OECD publishing

TOWARDS DIGITAL SAFETY BY DESIGN FOR CHILDREN

OECD DIGITAL ECONOMY PAPERS
June 2024 **No. 363**

OECD
BETTER POLICIES FOR BETTER LIVES

Foreword

This report provides an overview of how safety by design initiatives respond to the needs and vulnerabilities of children in the digital environment, with a view to supporting governments' responses and digital service providers' actions. It provides further context to the Recommendation of the OECD Council on Children in the Digital Environment (OECD, 2021[1]) and the OECD Guidelines for Digital Service Providers (OECD, 2021[2]).

This report was written by Andras Molnar and Lisa Robinson, under the supervision of Jeremy West. It incorporates feedback from delegates of the OECD Digital Policy Committee (DPC) and its Working Party on Data Governance and Privacy. This paper was approved and declassified by the DPC on 5 April 2024 and prepared for publication by the OECD Secretariat.

The support and feedback of the OECD's informal, international group of external experts on children in the digital environment is gratefully acknowledged. Clarisse Girot from the OECD Data Governance and Privacy Unit also provided valuable feedback. This report was informed, in part, by a Roundtable on Digital Safety by Design for Children convened by the OECD in July 2023. Contributions of the speakers and participants are likewise gratefully acknowledged. The authors wish to thank Melanie MacNeil and Aahil Sheikh for research and editorial support, and Andreia Furtado for her assistance with publication.

Note to Delegations:

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

DSTI/CDEP(2023)13/FINAL

Table of contents

TABLES

BOXES

Executive summary

The digital environment is an integral part of children's lives, offering them unprecedented opportunities but also presenting significant risks to their safety. As this environment evolves, the need for robust safety measures becomes more urgent. Digital safety by design for children is a means of keeping children safe in the digital environment whilst allowing them to seize opportunities to enjoy its benefits. Just as physical products and spaces for children are designed to prioritise their safety, so should be the digital products and spaces children use and inhabit.

However, a clear understanding of the concept of digital safety by design for children, both in theory and in practice, can be elusive. This paper provides an overview of digital safety by design for children, what it encompasses, and how it can respond to the needs and vulnerabilities of children in the digital environment. The key findings of the report are:

There are increasing international calls for proactive digital safety measures for children

A growing international consensus highlights the necessity of digital safety by design for children. This report reviews different international guidance materials concerning children online and shows that, whilst there is no uniform definition of digital safety by design across them, all the initiatives stress the importance of integrating safety considerations at the outset of product or service development. Other common themes are transparency and accountability, tailored user experiences and child-friendly service provision. Some documents highlight the importance of continually adapting services to respond to new risks, and all stress the importance of protecting children's privacy. The growing international consensus on the need for digital safety by design for children goes hand in hand with a global prioritisation of children's digital well-being and a focus on the responsibility of digital service providers to safeguard children.

More jurisdictions are enacting laws that encompass digital safety by design for children, but they bring a risk of international regulatory fragmentation

Across the OECD, a growing number of new regulatory bodies focus on digital safety and several jurisdictions have enacted digital oversight legislation that places children's safety at the forefront. These laws often require practical tools and measures, such as age assurance, accessible complaint mechanisms, and mandatory reporting on safety measures. At the same time, digital safety for children is a global issue, and should jurisdictions act in isolation there will be a risk of regulatory fragmentation. A rights-respecting approach, as well as multi-stakeholder dialogue and co-operation at a forum such as the OECD, can assist in avoiding such fragmentation and support effective digital safety policies that protect children and promote their well-being.

Other "by design" approaches offer insights for digital safety by design for children

Lessons can be drawn from other frameworks that incorporate safeguards at the level of design - both offline safety by design examples, and other digital by design frameworks (such as privacy by design and

security by design). Essential elements include proactive safety integration, user-centred design, continuous risk evaluation, and clear accountability and transparency.

