VOICE project

Speaking up for change

Children's and caregivers' voices for safer online experiences

ECPAT

Eurochild
Putting children at the heart of Europe

Terre des Hommes Netherlands

Down to Zero
Fighting sexual exploitation of children

The VOICE project is an initiative under the Step *Up the Fight Against Sexual Exploitation of Children (SUFASEC)* programme from the Down to Zero Alliance in partnership with the Dutch Ministry of Foreign Affairs.

ECPAT International and Eurochild acknowledge the project support of *Oak Foundation* for making this project possible.

The opinions expressed in this document belong solely to the VOICE partners. Support from donors and partners listed above does not constitute endorsement of the views expressed.

Speaking up for change

Children's and caregivers' voices for safer online experiences

INDEX

Introduction

Introduction

In today's rapidly evolving digital landscape, ensuring the safety and well-being of children online is paramount. However, achieving this goal requires more than just implementing top-down policies and technological solutions, it necessitates a profound understanding of children's experiences, perspectives, and needs. As laid down by the UN Convention on the Rights of the Child (UNCRC), children have the right to have their voices heard and taken into account in all policies affecting them. By actively listening to children's voices on safety online, we not only empower them as agents of their own development, but we also support more effective policies and interventions tailored to their specific needs.

The VOICE project was designed to listen, to understand, and bring the views of children and caregivers[1] into the policy debate around safety standards and policies in digital environments. Through this research, ECPAT International, Eurochild, and Terre des Hommes Netherlands, on behalf of the Down to Zero Alliance, engaged in collaborative, meaningful child focus group discussions and the co-creation of advocacy messages with children in 15 countries in Europe, Asia, and South America.

Digital environments offer children opportunities for connection, learning, and entertainment. For example, the internet is key for children to develop their civic identity and engage in political issues.[2] Moreover, online entertainment can support children in developing new interests in educational, informative, and social online experiences.[3] Similarly, the online environment offers opportunities for play that are beneficial for children's development, learning, self-expression, and sense of belonging.[4]

However, online spaces have also been shown to pose unique risks for children globally, which include exposure to cyberbullying, violent or harmful content, and negative mental health experiences online, all of which continue to be reported as major threats by children's helplines.[5] In addition, while established forms of sexual abuse online continue to grow (e.g., grooming, self-generated sexual material, and live-streamed child sexual abuse), new risks to children emerge online, such as artificial intelligence-generated sexual abuse material, risks associated with extended reality, and financial sexual extortion.[6] As online harms continue to pose substantial dangers to children,[7] we believe they constitute a fundamental violation of their right to be protected from all forms of abuse, their right to privacy, their right to development, and their right to participate.

[1] The term "caregiver" is used throughout to encompass both parents and other caregivers

[2] OECD (2022), *Companion Document to the OECD Recommendation on Children in the Digital Environment*, *OECD Publishing*, Paris.

[3] Growing up in a connected world, UNICEF Office of Research — Innocenti, Florence, 2019.

[4] Livingstone, S. & Pothong, K. (2021). *Playful by Design: A Vision of Free Play in a Digital World*. Digital Futures Commission.

[5] Child Helpline International (2022), *Voices of Children & Young People Around the World*.

[6] We Protect Global Alliance. *Global Threat Assessment 2023*.

[7] Slavtcheva-Petkova, V., Nash, V. and Bulger M. (2014). *Evidence on the extent of harms experienced by children as a result of online risks: implications for policy and research*. *Information Communication and Society 18*(1): 48–62.

Within this study, we also explore online safety measures designed to protect children from online risk while using the internet by either preventing harmful situations from manifesting or mitigating their impact when they do occur.[8] Within the current debate on online safety and privacy,[9] this report delves into how children understand and conceptualise privacy and online safety and what modalities of safety tools work for them. While the UNCRC, the most ratified convention in the world, equally[10] upholds the rights to privacy (Art. 16) and the protection of children from abuse (Art. 19), existing national and international digital regulation is insufficient, leaving children vulnerable to online harm and distressing situations.[11]

With this in mind, this report presents the insights provided by the children and caregivers who participated in the VOICE research, focusing on online child safety and, where possible, specifically providing insights on online safety from child sexual abuse and exploitation. In the next section, the methodology of the study is presented, along with the limitations of the methods chosen. Thereafter, the findings are presented, with an introduction to the general benefits and challenges children and caregivers face online. The presentation of the findings is structured around the three key messages that the children and caregivers had to share:

1. "We need to know more about online safety": the first section of the findings explores children's and caregivers' knowledge and awareness of online safety;

2. "We want both protection from harm and privacy": the second section explores attitudes towards online safety measures, including practices that children and caregivers use to stay safe online;

3. "We want to be part of the solution": the third section addresses how children and caregivers allocate and understand their own and others' responsibility for online safety.

[8] De Kimpe, L., Walrave, M., Ponnet, K., and van Ouytsel, J. (2019). *Internet Safety. The International Encyclopedia of Media Literacy.*
[9] Gwyn Jones, M. (2023, October 19). *Planned EU laws on child sexual abuse have sparked a bitter privacy row. Why?*. *Euronews*.
[10] United Nations, *Children*.
[11] Livingstone, S., Tambini, D., Belakova, N., Goodman, E., (2018). *Protection of children online, does current regulation deliver?* Media Policy Brief 21. London: Media Policy Project, London School of Economics and Political Science.

1.

Acknowledgements

1. Acknowledgements

We would like to express our sincere appreciation to our esteemed national implementing partners in all 15 countries (see Figure 1), whose local expertise and support played a pivotal role in ensuring the successful implementation of this research. Their commitment to the project enhanced its depth and relevance, contributing to a more nuanced understanding of caregiving experiences on a global scale.

A special acknowledgement must also be reserved for the children who shared their perspectives, enriching our study with their unique voices, which need to be heard. Their willingness to contribute to this research underscores the importance of amplifying the voices of those directly impacted by digital policies.

We would also like to thank Savanta for their expertise in survey administration, data collection, and management. They were fundamental in providing us with the insights of the caregivers. This extends to all individuals who participated in the survey, all of whom provided valuable data.

Lastly, we would like to express our heartfelt gratitude to ECPAT International (Amy Crocker, Andrea Varrella, Daniel Mulati, Julie Fuchs, Martina Fontana, and Sendrine Constant), Eurochild (Fabiola Bas Palomares, Federica Marra, Mieke Schuurman, and Lenia Kriki), and Terre des Hommes Netherlands (Carolina Savonitto, Eda Gözel, Eva Notté, Jean Elphick, Katherine Austin, Maisy Charles, Monique Janssens, Nathalie Meurens, Roza Modzelewska, and Salomé Habousha) for their dedication and invaluable contributions throughout the entire research process. We extend a special thanks to the Down to Zero Alliance, funded by the Ministry of Foreign Affairs in the Netherlands, for their support in conducting research in this area.

Figure 1. Overview of national implementing partners conducting research in each country.

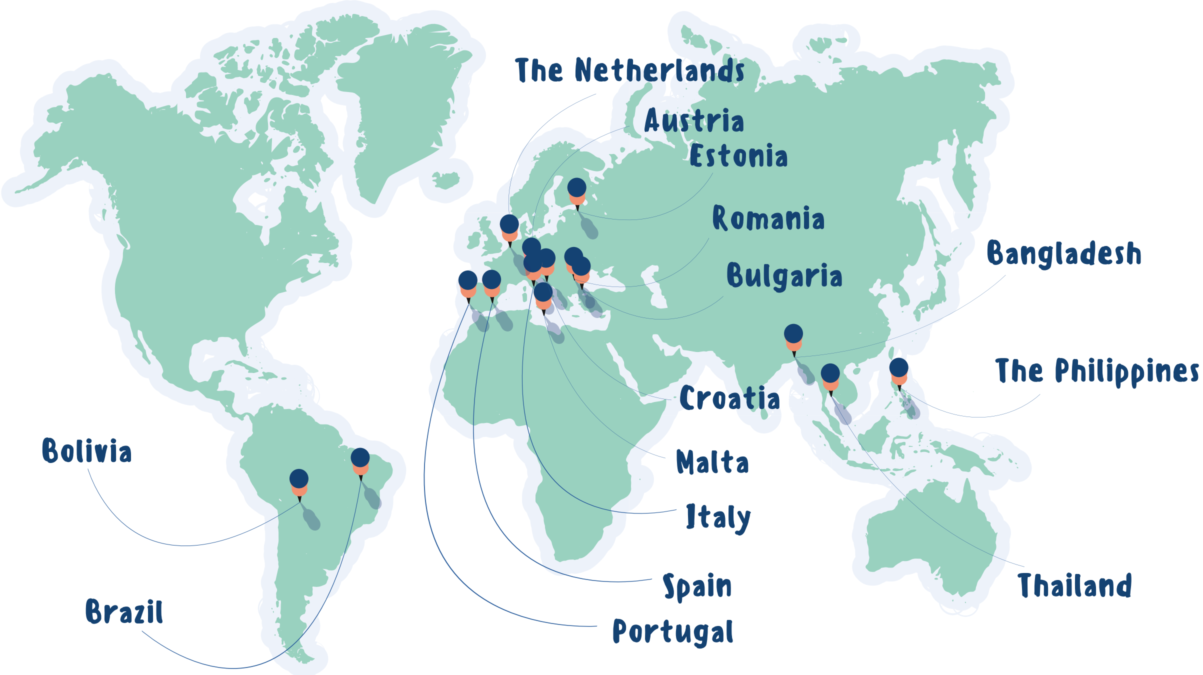
Country	National Implementing Partner
Austria	ECPAT Austria
Bangladesh	The Association for Community Development, Bangladesh Terre des Hommes Netherlands, Bangladesh Country Office
Bolivia	Fundación Munasim Kullakita
Brazil	ECPAT Brasil
Bulgaria	The National Network for Children
Croatia	Society Our Children Opatija
Estonia	Estonian Union for Child Welfare
Italy	Terre des Hommes Italia
Malta	Malta Foundation for Wellbeing Society
The Netherlands	Terre des Hommes Netherlands
The Philippines	The Center for Empowerment and Development (CoPE) ECPAT Philippines
Portugal	Instituto de Apoio à Criança
Romania	Terre des Hommes Lausanne, Romania Country Office
Spain	FAPMI
Thailand	The Life Skills Development Foundation

2.

Methodology

2. Methodology

To explore perspectives on safety from online child sexual abuse, the VOICE study used a **mixed-method approach** to collect primary data from children and caregivers. The methods included a literature review, surveys with caregivers, and participatory focus group discussions with children in 15 countries. The VOICE Steering Group partners drafted the research methodology and the accompanying tools for implementation by the national partners, each of which was an existing and well-established partner of one of the Steering Group Members (see the acknowledgement section). The data collection methods and samples are summarised below.



2.1 Literature review

A combination of different search engines (Google, Google Scholar, PubMed, ERIC, and JSTOR) were used to find relevant sources for the literature review. To enhance the research process, English key search terms, such as “online child safety”, “children’s perspectives”, and “online protection”, were developed by deconstructing each research question to pinpoint the key terms.

Sources that came up in the search engines were assessed according to the following criteria:

1. Relevance based on the scope and topic of the research;
2. Date, i.e., limiting sources to those published no later than 2016[12] to ensure current and up-to-date information;
3. Credibility based on a simplified version of the Authority, Accuracy, Coverage, Objectivity, Date, and Significance checklist.[13]

Sources originating from peer-reviewed articles, gray literature[14], government-commissioned outputs, European institutions, and international organisation outputs were all included. After filtering for author and date, titles were evaluated to determine their relevance to the research scope and their applicability to the research questions. Sources deemed relevant underwent a content analysis based on relevance and credibility criteria. Eventually, all sources conforming to the aforementioned criteria were incorporated in a Literature Review Tool to extract essential source information and the key themes addressed for further analysis.

2.2 Focus group discussions with children and young people

The VOICE team worked closely with national implementing partners to carefully select and prepare participants. In total, **483** children were engaged. In most countries, partners selected children through their existing programmes. Remaining partners engaged children through schools that were connected to their organisation. In each country, three participatory **Focus Group Discussions** (FGDs) were attended by around 11 children on average. To engage children at peak ages of online risk, children between the ages of 11 and 17 were selected. On average, children were **14.5 years old**, with a gender distribution of **53% girls, 44.7% boys, and 2.3% non-binary.** Caregiver consent and children’s assent were required for all children participating in the FGDs.

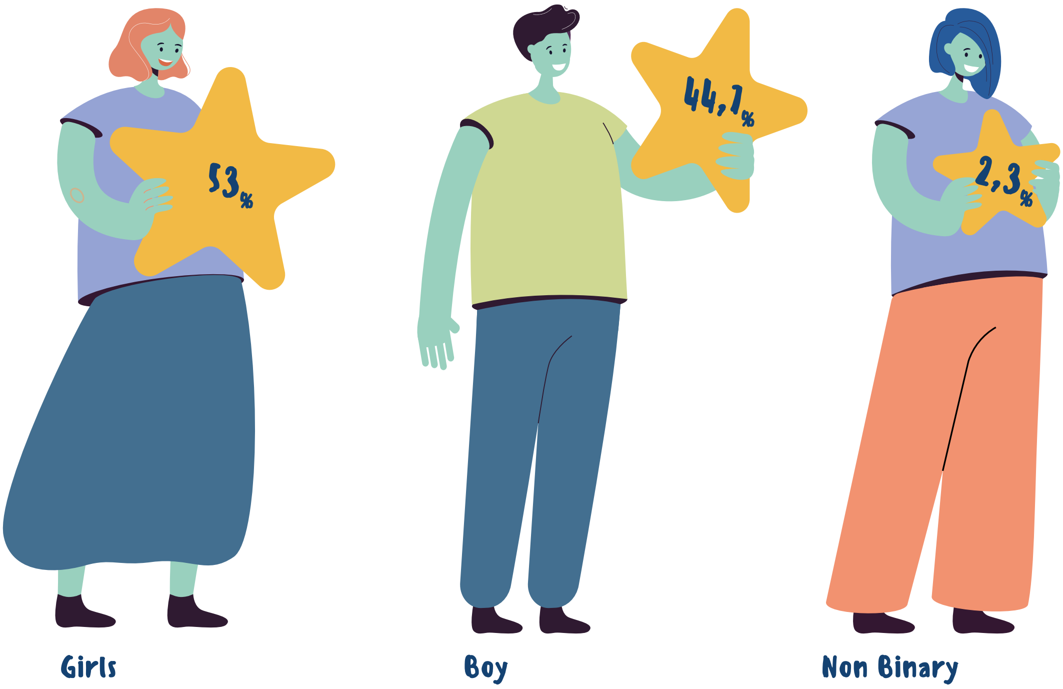
Each group of children and young people attended a participatory FGD, which lasted a maximum of three hours, to **discuss online safety**. These activities were first tested by the Eurochild Children’s Council[15] to check the relevance, engagement, and appropriateness and were adjusted according to their feedback. The focus group sessions started with a digital scavenger hunt, stimulating discussions on online safe behaviour. Afterwards, the facilitators showed a child-friendly video where online sexual harm was explained. With this topic in mind, in the next activity, children physically positioned themselves to indicate their response to statements,

[12] For key sources dating prior to 2016, exceptions were made to capture relevant insights.

[13] Tyndall, J. *AACODS Checklist*. Flinders University, 2010.

[14] Documentary material that is not commercially published or publicly available, such as technical reports or internal business documents (Oxford English Dictionary). Written material (such as a report) that is not published commercially or is not generally accessible (Merriam-Webster English Dictionary).

[15] *Eurochild Children’s Council* is a selected group of children supported by Eurochild members from different countries around Europe. They play an advisory role in relation to Eurochild’s key advocacy priorities, governance decisions, and events. The ECC has a mandate of two years. The current ECC was inaugurated in May 2022 and it is composed of 11 children.



Average age of children participating: 14.5

which prompted insightful discussions on perceived online risks. The last activity centred around dilemmas in online safety between protection and privacy, in which children took a stance and proposed solutions. Each FGD was facilitated by the national implementing partner in the most practical and commonly spoken local language. Additionally, participants filled out a small pre-session questionnaire to capture their views on online safety without any facilitator or peer mediation.

Data was collected by a note taker using a pre-made template in the national language. The information on the template, including direct quotes, was translated into English and carefully screened to ensure that it only captured anonymised information. No recordings were made of the FGDs and photos of participating children were always taken in such a way as to avoid the possible identification of any child. Children were presented with various ways in which to capture their messages, for instance, writing, making a poster, or recording a video message. Video messages required additional consent to ensure that children understood that the video would contain identifiable information. The consent form included an agreement stating that these videos would only be used in the context of this study and its dissemination.

The information in the template, together with the photos of the creative output, were analysed using the ATLAS.ti Web software and Google Sheets. ATLAS.ti was used to analyse the qualitative data with a codebook based on the research questions. Additionally, new codes were added to capture information derived from the data. The coding led to the identification of key themes that were compared to the findings of the literature review and survey outcomes. These findings were written out in the report, using the themes, quotes, and pictures of the outputs. In Google Spreadsheets, percentages of the quantitative data, as derived from the focus groups, were calculated, such as responses to statements and answers to the pre-FGD survey. The VOICE team shared the findings of the research with all facilitators involved in the FGDs to check whether the findings were consistent with their respective groups and adjustments were made accordingly.

2.3 Surveys with caregivers

The **survey methodology** involved engaging a diverse group of caregivers through a survey company, Savanta, to gain insights into their perspectives. The survey company specifically targeted respondents in the 15 selected countries that were caregivers. A total of 6,618 respondents participated in the survey, with representation from various countries and regions (see Figure 2).

Figure 2. Number of respondents per country.

The age distribution of the participants was an average of 42.2 years, with an average of 1.49 children.

Caregivers		Children	
Europe	4,600	Europe	324
Austria	508	Austria	39
Bulgaria	506	Bulgaria	33
Croatia	508	Croatia	24
Estonia	504	Estonia	25
Italy	507	Italy	41
Malta	49	Malta	25
the Netherlands	501	the Netherlands	32
Spain	509	Spain	34
Portugal	507	Portugal	32
Romania	501	Romania	39
Asia	1,262	Asia	88
Bangladesh	253	Bangladesh	34
the Philippines	508	the Philippines	27
Thailand	501	Thailand	27
South America	756	South America	71
Bolivia	250	Bolivia	30
Brazil	506	Brazil	41

To ensure a robust sample size, 500 caregivers were targeted in each country. Due to a smaller respondent market, lower numbers in Bangladesh, Bolivia, and Malta were agreed upon in order to complete the fieldwork in the required time.

The analysis of the raw survey data involved an examination of the provided spreadsheet, enabling a thorough understanding of caregiver perspectives across different regions and countries. Calculations were performed to derive distributions, presenting a comprehensive overview of responses. The team calculated data at both the regional and country levels, unveiling nuanced insights into caregiving experiences. The answers provided in the open text boxes underwent a thematic analysis, searching for recurring patterns for caregivers, as well as overlaps and differences between countries and regions.

Savanta managed the participant data, ensuring that privacy measures were in place, personal information was stored securely, and the raw data was anonymous before it was sent for analysis. The research team employed various statistical analyses, from basic percentages to T-tests using SPSS, to discern differences between caregiver groups. The findings from the caregiver perspectives were compared and integrated with the data obtained from the children in order to provide a comprehensive overview of the caregiving landscape.

2.4 Ethical considerations

All VOICE consortium partners are committed to **ensuring the safety** of respondents and multiple measures were taken to ensure that our research was safe. First, **the Scientific and Ethical Review Board** (VCWE) of the Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam (the Netherlands) reviewed the VOICE research methodology and determined that the methodology complied with the ethical guidelines of the faculty. Second, all members of the consortium partners signed a publicly available **Child Safeguarding Policy**. Third, all national implementing partners were required to have a **Child Safeguarding Focal Point** that was engaged before, during, and after the focus group discussions. An overall safeguarding focal point was assigned to oversee the research; this was the only person from the Steering Group with access to the consent forms. Additionally, facilitators, note takers, and safeguarding focal points received **methodology training**, including a briefing on child safeguarding protocols. No safeguarding concerns were raised with the in-country or overall child safeguarding focal points. Fourth, the methodology was carefully designed to avoid prompting the participants into any type of **disclosure** of personal experiences, thus ensuring the focus of the discussions was general in nature. Lastly, **data security and privacy standards** were upheld by adhering to general data management protocols and by carefully de-identifying all data.

