforts that could bring a multiplicity of benefits. Key to this is generating political will to open up funding though building consensus and consistency around terminology and understanding where TFGBV interacts with digital inclusion and safety agendas, alongside online child safety – for example the Global Digital Compact.[3]
- **The role of effective legal frameworks** on violence against children has facilitated multi-sectoral collaboration around online child safety, but there are significant challenges and limitations in their implementation, and this is particularly the case for TFGBV.
- **The role of convening partners** bringing stakeholders together across sectors to support understanding and trust building was found to be crucial in multisectoral collaboration in the online child safety space, and the same would apply to TFGBV. These convening partners

need to be well networked and respected across sectors, able to play a bridging role among diverse interest groups.

- **Sustainable and reliable funding** continues to be one of the biggest barriers to building and maintaining a prevention and response approach to addressing online child safety. The issue is even more acute for TFGBV, where funding for GBV prevention is already limited.
- **The role of the Trust and Safety teams** has been vital in elevating these issues within their companies, and their integration into core teams, and access to the executive decision makers within the technology industry has proved more effective in achieving the changes required. Continuing to engage them and lobby for their involvement and influence in this work, as well as advocating for their continued resourcing, is vital to the success of TFGBV.

## 1. Introduction

Online child safety is an increasingly urgent global priority, as the digital environment continues to grow at an unprecedented pace and children spend more time online. Because of its digital nature, online child safety is a constantly evolving area of work. It encompasses online child sexual exploitation and abuse (OCSEA), including financial sextortion, child sexual abuse material (CSAM), and online grooming, which can result in self-generated CSAM, as well as other non-sexual online harms such as cyberbullying and discrimination. This paper focuses primarily on CSAM. A glossary including key terminology mentioned throughout this report is included in Annex 1.

The number of children who experience online sexual exploitation and abuse is growing. According to the Childlight Into the Light Index, over 300 million children are affected by OCSEA each year.[4] Most recently, emerging technologies such as generative artificial intelligence (AI) are amplifying these risks and demonstrating the rapid evolution of the current threat landscape.[5]

Perpetrators of online harm include individuals known to children, including friends and family members, as well as strangers, of different age groups.[6] The design of platforms can also influence users' exposure to these risks.

In recognition of these harms, new approaches and solutions have emerged across advocacy, technology, partnerships, and evidence generation. This includes an increased recognition of the importance of Safety by Design, an approach that encourages technology companies to build platforms and services in such a way that online harms are anticipated, identified, and addressed before they occur.[7] Advancements in the field of online child safety can generate useful learning for other types of online harms, including TFGBV and its different forms.[8]

There are various important elements in addressing online child safety, including tackling deeply rooted social norms and behaviours that undermine children's safety in both the online and offline space, as well as the importance of addressing the wider issue of the 'gender digital divide' which is further exacerbated by the risk of violence online. This query focuses on four key areas, but it is important to recognise the need for comprehensive approaches that integrate approaches to tackle social norms and behaviour. This report explores lessons learnt in the online child safety space that can be applied to efforts to address TFGBV – the authors of the report use this as an overarching term to be inclusive of all forms of GBV that are facilitated online and through digital technologies, including those that do not make use of the internet. In recognition of the existing collaborative efforts in the space and their role in driving solutions forward, this report adopts a specific focus on multi-stakeholder, cross-sectoral engagement involving the technology sector. Action at the international level has largely focused on OCSEA over other forms of non-sexual forms of online harms against children, such as cyberbullying and discrimination. The focus on OCSEA is reflected in the partnerships and collaborations discussed in this report.

The report has a global focus, although it is important to note that there are existing gaps in the evidence base with respect to the inclusion of perspectives from low- and middle-income Countries (LMICs). This presents a significant opportunity for future research, to ensure that lessons learnt are inclusive of a diverse range of contexts. Additional details on limitations faced by this study are included in Annex 2.

The report is informed by eight key informant interviews (KIIs) with representatives of key multi-stakeholder, cross-sectoral initiatives involving the technology sector and active in the online child safety space. This primary data is complemented by a desk review. A more detailed methodology is included in Annex 2. To further contextualise this research, Section 2 provides an overview of the evolution of the online child safety space, and Section 3 presents three thematic case studies focusing on technology tools and innovation, data and evidence generation, and advocacy, legislation, and policy guidance development. These case studies are informed by both the KIIs and desk review conducted as part of this research. The report concludes by illustrating a checklist of elements to consider for multi-stakeholder collaboration on TFGBV, informed by learning from online child safety sector.