Digital safety by design for children requires several elements

The paper outlines key components for digital safety by design for children. Whilst the paper does not attempt to provide a comprehensive overview of what should be encompassed in digital safety by design initiatives for children, together, the following key components can contribute to a safer digital environment for children:

- ***Employing age assurance mechanisms***: To provide age-appropriate experiences, digital service providers need to know which of their users are children. Age assurance is important for this purpose, but consideration should be given to potential solutions' accuracy, usability, privacy preservation and risk proportionality.
- ***Implementing child-centred design:*** Child-centred design puts the evolving needs, preferences, and safety of children at the core of product and service development. This approach aims to ensure that digital products and services are not only accessible and engaging for young users, but inherently safe and beneficial, as well, and that they remain so as children mature.
- ***Detecting and preventing harm:*** Digital service providers can proactively identify and mitigate risks by implementing technical safety measures, such as advanced detection systems, default settings, content filters, and real-time monitoring tools. Specific attention should be paid to the risks posed by such measures and to their compliance with existing regulations.
- ***Protecting children's privacy and personal data:*** Breaches of children's privacy and the misuse of their data can directly affect their safety. Accordingly, when digital service providers prioritise privacy by design and offer clear privacy settings, it advances safety by design.
- ***Ensuring child-friendly information provision:*** Children need to understand the digital spaces they inhabit. Consequently, they need clear, timely, accessible, and age-appropriate information about how digital services work, the risks involved, and how they can be protected.
- ***Facilitating complaints and redress:*** Empowering children to voice concerns and seek remedies is crucial. By establishing clear, user-friendly, accessible and age-appropriate channels for reporting issues and ensuring timely and effective responses, service providers demonstrate a commitment to children's safety and help uphold trust in the digital ecosystem.
- ***Encouraging child participation and putting children at the centre of decision-making:*** Children are active digital citizens and both service providers and policymakers should involve children in discussions about online safety, design processes, and policy formulation. By giving children a seat at the table, stakeholders can help to ensure that the digital environment is shaped with children's best interests at heart.
- ***Promoting a culture of safety and well-being:*** Promoting a culture of safety and well-being is essential to developing responsible corporate culture that prioritises children's safety.

Finally, this **paper provides case studies that demonstrate diverse digital safety strategies across platforms, highlighting the need for tailored approaches based on unique risk profiles**. These real-world examples emphasise that a one-size-fits-all solution would be inadequate. Instead, promoting children's safety requires adaptable measures, especially given that merely labelling a service as "not for children" does not guarantee protection.

1 Introduction

Digital technologies have become central to children's well-being and development and the digital environment is an integral part of their lives, offering important opportunities for self-expression, learning, socialising, connecting with community and culture, and the enjoyment of their rights. It also presents a wide spectrum of risks to which children are more vulnerable than adults, including exposure to harmful or illegal content, cyberbullying, breaches of privacy, commercial exploitation and falling victim to abhorrent crimes like sexual exploitation and abuse.

As digital technologies advance, so do the possibilities for design flaws, unintended consequences, and intentional misuse, as well as the attendant risk of children suffering harm via the digital environment. In fact, the risk of exposure to well recognised harms is accelerating. The incidence of child sexual exploitation and abuse online, for example, is growing at an alarming rate (OECD, 2023[3]). New technologies also bring new potential for harm. For instance, generative Artificial Intelligence (AI) technologies have already given children dangerous advice and engaged them in adult conversations (e.g. regarding sex or alcohol) (Vosloo, 2023[4]), and there is increasing concern that children may be harmed not just by what they see or who contacts them online, but by the very design or use of certain services (Cortesi et al., 2019[5]) (US Surgeon General, 2023[6]).

As risks expand, stakeholders are paying greater attention to the responsibility of the digital service providers[1] who design and develop the digital spaces children inhabit to provide products and services that are safe. The OECD recognises digital service providers as having an essential role in providing a safe and beneficial digital environment for children (OECD, 2021[1]). Ideally, when designing digital spaces, digital service providers would find and implement effective ways to protect children from harm whilst providing them with opportunities to enjoy the benefits of the digital environment.