2.5 Limitations

When reading this report, it is important to consider several limitations of the study that may have impacted the results. Firstly, due to the complex nature of the topic, the **survey design** was at times challenging. This was especially true of question formulation; however, we did our utmost to design questions that were clear and understandable and facilitated comprehensive responses despite the inherent difficulties. Secondly, since two distinct data collection methods were used for children and caregivers, the two datasets obtained were not directly comparable and no relationship could be ascertained. Additionally, for child safeguarding reasons, the qualitative discussions with children were focused on online safety in general; whereas, the questions in the survey with caregivers were always specifically aimed at online safety from child sexual abuse. This was necessary to prevent the children from disclosing direct or indirect experiences of abuse. For this reason, the findings from the caregivers and children in the report had to be treated separately. Comparisons are only made when deemed appropriate. In relation to this point, the study's focus on safety from online child sexual abuse may have **biased discussions** towards the negative aspects of the internet, potentially resulting in an overrepresentation of negative remarks.

Moreover, the methodological choice of in-person focus group discussions, as opposed to more anonymous, individualised methods, could have influenced the **ease at which children discussed the sensitive topic** of sexual abuse online. Additionally, the discussions required a certain level of knowledge about topics, such as social media platforms and technology, which may have hindered the level to which (certain) children were able to engage. For instance, the topic proved to be very complex for the children in the lower end of our age range.

Thirdly, respondents were more heavily drawn from European countries than countries in Latin America and Asia. This means that the general findings and totals need to be read with this **geographical skew** in mind and that regional comparisons are difficult due to disproportionate representations in this regard. Fourth, **translation** could have influenced the data, as all research tools were provided in English, translated by national implementing partners during the research, and then translated back for analysis. This potentially gave rise to errors in the translations, which could have led to different wording and/or phrasing, and, as a result, interpretation. The quotes in this report appear how they were provided in the English country report, but the meaning could affect the accuracy of the conveyed sentiments. Findings were checked with facilitators, but due to time constraints and other practical factors, it was not possible to validate our findings with the children involved in the discussions. Lastly, while the sample included a diverse range of respondents, the methodology design did not always allow for the **disaggregation of findings** according to these various demographic characteristics. Findings are, therefore, mostly presented in general terms.

Results validation

Results validation took place with 12 of the 15 implementing partners. The facilitators present at the session confirmed the "raw" findings and the interpretation as outlined in the present report, while contributing some nuances from their own interpretations. Overall, there was a strong alignment with the insights presented below as facilitators rated the accuracy of the findings 4.3 on a scale of 5.

3.

Findings

3. Findings

Key introductory findings:

- Children acknowledged the various benefits of the internet, for activities such as communication, entertainment, and education;
- Despite the positive aspects, children also identified online safety issues such as consequences to their mental health and concerns about protection of their data and privacy;
- Children were especially worried about risks derived from contacts with unknown people (this is often referred to as “stranger danger”);
- With regards to online child sexual exploitation and abuse (OCSEA), children were mostly worried about sexual content and unknown people with bad intentions;
- As safety strategies to prevent and respond to OCSEA, children only interact with familiar contacts and make use of blocking and reporting features.

Children note the many benefits of the internet, such as communication, entertainment, and education

One of the primary objectives of the VOICE project was to understand the perspectives of both children and caregivers regarding online safety, as this is an important first step in understanding what they need in order to be safer online. During this process, the children frequently mentioned the **positive aspects of social media**. Moreover, in all countries, children reported the value of digital **communication**, emphasising the opportunity to meet and interact with people, for instance, those with similar interests or individuals from around the world.[16]

> "Thanks to the networks, I was able to get to know FreeFire, a game that I like a lot and where I met more people with whom I talk and we have the same musical tastes and we pass memes to each other." (Child from Bolivia)

Beyond the societal dimensions of the internet, children reported liking the **entertainment** value of social media, with games, music, streaming, and social media content, such as "funny videos", being mentioned.[17] Social media was considered a good way to "pass time".[18] **Educational aspects** were mentioned less frequently, but children in eight countries[19] mentioned appreciating the opportunity to learn something new. One child provided an example of how online platforms allow them to "cross borders" and explore other cultures:

> "I like the opportunity to meet with interesting people from all over the world and to cross borders. I can learn many things related to cultures, and habits of people living in a totally different context than me." (Girl from Bulgaria)

In the reasoning provided by children for their internet use, a strong sentiment emerged. They viewed the online realm as an escape from their reality or a space in which they could express their authentic selves more freely. Children said that apps allowed them to "immerse"[20] themselves in their own world, to be "more honest and free"[21], and that they served as a "distraction"[22] from their normal lives.

Children encounter many risks online, mostly fearing mental health issues, data protection issues, and engaging with people with bad intentions

However, the use of the internet and its benefits are inherently linked to risk. In a systematic review of empirical research studies of children, researchers noted that the more time children spend online, the more the likelihood that they are exposed to risk.[23] Children in the focus groups indicated that, on average, they spent **4.8 hours on social media every day.** In the focus group discussions, children reported knowing that what they experienced online was not always safe and appropriate for children and shared the following concerns.

[16] Similar findings, specific to TikTok, can be found in De Leyn, T., De Wolf, R., Vanden Abeele, M., & De Marez, L. (2022). *In-between child's play and teenage pop culture: tweens, TikTok & privacy*, *Journal of Youth Studies*, *25:8*, 1108–1125.
[17] Words used by children in Austria, Bulgaria, Croatia, and Italy.
[18] Words used by children in Austria, Estonia, Malta, and the Netherlands.
[19] Austria, Bolivia, Bulgaria, Croatia, Estonia, Malta, Romania, and the Philippines.
[20] Words used by a child in Italy.
[21] Words used by a child in Croatia.
[22] Words used by children in Austria, Bolivia, Brazil, and the Netherlands.
[23] Livingstone, S., Stoilova, M., & Nandagiri, R. (2019). *Children's data and privacy online: Growing up in a digital age: An evidence*, p. 29; Livingstone, S., & Helsper, E. (2010). *Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy.* *New Media & Society*, *12(2)*, 309–329.

Box 1. Online safety tips from children to their peers

Although not specifically asked, children in the focus groups shared multiple tips for their peers regarding online safety.

- Tips included creating strong passwords and using two-step verification to prevent hacking, account takeover, and identity theft. Additionally, the importance of regularly updating passwords and refraining from sharing them with anyone was emphasised. Other tips included activating **facial** and **fingerprint recognition** for logging in, avoiding public Wi-Fi networks, and installing antivirus software (Bolivia, Brazil, Romania, Thailand, and the Philippines);

- In addition, they recommended avoiding accepting requests from **unknown people** and **blocking profiles** that are thought to be fake, especially if unsolicited messages have been sent. It was also stated that profiles with no profile photos, full names, or other friends added could be "fake profiles" (Bolivia, Brazil, and Bulgaria);

- They also advised being **mindful about online activity**. This included using credible apps, being careful with online games, paying attention to what you watch and share online, and avoiding clicking on links that could be scams. Other tips included "being cautious about sharing" content and information about yourself online. Advice ranged from caution against sharing confidential or personal information or photos with "anyone" or simply avoiding sharing with "anyone you do not know" (Austria, Bangladesh, Brazil, Bulgaria, Bolivia, Spain, the Netherlands, and the Philippines);

- Being aware of the risk of **posting or sharing identifiable pictures** of yourself was also mentioned, as they become part of your digital footprint and you have limited control of how they will be used online. Children from Brazil were especially cautious, warning others not to "take half-naked photos";

- Children advised taking immediate action when suspecting something malicious. This included blocking, deleting, unfriending, and logging out, as well as using reporting mechanisms, possibly after saving screenshots as evidence. Other responses included "scolding" the person or "suing or annoying them in return" (Austria, Bangladesh, Bolivia, Brazil, Spain, and Thailand).

One of the first themes to come up was related to social media causing **mental health issues,** which was mentioned in all countries except Bangladesh, Portugal, and Thailand. In other research, social media was linked to mood and anxiety disorders, cyberbullying, and addiction.[24] These findings are largely consistent with the current study. Children in all countries except the Philippines and Portugal mentioned fearing cyberbullying or related concepts such as harassment, gossip, negative comments, and insults. These issues made children feel bad or insecure about themselves or their position in groups. Other studies also found cyberbullying to be a top worry for children.[25] Moreover, children mentioned that social media could be addictive and, as a consequence, isolating, with many children saying that they would like to spend less time online, but find it difficult to do so. When asked about the "least-liked" aspect of social media, one child remarked that:

> "They can create addiction and isolate you from society. [And] there's a risk of receiving bad information." (Girl from Italy)

Another big theme for children was related to **protection of information they share online.** Children were particularly concerned about their pictures, as they can be shared or viewed without their consent.[26] Being recognisable in the photos of others was another concern, as children wanted to be the ones deciding where they appear online. In many instances, children reported being cautious about sending out personal information that could reveal their "offline identity", with specific reference to sharing their location or schedule. Participating children mentioned that this could lead to in-person harm, such as "kidnapping"[27], "assault"[28], "fraud"[29], and "stalking"[30]. This indicates that children understand that sharing information can affect their physical safety. At the same time, these sentiments suggest that they place more weight on in-person risk than online risk.

The State of the World's Children 2017 Companion Report, which gathered data from adolescents in 26 countries regarding their perspectives on online risk, also found that children were very concerned about these **in-person consequences**. One of the children in a group setting in Uruguay stated that revealing your address could lead to being "killed or raped". The authors noted that jumping to "extreme scenarios" was especially prevalent in low-income contexts.[31] In our study, the worry about in-person harm as a consequence of online behaviour was mentioned in the two countries included in South America as well as in seven countries in Europe, as the following examples will illustrate. In Bolivia, certain children expressed fear about sharing their information with unknown individuals due to their awareness of Bolivian networks being used for recruiting and kidnapping young people. Similarly, in Croatia, children conveyed concerns about sharing information online, fearing potential consequences such as being located, kidnapped, or blackmailed.

[24] O'Reilly, M., Dogra, N., Whiteman, N., Hughes, J., Eruyar, S., & Reilly, P. (2018). *Is social media bad for mental health and wellbeing? Exploring the perspectives of adolescents. Clinical child psychology and psychiatry, 23(4)*, 601–613.

[25] Third, A., Bellerose, D., De Oliveira, J. D., Lala, G., & Theakstone, G. (2017). *Young and Online: Children's Perspectives on Life in the Digital Age* (The State of the World's Children 2017 Companion Report); Eurochild. (2023). *Paving the way to realise children's rights online in Europe. In Children's Rights: Political will or won't?*

[26] Third, A., Bellerose, D., De Oliveira, J. D., Lala, G., & Theakstone, G. (2017). *Young and Online: Children's Perspectives on Life in the Digital Age* (The State of the World's Children 2017 Companion Report).

[27] Mentioned by children in Bolivia, Brazil, Croatia, Romania, and Spain.

[28] Mentioned by children in Portugal.

[29] Mentioned by children in Brazil, Estonia, and Romania.

[30] Mentioned by children in Austria and Croatia.

[31] Third, A., Bellerose, D., De Oliveira, J. D., Lala, G., & Theakstone, G. (2017). *Young and Online: Children's Perspectives on Life in the Digital Age* (The State of the World's Children 2017 Companion Report), p. 66.

Box 2. How children perceive and handle concerns regarding online safety, including worries about sexual content and contacts

For safeguarding reasons, children were not prompted about online child sexual exploitation and abuse explicitly[32], but the topic was mentioned in all countries to varying degrees. Children mostly spoke about viewing explicit content and grooming. A big concern for children in Thailand was their pictures or videos being used for "pornographic purposes, [or] exploitation", for instance, their pictures being altered using artificial intelligence (AI).

In other countries, children were less explicit and mentioned being worried about people with bad intentions. A major identified theme was that children feel less safe if they do not know who is on the other side of the screen. Children mentioned the risk of interacting with people that pretend to be someone else, people that have bad or wrong intentions, or people they do not know. The overemphasis on unknown people was also evident amongst caregivers, who often mentioned that they warned their children about this risk. This is a clear example of how caregivers' perceptions of risk trickle down to children, something that has also been noted in other studies. In their research with 14 caregiver–children pairs from local community groups in Ontario (Canada), Zhang-Kennedy and colleagues noted that, from a young age, caregivers teach children about the dangers of speaking with unknown people, contributing to the children's awareness of this risk.[33]

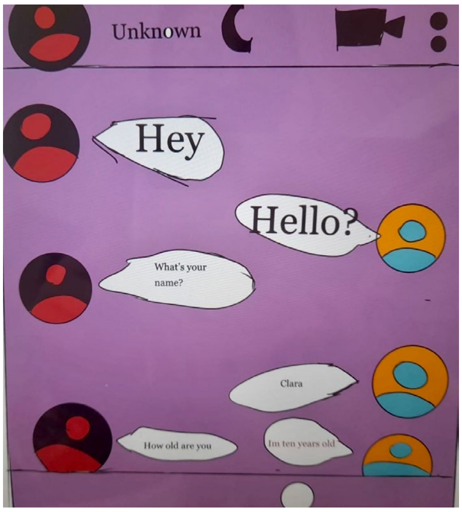


Figure 3. Poster created by children during the focus group discussion about a conversation with an unknown person (Malta).

[32] Please see methodology section for the specific limitations of the research.
[33] Zhang-Kennedy, L., Mekhail, C., Abdelaziz, Y., & Chiasson, S. (2016). *From Nosy Little Brothers to Stranger-Danger: Children and Parents' Perception of Mobile Threats.* IDC '16: Proceedings of the 15th International Conference on Interaction Design and Children, 388–399.

Despite grappling with these challenges, children reported using proactive strategies for self-protection. A principal strategy identified by the children was to only interact with people they know.[34] Upholding this stereotype of risk might lead to children missing signs of potential harm from people they know. This implicit assumption that children make about their relatives is visible in a question posed by one of them to the facilitator:

"How can we know if a person with bad motives pretends to be a family member and even his name is that of [a] relative? How would I know he is not my relative?" (Child from Bolivia)

Additionally, children mentioned reactive strategies, for example, in-app features such as blocking and reporting.

While children, thus, highlight the benefits of internet usage, their positive experiences are intertwined with significant risks. This underscores the urgency to explore what children and caregivers need in order to be safer online. The following sections outline the three key messages that children and caregivers shared in relation to developing effective strategies for online safety:

1. We need to know more about online risk and safety;
2. We want both protection from harm and privacy to be ensured when thinking about online safety measures;
3. We want to be part of the solution, making the internet safer together.

3.1 How can children and caregivers learn more about online safety with the support of schools, online platforms, and policy makers?

Key highlights:

- **Despite encountering various online risks, only a small percentage of children (10.1%) reported feeling unsafe online;**
- **A higher sense of safety was observed among the European children in the study as compared to those in South America and Asia;**
- **Children's tolerance for online risks appeared to be high, potentially stemming from desensitisation, normalisation, knowledge gaps, and/or factors linked to their age;**
- **Caregivers expressed confidence in their overall knowledge of online safety, but displayed less confidence in issues regarding online sexual abuse;**
- **There was a significant discrepancy between caregivers' perceived knowledge of their children's online behaviour and children's actual experiences, with caregivers overestimating their awareness;**

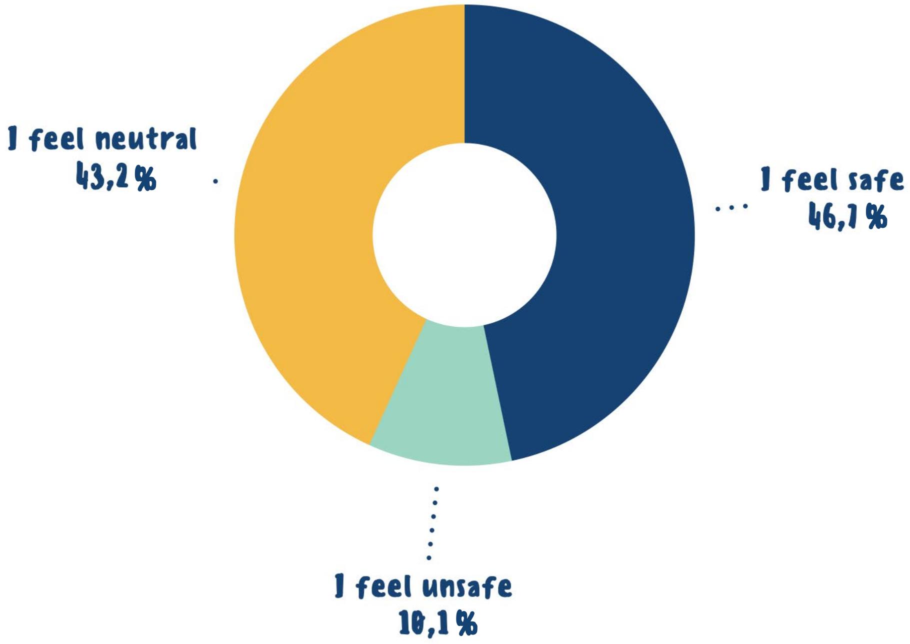
[34] WeProtect Global Alliance and Economist Impact. (2023). Global Threat Assessment 2023: Parents' perceptions of their children's exposure to online sexual harms; Zhang-Kennedy, L., Mekhail, C., Abdelaziz, Y., & Chiasson, S. (2016). From Nosy Little Brothers to Stranger-Danger: Children and Parents' Perception of Mobile Threats. IDC '16: Proceedings of the 15th International Conference on Interaction Design and Children, 388–399.

- **Both caregivers and children highlighted the need for more comprehensive online safety education, calling for greater involvement from schools, online platforms, and the government**

3.1.1 Children feel safe online, despite encountering risks

Although children mentioned many forms of online risk, only **a small percentage of children (10.1%) said that they felt unsafe online**. More commonly, children either expressed feeling safe online (46.7%) or that they were neutral about the statement (43.2%). When comparing the average answers from different countries[35], there appears to be a higher sense of safety amongst children in Europe (with an average answer of 3.6) than in the countries in South America (3.1) and Asia (3.1).

Figure 4. To what extent children feel safe online as a percentage.



Children displayed a high tolerance for risk, accepting this as part and parcel of being online. Some children seemed to be "**desensitised**"[36] **to being exposed to online risk and harm, normalising** its occurrence. In Malta, for instance, children said "you get used to it", referencing situations such as "random men who want to connect with them on social media". In other countries, children expressed the same high tolerance for risk, seeing it as part of being on social media. In some instances, children even viewed social media and safety to be mutually exclusive, with one child stating the following:

> "If you want to be safe online, you shouldn't be on social media!" (Girl from the Netherlands)

[35] Children answered this question on a Likert scale from 1 (Very unsafe) to 5 (Very safe). Averages closer to 3 indicate more children answering "3 — Neutral". Averages closer to 4 indicate more children answering "4 — Safe".
[36] Word used by focus group facilitators during the validation meeting.

Another explanation for the tolerance could be found in the developmental stage of some children in our dataset (which included children as young as 11 years old). Teenagers tend to underestimate risk and overestimate their own ability to cope with risk, which can lead to an attitude where risks or the consequences of risks are downplayed.[37] Lastly, the observed tolerance could be attributed to cultural nuances, knowledge gaps, or a lack of sexual education. In the discussions, children recognised that they needed to know more about technology such as AI[38], cookies[39], and algorithms[40] and revealed that they wished they knew more about online safety measures[41], as the following quote illustrates:

> "If we are not aware [of] how to protect ourselves... this can increase the risk." (Child from Bulgaria)

It may be, for instance, that children do not adequately identify risks that are presented to them; however, more information is needed to establish a strong causal interpretation regarding such a phenomenon. In any case, as previously discussed, children do fear risks originating from unknown people, but the following quote illustrates that it is hard to distinguish between safe and risky interactions:

> "How can we know who is safe and who to chat with? I heard we should not talk to strangers online, but what if they seem friendly?" (Child from Bangladesh)

The facilitators from the focus group discussion present at the results validation session supported the interpretation that the high tolerance to risk shown by children participating in this study may indicate a normalisation of harm and a general "desensitisation" to online risk.

Children did not reflect positively on the level of online safety education that they received at school.[42] While examples of relatively successful, up-to-date education on prevention and response to online harm exist, such as those demonstrated in England[43], Australia[44], and Finland[45], children from the countries in this study perceived these to be outdated[46], inconsistent[47], or non-existent[48] in their context.