## 2. Evolution of the Online Child Safety Space

**In response to the increased uptake of technology, the landscape of online harms continues to expand at an unprecedented rate**. Over the course of 2024, it is estimated that 1 in 8 children faced non-consensual image offences (12.6%) and online solicitation (12.5%).[9] The rapid advancement of technology is increasing the complexity of the issue. For instance, the evidence reviewed for this paper found an increase in reported cases of perpetrators using generative AI.[10] Over the course of 2023, the National Centre for Missing & Exploited Children (NCMEC) received 4,700 reports of CSAM generated by AI. Although this still constitutes a small portion of OCSEA reports, the evidence suggests that these numbers are steadily increasing.[11] The unique challenges posed by the rise of generative AI have encouraged the development of new technologies in the context of both OCSEA and TFGBV to counter this threat (see Section

4.2). Other emerging technologies, such as extended reality (XR), which include virtual and augmented reality, create immersive digital spaces that can expose children to additional threats such as grooming, exploitation, and harmful content.[12]

**The legal environment has evolved over time in response to the rise of threats to online child safety.** A substantial momentum in regulation has emerged over the last two decades, starting with the 2007 Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse. The latest CSAM Legislation Review found that only 10 countries globally do not have any legislation that specifically addresses CSAM, with 85 countries having adopted some form of legislation since 2006.[13] Recent legislative milestones regulating online harms directed at children among other digital concerns include Australia's Online Safety Act (2021), the EU's Digital Service Act (2022), and the UK's Online Safety Act (2023), as well as the Computer Misuse Act (2018) and Sexual Offences Act in Kenya (2006), along with similar legislation in India and Brazil.

**Beyond legislation, governments have come together through partnerships and initiatives aiming to combat OCSEA**. For example, the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse, which aim to inform and drive collective industry action, were first developed in 2020 and recently updated in 2022 by the governments of Australia, Canada, New Zealand, the US and the UK  in partnership with sector experts and technology companies.[14] In 2023, 71 UN member states joined together to issue a call to action in favour of urgent efforts to remove and combat online CSAM.[15] In addition, governments have been taking action under the Model National Response developed by the WeProtect Global Alliance as a pathway for countries to build holistic responses to preventing and responding to OCSEA.[16]

**Growing legislative scrutiny has translated into increased collaborative action and effort across sectors**, as described throughout this report. This action has largely focused on OCSEA over other forms of non-sexual forms of online harms against children. In this context, a number of cross-sector, multi-stakeholder partnerships and initiatives involving the tech industry have also emerged to accelerate action to combat OCSEA. These include the following:

- **Tech Coalition** – A global alliance of technology companies working together to advance technology, drive collective action, promote industry-level transparency and accountability, and share best practices to combat OCSEA. It was founded in 2006 as a voluntary group of industry professionals aiming to create a space to collaborate and share expertise on the topic. Today its membership includes leading platforms such as Google, Meta (formerly Facebook), Microsoft, and TikTok. Recent initiatives emerging out of this collaboration include: the Trust Framework, a voluntary framework providing guidance on transparency reporting to technology companies; Pathways, providing access to key resources and support to license child safety technology.
- **WeProtect Global Alliance** – A global movement bringing together governments, the private sector, civil society, and international organisations with the aim developing

solutions to end child sexual abuse and exploitation online. Initially established in 2016 by the UK Government and relaunched in 2020 as an independent organisation, the alliance organises biennial summits, and regular meetings to collaboratively explore solutions, and contributes to evidence generation through its biennial Global Threat Assessment.

## 3. Case studies

The tech industry is engaging in a number of innovative ways in multi-sectoral collaborations, which interviewees identified as being crucial to establishing an effective response (see box below). Below we pull out some common lessons from across the different stakeholders that relate to the overarching theme of industry engagement, before we present three specific case studies related to i) technology tools and innovation, ii) evidence generation and sharing, and iii) advocacy, legislation and policy guidance.

All of these examples shared similar lessons in relation to this engagement including the following:

- **Convening role.** A recurring theme among stakeholders was the importance of having an organisation playing a convening role bringing together different stakeholders. This role was crucially played by Safe Online and We Protect Global Alliance, both of whom were able to mobilise a diverse group of high-profile actors from across sectors. Knowing who to invite into the room and how to establish a safe space where trust and alliances can be built has been crucial for fostering this type of cross sectoral engagement.
- The role of the **Trust and Safety teams** is seen as instrumental in securing engagement from the technology industry. Over time their influence and capacity has been declining, as well as their funding, and they can be quite isolated and under-resourced, making this cross-sectoral collaboration more challenging. The T&S teams have a crucial role to play in elevating these issues to the wider company, and where they are integrated as members of core teams this has worked better.
- **Legislative scrutiny is now greater** which makes it harder to have open conversations and has changed the rules of engagement. The required rules of engagement for technology industry may be unfamiliar to the voluntary or academic sector and is often guided by non-disclosure agreements and closed-door discussions, requiring assurances that there would be no policy officials or legal offices present. Only then could there be a free and open brainstorming discussion, which is essential to enable this type of innovation to flourish.

### 3.1 Technology tools and innovations

There are several existing technology solutions to prevent and detect OCSEA. These include settings that can be enabled to enhance user safety, as well as features that are embedded in the design of content-hosting platforms to ensure they are safe by default, commonly referred to as a Safety by Design approach.[17] Table 1 presents examples of technological tools cited in the reviewed evidence. This does not constitute an exhaustive list of the wide-ranging technologies currently utilised to address the issue of online child safety.