Digital safety by design for children is emerging as a concept that could point the way towards innovation that is both beneficial and responsible, placing children's safety and well-being at the forefront. Designing technology for safety is not a new idea. It is widely expected that safety is designed into physical products that are made for children, such as car seats, strollers and toys. Indeed, when such products are found not to be safe, they are recalled (OECD, 2023[7]). Applying the same philosophy to digital products and services can help address the evolving and widespread risks of the digital environment, whilst ensuring that it can remain inclusive and beneficial.

Nonetheless, a clear understanding of what digital safety by design for children comprises – both at the level of theory and at the level of practice – is still developing. In July 2023, the OECD convened an Expert Roundtable on Digital Safety by Design for Children (the Roundtable) for the purposes of discussing the concept in depth and informing this report (OECD, 2023[8]) (OECD, 2023[9]). Experts at the Roundtable stated that digital safety by design for children means:

- safety is a paramount consideration through the whole life cycle of a digital product or service's design, development and deployment;
- the burden of safety does not lie with children (or their parents[2]);
- laws and policies should be outcome-based and technology-neutral;
- safety forms a key part of corporate responsibility; and

- decision making is child-centred.

As one expert put it, digital safety by design for children requires all stakeholders involved in the design, development and delivery of services and products to ask themselves "*what would be done differently if it was known that the end user is a child?"* (OECD, 2023[9]).

At the OECD, the importance of responsible innovation and putting child safety and well-being at the forefront during the development phase of digital technologies and services is set out in the Recommendation on Children in the Digital Environment (the OECD Recommendation) (OECD, 2021[1]), as well as in its accompanying Guidelines for Digital Service Providers (the OECD Guidelines) (OECD, 2021[2]), which both call for child safety by design. A brief background to these documents and the OECD's work on children in the digital environment is set out in Box 1.1.

Box 1.1. OECD Work on Children in the Digital Environment

The OECD's work on children in the digital environment commenced with the 2008 Seoul Ministerial Declaration, which called for a collaborative effort "between governments, the private sector, civil society and the Internet technical community in building an understanding of the impact of the Internet on minors in order to enhance their protection and support when using the Internet". In 2011, the OECD released a detailed report that analysed the risks children faced in the digital environment at that time, and in 2012, the OECD adopted the Recommendation on the Protection of Children Online.

Subsequently, in 2020, work undertaken by the OECD revealed that there were fragmented legal and policy responses to the needs of children in the digital environment, as well as an evolving and complex risk landscape. It found that parents need support in helping their children navigate the digital world, and children themselves must be consulted and recognised as key stakeholders and individual rights holders in the development of policies. Reliance on co- or self-regulation of digital service providers was not meeting the needs of children, and digital service providers were seen to have an essential role in ensuring a safe and beneficial digital environment for children.

On 31 May 2021, in view of these findings, as well as the technological, policy and legal advances since 2012, the OECD Council adopted the Recommendation on Children in the Digital Environment (revising the 2012 version). The OECD Recommendation sets out key principles and provides concrete guidance to governments and other actors for achieving a safe and beneficial digital environment for children. Adopted alongside the OECD Recommendation (as an addendum), the OECD Guidelines for Digital Service Providers seeks to support digital service providers in taking actions that respect and protect the rights, safety, and interests of children. A Companion Document accompanies the OECD Recommendation, providing guidance and context for governments and stakeholders, to support their implementation efforts.

Source: (OECD, 2008[10]) (OECD, 2011[11]) (OECD, 2021[1]) (OECD, 2017[12]) (OECD, 2020[13]) (OECD, 2021[14]) (OECD, 2021[2]) (OECD, 2022[15])

Outside the OECD, at both the international and national levels, stakeholders are increasingly calling for digital safety by design to be embedded in digital products and services (see further sections 3 and 4). However, along with these calls come different attempts to define and expand upon the concept of digital safety by design for children. Consequently, there is a risk that the concept and efforts to implement it will diverge and be rendered unclear for stakeholders including governments, regulators, businesses, parents, and children themselves. Developing a common understanding of digital safety by design for children can help to form a basis for a constructive and collaborative dialogue between the stakeholders who aim to foster trust whilst ensuring a safe and beneficial digital environment for children.