> "At school, they don't give us much information about how to use social media, [...] about who we should go [to] if we are victims of cyberbullying or [when] strangers write to us." (Child from Bolivia)

[37] Down to Zero Alliance. (2023). *Child safety by design that works against online sexual exploitation of children*.
[38] Mentioned by children in Malta.
[39] Mentioned by children in Malta.
[40] Mentioned by children in Spain.
[41] Mentioned by children in Austria.
[42] Mentioned by children in Croatia, Italy, Netherlands, Romania, Spain, and Thailand.
[43] UK Department for Education. (2023). *Keeping children safe in education: Statutory guidance for schools and colleges*.
[44] Australian Curriculum (n.d.), Assessment, and Reporting Authority on *Online safety*.
[45] Lavonen, J. (2020). *Curriculum and Teacher Education Reforms in Finland That Support the Development of Competences for the Twenty-First Century*. In: Reimers, F.M. (eds) *Audacious Education Purposes*.
[46] Mentioned by children in Malta.
[47] European Commission. (2023). *Media literacy and safe use of new media. In Youth Wiki: Encyclopedia of National Youth Policies (Netherlands).*
[48] European Commission. (2023). *Media literacy and safe use of new media. In Youth Wiki: Encyclopedia of National Youth Policies (Croatia).*

Although we did not specifically ask children to what extent they learned from their caregivers, **some children indicated that they were not positive about caregiver guidance** either. Children said they hardly went to their caregivers for advice. The lack of experience with modern technologies made children feel uncertain as to whether their caregivers would be able to help them with online safety matters.[49]

> "In their time, there were no such technologies and they [the caregivers] think everything is bad." (Child from Bolivia)

Children identified the following characteristics as contributing to vulnerability to online risk. Besides characteristics such as being a (young) child, or a girl, children specifically mentioned that a lack of awareness contributes to online risk. In Estonia and the Philippines, children talked about gaps in technological and safety awareness, suggesting that this could lead to risky behaviour, such as using dangerous apps, "over-sharing" information, or spending a lot of time online. According to children in the Netherlands and the Philippines, **the home environment of a child** plays a crucial role in contributing to knowledge, with children not "adequately guided by adults"[50] being identified as more susceptible to online risks.

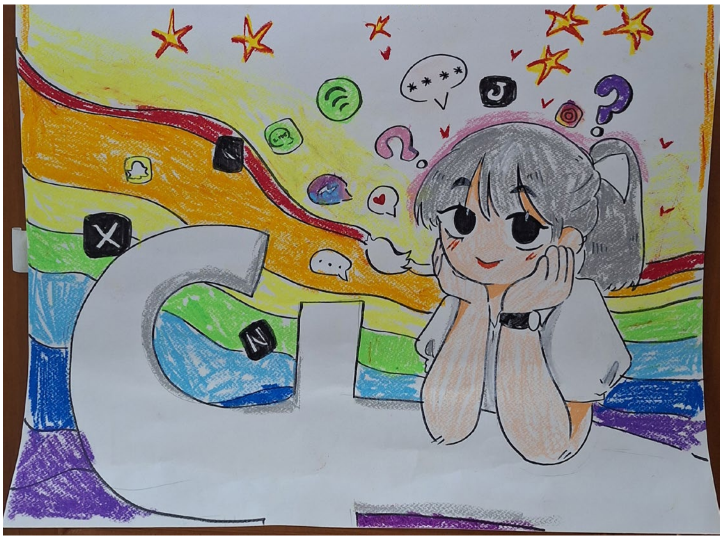
The literature suggests that children benefit most from interactive and positive learning experiences. Additionally, children prefer positives over negatives, for example, **concrete advice rather than vague warnings**.[51] Many caregivers in our dataset, however, indicated using general warnings, often remaining abstract and using negative phrasing. The following quote represents a common approach:

> "We talk about it… I always try to warn her of the dangers of abuse wherever and however it is, so that she always tells me everything and never lets anyone abuse her in any way." (Caregiver from Portugal)

When asked where they learned about online safety, the children identified information from the "news and media"[52], within applications, for example, in "videos about privacy settings [o]n Facebook, TikTok, and Line apps"[53], and from influencers such as "Sir Geyben" and channels like "Wolfgang's Channel"[54]. Children in Italy shared a government-funded video and handbook called "Fatti SMART"[55], which was designed to build awareness around practical steps to ensure personal data protection on phones and internet-enabled devices.

[49] Mentioned by children in Bolivia, Croatia, and Malta.
[50] Words used by a child in the Philippines.
[51] Hartikainen, H. (2017). *Malice in Wonderland: Children, online safety and the wonderful world of Web 2.0.*
[52] Mentioned by children in Thailand.
[53] Words from a child in Croatia.
[54] Mentioned by children in Austria and the Philippines.
[55] Garante Per La Protezione Dei Dati Personali (n.d.). *Fatti smart! Le indicazioni del Garante per tutelare la tua privacy quando usi smartphone e tablet.*

Figure 5. A child user surrounded by apps, chats, and question marks (children from Thailand).



Case study: navigating Bolivia's online world with limited risk awareness

Children from all three focus groups in Bolivia described unsafe experiences online, from being "bothered" by cyberbullies, to having photos or accounts "hacked", to being groomed. One child, for example, recounted: "I was teased for a while on TikTok. He commented dirty things about what I posted. He sent me his number and asked me to send him photos and that he was going to pay me if I sent them to him." These encounters were described as being with "bad", "ill-intentioned" people and "fake profiles". The children described being offered "a lot of money", jobs and offers to "work abroad", and "cards to recharge my credit" online, but said they were aware of the risks. These included being "deceived", having photos shared or used in advertising, and even being "kidnapped", trafficked, or forced to "carry drugs".

Despite being one of the countries that was most outspoken about experiences around online safety issues, it was also apparent that the level of safety awareness in Bolivia was quite low. From the summary report of the focus group discussion, the researchers reported that the participants were very interested in the topic and asked for guidance on topics such as whether it is advisable "to have social media from the age of 13", "to upload photos", or "how many friends" to have "to be safe". This illustrates the significant gap between knowledge and experience that needs to be bridged.

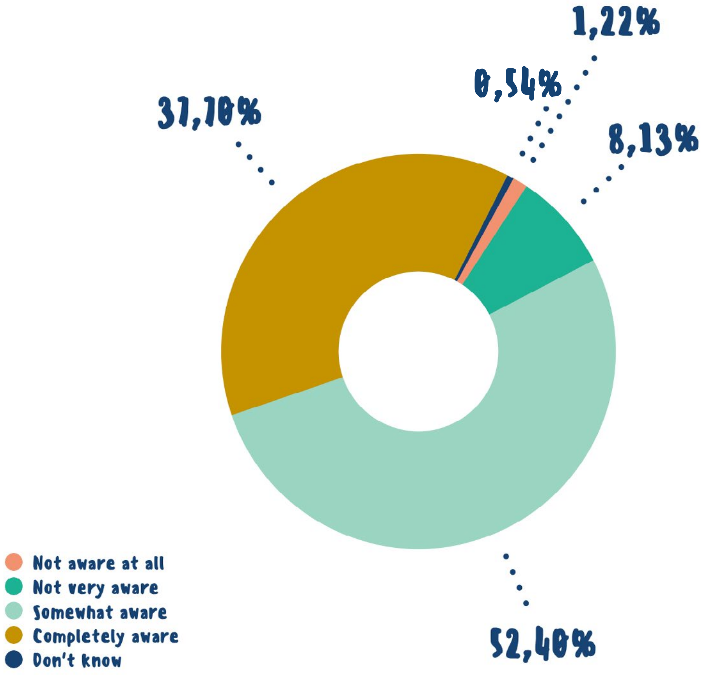
3.1.2 Caregivers are confident in their knowledge on online safety, but are less confident regarding specific issues around online child sexual abuse

The caregivers surveyed were very confident about their knowledge on how to keep children safe from online sexual abuse, rating themselves **8.3 out of 10 on average**. When looking at regional variations, caregivers in the three countries in Asia (8.6) and two in South America (9.1) consistently rated their knowledge higher than the caregivers in Europe (8.1). Previous studies demonstrated that caregivers were confident about their online safety knowledge and capacity to keep their children safe online.[56] From a national survey of 1,000 caregivers in the United States, for instance, more than eight out of ten caregivers said they felt confident that they could protect their children from exploitation online.[57] In a 2023 study involving self-reports from 2,946 adolescent–caregiver pairs, the authors discuss the **Dunning–Kruger effect** in this context**.** This phenomenon suggests that caregivers with a high confidence in preventing online risks may lack awareness of their children's actual experiences.[58] This effect holds true for the caregivers in our study, as caregivers also positively assessed their awareness of children's online activities, with **90.1% of caregivers saying they are somewhat or completely aware of their children's behaviours online**. In Bangladesh, Brazil, Malta, Portugal, Romania, and the Philippines, caregivers most commonly answered that they are completely aware of their children's online whereabouts.

If we look at our data from children, many children indeed reported that their caregivers do not know what they are doing online, nor did they want their caregivers to know everything. This discrepancy between the perspective of caregivers and children has been observed in many other studies. For example, a study that aimed to compare caregivers' and children's perceptions of online risks through web-based diaries revealed this same discrepancy: children were more likely to express that they do not disclose their online experiences to caregivers, while caregivers were more inclined to assert that their children do, in fact, share such information.[59]

[56] Kuldas, S., Sargioti, A., Staksrud, E., et al. (2023). *Are Confident Parents Really Aware of Children's Online Risks? A Conceptual Model and Validation of Parental Self-Efficacy, Mediation, and Awareness Scales*. *International Journal of Bullying Prevention*; Saeed, S. (2020). *Online Safety: A Parent's Perspective*.
[57] The Rape, Abuse & Incest National Network. (2023). *2023 Sexual Assault Awareness and Prevention Month (SAAPM) National Survey of Parents*.
[58] Kuldas, S., Sargioti, A., Staksrud, E., et al. (2023). *Are Confident Parents Really Aware of Children's Online Risks? A Conceptual Model and Validation of Parental Self-Efficacy, Mediation, and Awareness Scales*. *International Journal of Bullying Prevention*.
[59] Wisniewski, P., Xu, H., Rosson, M. B., & Carrol, J. M. (2017). *Parents Just Don't Understand: Why Teens Don't Talk to Parents about Their Online Risk Experiences*. *The 2017 ACM Conference*.

Figure 6. Extent to which caregivers say they are aware of their children's online behaviour and activities.

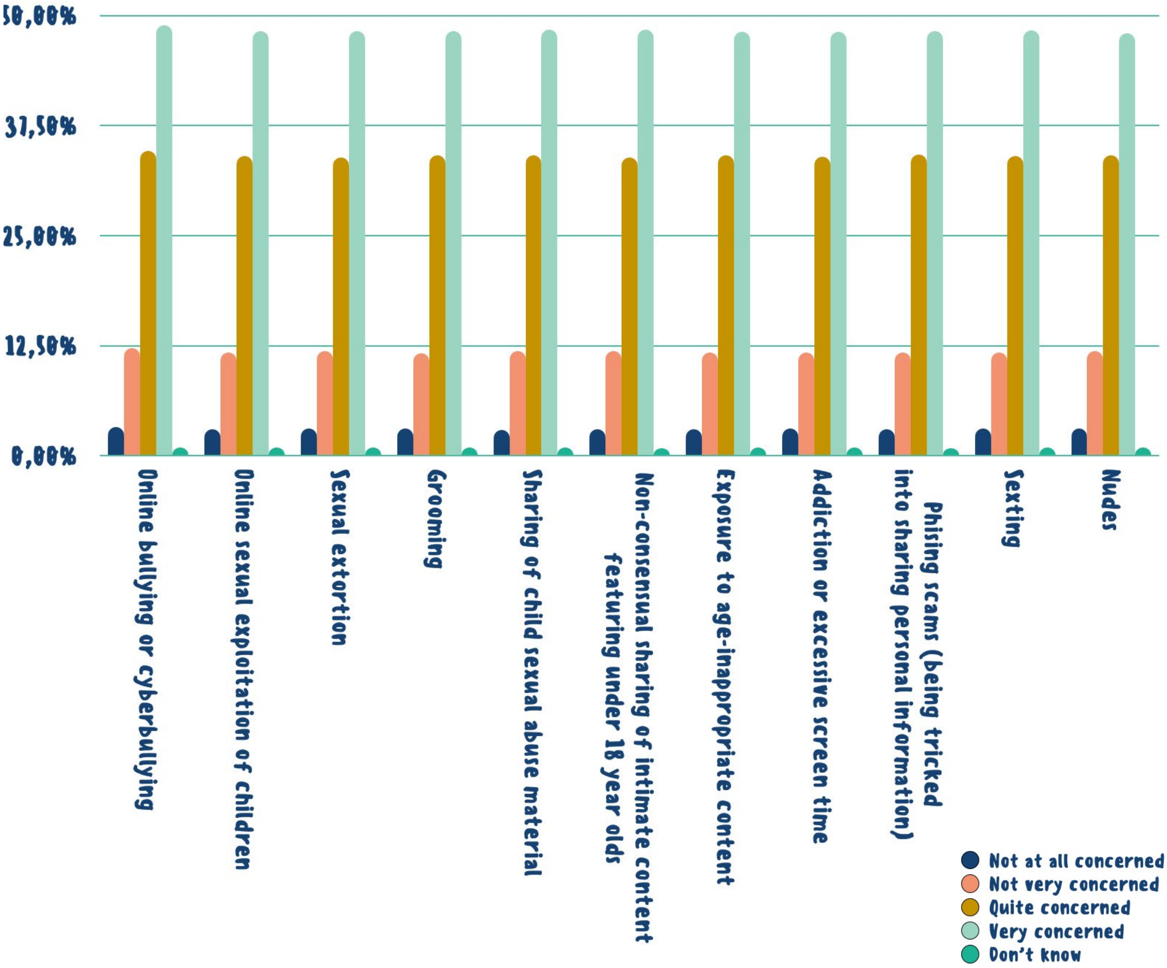


Caregivers being overconfident was also identified as a risk factor because it can point towards an **underestimation of different types of online safety risks.**[60] However, the caregivers in our study displayed a **high level of concern** towards all of the online safety issues that we prompted them about (see Figure 7), with over 80% of caregivers being quite to very concerned being consistently reported. Their concern does not, however, automatically mean that they will correctly assess risks. From our data, the caregivers surveyed seemed to have a more positive perception of how safe children feel. Although not precisely comparable, almost **three out of four (73.6%) caregivers believed that their children feel safe online**, whereas less than half of the children engaged in our focus groups reported feeling safe online.[61]

[60] Geržičáková, M., Dedkova, L., & Mýlek, V. (2023). *What do parents know about children's risky online experiences? The role of parental mediation strategies. Computers in Human Behavior*; Kuldas, S., Sargioti, A., Staksrud, E., et al. (2023). *Are Confident Parents Really Aware of Children's Online Risks? A Conceptual Model and Validation of Parental Self-Efficacy, Mediation, and Awareness Scales. International Journal of Bullying Prevention.*

[61] See figure 4: 46.7% of the children said to "feel safe", 43.2% said to "feel neutral" and 10.1% said to "feel unsafe".

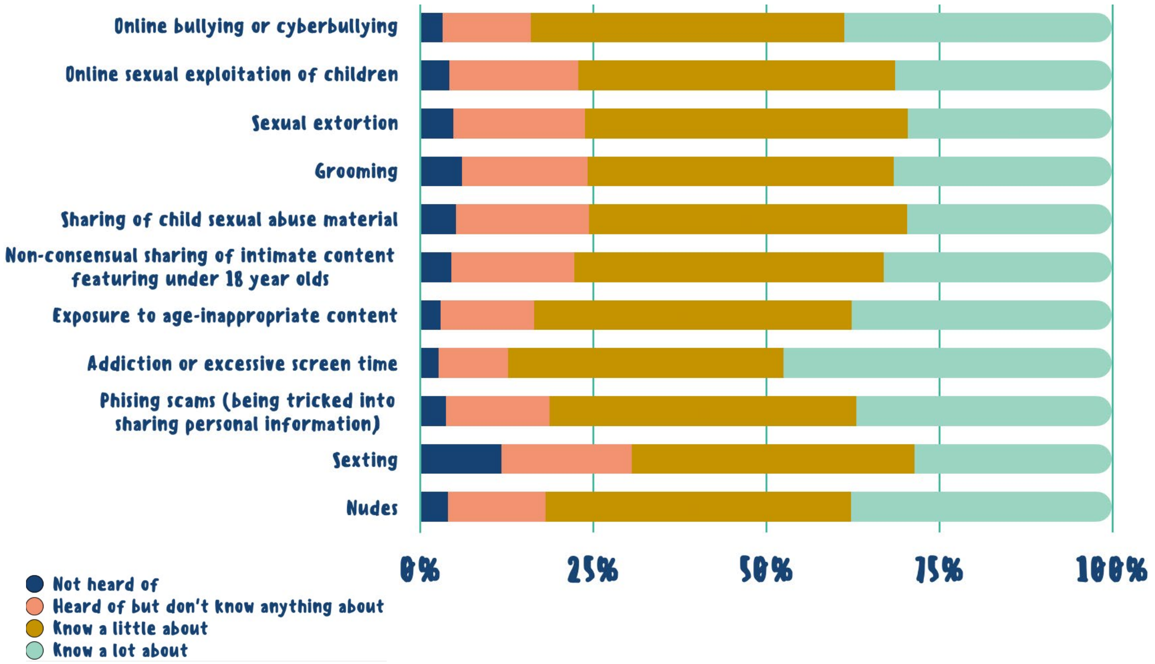
Figure 7. Extent to which caregivers are concerned about online safety issues.



Interestingly, when the caregivers were asked to indicate their level of knowledge regarding the prompted online safety issues for children, they displayed a **lower level of confidence**. Caregivers displayed the highest level of familiarity with the topic of addiction and excessive screen time, with almost 48% of caregivers "knowing a lot about it". For the other topics, most caregivers confessed to knowing a little about the other topics, backtracking on their earlier expressed confidence. In other studies, such as a study based on interviews with 25 caregivers in Norway, the knowledge gap was found to be the main challenge to ensuring cybersecurity at home.[62] Similarly, in a survey among 2,360 caregivers in Australia, researchers found a need for more information on online safety.[63] It is unclear as to where this drop in confidence originates in our data. It may have to do with an unfamiliarity with terminology or technological developments. For example, child sexual abuse material might be better known under the problematic label "child porn".

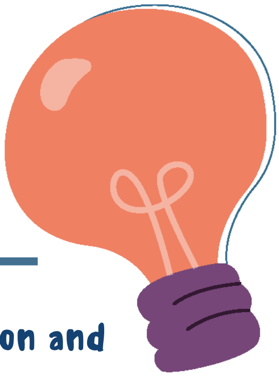
[62] Quayyum, F., Bueie, J., Cruzes, D., Jaccheri, L., & Torrado, J. C. (2021). *Understanding parents' perceptions of children's cybersecurity awareness in Norway*. Proceedings of the 53rd Hawaii International Conference on System Sciences.
[63] Office of the eSafety Commissioner of Australia. (2016). *Digital Participation Research, Parent Views*.

Figure 8. Extent to which caregivers are familiar with online safety issues.



Box 3. Language around Online Child Sexual Exploitation and Abuse (OCSEA) matters

Language is a powerful tool that shapes perceptions, attitudes, behaviours, and action. For this reason, our choice of words when discussing OCSEA is vitally important. For example, referring to child sexual abuse material as "child pornography" detracts from the exploitation and abuse experienced by the children in the images and videos, and the crime perpetrated when people make or consume them. Instead, using the term "child sexual abuse material" (CSAM) accurately reflects the criminal nature of these materials and emphasises the exploitation and harm inflicted upon children. By adopting this terminology, we acknowledge the gravity of the issue and reinforce the message that such content is not a product of legitimate or harmless activity, but rather a form of sexual abuse that must be stopped. Additionally, using precise language like CSAM helps to reduce stigma surrounding victims and survivors, promotes more effective communication within legal and advocacy frameworks, and underscores the urgent need for prevention, intervention, and justice.