**Table 1. Examples of commonly adopted online child safety technologies**

| **Purpose of the technology** | **Online safety technology** | **Definition and use in the context of online child safety** | **Current use in the context of TFGBV** |
|---|---|---|---|
| Detection | Hash-based detection | Hashing is a technology that converts images and videos into a unique string of numbers. These 'digital fingerprints' can be compared to existing hash lists of known CSAM without needing to access the content itself. This technology can help identify victims and reduce instances of revictimization by preventing CSAM from being continuously shared across platforms.[18] | Hash-based technology is being applied to detect video and image-based TFGBV, including by big technology companies.[19] However, the evidence suggests that there is a more limited number of shared hash lists in the context of TFGBV.[20] One example is the StopNCII.org project, which generates hashes from intimate images and videos selected from an individual on their device, and these hashes are then shared with participating companies who can look for matches. |
|  | Classifiers | Classifiers are algorithms that use machine learning to sort data into predefined categories. These models can be trained to detect new or unknown CSAM in both images and videos, as well as to identify text-based child sexual exploitation, including grooming and sextortion.[21] | The literature indicates that algorithms have been developed to identify gender misinformation and discrimination taking place online, but information on their uptake across content-hosting platforms is more limited.[22] |
| Prevention | Age assurance | Age assurance technologies are used to check the ages of individuals interacting with online services. This can be done via self-declaration, personal documents or other proofs of identity, or AI-based facial recognition.[23] | Although those impacted by TFGBV are often adults, age assurance technologies could be relevant to better identify minors experiencing TFGBV and inform tailored responses. This is also important as exposure to violent or abusive material at an early age is a known risk factor for future perpetration. |

| | | | |
|---|---|---|---|
| | Deterrence messaging | Deterrence messaging is a preventative technology that displays messages to users that attempt to search for CSAM. These messages commonly clarify the law and the harm done to children in the creation and viewing of such material.[24] | Although there may be instances where deterrence messaging is being used to prevent TFGBV from taking place, this report found limited evidence on the uptake of this technology in the context of TFGBV. |

It is commonly understood that perpetrators of OCSEA operate simultaneously across multiple domains.[25] For this reason, as well as to avoid duplication and optimise resources, the reviewed evidence highlights the importance of cross-industry collaborative approaches to developing, implementing, and expanding technological responses to OCSEA. Promising examples of such initiatives are described in the box below.

**Box 2. Examples of collaborative initiatives for the development and sharing of technology tools and innovations**

**Lantern (Tech Coalition)**. This collaborative initiative brings together technology companies to share key information linked to the perpetration of OCSEA. This includes information on perpetrators' accounts such as email addresses and usernames, as well as CSAM hashes, or keywords used to groom and solicit CSAM. This information is uploaded to Lantern and can be used by participating companies to checks if signalled accounts and behaviours are present on their platforms (Tech Coalition 2023).

**Safer (Thorn)**. Founded in 2012, Thorn transforms how children are protected from sexual abuse and exploitation in the digital age through research and innovative technology solutions. In 2015, Thorn was approached by a law enforcement agency that was struggling to determine the identity and location of a child whose abuse was being circulated online (TED 2019). At that point in time, Thorn had not previously developed solutions to address CSAM but following consultations with technology companies, the organisation realised that there was room to improve coordination and develop cross-platform solutions to the issue. Safer was launched in 2019, informed by several years of consultations with technology organisations and an initial piloting phase with Flickr (AWS 2021). The launch and scale-up of the tool was supported by a grant from the Audacious Project, received in 2019.

This software uses hash-matching (Safer Match) and AI classifiers (Safer Predict) to identify known and unknown CSAM, as well as detect text-based conversations that could lead to child exploitation. More recently, Safer has integrated classifiers that, when used in conjunction with CSAM classifiers, are able to predict whether an image or video is AI- or self-generated. To support cross-platform detection of harmful content, the software also enables participating companies to share and access CSAM hashes found on other platforms (Thorn n.d.). Safer has currently been adopted by around 50 companies including Bluesky, Slack, Patreon, and Vimeo (Thorn n.d.).

**Video Interoperability Alpha project (Tech Coalition)**. As mentioned above, hash matching is a technology commonly used to detect image and video-based CSAM. However, companies currently use a range of different hash formats, limiting the interoperability of hash lists across platforms. The project is working to rehash existing CSAM videos into multiple formats with the ultimate aim of increasing the effectiveness of industry-wide hash-based detection (Tech Coalition 2022).

**DevOps (Interpol) and Initiate (Tech Coalition).** Through its annual meetings, Interpol DevOps Group brings together law enforcement, NGOs, academia and technology companies to ideate and co-develop innovative tools and technologies to improve online child safety (Safe Online n.d.). Similarly, the Tech Coalition organises annual hackathons and working sessions with the aim of driving innovation and sharing resources to combat OCSEA (Tech Coalition 2024).

The mandate of both Tech Coalition and Thorn focuses on OCSEA. Desk research and consultations did not find indications of current plans to expand the above tools and

**The evidence suggests that the existing technology developed within the online child safety space is highly adaptable to different contexts**. However, a significant enabler for the development of effective algorithms that can be deployed to detect CSAM is the presence of a large, centralised database of information. U.S.-based companies are legally required to report suspected cases of OCSEA to the National Centre for Missing & Exploited Children (NCMEC). In addition, an increasing number of non-U.S.-based companies voluntarily choose to also report to NCMEC, creating a large data repository of OCSEA reports. On the other hand, key informants indicated that reporting and data storage are likely to be more fragmented in the context of TFGBV as there currently is no global repository of data.