Call to action[64]

Caregivers and children urge schools, platforms, and government to provide more information on online safety

This study found that while caregivers feel confident and aware of what their children do online, they show a high level of concern about online risk. The responses from both children and caregivers point to multiple gaps in knowledge, such as how safe children are online, what tools are available to ensure safety, and how safety measures work. The respondents, therefore, clearly emphasised that they need to know more about how to stay safe and asked for help from schools, online platforms, and the government. Additionally, they expressed a preference for diverse and innovative educational content, noting concerns that current resources on online risks may rely on outdated concepts, emphasising the need for up-to-date and relevant educational materials.[65]

To begin with schools, many children felt that online safety education was “essential”[66], for instance, by “implementing online safety and privacy protection in curricula at school”[67]. In Spain and the Netherlands, there was a specific call to start this education from a young age. In Austria and Malta, teenagers were identified as the main target. Children called for information that went beyond the basics, such as “password” security, and “don’t communicate with strangers”[68]. An example was provided in Spain, where discussions about grooming have found their way into classrooms, with children valuing educational videos on the subject. However, they expressed scepticism about videos that predominantly feature older men, pointing out a potential mismatch with real-life scenarios.

Concrete suggestions included teaching children about “appropriate online behaviour and how to protect yourself”.[69] More than half of caregivers (58.4%) shared this call, strongly indicating that schools should communicate with children about online safety. This resonates with other studies that identify schools as crucial educational hubs for

[64] The ideas and messages herein were voiced directly or indirectly by children (in the form of posters and drawings) and caregivers (through the open text options of the survey) during the consultations. The authors collected and summarised them to reflect the conclusions presented in this box.
[65] Mentioned by children in Malta.
[66] Word used by a child from Bangladesh.
[67] Quote from a child from Italy; a similar sentiment was noted in Romania, Spain, and Thailand.
[68] Words used by a child in Spain.
[69] Words used by a child in Croatia.

enhancing digital literacy in children.[70] Caregivers in nine countries[71] emphasised the role of the **government** in ensuring that there are policies to support better education.

Online platforms were given an important role as well, as children and caregivers suggested that apps could incorporate features to raise awareness about risks and safety, but also how to report and access support.[72] To achieve this, children in Bulgaria suggested that videos and "brochures" could be integrated into the "main social media platform" to "inform users what kind of support people can receive". Children expressed the desire for more engaging education materials, for instance, using "fun, story-like explanation[s]" that are "better than a boring, simple explanation"[73], or "in a game format with simpler wording and colourful formatting"[74].

In addition, many online platforms fall short in providing adequate support for languages other than English, resulting in a significant language barrier for non-English speakers. This reliance on English hampers accessibility and inclusivity, particularly for communities where English proficiency is limited. The lack of support for other languages exacerbates the already prevalent issue of misunderstanding and misinterpretation of critical topics, including those related to online safety and child protection. Therefore, there is an urgent need for platforms to prioritise and enhance their support in multiple languages, so that all users can access essential support and information.[75]

3.2 Online safety measures must ensure both privacy and protection from harm

Key highlights:

- **Children mostly understood online safety as a matter of ensuring personal data security. They also defined privacy in a similar way, associating it with being protected from data leaks, although sometimes also linking it to a lack of caregiver supervision;**
- **Children were aware of the harmful and age-inappropriate content that can be disseminated online, and understood the importance of online safety measures intended to protect them online;**
- **Most caregivers did not believe that current online safety measures are sufficiently protecting children from OCSEA;**
- **Only a small minority of caregivers believed that safety measures do not infringe on online privacy; however, the majority of caregivers would prioritise child protection**

[70] Throuvala, M. A., Griffiths, M. D., Rennoldson, M., & Kuss, D. J. (2021). *Policy Recommendations for Preventing Problematic Internet Use in Schools: A Qualitative Study of Parental Perspectives*. *International Journal of Environmental Research and Public Health*, 18(9), 4522; Peter J. R. Macaulay, Michael J. Boulton, Lucy R. Betts, Louise Boulton, Eleonora Camerone, James Down, Joanna Hughes, Chloe Kirkbride & Rachel Kirkham (2020) *Subjective versus objective knowledge of online safety/dangers as predictors of children's perceived online safety and attitudes towards e-safety education in the United Kingdom*, *Journal of Children and Media*, 14:3, 376–395.
[71] Mentioned by caregivers in Bangladesh, Bolivia, Brazil, Croatia, Estonia, Portugal, Romania, Spain, and Thailand.
[72] Mentioned by children and caregivers in Bangladesh, Bolivia, Brazil, Croatia, Estonia, Portugal, Romania, Spain, and Thailand.
[73] Words used by a child in the Netherlands.
[74] Words used by a child in Malta.
[75] Recommendation made by focus group facilitators during the validation meeting.

from OCSEA over privacy;

- **Children exhibited a preference for a balance between privacy and protection and for privacy-preserving technologies to detect OCSEA;**
- **While more than half of the children were in favour of age-verification systems and believe these are necessary to allow for age-appropriate experiences and connections, some children were concerned about a misuse of their data, and others were opposed to such systems as they may limit their online experience;**
- **The lack of safety-by-design measures was commonly identified by children as an element contributing to a decreased feeling of safety online. Often children felt overwhelmed by safety settings that are not user-friendly and difficult to navigate.**

3.2.1 Children find it difficult to conceptualise online safety and privacy

Livingstone and colleagues (2019) make the distinction between three privacy contexts: 1) Interpersonal privacy; 2) institutional privacy; and 3) commercial privacy. **Interpersonal privacy** applies to relationships between individuals as well as groups. It relates to the privacy decisions and practices in the online environment that are most relevant to online safety.[76] Children's online privacy decisions are **influenced by gender**, **caregivers**, **peers**, their interpretation of the social situation, their attitude towards privacy, prior negative experiences, their social media use, their **digital literacy** in navigating privacy features, and the design of the online environment.[77]

Analysing children's answers showed their struggles in making sense of the concept of online safety, often defining it through **synonyms** such as "personal safety"[78], "being safe"[79], and "security"[80]. In some instances, children indeed seemed to **echo warnings** they might have heard from caregivers, such as "once on the internet, always on the internet"[81], pointing to the influence caregivers have on their children's online safety understanding. In a different study conducted on perceptions of privacy and security online involving 66 children and their families, children showed a basic awareness of certain elements of privacy, such as the actors involved and the types of information at play (e.g., one's home address being more sensitive than one's favourite ice-cream); however, most children did not understand that sharing information online could involve various privacy concerns.[82]

Most often, children seemed to tie online safety to **ensuring personal data security,** for instance, by preventing their information or pictures from being shared without consent. They emphasised being protected against unauthorised access as utmost important, emphasising having strong passwords and other data safeguards in place. When directly asked "what does feeling safe online mean to you?" multiple children equated it to information security:

[76] Livingstone, S. Stoilova, M. and Nandagiri, R. (2019) *Children's data and privacy online: Growing up in a digital age. An evidence review.* London: London School of Economics and Political Science.
[77] Livingstone, S. Stoilova, M. and Nandagiri, R. (2019) *Children's data and privacy online: Growing up in a digital age. An evidence review.* London: London School of Economics and Political Science.
[78] Words used by a child in Bangladesh.
[79] Words used by a child in Croatia.
[80] Words used by a child in Brazil.
[81] Words used by children in Malta and the Netherlands.
[82] Kumar, P., Naik, S. M., Devkar, U. R., Chetty, M., Clegg, T. L., & Vitak, J. (2017). *No Telling Passcodes Out Because They're Private': Understanding Children's Mental Models of Privacy and Security Online*. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 1–21.

> “Feeling safe online means to me that all my information is safe.” (Child from Bangladesh)

Similar answers were given when children were prompted about what **privacy** meant to them. Children often linked having privacy to having strong passwords[83], not sharing personal information such as their location[84], and being protected from data leaks[85] and from being hacked[86]. **Their conception of a privacy issue, therefore, seems to relate to the fear of non-consensual dissemination of personal information and content, which is close to how they conceptualise online safety**. This formulation is more related to data protection as it relates to how data is being used, stored, and collected by online platforms and third parties, rather than privacy as such. The answers of children mostly emphasise knowing what is happening with their data and exercising control over who has access to it. This is in line with the findings of a 2022 study conducted with 40 children aged 8–18 and one of their caregivers (with 80 participants in total) in which the children defined privacy as protecting their personal information and privacy online.[87] In the present study, one participant from the Netherlands defined privacy as “keeping things to yourself, and others can’t see your data”.

In many cases, children associated privacy with not having any adults monitoring them and using the internet according to their wishes, or not having anyone look at their mobile phone while they are using it, as it is a “personal item”[88]. A study analysing 736 Google Play reviews of 37 mobile safety apps that were publicly posted and written by children and young people (8–19) found that, often, children thought online safety apps violated their privacy and were a form of caregiver stalking, stressing how this negatively impacted the trust relationship they had with their caregivers.[89]

Some children understood privacy as ownership over their own data, as is exemplified in the following:

> “Personal data belongs to the owner who has the right to or to not disclose his/her/their data.” (Child from Thailand)

[83] Mentioned by children in Bolivia.
[84] Mentioned by children in Bolivia, Croatia, the Netherlands, and Spain.
[85] Mentioned by children in the Netherlands.
[86] Mentioned by children in Austria, Bolivia, Croatia, the Netherlands, the Philippines, Spain, and Thailand.
[87] Murphy, O., Choong, Y. Y., & Buchanan, K. (2022). *Challenges to building youth’s online safety knowledge from a family perspective: Results from a youth/parent dyad study.* *In Proceedings of the 18th Symposium on Usable Privacy and Security*, (conference paper, p. 2).
[88] Words used by a child from Thailand.
[89] Ghosh, A. K., Badillo-Urquiola, K., Guha, S., LaViola Jr, J. J., & Wisniewski, P. J. (2018). *Safety vs. Surveillance: What Children Have to Say about Mobile Apps for Parental Control.* CHI ,18: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Paper No. 124, 1–14.

Box 4. Focus on Online Child Sexual Exploitation and Abuse (OCSEA) - Children rarely mention OCSEA when discussing privacy concerns

It is interesting to observe that **online child sexual abuse issues**, such as non-consensual sharing of intimate photos[90], child sexual abuse material[91], and grooming[92], which violate values that children mention in their online safety and privacy conceptualisation, were only mentioned explicitly in a few cases. Even after showing a video where harm of OCSEA was specifically explained, children **rarely "spontaneously"** mentioned one of these issues. It could be that this topic is uncomfortable for children, that they might not know how to refer to it or identify it very narrowly, or it may be related to a cultural bias or insufficient sexual education. We cannot infer, therefore, that the use of **euphemisms** or the lack of proactive references to OCSEA is related to a lack of acknowledgement of the risks of online abuse. Indeed, many children used euphemisms, such as things being "weird"[93], "strange"[94], "unpleasant"[95], or "bad"[96], which possibly cover a wide range of serious experiences including OCSEA, as in the following example:

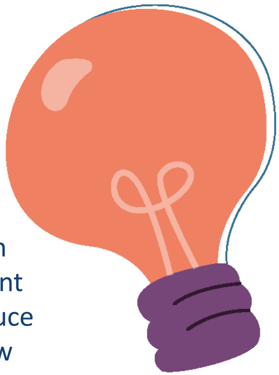
"Some people I knew felt insecure because they were tricked by using their photos (in a bad way)." (Boy from Thailand)

3.2.2 Children struggle to see how technology prevents online harm, while caregivers think current measures are insufficient

Although children understood the importance of online safety measures, in most cases, they were unsure how to define them, but seemed to understand the concept as technology that "can protect children who are not aware of online risks"[97] and demonstrated awareness of its application by reference to practical examples. Most children mentioned age-verification and parental-control technologies, while many others talked about reporting and blocking tools, or child-friendly versions of existing apps, such as YouTube Kids. Other children referred to the mechanisms used by Instagram and other social media platforms to blur or hide content that might be harmful through their so-called "Sensitive Content Control" system, while some discussed detection technologies capable of detecting harmful content.

[90] Children in Austria and Croatia only brought up the topic of sending photos, while in Brazil and Thailand, they specifically mentioned requesting photos. In the Philippines, both sending and requesting photos were mentioned.
[91] Only mentioned by children in Brazil and Romania.
[92] Only mentioned by children in Austria, Bolivia, Bulgaria, Estonia, and Romania.
[93] Words used by children in Bolivia, Croatia, Estonia, the Netherlands, and Romania.
[94] Words used by children in Bolivia, Brazil, Bulgaria, Croatia, Estonia, Italy, and Spain.
[95] Words used by children in Bolivia, Estonia, and Romania.
[96] Words used by children in Bangladesh, Bolivia, Brazil, Bulgaria, Croatia, Estonia, Malta, the Netherlands, the Philippines, Spain, and Thailand.
[97] Words used by a child in Bangladesh.

Children confessed that they did **not fully understand the technology that underlies online child safety measures to prevent OCSEA** or how platforms ensure child users' safety from OCSEA. While this technical knowledge gap was expected, especially from the youngest cohort of participants, children showed a certain level of intuition regarding some features promoting their safety online. They often referred to words like "privacy", "detection", "artificial intelligence", and "filters", but demonstrated limited understanding. While children mentioned some measures, like age verification[98], privacy settings[99], and keyword or hashtag blockers[100], again they did not make reference to specific measures that might identify potential online grooming or detect CSAM. Artificial intelligence, for example, was often referred to but could not be explained in any depth.[101] Sometimes children recognised that "bots"[102] were used by offenders to make fake profiles, but they also referred to detection tools that "recognised" and "blocked" content associated with "bots".[103]



A case example of detection technology is **CSAM detection technology**, which encompasses existing and evolving technologies that aim to detect and prevent the distribution of child sexual abuse material. Many service providers introduce their own detection tools to either identify known CSAM and/or to detect new material. Research shows that the best results are achieved when multiple methods are combined and a collaboration between online service providers and law enforcement is established.[104]

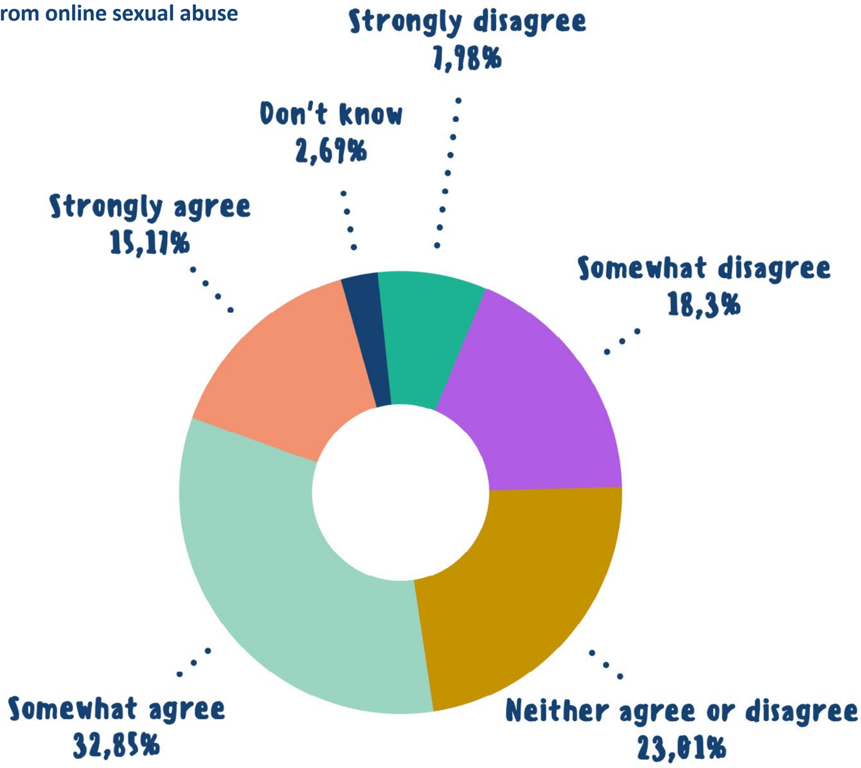
The caregivers surveyed were asked to what extent they believed current safety measures are protecting children from online sexual abuse. Figure 9 shows that less than half of caregivers thought that such measures are sufficiently protecting children from OCSEA. Among those who disagreed, some of the common concerns included the inconsistent effectiveness of certain online safety measures, the risk of both offenders and children finding a way to bypass such measures, and concerns related to data security and hacking. Many caregivers pointed to the fact that many instances of OCSEA occur as proof that current online child safety measures are insufficient and ineffective:[105]

> "Even with these measures there are still children who suffer from these abuses."
> (Caregiver from Brazil)

[98] Mentioned by children in all 15 countries.
[99] Mentioned by children in all countries except Croatia, Bulgaria, Estonia, and Portugal.
[100] Mentioned by children in Austria, Bangladesh, Croatia, and Romania.
[101] Mentioned by children in Austria, Bulgaria, Estonia, Italy, Malta, the Netherlands, Portugal, Spain and Thailand.
[102] Mentioned by children in the Philippines.
[103] Mentioned by children in Austria, Bulgaria, Estonia, and the Philippines.
[104] Lee, H.-E., Ermakova, T., Ververis, V., & Fabian, B. (2020). Detecting child sexual abuse material: A comprehensive survey. *Forensic Science International: Digital Investigation*.
[105] Mentioned by caregivers from all 15 countries.

Figure 9. Percentage of caregivers who agree or disagree that current overall safety measures are sufficiently protecting children from online sexual abuse.

Current safety measures are sufficiently protecting children from online sexual abuse



Box 5. Focus on Online Child Sexual Exploitation and Abuse (OCSEA) - Children dispute the accuracy of technology to prevent and respond to OCSEA

Children wanted measures to detect and remove CSAM or OCSEA-related activities, but they did not fully trust the accuracy of detection technology despite the ability of some tools to identify harmful behaviour and content online (like CSAM[106]) with a high degree of sensitivity. Machine learning detection technologies, such as perceptual hashing and predictive models, play a crucial role in identifying and combating CSAM[107], but children and caregivers from multiple countries[108] suggested that they made "**mistakes**"[109]. This fear might stem from some of the children's experiences whereby technological

[106] EU Commission. (2023). *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*. p. 283.
[107] Gözel, E. (2022). Safeguarding Cyberspace for Children: Navigating End-to-End Encryption's Effects on Online Child Sexual Abuse through the Lens of Routine Activity Theory. p.15.
[108] Austria, Brazil, Bulgaria, Italy, Malta, the Netherlands, Spain, and Thailand.
[109] Words used by a child in Bulgaria.

measures mistakenly blocked content that they felt was not harmful, or their frustration about measures taken by applications and social media to address hate speech.[110] The low level of trust in the accuracy of detection technologies may also be explained by a lack of knowledge on how CSAM detection technologies truly work or by low exposure to child sexual abuse experiences by the youngest participants.

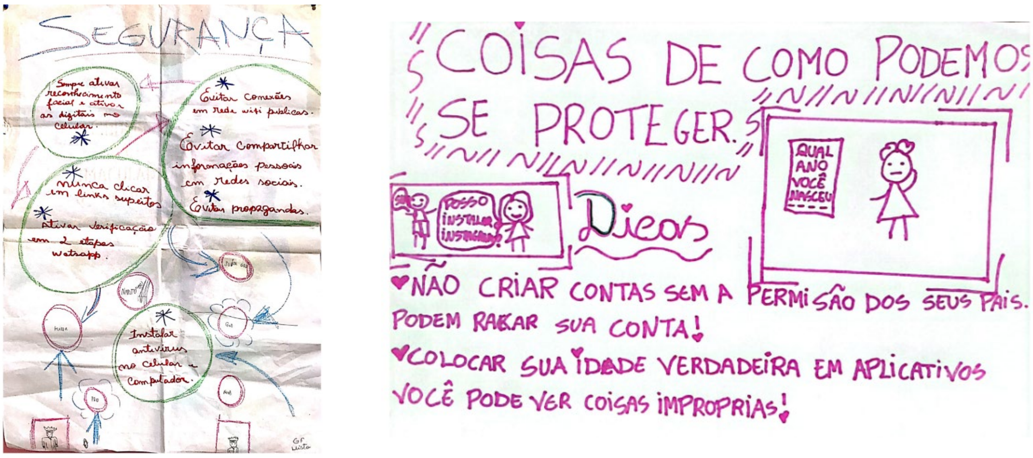
In reality, however, new CSAM detection technology has a very high accuracy rate, with some tools — such as Safer (Thorn) — achieving a **99% precision rate** for both known and new CSAM, with only a 0.1% false positive rate.[111]

The examples children shared were primarily limited to blocked "trigger words"[112], not blocked pictures. Children from several countries[113] found it annoying that platforms like TikTok[114] flag "offensive" or "bad words" and reportedly misinterpreted jokes. One child reported that he encountered the issue multiple times:

"I have had my account closed for offensive words continuously… sometimes I say them as a joke and sometimes in anger, but it is normal for them to close it when it is a joke." (Boy from Spain)

In another example, the children thought ketchup could be mistaken for blood in a video that might be wrongly subjected to an age restriction.[115] No examples related to grooming conversations or CSAM were shared.

Figure 10. Posters designed by children urging others to take measures to secure their accounts, not create accounts without "parents' permission", and to use "your real age on apps" to avoid seeing "inappropriate things" (children's focus groups, Brazil).


SEGURANÇA
Sempre ativar reconhecimento facial e ativar as digitais no celular.
Evitar conexões em rede wifi publicas.
Evitar compartilhar informações pessoais em redes sociais.
Evitar propagandas.
Nunca clicar em links supeitos
Ativar verificação em 2 etapas watsapp.
Instalar antivirus no celular e Computador.
¿COISAS DE COMO PODEMOS SE PROTEGER?
POSSO INSTALAR INSTAGRAM?
Dicas
QUAL ANO VOCÊ NASCEU
NÃO CRIAR CONTAS SEM A PERMISSÃO DOS SEUS PAIS.
PODEM RAKAR SUA CONTA!
COLOCAR SUA IDADE VERDADEIRA EM APLICATIVOS
VOCÊ PODE VER COISAS IMPROPRIAS!

[110] These measures were implemented by platforms like Instagram, Snapchat, and TikTok in compliance with the EU Commission Code of Conduct on countering illegal hate speech online between the EU and companies. See EU Commission. (2019). *The EU Code of conduct on countering illegal hate speech online*.
[111] EU Commission (2023). Report from the Commission to the European Parliament and the Council on the implementation of 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 communications services for the processing of personal and other data for the purpose of combating online child sexual abuse p 29.
[112] Words used by a child in Estonia.
[113] Brazil, Bulgaria, Malta, Spain, and Thailand.
[114] Keenan, C. (2022). *More ways for our community to enjoy what they love*. Safety.
[115] Mentioned by a child in Bulgaria.

3.2.3 Children and caregivers value privacy and protection from harm

The children and caregivers included in the analysis seemed to find both privacy and protection to be important. After receiving an explanation of how detection tools work and their role in detecting potential CSAM, caregivers were asked whether they saw this technology as an infringement of their online privacy. Figure 11 shows that about a quarter of the caregivers surveyed were neutral, while even less disagreed. Interestingly, we observed differences across regions, with a higher percentage of caregivers agreeing that they would see the use of detection technology as an infringement of their online privacy in the two Asian countries, with a total of 68.1% in agreement, and in the two countries in South America, with 57.4% in agreement, while less than half (46.9%) of the European caregivers agreed (Figure 12).

Figure 11. Extent to which caregivers agree that safety measures, such as these detection tools, can infringe on privacy.

To what extent do you agree or disagree that safety measures, such as these detection tools, can infringe on your privacy?

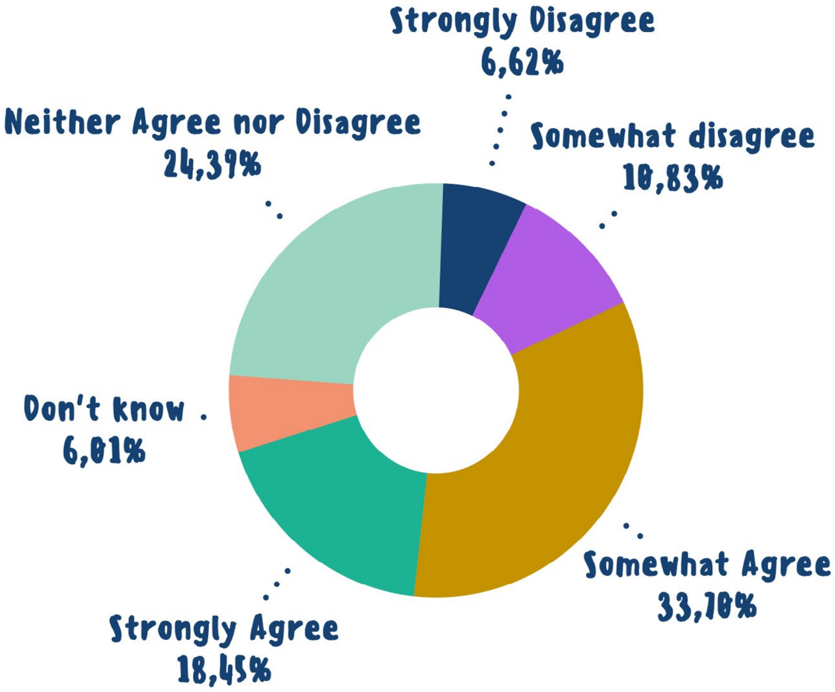
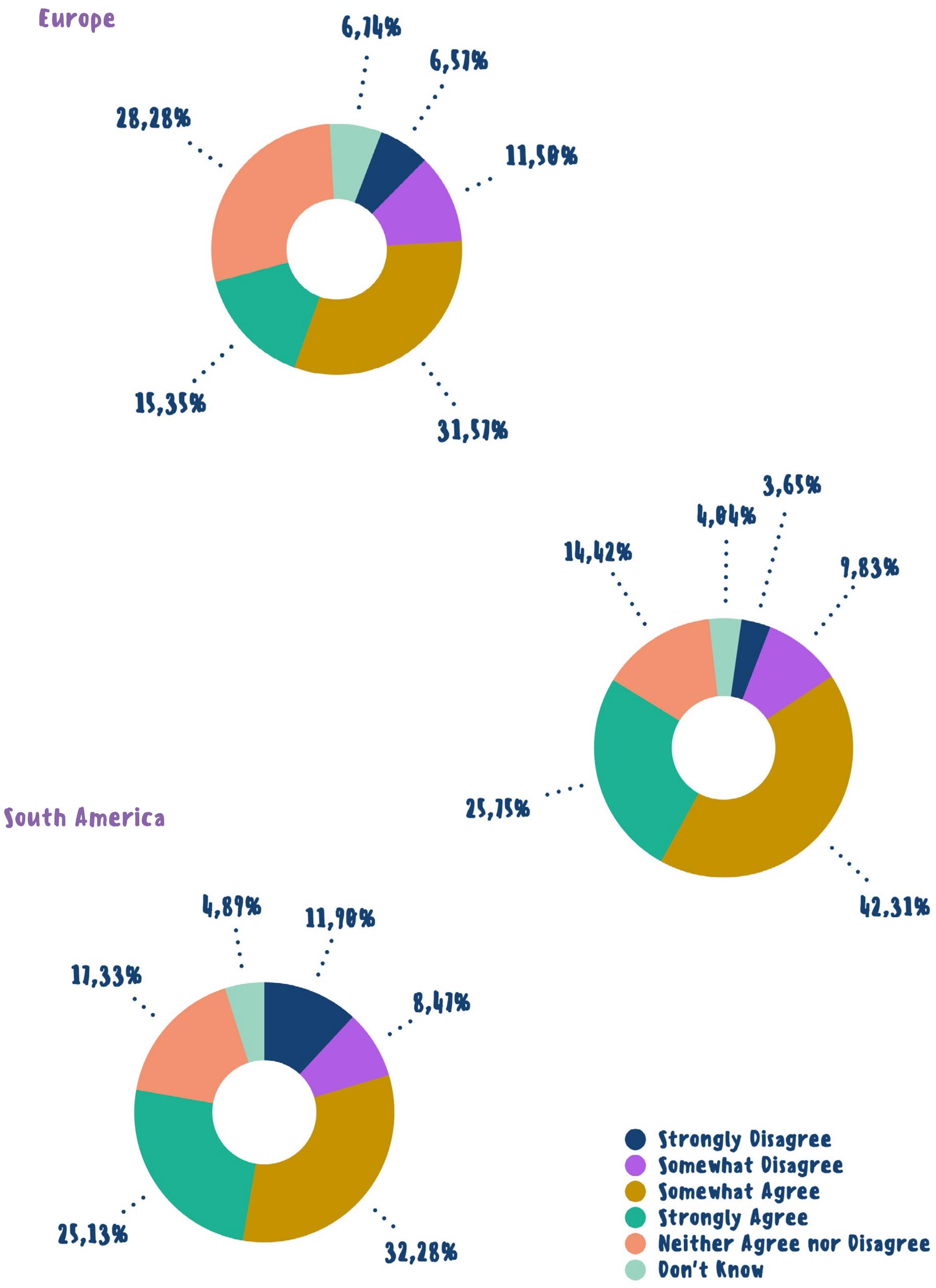
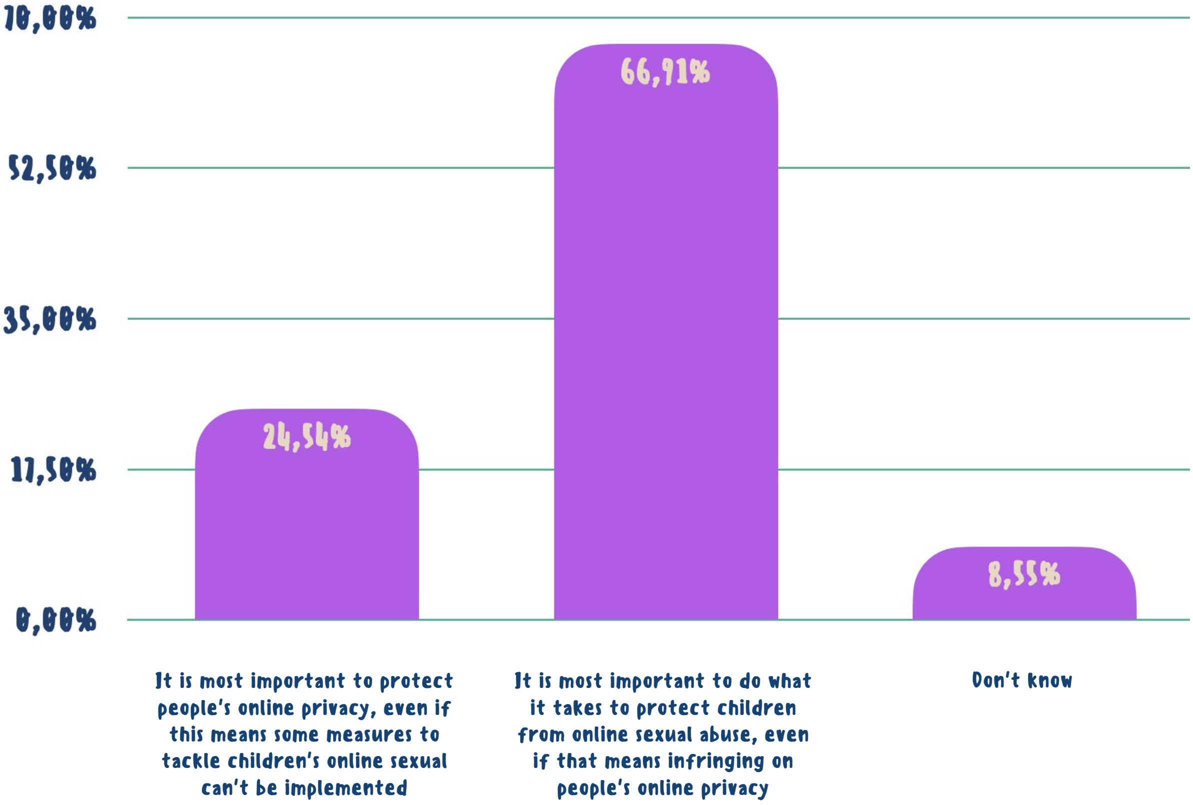


Figure 12. Extent to which caregivers agree that safety measures can infringe on privacy across three regions.



When asked which one to prioritise, the majority of the caregivers surveyed said that they find child protection from online sexual abuse more important (see Figure 13). Regional differences exist, with caregivers from the two countries in South America being most likely to prioritise child protection online (69.2%), closely followed by the European caregivers (68.7%), and a smaller percentage of caregivers from the two Asian countries (59.2%). According to a survey conducted in Europe by Defence for Children — ECPAT in 2021, only 7% of adults believed that detecting signs of OCSEA is less important than online privacy.[116] Adults mostly believe that regulating online spaces is essential to ensure online child safety, and to do so they are willing to give up some of their privacy.

Figure 13. Percentage of caregivers agreeing with statements about prioritising protection or privacy (statements are presented in the graph).

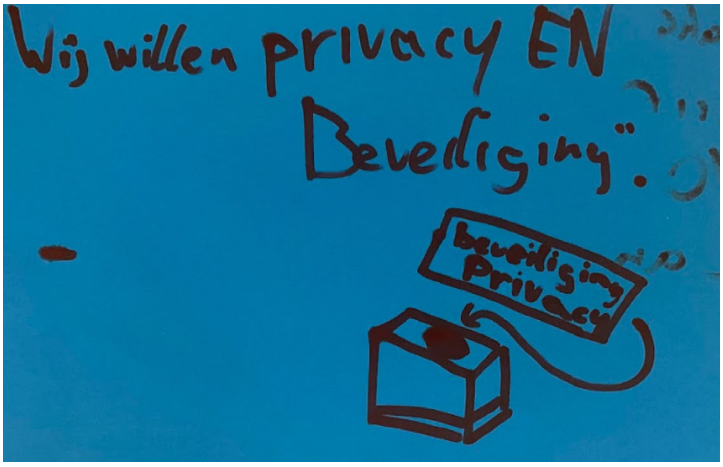


Children hinted at a preference for a balance between privacy and protection. They suggested that detection technology should be privacy-preserving, but conversely supported being shielded online when it was in their best interest. In 6 of the 15 countries sampled[117], children explicitly stated that both their privacy and safety online should be protected and, in most countries, this balance was palpable in how children interlinked the notions of safety and privacy.

[116] Defence for Children ECPAT (2021). *What do EU Citizens think of the balance between online privacy and child protection?* p. 9.
[117] Bangladesh, Croatia, the Netherlands, the Philippines, Romania, and Spain.

> “We do not think that there is a debate between online safety and privacy. Both of them are important and should be protected. Privacy is important but when we talk about criminal acts or prevention of [criminal acts], it should be considered as less important. If you agree to use a certain platform, you should accept that online safety is more important than privacy.” (Child from Bulgaria)

Figure 14. The text in the poster translates to “we want privacy AND protection” (the Netherlands).



While both were prioritised, some felt that protective measures were especially important when there was a “risk of abuse”[118] and children were in “danger”. An example of a practical protective measure can be found in boxes 6 and 7. Children especially supported **age and identity verification**[119], as exemplified in the following:

> “I think that Snapchat should restrict the access of children. Many children and young people are using it because of the picture effects and as a result of that many paedophiles have an access to such a young audience. They are sending harmful sexual messages and I think they should be blocked.” (Child from Bulgaria)

Some children were willing to “sacrifice a bit of our privacy”[120], stating that safety was a prerequisite for privacy.[121] As one child from Austria put it: “both privacy and safety are important, but safety always comes first.” In contrast, some children were more measured, supporting enhanced detection technologies but still wanting a sense of privacy in “personal chats with another person”[122] and favouring end-to-end encryption.[123]

[118] Words used by a child in Spain.
[119] Mentioned by children in Austria, Brazil, Italy, Malta, the Philippines, Portugal, Spain, and Thailand
[120] Words used by children in Italy.
[121] Mentioned by children in Italy, Portugal, and Romania.
[122] Words used by a child in Bulgaria.
[123] Mentioned by children in the Netherlands.

Box 6. Practical example illustrating how default safety measures that protect children from Online Child Sexual Exploitation and Abuse (OCSEA)

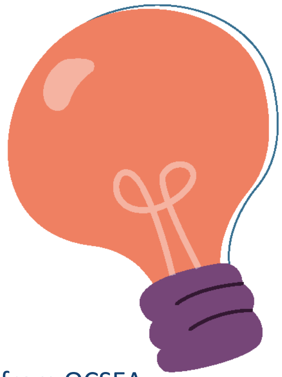
In January 2024, *Meta* implemented default settings for users under 16 that entail blocking messages from people they do not know and limiting communication to those they follow or are connected with[124]. Although these are positive steps that will better protect children's profiles, this measure is not flawless. Studies with children suggest that settings can be changed and measures can be circumvented. This is especially true for measures that prevent children from accessing parts of social media that they most like. Popularity metrics and an urge for social acceptance could lead to children changing these settings in order to increase their connections.[125] The combination of a lack of awareness of online mechanisms and a sense of individual responsibility for their safety normalises the notion that it is the fault of the child or caregiver if something unsafe occurs and that it is their own responsibility to take care of these matters. Even during discussions concerning policy solutions for online safety, a child claimed that:

"Every user is responsible for their online safety. It depends on us — whether we are aware of the measures which can be taken into consideration or not." (Child from Bulgaria)

Children from almost every country in the study were concerned about being exposed to "inappropriate"[126] content. They, therefore, supported measures to prevent them from seeing such content. This included expanding the use of "blurring of sensitive content"[127] or other preventative measures. For example, instead of receiving inappropriate friend suggestions, one child proposed "filtered friend requests, so that requests from accounts that don't have the same interests as you… does not reach you"[128]. They suggested this would protect them from "dirty old men who talk to girls"[129].

Other children liked pop-up blockers[130], but children in one of the discussion groups in Estonia were concerned that technology might filter out "content that actually interests you".[131] The example they shared was that a platform might see them as young children and "filter out extreme sports and fight scenes," which they might like. They suggested "optional scanning" so that they would still get the option of seeing things that might be filtered out.[132]

[124] Meta (2024). *Introducing Stricter Message Settings for Teens on Instagram and Facebook*. Accessed on 2 February 2024.
[125] Down to Zero Alliance. (2023). *Child safety by design that works against online sexual exploitation of children.*
[126] Mentioned by children in Austria, Bolivia, Brazil, Bulgaria, Croatia, Italy, Malta, the Philippines, Portugal, Romania, Spain, and Thailand.
[127] Words used by a child in Italy.
[128] Words used by a child in Spain.
[129] Words used by a child in Spain.
[130] Mentioned by children in Malta.
[131] Words used by a child in Estonia.
[132] Mentioned by children in Estonia.



Box 7. Digital nudging: a strategy to safeguard children online

Online platforms are rolling out new technologies meant to protect children from OCSEA through a system of **digital nudging**. Digital nudging entails guiding the behaviour of users online, through warnings, design, and information, without restricting the individual's freedom of choice.[133]

Apple recently started using digital nudging in its new Communication Safety System. If a child receives or tries to send photos or videos that the Communication Safety System determines to contain nudity, the system blurs the photo/video and displays a warning, offering ways to help. These include leaving the chat, blocking the contact, leaving a group chat, and accessing online safety resources, while reassuring the child that it is okay for them to leave or not want to view the content. If a child is under 13 years old, the system prompts them to message a caregiver or another adult they trust for help. Should a child decide that they still want to view the content, the system double checks with them and offers them alternatives, while continuing to reassure them. The Communication Safety System can be turned on or off by caregivers on the child's device through the parental-control screen time settings.[134]

Apple is not the only company integrating digital nudging. *Privately*, a company providing "privacy-preserving, smart tech solutions" to help businesses provide safe online environments for children, offers both nudging for sensitive images and videos as well as for sensitive texts.[135] Their services are compliant with the European Union General Data Protection Regulation (GDPR).[136]

Interestingly, some participating children also expressed that caregiver control is needed. Simultaneously, they asked for reasonable boundaries to that control, as a child from Malta explained in the following quote:

> "Parents need to monitor what children are viewing online to ensure safety but, on the other hand, too much monitoring from parents may make children uncomfortable." (Child from Malta)

Whilst there was a range of ages until which children suggested caregivers should have some oversight of online activities (for example, 12 years old in Bulgaria and 16 in Spain), children in six countries[137] explicitly stated appreciating some type of caregiver control. Children in Malta liked "adequate and balanced" caregiver monitoring but also suggested "automatically installed

[133] Jesse M. and Jannach D. (2021). *Digital nudging with recommender systems: survey and future decisions*.
[134] Apple (2023). *About Communication Safety on your Child's Device*.
[135] Privately. *Online Child Safety*.
[136] Privately. *Online Child Safety*.
[137] Bolivia, Brazil, Italy, Malta, Portugal, and Spain.

caregiver controls". Some felt strongly that all platforms should have "kid versions" and shield young children from harm:

> "We believe that the detection mechanism should be used because some children with the age below 13 years old can make accounts on social media and they could be influenced by adults to [participate] in sexual affairs and to receive inappropriate photos." (Child from Romania)

Case study: age assurance

In the focus groups, children frequently initiated discussions about their sentiments regarding age verification online, bringing it forward as a salient and relevant topic, even when not directly prompted.