There are also a number of persisting challenges to the advancement of technological solutions to preventing and detecting both OCSEA and TFGBV. These include:

- Online safety technologies, such as text-based AI classifiers, are commonly trained on English datasets, and are limited in their ability to ascertain cultural and contextual nuances in language.[26][27]
- Many of the technologies described in this sub-section do not currently function on platforms that apply end-to-end encryption (E2EE) to their messaging services. However, a number of additional solutions are being developed and proposed to detect abuse that is directed at children and/or gender-motivated in E2EE environments.[28]

### 3.2 Data and evidence generation

Governments, law enforcement, technology companies, and civil society, all have a role in improving the safety of the digital world for children. To inform action, it is important for these actors to develop a comprehensive understanding of the landscape of OCSEA threats, trends and solutions. Stakeholders working to address online child safety also reported that there is a disparity in terminology and understanding of key concepts among different actors working in different sectors, as well as within the same sector. So it will also be important to ensure words and terms are unpacked and explained.

The evidence reviewed emphasises that a robust independent research ecosystem can complement internal research by the technology industry to better contextualise trends in OCSEA and help inform the direction of industry practices. On the other hand, collaboration with the private sector provides researchers with access to industry insights and additional datasets.

Motivated by the need for joint action on evidence generation, the Tech Coalition Safe Online Research Fund was first launched in 2020 as a collaborative initiative funding research projects on OCSEA prevention, with close involvement from technology industry practitioners. Since its inception, the partnership had invested USD 2.5 million in 13 research projects across three funding rounds.[29]

Interviewees highlighted the following good practices for bridging the gap between these stakeholder groups.

- **Close engagement throughout the research project**. Interviewees emphasised the importance of cross-sector approaches being embedded from the onset of collaborative initiatives. Although engagement is often limited to the dissemination phases of research, stakeholders from the technology industry can play an important role in shaping the research agenda to respond to emerging industry needs. Ongoing contact can support the process of building trust between different stakeholder groups, streamline joint ways of working, and ensure that outputs are relevant and usable for the industry at the frontline of online child safety technologies.
- **Alignment of language and definitions**. There is currently limited conceptual clarity and shared definitions that are responsive to the evolving threats in the child safety ecosystem. Initiatives such as INHOPE's Universal Classification Schema, and Save the Children and UNFPA's work to develop a framework for TFGBV aimed at children and adolescents[30], are currently working to harmonise language to improve the comparability of evidence generated across different sources. When bringing together research and the technology industry, addressing inconsistencies in the use of language can help increase the utilisation and uptake of research products and findings. There is also a need to unpack terminology and language that may be more familiar to academics and less so to technology experts.
- **Actionable messages in digestible formats**. To support the application of independent research to product and service development, key informants emphasised the importance of distilling findings into digestible products, which clearly outlined opportunities for action. These messages can be further strengthened by clarifying the industry implications of not considering the risks to children and other vulnerable groups that exist within the online space.
- **Working through existing networks.** Bridging the gap between different groups of working requires trust to be built between the parties involved. Making the most of existing networks and alliances of stakeholders can facilitate and speed up this process. For instance, in the context of the Tech Coalition Safe Online Research Fund, an important facilitating factor for bringing together researchers and industry was the convening role of Tech Coalition and Safe Online, which contributed their existing networks of technology companies, and academia and civil society respectively.

The above lessons learned are adaptable to efforts around generating research on TFGBV that bridges the gap between the needs, approaches, and ways of working of researchers and the technology industry. The recently completed TFGBV Research Priority Setting exercise led by the Sexual Violence Research Initiative (SVRI) and informed by stakeholders from civil society, the research field, and the technology industry, identifies opportunities for collaborative action around shared research priorities.[31]

## 3.3 Advocacy, legislation and policy guidance

> *"Women-led and women's rights organisations and advocates around the world have been at the forefront of efforts for TFGBV to be recognised as a form of discrimination, a human rights violation, and as part of the continuum of violence that women and girls in all their diversity experience throughout their lives. This includes collecting data and evidence to support advocacy at the international level."*[32]

Numerous examples highlight the crucial role civil society plays, alongside the technology industry, in preventing and responding to online child sexual exploitation, abuse and other harms which are vital to global efforts to end TFGBV. These efforts are spearheaded by a diverse range of civil society organisations, including youth-led and survivor-led groups, as well as those representing children most at risk of harm. Their contributions have been instrumental in ensuring that interventions are well grounded in survivor experiences, and that strategies are both appropriate and relevant.  Key areas of collaboration include:

**Advocacy.** Civil society organisations can advocate for prevention and response measures to address online violence and raise awareness of the harms for children, women and girls, and society at large. They can engage with policy and decision makers from the technology industry and government using up-to-date research and information to support their efforts.

Australia's eSafety Commissioner (eSafety) explained that its remit had expanded from children's online safety, to include online safety for all Australians with a focus on women and gender-based violence, in response to both increased evidence of online harm and violence, as well as strong public advocacy and communications campaigns by civil society. Research from the sector on the scale of the problem and survivors' experiences also played a significant role. Interviewees noted that such high-profile advocacy has led to a culture shift, highlighting the importance of digital safety and the role of regulation in supporting that.