Age-verification technologies, along with **age-estimation technologies**, fall under the umbrella of "**age-assurance**" mechanisms, which aim to verify whether an internet user is of the required age to access age-restricted content and services.[138] The most common age-assurance tool is self-certification, where users confirm their own age simply by ticking a box. Aside from self-certification, common age-assurance mechanisms can be linked to a financial profile by requesting credit card or e-payment service (such as PayPal) information. Other forms of age assurance depend on either manual scanning of identity documents or the use of publicly available information by data agencies. Age-assurance mechanisms are constantly evolving and have lately started to include biometric scans, face recognition, voice recognition, and profiling.[139] While age-assurance technologies are often associated with privacy concerns, they are an implicit obligation for online service providers under the European Union[140] GDPR.[141] The latest age-assurance technology makes the verification on the device level, with no data saved or sent to the platforms, making it privacy safe.[142]

Over half (58.4%) of the surveyed children in our focus groups displayed agreement with the question of whether it is okay for a platform to ask for your age and verify it, while 13.9% expressed disagreement, and 27.7% remained neutral. They presented us with the following arguments in favour and against age assurance.

[138] 5 Rights Foundation (2021). But how do they know it is a child? Age assurance in the digital world.

[139] Vander Maelen, C. (2019). The coming-of-age of technology : using emerging tech for online age verifications. *Interdisciplinary review of emerging technologies*.

[140] Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2017 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[141] van der Hof, S., & Ouburg, S. (2022). We Take Your Word For It' — A Review of Methods of Age Verification and Parental Consent in Digital Services. *European Data Protection Law Review.*

[142] See, for example, Yoti.

✓ Asking for the user’s age will allow for age-appropriate experiences.

✓ It is important to know the age of users, so you know who you are talking to.

✗ It requires sharing data that might be misused or leaked.

✗ It is not necessary, because it limits our experiences online and children want to use platforms that they are not allowed to use.

Participating children seemed to understand why age assurance could be used as a protective measure, but expressed concerns that it could lead to data protection issues or limitations online. In fact, many children admitted to providing a different age during registration processes on specific platforms in order to access a broader range of content, as they believe that “if we only watch films our age, nothing will be cool”.[143] However, in the case of the Philippines, children explicitly stated the need for stronger age-verification mechanisms that account for the fact that children lie about their age. Facebook is the most commonly used social media platform in the Philippines and the FGD participants noted that its method of asking for the user’s date of birth is easily circumvented as it requires no factual verification.

Some children indicated that whether age assurance is okay or not depends on the specific situation. It could depend on the type of platform, where an explicit distinction was made between social media and games. Children in Malta and the Netherlands did not understand why games would be subjected to an age verification.

Many children pointed out that young people nowadays are more mature than adults think. The UN Committee on the Rights of the Child (UNCRC) also acknowledges the changing position of children and their agency in the modern world, and emphasises the importance of promoting awareness among caregivers about the need to respect the evolving autonomy, capacities, and privacy of children.[144]

As the following quote illustrates, it was often observed in the focus groups that children were confident in their own ability to handle risk, but shared concerns about the vulnerabilities of their peers. Some children, therefore, concluded that age verification could be a good tool, but should not apply to them, as they were old and wise enough to decide what is good for them.

[143] Words used by a child in Brazil.
[144] General Comment 25 UNCRC, paras 19–21.

"For myself, I sometimes find age verification unnecessary, but imagine having a child. Everyone can post anything on the internet and I would be devastated if my child could see everything because that is just unsafe. I think verification is good because people can easily lie about their age." (Girl from the Netherlands)

The Yoti example: age estimation and data protection

Yoti is a digital identity company that offers age-verification services emphasising a "privacy-first" approach. With a true positive rate of 69.99% of 6–11-year-olds correctly estimated as being under 13, Yoti's age-estimation technology can accurately estimate a person's age by simply looking at their face. Yoti's privacy-by-design system does not require any personal details or documents, ensuring that users are not individually identifiable, and instantly deletes any information once a user's age has been estimated, thus no data is ever viewed by a human. Yoti is an example of how age verification can be safely carried out without jeopardising the user's privacy.[145]

Call to action[146]

Children and caregivers demand action: governments and online platforms should be held accountable for safety

The relative mystery surrounding online safety measures resulted in children and caregivers largely feeling the need to fend for themselves. Whilst feeling like they carried the heaviest responsibility for their children's safety online, caregivers indicated that Internet service providers, digital platforms, and social media companies were other actors that are "most responsible" for preventing online sexual abuse. There seems to be an underestimation of the role of others and, in particular, online platforms and governments, suggesting that children and caregivers need more awareness of this shared responsibility.

Children and caregivers valued both the safety and protection of children online, as well as their privacy. To ensure both, online platforms and governments need to take more responsibility to make this happen.

[145] Yoti (2023). Yoti Facial Age Estimation.

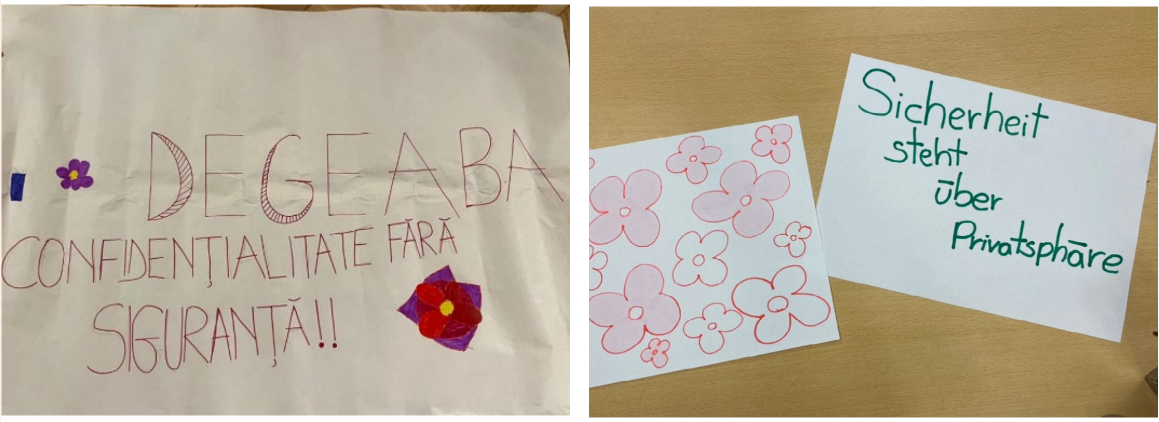
[146] The ideas and messages herein were voiced directly or indirectly by children (in the form of posters and drawings) and caregivers (through the open text options of the survey) during the consultations. The authors collected and summarised them to reflect the conclusions presented in this box.

For governments, children and caregivers called for effective sanctions[147], “mandatory implementation of detection technologies for every platform”[148], “improved internet control”[149], providing online safety education that goes beyond “only the basics”[150], and implementing “the same restrictions and regulations” among all governments[151]. Children in Portugal formulated this as follows:

“Create legislation in Europe in which all sites are safe and punish those that don’t comply — DeepWeb shouldn’t exist.” (Children[152] from Portugal)

Caregivers in Austria, Italy, the Netherlands, and Bulgaria asked for **better legislation** to regulate online safety measures for children. Moreover, the children in Portugal recommended the establishment of a cybersecurity team, suggesting that this specialised force should have a VIP account and be able to access personal data, indicating a need for accountability and enforcement.[153] Similarly, other children said that law enforcement should have increased access to private content (e.g., to the private conversations of people who have a criminal past)[154] if it would prevent “violent crimes”[155] like “torture, gender-based violence, or child abuse”[156]. When asked to choose, caregivers clearly favoured protection and safety, while children advocated for a balanced approach.

Figure 15. Text in posters translates to “No use for privacy without safety!!” (Romania) and “Security comes over privacy” (Austria).



[147] Mentioned by children in Austria and Thailand.
[148] Words used by a caregiver in Bangladesh.
[149] Words used by a caregiver in Bulgaria.
[150] Words used by a child in Malta.
[151] Words used by a child in Portugal and Malta.
[152] When “children” are quoted, it indicates that the quote comes from a group output or is attributed to a group discussion rather than an individual child.
[153] Mentioned by children in Portugal.
[154] Mentioned by children in Bulgaria.
[155] Words used by a child in Portugal.
[156] Words used by a child in Bulgaria.

For **online platforms**, children and caregivers called for measures to safeguard children from harm and empower them to make safe choices and access safe settings. In the eyes of the children, simply having a children's version is not enough. Therefore, they advocated for differentiating features according to their age and maturity.[157] Children in the Netherlands suggested child-friendly information, with simple prompts that make it easy for them to work out how to block or report a specific user. Children in Italy shared the idea of pop-up prompts that lead them through a process by asking them simple questions, eventually leading them to the right place. Other children suggested awareness videos upon downloading new apps.[158]

Children also urged platforms to take **concrete actions to reduce the risk of them being contacted, befriended, and groomed** by "fake profiles".[159] Regrettably, the concerns of children regarding fake accounts and people with "bad intentions" are justified. One study listed recent cases in which fake accounts in apps that allow private messaging were used in online child sexual abuse (OCSA) offences.[160]

Children offered plenty of suggestions regarding how to make assurance measures less easy to "fool".[161] Similar to the children in a Southeast Asian study, the children proposed more rigorous identity verification methods, such as identity cards and birth certificates, while being hesitant about their need for privacy.[162] The proposed measures included facial recognition technology[163] and eye scanning[164], personal questions[165], requiring an identification card to register[166], and only allowing an email address to register one profile on each platform.[167] In their messaging, children called for policy solutions compelling platforms to put better measures in place to ensure they could engage online with others whose ages and identities were assured, whilst also being free of concerns that platforms were misusing their personal data.

To prevent the **non-consensual usage** of their photographs, children suggested that a "double confirmation" of consent should be required before posting a photo[168] or that the "photo owner" should receive a notification when their photo is being used.[169] Other ideas included providing an option to prohibit screenshotting of Instagram stories,[170] sharing the sentiment that children wished for increased user control over who has access to

[157] Mentioned by children in Malta and Portugal.
[158] Mentioned by children in Estonia, the Netherlands, and Spain.
[159] Mentioned by children in Bolivia, Brazil, Malta, Portugal, Spain, and Thailand.
[160] Gözel, E. (2022). Safeguarding Cyberspace for Children: Navigating End-to-End Encryption's Effects on Online Child Sexual Abuse through the Lens of Routine Activity Theory. p.20.
[161] Words used by a child in Thailand.
[162] Lala, G., Chandra, S., Ogun, N., Moody, L., & Third, A. (2022). *Online safety perceptions, needs, and expectations of young people in Southeast Asia: Consultations with young people in Indonesia, Malaysia, Thailand, and Vietnam.* Young and Resilient Research Centre, Western Sydney University.
[163] Mentioned by children in Bolivia, Brazil, and Portugal.
[164] Mentioned by children in Thailand.
[165] Mentioned by children in Brazil, Italy, and Portugal.
[166] Mentioned by children in Austria, Brazil, and Portugal.
[167] Mentioned by children in Thailand.
[168] Mentioned by children in Italy and Spain.
[169] Mentioned by children in Thailand.
[170] Mentioned by children in Italy.

their content.[171] One group proposed the creation of an application that would limit the communication functions:

"We propose an application that only allows us to talk to known people and does not allow us to share photos with anyone." (Children from Spain)

3.3 Making the internet a safer place together - children and caregivers want to be part of the solution

Key Highlights:

- **Caregivers overwhelmingly perceive themselves as those most responsible for protecting children from OCSEA, with a smaller role attributed to online platforms and governments;**
- **Children take on a high level of responsibility for their own online safety, emphasising personal behaviour and measures;**
- **Both children and caregivers underestimate the potential for platforms to be designed in a manner that prioritises safety, while also being subjected to government accountability;**
- **Children primarily rely on online support tools and only turn to caregivers as a secondary option;**
- **Caregivers mostly rely on parental-control apps and fostering a supportive environment so children feel free to come to them for help;**
- **Children and caregivers face difficulties in talking to each other about online safety;**
- **Shared responsibility is needed with governments, online platforms, caregivers, and children to enhance online safety;**
- **Children prefer safety measures that promote their agency and would like to be involved in decision-making around online safety.**

3.3.1 Caregivers see themselves as those most responsible for children's online safety

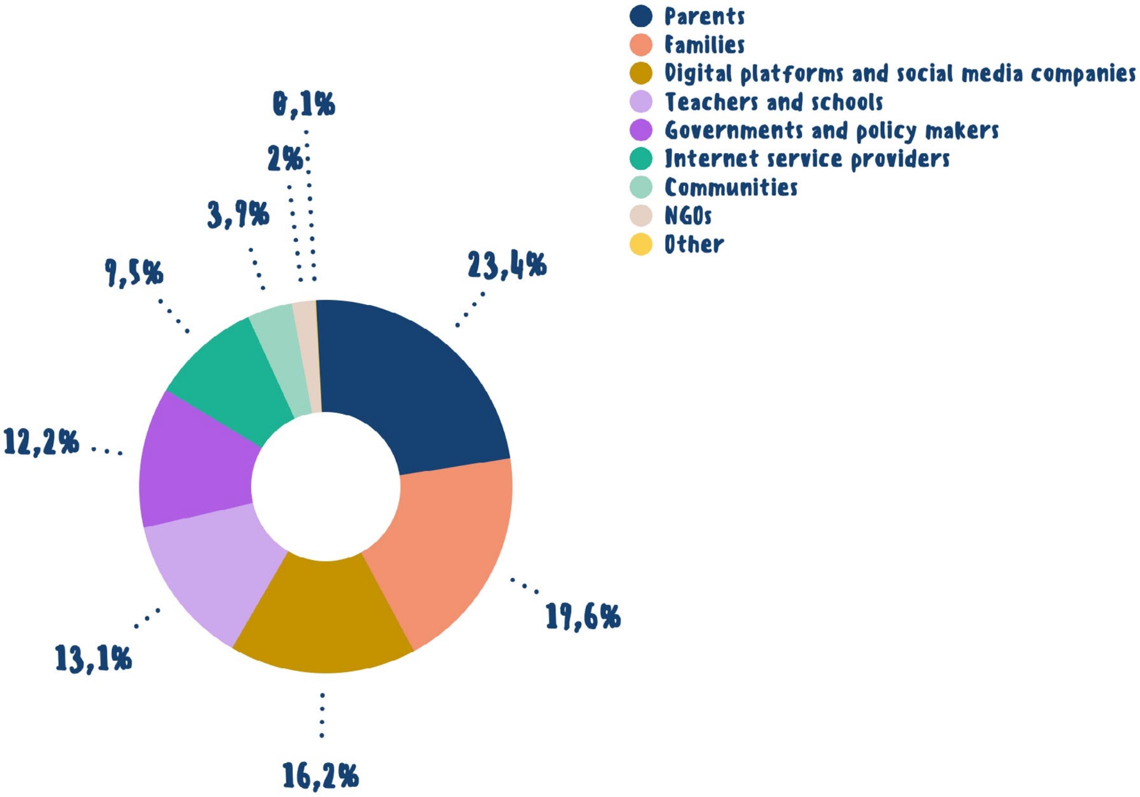
While caregivers are the ones primarily responsible for their children[172], the UN General Comment No. 25 on children's rights in the digital environment recognises the responsibility of governments to create a supportive legislative and policy environment that fosters compliance with the full spectrum of children's rights online.[173] At the same time, the Global Commission on Internet Governance recognises the fast evolution of the online world, limiting governments' capability to create such power. Therefore, the Commission calls on internet companies, such as social media platforms, to secure children's rights online.[174] In our data, caregivers and children did not seem

[171] Mentioned by children in Bolivia, Brazil, and Malta.
[172] See Article 3 and 18 of the United Nations Convention on the Rights of the Child.
[173] UN General comment No. 25 (2021) on children's rights in relation to the digital environment.
[174] Global Commission on Internet Governance. One in Three: Internet Governance and Children's Rights.

to recognise this shared responsibility, underestimating the extent to which platforms could, or maybe should, be designed in such a way that safety is upheld in order to reach a balance between caregiver and platform responsibility; the latter being held accountable by governments. A 2016 Finnish study on children's online safety concluded that, in the public discourse concerning mediating children's online safety, discussions were centred around the responsibility of society, while the responsibility of the relevant industries was given little attention.[175]

In our data, caregivers reported feeling the responsibility of keeping children safe online. This is consistent with the findings of the 2023 Global Threat Assessment, in which nearly two out of three caregivers felt it to be their responsibility to keep children safe online since platforms were not providing sufficient protection.[176] In our study, **caregivers most often** (23.4%) **ranked themselves as being those most responsible** for ensuring the safety of children from online sexual abuse, followed by families (19.6%), digital platforms and social media companies (16.2%), teachers and schools (13.1%), and governments and policy makers (12.2%) (see Figure 16). These trends were consistent across the three regions surveyed.

Figure 16. Percentage of how often caregivers rank certain actors as being most responsible for ensuring child safety from online sexual abuse.



[175] Hartikainen, H., Iivari, N., & Kinnula, M. (2016). Should We Design for Control, Trust, or Involvement?: A Discourses Survey about Children's Online Safety. In *Proceedings of the 15th International Conference on Interaction Design and Children* (IDC '16), June 2016 (pp. 367–378).

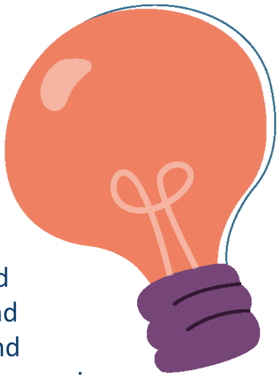
[176] WeProtect Global Alliance and Economist Impact. (2023). *Global Threat Assessment 2023*.

3.3.2 Children take on a lot of responsibility for their own safety but face challenges in choosing the right measures

Children, too, take on a high level of "**user responsibility**".[177] When analysing the children's answers regarding what elements contribute to an increased or decreased feeling of safety, we found that the children often talked about **external sources** when talking about decreased safety factors. These mostly concerned platform design factors, such as whether they are able to view inappropriate content, not being certain how their information and content is being protected, and chatroom functions. For elements that increase their sense of safety, children mentioned things that were mostly related with their **own behaviour**, such as censoring themselves, what they share online, and the extent to which they adopt the right safety settings, such as the following quote illustrates:

> "Every user is responsible for their online safety." (Child from Bulgaria)

A high number of child respondents felt that they are the ones responsible for protecting themselves if they encounter risk online, regardless of whether they understand how social media platforms work. However, the tendency of children to assume responsibility for their own safety online may not be a deliberate choice, but may rather originate from an inability to envisage alternatives, thus requiring them to focus on their own behaviour to stay safe.[178]



Child safety by design is an approach to address online risks by proactively anticipating potential harms and incorporating preventive measures. This method emphasises embedding safeguards into the design, development, and deployment of online and digital services and products in order to mitigate and avoid risks.[179] It consists of "taking preventative steps to ensure that known and anticipated harms have been evaluated in the design and provision of an online service; that user empowerment and autonomy are secured as part of the in-service experience; and that organisations take ownership and responsibility for users' safety and well-being, and are clear about the steps required to address any issues".[180] Child safety by design is deemed necessary for the full protection of children's rights and should be integrated in the design of digital services and products children use.[181] Examples of child safety-by-design measures include age-appropriate content filters, user-friendly interface and reporting tools, and separate platforms for children.

[177] Words used by a child from the Philippines.

[178] Reflection made by focus group facilitators during the validation meeting.