Another partnership that was explored was between the non-profit technology company Tech Matters and Child Helpline International, a voluntary sector network of helplines. This partnership came about as a result of a meeting between the founders of the two organisations, who identified a problem that needed a tech solution. This led to the development of Aselo, a contact centre to support helplines to streamline their communications, client management, data collection, quality assurance and reporting processes.  One of the areas Aselo wanted to help with was around advocacy by the helpline movement, key to that was ensuring that data was available to support the messages and calls to action. This shared understanding of advocacy goals between Tech Matters and Child Helpline International was vital to this partnership. The technology sector can provide crucial data to support civil society's advocacy efforts with decision makers, underscoring the importance of reliable data for ensuring a thorough and appropriate response.

The Brave movement, a survivor-centred global movement are currently running a campaign to commit stakeholders to prevent and end the sexual exploitation and abuse of children online and create a safer digital future for every child. Survivor voices are a key part of this campaign.

Child protection organisations, youth-focused and women's rights organisations all have a central role to play in supporting victim-survivors of online child sexual exploitation and abuse and TFGBV. These civil society organisations offer vital survivor-centred responses and essential connections to support and redress for survivors. Helplines are particularly useful in this regard, as well as providing essential evidence related to trends and prevalence of technology facilitated abuse.

**Involvement in drafting legislation.** Another crucial area for cross sectoral engagement is in ensuring the voices of users and survivors are considered in the legislation and design of products and services. An example highlighted how civil society, particularly those active in child protection and online safety, provided greater legitimacy to regulatory authorities by offering evidence and real-life examples of the harm caused by inadequate regulation and protections.

UNICEF guidelines on developing legislation to protect children from OSEA recommend children and young people's views are considered as key element in the development of legislation, alongside that of civil society, industry and academia. [33]

**Consultation and guidance.** Several stakeholders provided examples of civil society's significant role in cross sectoral collaborations. For example, WeProtect Global Alliance have reference groups that bring members together from the private sector and civil society. They emphasise the importance of bringing these diverse groups together, especially when there are areas of tension or potential distrust. These groups serve as vital spaces for constructive challenge, addressing tensions, and differing viewpoints to avoid complacency and develop meaningful solutions. Involving people with lived experience of TFGBV, GBV, or C/SEA in the design of products and services is crucial to address the potential disconnect among technology companies.

Raising community awareness and providing education are crucial aspects of this work, requiring contextualised and tailored approaches for different stakeholders. eSafety collaborates closely with various stakeholders including civil society through community outreach and training, partnering with NGOs through grants and the establishment of a network of Trusted eSafety Providers, who provide online safety education in schools and other organisations. These efforts aim to enhance expertise in online safety and digital literacy. Additionally, civil society can significantly contribute to building the capacity of other stakeholders, such as law enforcement, policymakers, and online service providers, to effectively understand and address online violence.

Consultation and collaboration between eSafety and the technology sector played an essential role in developing guidance on Safety by Design. The approach to developing the Safety by

Design guidance was unique, as eSafety primarily serves as a regulator of the tech sector, enforcing adherence to mandatory codes and standards. However, engagement on the Safety by Design initiative fostered a more collaborative relationship, bringing these companies and the regulator together to agree to key principles, and develop guidance. Recently, eSafety published an industry guide on TFGBV: SafetyByDesign Technology facilitated gender-based-violence Industry Guide.

The partnership between e-Safety and the Technology Sector played a essential role in developing guidance on Safety by Design. Unlike some of the other partnerships discussed above, this initiative operated differently, as e-Safety primarily serves as a regulator of the Tech Sector, enforcing adherence to mandatory codes and standards. However, the launch of the 2018 to design Safety by Design initiative fostered a more collaborative relationship, bringing these companies together to develop this guidance.

## 4. Checklist of what to consider for multi-stakeholder collaboration on TFGBV based on learning from online child safety sector

- ✓ **Ensure terminology and definitions are aligned.** Before embarking on any multi-stakeholder collaboration, words and terms need to be unpacked and explained, as technology industry, government, academia and civil society use different language, and interpret the definition of these concepts differently. In order to do this stakeholders should make use of existing, and developing, classification systems and frameworks, such as those developed by IHOPE and Save the Childre and UNFPA.[34] It is important to recognise that forms of TFGBV are wide and varied and that many of the characteristics of TFGBV are shared with other forms of GBV. TFGBV must be situated within the contexts of both GBV and digital exclusion, both of which are underpinned by structural gender inequality. **Identify entry points aligned with motivations and incentives for each stakeholder.** It is important to consider the different incentive structures and motivators for each stakeholder. Industry decision making may be more influenced by factors such as regulatory, reputational or revenue risks, considering these drivers could be a more effective approach.
- ✓ **Identify the national, regional and international legal frameworks that are applicable to TFGBV.** Promote awareness of those, and identify ways to support survivors to use these existing human rights and women's rights laws to enhance cross-sector, multi-stakeholder collaboration
- ✓ **Consider ways to improve to improve coordination/partnership between law enforcement agencies to improve algorithm detection.** This could include standardising signal sharing approaches, and ensuring for example that data is being hashed in the same format and that hashes are being stored in the same datasets, which could then be made available to companies/platforms involved in taking down this content
- ✓ **Clarify expectations and ways of working** from the outset to mitigate any challenges stemming from different organisational cultures and ways of working.
- ✓ **Consider having a convening partner** to bring stakeholders together from across sectors as a way to support building common ground and trust in the relationships. The different organisational cultures and ways of working can create attitudes of scepticisms and antagonism on all sides that needs to be acknowledged and addressed. This convening partner should possess strong networks and relationships across sectors, along with the legitimacy and trustworthiness needed to unite diverse stakeholders.
- ✓ **Be open to addressing and responding to opportunistic interactions**, real time contextual issues and global moments to leverage support and engagement on the

issues.  This requires flexibility and nimbleness on the part of all stakeholders to be able to be in the right place at the right time.