[179] Down to Zero Alliance. (2023). *Child safety by design that works against online sexual exploitation of children.*

[180] UNICEF (2021) *Digital Age Assurance Tools and Children's Rights Online across the Globe.*

[181] UN CRC (2021) *General comment No. 25 (2021) on children's rights in relation to the digital environment.*

It was recognised that some platforms made it more difficult to implement safe settings. Children from many countries[182], for example, expressed that it was important to make use of "**strong privacy settings**" to limit who can see their photos or videos and "unwanted interactions".[183] Children in some countries knew how to do this in specific apps, but others were very unsure (specifically children in Bangladesh and Brazil) and found navigating privacy settings sometimes overwhelming. For example, one child stated that they "didn't understand much about privacy rules" because they were intimidated by windows with "too many letters".[184] Another child relayed how, when she wanted to block someone on a new social media platform, she could not work out how to do it within the application so she needed to "Google on how to do it and how to change the privacy settings".[185] This is consistent with reports stating that some social media platforms were designed to make it "complex and time-consuming [for users] to opt for stronger privacy settings".[186]

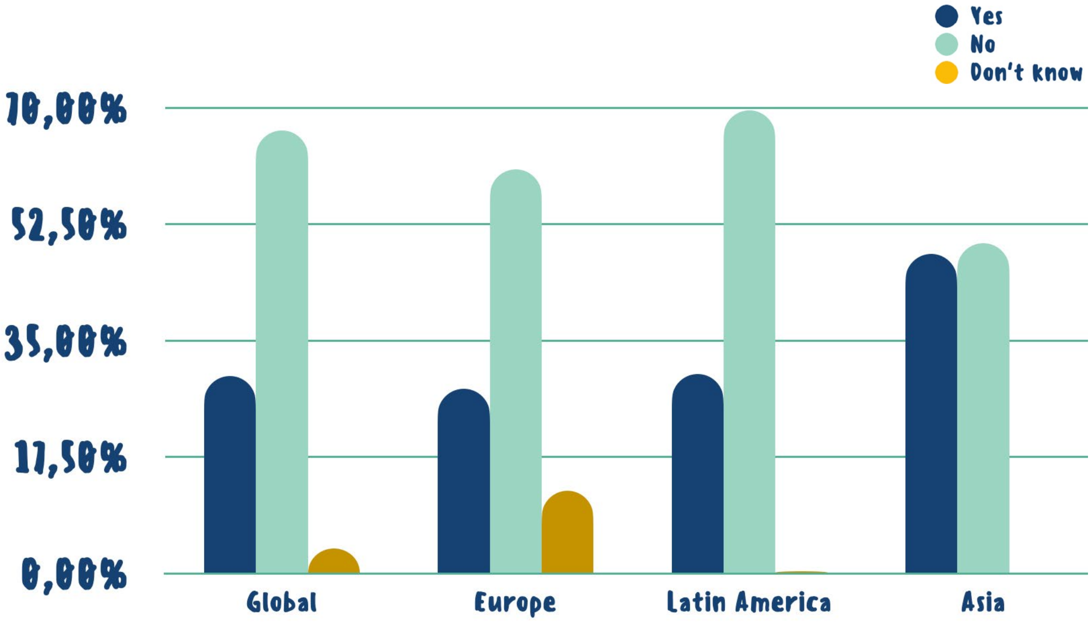
3.3.3 Bridging the gap between parents' and children's online safety strategies

In our data, we found that the caregivers surveyed used many strategies to engage with their children on the topic of online safety from child sexual abuse. A first strategy was the use of parental controls. In our data, caregivers often mentioned that parental-control apps are necessary for child protection. However, two out of three caregivers (66.6%) said that they did not use parental-control apps. This finding is consistent with another study that demonstrated that caregivers had limited utilisation of technology specifically designed for restricting child usage[187]. Looking at the differences across the regions, Figure 17 shows that, in the two countries in South America and in Europe, a little under one in three caregivers used parental-control apps (29.95% and 27.73%, respectively), while in the three Asian countries, almost half of caregivers (48.03%) used them. The Philippines proved to be an outlier, with 62.01% of caregivers using parental-control apps, accounting for a big part of the difference.

[182] Bangladesh, Brazil, Italy, Malta, the Netherlands, the Philippines, and Thailand.
[183] Words used by a child in Bangladesh.
[184] Words used by a child in Bolivia.
[185] Words used by a child in Malta.
[186] Down to Zero Alliance. (2023). *Child safety by design that works against online sexual exploitation of children.* p. 60.
[187] Alqhatani, A., & Lipford, H. (2018, January). Exploring Parents' Security and Privacy Concerns and Practices. In *Workshop on Usable Security*, p. 3.

Figure 17. Percentage of caregivers using parental-control apps, globally and per region.

Do you use a parental control application to monitor the safety of your child/children online?



Another strategy used by many of the caregivers surveyed is communication. Only 5.1% of caregivers indicated that they never engaged in this type of conversation, with a minimum of 0.0% in Brazil and maximum of 10.0% in Bangladesh. The differences across the regions are noticeable, as caregivers in the countries in Europe most often indicated talking to their children every few months about this topic, and the caregivers of the three countries in Asia and two in Latin America most often said that they engaged with their children every time their child went online.

Communicating about this topic was deemed to be important, as many caregivers across several countries mentioned it as a way to better protect children from online risk. Caregivers indicated initiating conversations about safety from online sexual abuse at an average age of 10.0, and the children in our dataset said they went **on social media for the first time at age 9.6**. With the minimum age on many social media platforms, such as TikTok and Facebook, being 13 years old, this means that children are often on platforms that are not appropriate for their age for almost three and a half years. Figure 18 gives an overview of the regional differences in the age the child first enters social media and the age when caregivers first talk about online safety from online sexual abuse. It shows that, on average, only in South America do caregivers have conversations about this topic before children go on social media.

Figure 18. Total and regional information about conversations between caregivers and children.

	Age of first conversation about online safety from online sexual abuse	Social media entrance age
Average in Europe	9.7	9.0
Average of three Asian countries	10.5	10.0
Average of two South American countries	8.5	9.8
Average all countries	**10.0**	**9.6**

Caregivers mostly utilise the **news or experiences they hear of others** to talk to their children about online safety issues. Knowing that caregivers do not always possess adequate knowledge of many of the online safety issues, they might need prompts like these for their own awareness and that of their children. Additionally, caregivers highlight fostering a **supportive environment** to discuss online safety, where children should feel safe to “easily share anything”[188] with them. In all countries, except Bulgaria, some caregivers even spoke in absolute terms as “we discuss everything”. Other caregivers were less positive, mentioning challenges such as the balance between controlling what children do online and giving them space to learn and make mistakes. Caregivers highlighted the difficulty of communicating on the topic of online activities as they felt that children were not willing to share with them:

> “Children don’t want to talk about what they’re actually doing online. They say partial things, but unfortunately not everything. Today’s children are no longer open and keep everything to themselves. Whether it’s a good thing or a bad thing.” (Caregiver from Estonia)

Approximately **three out of four children[189] said that they knew what to do when they are bothered online**. Children mostly said to firstly make use of **online support tools** on the platform to deal with these issues. Children in all countries used blocking, and reporting the issue to a platform was a close second and was noted in all countries except Malta. Other online options that children mentioned were adjusting settings on their profiles[190] and taking screenshots to document evidence of what happened[191].

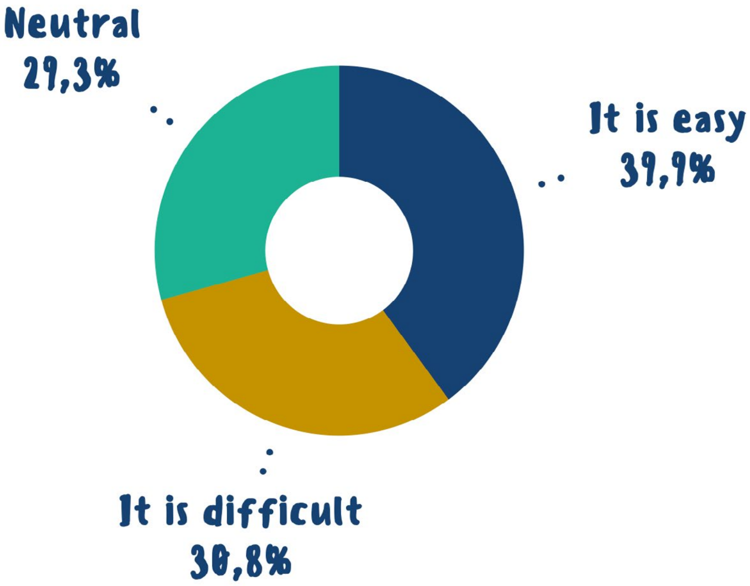
[188] Words used by caregivers in Bangladesh.
[189] Minimum of 52.9% of children in Thailand, maximum of 100% in Austria, Bangladesh, Bulgaria, and Croatia.
[190] Mentioned by children in Bangladesh, Malta, the Netherlands, and Romania.
[191] Mentioned by children in Bangladesh, Bolivia, and Bulgaria.

Even though these online support options were the first things that children mentioned doing when bothered online, **children felt mixed about the effectiveness** of these mechanisms. These feelings mostly stemmed from the possibility that people can make a different account after blocking[192], the platform doing nothing with their reports[193], and feelings that blocking and reporting are inappropriate when the one bothering you is someone you know[194]. This was highlighted during the discussion on what children do when they feel uncomfortable about something that happens online, while answers centred on blocking, some children stated that:

> "I am not very sure what to do because someone might create multiple accounts or use different phones to bother us." (Child from Italy)

As a secondary measure, children said that they know **in-person support** is available**,** for instance, from their caregiver, many acknowledging the benefits of talking to caregivers about online safety. However, **only 39.9% of children indicated that it is easy to talk to caregivers about online safety** (see Figure 19).

Figure 19. Percentage of children thinking it's easy or difficult to talk to their caregivers about online safety.



[192] Mentioned by children in Bolivia, Italy, and the Netherlands.
[193] Mentioned by a child in the Netherlands.
[194] Mentioned by children in Malta and Romania.

Children expressed **facing many barriers in relation to speaking about online safety issues with their caregivers**, with common reasons including feeling uncomfortable[195], fearing potential consequences that would restrict their access to the online world[196], fearing the reaction of their caregivers[197], and thinking their caregivers would not understand[198]. A 2021 study by Rutkowski and colleagues confirmed that children perceive restrictive measures, such as limiting their access to the online world, negatively. They instead value positive emotional states as conducive to better communication about online safety.[199] It is important to note that the level of difficulty children might feel about speaking with their caregivers is connected to the wider relations in the family and is not limited to the issues of online safety. While the matter of online safety related to sexual matters might increase the discomfort in communication, children would have a similar reaction to speaking about those matters in an offline context. Therefore, the fears and concerns expressed by children are likely to reflect the general state of their relationship with their caregivers and should not be seen as limited to the issues of online safety.

Children said that they know that they can go to their caregivers for help, but their statements made this seem **conditional on the nature and severity** of what happened online. For instance, participating children stated that they would not go to their caregiver for “small stuff”[200], and would only ask for help when “big things happen online”[201]. Other factors mentioned by the children were their relationships with their caregivers and an assessment of the consequences after, as the following quote illustrates:

> “I feel mixed. It is important to share with the family, but I am also fearful of potential consequences, such as having my cell phone confiscated.” (Child from the Philippines)

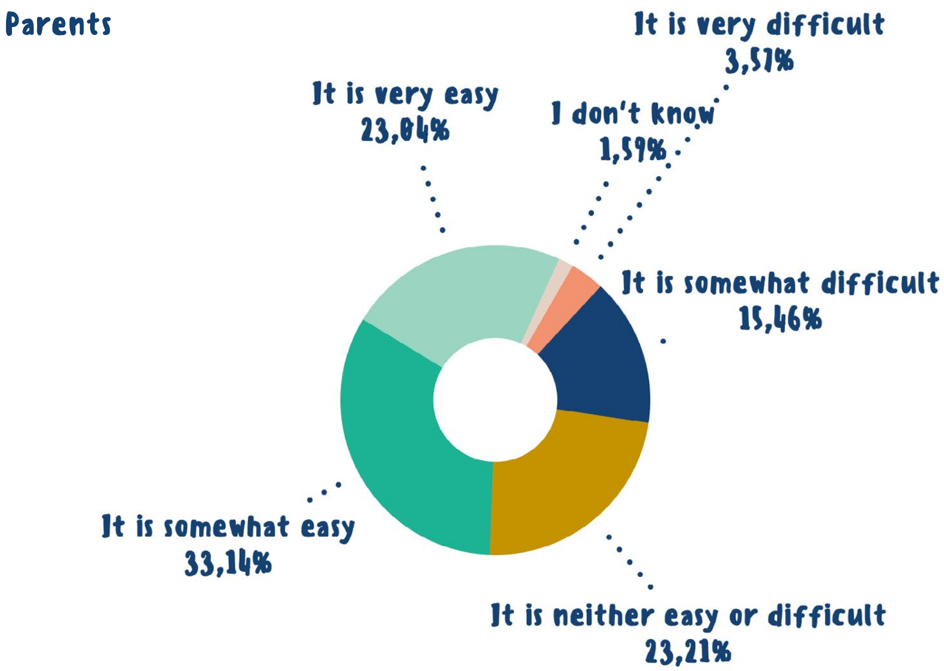
Children in multiple countries[202] said they would filter what they were telling their caregivers, indicating that they would not tell them everything. Some children expressed that they were more likely to turn to siblings[203], teachers[204], or friends[205] instead of their caregivers to talk about online safety.

[195] Mentioned by children in Malta, Romania, and the Philippines.
[196] Mentioned by children in Bolivia, Bulgaria, the Netherlands, and the Philippines.
[197] Mentioned by children in Bulgaria, Italy, and the Netherlands.
[198] Mentioned by children in Bolivia, Bulgaria, Croatia, and Romania.
[199] Rutkowski, T. L., Hartikainen, H., Richards, K. E., & Wisniewski, P. J. (2021). Family Communication: Examining the Differing Perceptions of Parents and Teens Regarding Online Safety Communication. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), Article No. 373, 1–23.
[200] Words used by a child in the Netherlands.
[201] Words used by a child in the Netherlands.
[202] Austria, , Croatia, Italy, Netherlands, Spain, and Thailand.
[203] Mentioned by children in Bolivia Bulgaria, Croatia, Italy, and Malta.
[204] Mentioned by children in Croatia, Italy, Malta, and Romania.
[205] Mentioned by children in Bolivia and Croatia.

Box 8. Difficulties in discussing online safety from child sexual abuse with caregivers

Talking about sensitive topics such as online child sexual abuse can be difficult. In one of the questions, more than half of caregivers thought their child found it somewhat easy (33.1%) to very easy (23.0%) to talk to them about safety from online sexual abuse (see Figure 20).

Figure 20. Percentage of caregivers that believe talking to children about online safety is easy or difficult.

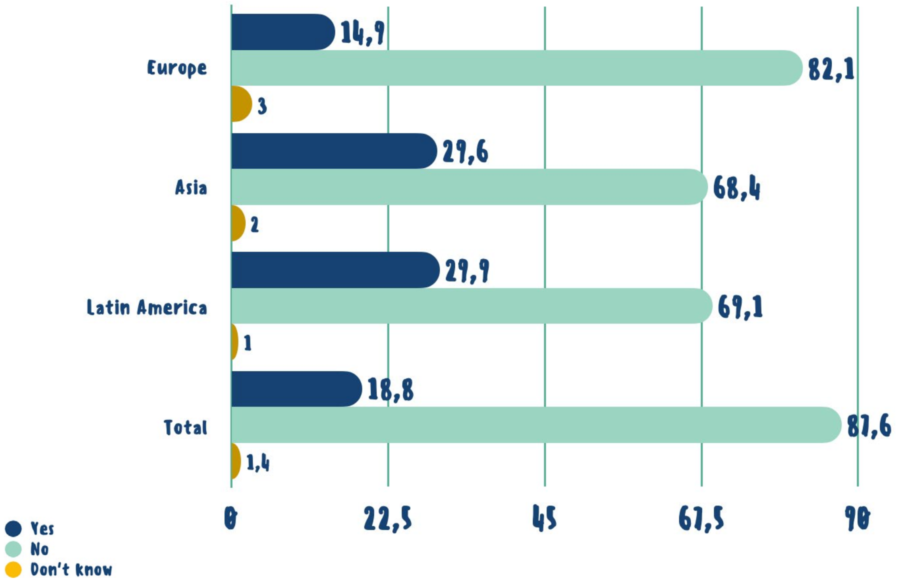


A notable gap exists between how easy caregivers think it is for their children to come to them and how often children do this in practice. Less than **one in five caregivers (18.8%) indicated that their child reported an online child sexual abuse concern to them**. This is significantly lower than the 55% reported in the Global Threat Assessment of 2023, in which perceptions of caregivers were gathered concerning online sexual harms.[206] Regional

[206] WeProtect Global Alliance and Economist Impact. (2023). *Global Threat Assessment 2023: Parents' perceptions of their children's exposure to online sexual harms*.

differences do exist, with a higher prevalence of children reporting online child sexual abuse concerns to caregivers in the Asian (29.6%) and South American countries (29.9%) as compared to those in Europe (14.9%), but these percentages are still much lower than those reported in the Global Threat Assessment of 2023.

Figure 21. Percentage of online child sexual abuse concerns reported by children to caregivers.



Some caregivers mentioned that having conversations about online safety from child sexual abuse can be difficult. For teenagers specifically, caregivers highlighted the challenge of children **hiding parts of their lives** in general. This may be exacerbated by the fact that, in all but three focus groups[207], the children called talking about child sexual abuse **uncomfortable, sensitive, or embarrassing**. Caregivers indicated the difficulty of bringing up such a heavy topic or not wanting to upset the children.

In one of the focus group discussions in Malta, a psychologist was present to oversee the safeguarding of children during the activities.[208] In their notes, they discussed how “intriguing” it was to see the “ambivalence displayed by children towards discussing online challenges with their caregivers or guardians”. They shared the insight that this ambivalence could be seen as “a reflection of the **typical developmental challenges** faced during adolescence, such as the tension between seeking independence and needing guidance”. The psychologists recognised that this is a “complex period” to navigate, underscoring the mentioned caregiver dilemmas where “trust, autonomy, and vulnerability are continuously negotiated”. This message was also reported in another study, where the developmental stage in which children, and especially teenagers, find themselves can lead to seeking “**digital independence**”.[209]

[207] Bolivia, Brazil, Croatia, Estonia, Italy, Malta, the Netherlands, the Philippines, Portugal, Romania, Spain, and Thailand.
[208] In all focus group discussions, a child safeguarding focal person was appointed to oversee the safety and well-being of participants. In Malta, this person was a psychologist.
[209] Down to Zero Alliance. (2023). *Child safety by design that works against online sexual exploitation of children*.

Call to action[210]

Children want safety features that enhance their agency and to be involved in their design

Children expressed a strong preference for online safety measures that promote user agency and stressed the importance of including children in the design of such measures.

Children in eight countries[211] specifically asked for technological solutions that provided them with **agency**, for instance, providing optional safety settings, where children can “decide for themselves”[212] how strict the measures need to be. Other ideas centred around warning messaging, such as user-driven double-checks[213], for example, by flagging potentially harmful chat messages to children and asking if they are having a problem or not[214], as described below:

“Our solution is that instead of the app reporting [bad] people, the app should send a message to both parties (sender and receiver of flagged message) with ‘Hey! There is something unsafe detected’. Then if either party says, ‘Yes, it is something unsafe’ then the app should report it.” (Children from the Netherlands)

Pop-up warnings are preferred by children, as they effectively present children with **choices** while raising their awareness about online safety. A specific example related to sexual abuse was raised in Spain, where children suggested having a “visible announcement” to children before they share explicit pictures or videos online. The children described how this might help deter non-consensual sharing of such content. A recently implemented example can be found in Box [x] on digital nudging. This preference for agency is in line with previous research on the subject as can be seen in Badillo-Urquiola’s work on co-designing safety features for social media apps.[215]

Aside from these concrete suggestions for platforms and policy makers, children were also largely in favour of being **involved**, believing that increased cooperation and inclusion could facilitate better child safety online. As users of online platforms, children raised their voices as willing collaborators in designing child-friendly and safe online spaces. During the focus group discussions, they articulated:

[210] The following ideas and messages herein were voiced directly or indirectly by the children (in the form of posters and drawings) and caregivers (through the open text options of the survey) during the consultations. The authors have collected and summarised them to reflect the conclusions presented in this box.

[211] Mentioned by children in Bulgaria, Croatia, Italy, Malta, the Netherlands, Portugal, Romania, and Thailand.

[212] Words used by a child in Bulgaria.

[213] Mentioned by children in Bulgaria, Italy, and the Netherlands.

[214] Mentioned by children in Italy and the Netherlands

[215] Badillo-Urquiola, K., Smriti, D., McNally, B., et al. (2019). *"Stranger Danger!" Social Media App Features Co-designed with Children to Keep Them Safe Online*.

“We should be included in designing a platform.” (Children from Croatia)

For existing platforms, they were able to critique specific applications, and called for **opportunities to provide feedback in applications**. For example, some found that existing report “buttons” were insufficient[216] or needed to be better designed to be easy-to-use[217].

Some wished to contribute to making complicated privacy rules and complex guidelines clearer and more digestible for their peers.[218] Furthermore, they also discussed how their insider knowledge might be helpful for adults designing detection tools, for example:

“People of our age should participate in developing technology because some bad words are overlooked by older people.” (Children from the Netherlands)

[216] Mentioned by children in Bolivia and Bulgaria.
[217] Mentioned by children in Malta and Thailand.
[218] Mentioned by children in Bolivia, Italy, and Malta.

Conclusion

Conclusion

The objective of this study was to meaningfully engage children and caregivers on the topic of online safety. Through engaging with 483 children and 6,618 caregivers, we identified three critical issues: children and caregivers expressed a **need for more information** on online risk and protection; moreover, there was a call to **ensure both privacy and protection** in online designs, and for **shared responsibility**.

Children and caregivers highlighted multiple **gaps in knowledge** from their points of view. As children specifically mentioned that a lack of awareness contributes significantly to online risk, they called for more information on online safety measures, particularly from schools and online platforms. The caregivers surveyed expressed confidence in their knowledge on overall online safety, but were less confident regarding the specific issues around online sexual abuse of children. Children and caregivers called attention to the fact that they take on the **highest share of responsibility for safeguarding** even without sufficient knowledge and tools. In fact, children emphasised their own personal behaviour as a means to stay safe online, underestimating the potential for platforms to implement safer-by-design approaches. The findings support the need to acknowledge the shared responsibility and accountability of all actors involved, including governments and online service providers.

The conversations with children also revealed a trend of normalisation of online risk, with **children acknowledging potential risks without necessarily expressing feeling unsafe**. This finding reflects the fact that perceived safety does not always correspond to actual risks for children. Children identified being most fearful of unknown people and offline risks in the digital environment. This link is particularly visible in relation to children's fear of OCSEA, which, despite not often being brought up by the children, was mentioned in the context of the misuse of their personal information (e.g., images) and fears regarding unknown users with malicious intentions (e.g., grooming).

Responding to questions around the debate of privacy and online protection, children showed a nuanced understanding of privacy, which was contextualised in relation to the fear of the non-consensual dissemination of personal information and content. Their **understanding of online protection and privacy** seems to be interconnected. While many children supported strong online safety measures when at risk, overall, they called for a mutual reinforcement of protection and privacy mechanisms. In fact, **children tended to favour safety features that promote their agency as users**, allowing them to play an active role in their own safety online. For this, children and caregivers are asking to be part of the solution: policy-makers and online platforms must listen to children to be able to cater for their needs and to ensure that protection mechanisms are working effectively for them.

Overall, children and caregivers conveyed a strong call for safer digital spaces that **empower children to make safe choices and behavioural decisions online**. The findings call for a collaborative approach that expands the knowledge and agency of children, empowering them to exercise their rights online through safe spaces. Children are echoing the key importance of being consulted and listened to on issues related to online safety. To ensure targeted and empowering interventions to protect children's rights online, we must meaningfully involve them in decision-making processes. Child participation is indispensable for effective child protection.

Policy Recommendations

Policy Recommendations

The present study aimed to gather children's views on online safety to better inform digital policy at multiple levels, including EU institutions, national governments, regional and local administrations, online service providers and platforms, and other regulators. Children and caregivers provided advice and concrete recommendations on online safety in the VOICE research, the policy implications of which are summarised below.

For governments and regulators

Regulators should:

In the realm of online child safety, both caregivers and children bear a heavy responsibility. However, they highlighted a critical need for further support. Urgent regulatory action is essential to create a culture of shared responsibility and ensure the safety of all children online, which includes:

- Harmonised legal obligations across all countries and all platforms with appropriate sanctions for non-compliance;
- Increased regulation to prevent harmful platform designs and foster safety-by-design approaches, drafted in consultation with children;
- Ensuring a transversal approach to digital policy that considers children's rights and appointing dedicated competent authorities to monitor and enforce it (e.g., safety commissioners at the national level);
- Consulting children throughout the drafting and implementation of digital policy.

In addition, national governments should:

Ensure compliance of existing regulation and support new actions within their national competences, such as:

- Providing guidance and developing technical solutions via the relevant national data protection and audio-visual regulating authorities.

Implement measures that improve children's mental health as a result of their interactions

In addition, the EU should:

As part of the Better Internet for Kids+ Strategy[219], implement initiatives focusing on procuring safer digital experiences and empowering all children, especially the most vulnerable, and promoting active participation:

- This should include the elaboration of an EU Code of Conduct on Age-Appropriate Design.

[219] The *strategy* for a better internet for kids (BIK+), adopted on 11 May 2022, will ensure that children are protected, respected and empowered online in the new Digital Decade. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+), COM/2022/212.

online by:

- Establishing or extending well-resourced awareness-raising campaigns, which must be informed by emerging online issues (e.g., artificial intelligence and virtual reality);
- Better integrating the dimensions of online harm in policies and investment in mental health support services, while ensuring practitioners that work with children are well resourced and skilled in addressing these issues.

Develop and strengthen online safety education provided at schools by:

- Updating (or developing, when needed) online safety education programmes to include innovative techniques, e.g., gamification, and the latest information regarding the real manifestations of the online risks children face. For this, the co-creation of such programmes with children themselves should be encouraged;
- Designing community-level interventions to support children's and caregivers' understanding of online risks and resilience to online harm. These will also act to improve caregiver digital literacy levels and facilitate dialogue between caregivers and children on online safety.

Strengthen the accountability of online platforms for keeping children safe online by:

- Ensuring the implementation of the Digital Services Act[220], especially Articles 28 and 25, which ensure safer digital environments through design;
- Guaranteeing EU-level regulation that mandates the prevention, detection, and removal of all child sexual abuse online across platforms;
- Providing relevant guidelines and support to Member States for the application of national and EU law (e.g., guidelines on the General Data Protection Regulation (GDPR)[221] related to children's data protection);
- Ensuring a coherent framework of minimum standards for the provision of age-verification mechanisms and age assurance;
- Promoting the development of standards that ensure good practice and self-regulatory frameworks, all of which advance issues related to online child safety (e.g., child rights impact assessments for online designs).

[220] The Digital Services Act, adopted in October 2022, is a regulation aimed at fostering a safe, predictable, and trusted online environment by laying down rules directed at providers of intermediary services. Regulation (EU) 2022/2065 of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC, OJ L 277, 27 October 2022.

[221] Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 2017 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

For online platforms

Online platforms should:

Provide more information on online risks and safety measures by:

- Using child-friendly and age-appropriate language in the information and support mechanisms offered to children, including the Terms and Conditions agreement and reporting mechanisms of the service;
- Ensuring information is accessible in a user-friendly way. For example, children emphasised the need for this information to be embedded within the main platform, with information appearing for them when relevant (e.g., through a pop-up). Information and support should also be made available in languages other than English.

Create secure digital environments for children. In response to the children demonstrating an understanding of online safety that is closely interlinked to the protection of their privacy, online platforms should:

- Foster a balance between protection and privacy in all features directed at children, taking into consideration the best interests of the child. A good practice in this sense would be to develop child rights impact assessments for such tools;
- Establish effective age-assurance and other online safety measures, including the detection of child sexual abuse, in a privacy-preserving way and with a focus on the best interests of the child;
- Prioritise the best interests of the child in the design of their services, especially in terms of avoiding the use of persuasive design for children (e.g., addictive features such as auto-play should be deactivated by default)[222];
- Provide age-appropriate experiences that ensure the highest standards of built-in privacy, safety, and security by design for children[223]. This includes ensuring children's data minimisation, a high level of privacy and security default settings, and the deactivation of detrimental profiling and recommender systems, among others.

Implement a safety by design approach, noting that children associate its absence with a decreased feeling of safety online. In this regard, online platforms should:

- Ensure that safety and privacy settings are accessible and user-friendly for children. This includes improving the features of reporting and blocking, along with features preventing children from being connected to, befriended by, and groomed by fake profiles;
- Prioritise online safety measures that develop the agency of children, e.g., those that present children with choices and appropriate information to make informed choices;
- Personalise safety features according to children's age and maturity, accounting for the evolving capacities of children.

Involve children in the design of their services and their safety features, in order to facilitate effective protection online by:

- Consulting children in the design of online safety measures and new features deployed for children in a meaningful manner, which should include receiving their feedback after implementation;
- Incorporating children in the design of child-friendly information and age-appropriate content;
- Providing more opportunities for children to give feedback on the online safety measures at their disposal in a meaningful and compelling way.

[222] For online platforms providing services to children in the EU, in compliance with Art. 25 of the Digital Services Act.
[223] For online platforms providing services to children in the EU, as mandated by Art. 28 of the Digital Services Act.

Children and caregivers highlighted a need for more awareness and information, privacy-preserving online safety measures, and participation and inclusion. Two parallel outcomes arise from this set of recommendations: (i) **children's and caregivers' knowledge and resilience online is enhanced** through better education and information on online safety; (ii) **governments and regulators hold online platforms responsible for ensuring that their services do not facilitate harm to children**, as part of their accountability to upholding children's rights. In order to ensure this is achieved, the full spectrum of children's rights must be considered in policies and legislation, balancing provision, protection, privacy, and participation rights, while respecting the best interests of the child.

Child protection organisations should actively listen and engage with children and bring their voices into policy debates by continuing to do research and implementing project activities that work with and for children. As ECPAT International, Eurochild, and Terre des Hommes Netherlands, we call on all stakeholders to take the children's opinions expressed here into account in order to make the internet a better place for kids.

Glossary

Account hacking: Account hacking refers to the unauthorised access, manipulation, or compromise of user accounts, typically on online platforms, websites, or computer systems.

(Having) Agency: Agency refers to an individual's capacity and ability to make independent choices, take intentional actions, and exert control over their own life circumstances. Agency helps individuals to weigh up their options, make decisions, and choose how to act.

Age-appropriate usability and content: Age-appropriate usability and content refers to designing digital platforms, applications, and online content in a manner that aligns with the cognitive, emotional, and developmental abilities of users within a specific age group.

Age-verification tool: An age-verification tool is a mechanism or technology designed to confirm the age or date of birth of an individual accessing certain online content, services, or platforms.

Age assurance: Age assurance refers to the process used to estimate and verify the ages of children and users on online platforms, services, and activities. It encompasses tools such as self-declaration, AI and biometric-based systems, technical design measures, tokenised age checking using third parties, and hard identifiers like passports.

Artificial intelligence (AI): Artificial intelligence, or AI, refers to the capacity of computers or other machines to exhibit or simulate intelligent behaviour and
the field of study concerned with this.

Algorithm: An algorithm is a step-by-step procedure for solving a problem or accomplishing some end. Algorithms are commonly used nowadays as the set of rules a machine (and especially a computer) follows to achieve a particular goal.

Blocking features on online platforms: Blocking features on online platforms refer to functionalities or tools that enable users to restrict or limit interactions with specific individuals or content.

Caregivers: In relation to children's rights, the term "caregivers" refers to adults who care for an infant or a child.

Child: A child refers to any person under the age of 18 years.

Child-friendly content: Child-friendly content refers to content that is welcoming to or suitable for children; and is designed with the needs, interests, or safety of children in mind.

Child protection: Child protection refers to the measures and systems implemented to safeguard the well-being, rights, and safety of children from any form of harm, abuse, exploitation, or neglect. It is part of the safeguarding process and focuses on protecting individual children identified as suffering or likely to suffer significant harm.

Child safeguarding: Child safeguarding refers to the collective efforts and measures taken to protect and ensure the well-being of children, particularly from any form of abuse, exploitation, harm, or neglect. Child safeguarding is the responsibility that lies with organisations and has the purpose to promote the welfare of children.

Child sexual abuse: Child sexual abuse refers to any non-consensual involvement of a child in sexual activities that are inappropriate for their age, developmental stage, and understanding. It is also often referred to as the **sexual exploitation of children**, which entails any non-consensual or coercive use of minors for sexual purposes. What distinguishes the concept of child sexual exploitation from other forms of child sexual abuse is the underlying notion of exchange present in exploitation.

Child Sexual Abuse Material (CSAM): The term “child sexual abuse material” refers to any material that depicts and/or that documents acts that are sexually abusive and/or exploitative of a child.

Consent: Consent refers to the compliance in or approval of what is done or proposed by another. Within the EU, consent is one of the six legal bases provided by the General Data Protection Regulation (GDPR) for data processing. The basic requirements for the effectiveness of a valid legal consent are defined in Article 7 and specified in Recital 32 of the GDPR.

Cookies: A cookie refers to a small file or part of a file stored on an internet user's computer, which is created and subsequently read by a website server and contains personal information (such as a user identification code, customised preferences, or a record of pages visited).

Cyberbullying: Cyberbullying is a form of harassment or intimidation that takes place online, typically through electronic communication channels such as social media, messaging apps, or other digital platforms. It involves the use of technology to deliberately and repeatedly harm, threaten, or humiliate an individual or group.

Data leak: A data leak refers to the unintentional or accidental disclosure of information to an unauthorised party.

Deep web: Deep web refers to the set of web pages on the world wide web that are not indexed by search engines but that may be viewable in a standard web browser (by logging onto a website, for example).

Detection technologies: Detection technologies refer to a set of tools, methods, and systems designed to identify and recognise specific conditions, behaviours, or activities online. They are widely used to find and remove illegal content online, such as child sexual abuse.

Digital footprint: A digital footprint refers to a trace or the traces of a person's online activity that can be recovered by electronic means. It is the information about a person that exists on the internet as a result of his or her online activity.

Digital literacy: Digital literacy refers to the ability to use and navigate digital technologies.

End-to-end encryption (E2EE): End-to-end encryption is a method of secure communication that prevents third parties from accessing data while it is transferred from one end system or device to another.

Facial recognition: Facial recognition refers to the identification of human faces by means of visible characteristics. Facial recognition refers to using computer-aided identification of faces and is especially common for security purposes.

Fake profiles: Fake profiles or accounts refer to online identities that are created with the intention of deceiving others about the true identity of the person behind the profile.

Focus group discussion (FGD): A focus group discussion is a qualitative research method that involves a small group of participants who share their thoughts, opinions, and experiences on a specific topic under the guidance of a facilitator.

Gender-based violence: Gender-based violence refers to any harmful act that is perpetrated against a person's will and that is based on socially ascribed (i.e., gender) differences between males and females. It includes private or public acts "that inflict physical, sexual, or mental harm or suffering, threats of such acts, coercion, and other deprivations of liberty".

Grooming: Grooming refers to the action of gaining the trust of or influence over a child, now often via the internet, as preparation for sexual abuse, exploitation, or trafficking.

Harassment: Harassment refers to the creation of an unpleasant or hostile situation especially by uninvited and unwelcome verbal or physical conduct. Harassment is legally prohibited in most domestic law but may vary across jurisdictions.

Hashing: Hashing refers to assigning a numeric or alphanumeric string to a piece of data by applying a function whose output values are all the same number of bits in length.

Human moderation: Human moderation refers to the process of overseeing and managing content on online platforms, websites, or social media through the direct intervention of human moderators.

Identify assurance: Identify assurance refers to the process of ensuring and verifying the legitimacy and accuracy of an individual's identity.

Identity theft: Identify theft refers to the illegal use of someone else's personal information.

Inclusion: Inclusion refers to aiming to provide equal access to opportunities and resources for people who might otherwise be excluded or marginalised, such as those with physical or intellectual disabilities or those that belong to other minority groups.

Machine learning technology: Machine learning is a computational method that is a subfield of artificial intelligence and that enables a computer to learn to perform tasks by analysing a large dataset without being explicitly programmed.

Mental health: Mental health refers to the condition of being sound mentally and emotionally that is characterised by the absence of mental illness and by adequate adjustment especially as reflected in feeling comfortable about oneself, positive feelings about others, and the ability to meet the demands of daily life.

Mixed-method approach: “Mixed method” is a term that is usually used to designate combining quantitative and qualitative research methods in the same research project.

Non-consensual sharing of online content: Non-consensual sharing of online content refers to the act of distributing intimate or explicit images or videos of an individual without their explicit consent.

Online child sexual exploitation and abuse (OCSEA): Online child sexual exploitation and abuse refer to the use of digital platforms and the internet to harm children sexually. Child sexual abuse also takes on an online dimension when, for instance, acts of sexual abuse are photographed or video-/audio-recorded and then uploaded and made available online, whether for personal use or for sharing with others.

Online harm: Online harm refers to a behaviour online that may hurt a person physically or emotionally.

Online platform: Online platform refers to an application or website that serves as a base from which a service is provided.

Online privacy: Online privacy is the ability to control one’s own identity and personal information in the online environment.

Online risk: Online risk refers to the potential threats, dangers, or adverse outcomes that individuals may encounter while using the internet or engaging in various online activities. Online risks are classified into four categories known as the 4Cs: Content, Contact, Conduct, and Contract.

Online safety measures: Online safety measures are a series of mechanisms and technologies designed to protect children from online risk while using the internet by either preventing harmful situations from manifesting, or mitigating their impact when they occur. They are meant to foster a safe digital environment for children.

Parental app control: Parental app control refers to the use of software or tools by caregivers or guardians to monitor, manage, and restrict their children's access to various applications on digital devices.

Personal data: Personal data is any information that is related to an identified or identifiable natural person.

Phishing: Phishing refers to the practice of tricking internet users (through the use of deceptive email messages or websites) into revealing personal or confidential information, which can then be used illicitly, for example, by taking money out of their bank account.

Pop-up blockers: A piece of software that prevents adverts, pop-ups, etc., from appearing on a web page.

Privacy settings: Privacy settings refer to the part of a social networking website, internet browser, piece of software, etc. that allows you to control who sees information about you.

Scams: A scam is a fraudulent or deceptive act or operation.

Self-generated child sexual abuse material: Self-generated child sexual abuse material is sexually explicit content created by and featuring children below the age of 18. These images can be taken and shared intentionally by minors, but are in many cases a result of online grooming or sexual extortion.

Sensitive content control: Sensitive content control refers to the management and regulation of content that is deemed sensitive, inappropriate, or potentially offensive. It involves the implementation of measures and tools to monitor, filter, or restrict access to content that may be considered harmful, explicit, or may violate certain guidelines or policies.

Service provider: A service provider refers to a company, organisation, or individual that offers services to others. These services can encompass a wide range of offerings, including professional, technical, or support services.

Sexual extortion: Sexual extortion, often referred to colloquially as “sextortion”, is a form of blackmail where someone threatens to share a nude or sexual image or video unless their demands are met.

Sexual harassment: Sexual harassment refers to unwelcome, unwanted, and inappropriate behaviour of a sexual nature that creates a hostile and intimidating environment. It can include unwanted sexual advances or gestures, requests for sexual favours, or any other physical or verbal actions. Sexual harassment is generally characterised by an imbalance of power between the offender and the victim.

Social media: Social media refers to any websites and applications that enable users to create and share content or to participate in social networking.

Sugar daddy: The term “sugar daddy” refers to a well-to-do man who is usually older and supports or spends lavishly on a mistress, girlfriend, or boyfriend.

Main sources used for the glossary: The Oxford English Dictionary, the Cambridge Dictionary, and the Merriam-Webster Dictionary. For terminology use and for terms and concepts that all actors understand and consider respectful, see the *Luxembourg guidelines* adopted by the Interagency Working Group in Luxembourg, 28 January 2016.

ECPAT International, Eurochild & Terre des Hommes Netherlands. (2024).
Speaking Up for Change: children's and caregivers' voices for safer online experiences.