- ✓ **Be creative in terms of funding models**. How can stakeholders leverage membership fees, donor funds, or other funds for innovation to support this work?
- ✓ **Make sure the role of the Trust and Safety teams is clearly defined,** and they are able to access decision makers and leaders within these companies.  Funding for these teams has been declining, and their role has been instrumental in advancing this work and agenda both internally and externally.

## 5. Conclusion

This rapid review of eight case studies involving multi-stakeholder, cross-sectoral engagement, particularly with the tech sector, around promoting child online safety identifies several lessons that are highly relevant to efforts to end TFGBV. The report highlights the importance of ensuring a common understanding of gender-based violence as an online/ offline continuum, with TFGBV having the same structural drivers, inequalities, beliefs and norms as other forms of GBV. The rapid, evolving nature of online harms requires innovative and flexible responses, which are better addressed by multi-sectoral stakeholders and partnerships, who are often brought together by a convening partner adept at navigating the different organisational cultures and ways of working. Innovations emerge for example co-creating Safety by Design principles, developing new and novel ways to research for adaptation, and developing a range of collaborative efforts among diverse coalitions of unconventional partnerships.

There remain key gaps around understanding how to address specific behaviours and social norms to address online child safety as well as TFGBV, particularly in low- and middle-income countries. However, it is clear that coordination across the entire ecosystem is important to design and implement effective interventions to address TFGBV.

## Annex 1. Glossary

**Algorithm –** A process or set of rules to follow in order to complete a task or find a solution to a problem in a finite number of steps (GBV AoR Help Desk 2023).

**Artificial intelligence (AI)** – There is currently no universally recognised definition of AI. The term commonly used in reference the theory and design of computer systems that can perform tasks requiring some degree of human "reasoning", such as perception, association, prediction, planning, motor control, as well as systems that can learn from applying algorithms to large amounts of data. It can refer to varying levels and kinds of big data and algorithmic innovations, including machine learning (ML) and deep learning (DL) (GBV AoR Help Desk 2021).

**Child Sexual Abuse Material** (CSAM) – Any visual or audio content of a sexual nature involving a person under 18 years old, whether real or not real (AI-generated). This is also often referred to as Non-Consensual Intimate Image Abuse (NCII) involving children.

**Cyberbullying – "**Defined as bullying with the use of digital technologies, which can take place on social media, messaging platforms, gaming platforms and mobile phones" (UNICEF, n.d.).

**Extended reality (XR)** – "XR includes virtual reality (VR), augmented reality (AR), and mixed reality (MR), and is also referred to using the umbrella term 'immersive tech'" (WeProtect Global Alliance 2023).

**Financial sextortion** – Threatening to expose sexual images of someone if they do not respond to demands for money (Thorn and NCMEC 2024). **Online Child Sexual Exploitation and Abuse** (OCSEA) – Child sexual abuse refers to "the involvement of a child in sexual activity that he or she does not fully comprehend, is unable to give informed consent to, or for which the child is not developmentally prepared and cannot give consent" (WeProtect Global Alliance 2021). Child sexual exploitation "is a form of child sexual abuse that involves any actual or attempted abuse of a position of vulnerability, differential power or trust" (WeProtect Global Alliance 2021). OCSEA refers to child sexual exploitation and abuse that is partly or entirely facilitates by technology (WeProtect Global Alliance 2021).

**Online grooming** – "A term broadly used to describe the tactics that abusers deploy through the internet to sexually exploit children" (Thorn 2024).

**Technology-Facilitated Gender Based Violence** (TFGBV) – There is currently no globally agreed definition of TFGBV. However, this report is informed by the current working definition of TFGBV set out by the Joint Programme on Violence against Women (VAW) Data: "Technology-facilitated gender-based violence (TFGBV) is any act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political, or economic harm, or other infringements of rights and freedoms."

## Annex 2. Methodology

The research design was informed by a reference group made up of representatives from the UK Home Office, the UK Department for Science, Innovation and Technology (DSIT) and Safe Online.

This report was informed by eight key informant interviews with representatives from cross-sectoral and multi-stakeholder initiatives involving the technology industry. Key informants were selected on the basis of discussions with the reference group. The interviews were semi-structured and guided by a topic guide covering the following areas: (a) data, measurement, and evidence generation; (b) Industry engagement, partnerships and collaboration; (c) innovations in technology; (d) advocacy with decision-makers. A list of organisations consulted is included below.

| Organisation | Relevant initiative |
|---|---|
| Safe Online | Tech Coalition Safe Online Research Fund |
| Interpol | DevOps |
| University of Kent | Tech Coalition Safe Online Research Fund |
| eSafety Commissioner | Online Safety Act |
| NCMEC | Take It Down |
| WeProtect Global Alliance | WeProtect Global Alliance |
| Tech Matters | Aselo |
| Thorn | Safer |

The research team also conducted a rapid review of the following documents:

- Tech Coalition (2022) *Initial Results of the Tech Coalition Video Hash Interoperability Alpha Project*. project (Accessed 4 November 2024).
- WeProtect Global Alliance (2021) *Guide for tech companies considering supporting the "Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse*.
- eSafety Commissioner, Australian Government (2023) *Basic Online Safety Expectations: Summary of industry responses to mandatory transparency notices*. (Accessed 4 November 2024).
- Thorn, *Proactive CSAM Detection Solutions Built by Experts in Child Safety Technology*. (Accessed 4 November 2024).
- Tech Coalition (2024) *Trust: Voluntary Framework for Industry Transparency*. (Accessed 29 October 2024).
- Tech Coalition (2024) *Online Grooming: Considerations for Deetection, Response, and Prevention of Online Grooming*. (Accessed 29 October 2024).
- Tech Coalition (n.d.), *Developer Good Practices: Combating Online Child Sexual Exploitation and Abuse*. (Accessed 29 October 2024).

- Thorn and National Centre for Missing and Exploited Children (NCMEC) (2024) *Trends in Financial Sextortion: An investigation of sextortion reports in NCMEC CyberTipline data*. (Accessed 29 October 2024).
- United Nations Children's Fund (2022) *Legislating for the digital age: Global guide on improving legislative frameworks to protect children from online sexual exploitation and abuse*, New York: UNICEF
- WeProtect Global Alliance (2022) *Framing the future: How the Model National Response framework is supporting national efforts to end child sexual exploitation and abuse online*. (Accessed 29 October 2024).
- Wilton Park (2023) *Building a shared agenda on the evidence base for Gender-Based Online Harassment and Abuse*. (Accessed 22 October 2024).
- eSafety Commissioner, Australian Government (2023) *Tech Trends Position Statement: Generative AI*. (Accessed 22 October 2024).
- UNICEF Innocenti – Global Office of Research and Foresight (2023). *The Role of Social Media in Facilitating Online Child Sexual Exploitation and Abuse*. Disrupting Harm Data Insight 7. Safe Online. End Violence Partnership, UNICEF, 2023. (Accessed 22 October 2024).
- ECPAT International (2022). *Promising Government Interventions Addressing Online Child Sexual Exploitation and Abuse. Disrupting Harm Data Insight 5. Global Partnership to End Violence Against Children*. (Accessed 22 October 2024).
- Ciardha, C. Ildenniz, G. Ruisch, B. *Message content and framing influences perceived effectiveness of warning messages for child sexual abuse material.* University of Kent

## Annex 3. Limitations

The research faced a number of limitations, which are described below:

- The research found that currently there is limited direct engagement and shared initiatives between the online child safety and TFGBV spaces. Most informants consulted for this research had no experience of engaging with the issue of TFGBV, and there was limited crossover in the documents reviewed. Although this exemplifies the value of further research at the intersection between these two areas of work, this evidence gap makes it difficult to reflect on the applicability of lessons from the online child safety to TFGBV. The significance of lessons learned from online child safety to efforts to combat TFGBV is described where possible based on the available evidence.
- This report aims to identify promising examples of impact, informed by desk research and KIIs. Due to the nature of the query and the limited scale of its data collection approach, the report does not measure the level of impact of the initiatives described.
- There are currently gaps in the evidence base with respect to the inclusion of lessons learnt from efforts to address online child safety in low- and middle-income Countries (LMICs). This presents a significant opportunity for future research, to ensure that lessons learned are inclusive of a diverse range of contexts.

# Endnotes

[1] See for example IHOPE Universal Classification Schema and the Framework for TFGBV being developed by Save the Children and UNFPA.

[2] Global Partnership for Action on Gender-based Online Harassment and Abuse. Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis (2023)

[3] The Global Digital Compact, UN Office for Digital and Emerging Technologies

[4] Childlight – Global Child Safety Institute (2024) *Into the Light Index on Child Sexual Exploitation and Abuse Globally: 2024 Report.* Edinburgh: Childlight.

[5] Jankowicz, N., Gomez-O'Keefe, I., Hoffman, L. and Vidal Becker, A. (2024) It's Everyone's Problem: Mainstreaming Responses to Technology-Facilitated Gender-Based Violence, IGP/Vital Voices https://igp.sipa.columbia.edu/sites/igp/files/2024-09/IGP_TFGBV_Its_Everyones_Problem_090524.pdf; FCDO (2024) 'Global Partnership for Action on Gender-Based Online Harassment and Abuse on the interlinkages between technology-facilitated violence against children and TFGBV' https://www.gov.uk/government/publications/global-partnership-for-action-on-gender-based-online-harassment-and-abuse-on-the-lifecycle-of-online-violence-from-childhood-to-adulthood/global-partnership-for-action-on-gender-based-online-harassment-and-abuse-on-the-interlinkages-between-technology-facilitated-violence-against-childre.

[6] UNICEF Innocenti – Global Office of Research and Foresight (2023). Who perpetrates online child sexual exploitation and abuse? Disrupting Harm Data Insight 8, Safe Online.

[7] eSafety Commissioner (n.d.) 'Safety by Design' https://www.esafety.gov.au/industry/safety-by-design.

[8] There is currently no globally agreed definition of TFGBV. However, this report is informed by the current working definition of TFGBV set out by the Joint Programme on Violence against Women (VAW) Data: "Technology-facilitated gender-based violence (TFGBV) is any act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political, or economic harm, or other infringements of rights and freedoms."

[9] Childlight – Global Child Safety Institute (2024) *Into the Light Index on Child Sexual Exploitation and Abuse Globally: 2024 Report.* Edinburgh: Childlight.

[10] WeProtect Global Alliance (2023) *Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response*. London: WeProtect Global Alliance.

[11] Ibid.

[12] Ibid.

[13] International Centre for Missing & Exploited Children (ICMEC) *Child Sexual Abuse Material: Model Legislation & Global Review*. Alexandria: ICMEC.

[14] WeProtect Global Alliance (2020) 'Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse'https://www.weprotect.org/resources/library/voluntary-principles-to-counter-online-child-sexual-exploitation-and-abuse/; UK Home Office (2022) 'Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse: formal letter' https://www.gov.uk/government/publications/voluntary-principles-to-counter-online-child-sexual-exploitation-and-abuse/voluntary-principles-to-counter-online-child-sexual-exploitation-and-abuse.

[15] UK Mission to the UN in Vienna (2023) 'Removing child sexual exploitation and abuse materials: call to action' https://www.gov.uk/government/news/call-to-action-removing-child-sexual-exploitation-and-abuse-materials

[16] WeProtect Global Alliance (2024) 'Model National Response to end child sexual exploitation & abuse online' https://www.weprotect.org/resources/frameworks/model-national-

response/#:~:text=The%20Model%20National%20Response%20includes,online%20child%20exploitation%20and%20abuse.

[17] ESafety Commissioner (2024) 'Safety by Design puts user safety and rights at the centre of the design and development of online products and services' https://www.esafety.gov.au/industry/safety-by-design

[18] Thorn (2023) 'How Hashing and Matching Can Help Prevent Revictimization' https://www.thorn.org/blog/hashing-detect-child-sex-abuse-imagery/

[19] Tony Blair Institute for Global Change (2022) 'A Ten-Point Strategy Towards Ending Technology-Facilitated Gender-Based Violence in Africa' https://institute.global/insights/tech-and-digitalisation/ten-point-strategy-towards-ending-technology-facilitated-gender-based-violence-africa

[20] Tech Coalition (2023) *Annual Report 2023: Advancing Industry's Collective Efforts*.

[21] Thorn (2023) 'How Thorn's classifiers use artificial intelligence to build a safer internet' https://www.thorn.org/blog/how-thorns-csam-classifier-uses-artificial-intelligence-to-build-a-safer-internet/

[22] Ward, J., Spencer, S., and Kasi, K. (2023) *Gender-Based Violence and Artificial Intelligence (AI): Opportunities and Risks for Women and Girls in Humanitarian Settings*. GBV AoR Helpdesk.

[23] WeProtect Global Alliance (2022) The role of age verification technology in tackling child sexual exploitation and abuse online' https://www.weprotect.org/resources/library/the-role-of-age-verification-technology-in-tackling-child-sexual-exploitation-and-abuse-online/#:~:text=One%20approach%20to%20tackle%20this,digital%20products%20safe%20by%20design.

[24] WeProtect Global Alliance (2021) *Global Threat Assessment 2021: Working together to end the sexual abuse of children online*. London: WeProtect Global Alliance.

[25] Tech Coalition (2023) 'Announcing Lantern: The First Child Safety Cross-Platform Signal Sharing Program' https://www.technologycoalition.org/newsroom/announcing-lantern

[26] WeProtect Global Alliance (2023) *Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response*. London: WeProtect Global Alliance. ; Meedan (2021) 'Claim matching global fact-checks at Meedan' https://meedan.com/post/claim-matching-global-fact-checks-at-meedan

[27] Including for example new initiatives that are working on new languages like the Lacuna Fund - dataset or Masakhane

[28] WeProtect Global Alliance (2023) Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response. London: WeProtect Global Alliance.

[29] Safe Online (n.d.) 'Tech Coalition Safe Online Research Fund' https://safeonline.global/tc-safe-online-research-fund/

[30] A framework for technology facilitated gender-based violence aimed at children and adolescents: preliminary findings to understand and address violence against young people. SVRI Forum, 2024

[31] Sexual Violence Research Initiatives (SVRI) *Technology Facilitated Gender-Based Violence: Developing a shared research agenda*. Pretoria: SVRI.

[32] Global Partnership (2022) Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis

[33] UNICEF (2022) Legislating for the digital age Global guide on improving legislative frameworks to protect children from online sexual exploitation and abuse. *New York: UNICEF.*

[34] See INHOPE - *Association of Internet Hotline Providers | What is the Universal Classification Schema?* Connor-Roth-PS3-1.pdf for details.