This draft report seeks to fill this need by providing a high-level and horizontal overview of digital safety by design for children, including how it is being approached at the international and national levels, and suggests practical actions to support it. Key aspects are examined and analysed as a means of contributing to a shared understanding of the concept of digital safety by design for children, and in this way the report seeks to aid the development and implementation of appropriate policy responses. It is acknowledged that a broad range of actions are important for fostering a safe and beneficial digital environment for children, including (e.g.) parental mediation and digital literacy. However, the main focus of this report is the role of digital service providers in deploying digital safety by design, as well as the oversight role held by governments and regulators.

The draft report is divided into two main parts. The first part provides an overview of the policy and legal landscape for digital safety by design for children, considering in turn: lessons that can be learnt from other by design approaches (section 2); international initiatives calling for digital safety by design for children (section 3); existing and emerging laws and policies (section 4) and the role of digital safety regulators (section 5). The second part considers practical aspects of digital safety by design for children, suggesting key components for embedding safety for children in digital products and services (section 6) and then applying these key components to a set of case studies (section 7). Section 8 concludes and sets out potential next steps.

2 Existing by design approaches

The very concept of a safe space for children consistently evolves. As noted at the Roundtable, a playground designed in the early 20th century, with concrete or asphalt surfaces and hard edges, would not be considered safe for children today. As safety expectations have changed, so have the ways physical spaces for children are designed. Playgrounds are now designed with safety in mind, featuring impact-absorbing materials and soft edges to minimise the risk of injury during play. Children's playgrounds are no longer just physical spaces, though. They are also found on children's connected devices, which are more than just playgrounds but also spaces for education and other non-play leisure activities (such as hobbies) (OECD, 2023[16]). It follows that safety by design concepts and standards should be applied in digital spaces, too.

This section reviews different examples of by design approaches, looking first at safety by design offline and then at other digital by design approaches. Finally, it considers the lessons that may be drawn from these existing approaches.

Designing for safety offline

Embedding safeguards into products and services at the level of design is not a new consideration. In the offline world, designing for safety is widely expected from physical products and services. For instance, features in cars which are today commonplace - such as seat belts, air bags and anti-lock brakes - are all examples of safety by design (INHOPE, 2021[17]).

In the offline context, designing for safety includes a number of different actions. For instance, rigorous testing before a product is released to market, such as clinical trials for pharmaceuticals or the testing of aircrafts for safety and reliability (Federal Aviation Administration, 2023[18]) (European Union Aviation Safety Agency, 2023[19]) (US Food & Drug Administration, 2023[20]) (European Medicines Agency, 2023[21]). It also includes ensuring that there is clear and accessible safety information for consumers, such as labels on medicines so that they can be clearly identified and conditions for safe use understood (UK Government, 2020[22]). Lastly, it involves independent oversight – for example, pharmaceuticals are released only after they have approvals from relevant authorities and airplanes must comply with the standards of authorities (US Food & Drug Administration, 2023[20]) (European Medicines Agency, 2023[21]) (European Union Aviation Safety Agency, 2023[19]).

For children, specific safety features are routinely required for both products that are directed at children (e.g. toys), as well as for products that are not directed at children but could cause them harm. An example of the latter are child resistant caps, which are commonly mandated and are designed to prevent children accidentally ingesting poisonous substances (White and Kibalama, 2018[23]). The European Union's Toy Safety Directive is an example of the former, and it includes obligations for mitigating safety and health risks to children, such as setting maximum noise values to avoid impairing children's hearing, pre-release safety testing, clear communication of risks with labels and warnings, and providing instructions and safety information in a language that can be easily understood (European Union, 2009[24]).

Digital by design concepts

The examples of security by design and privacy by design

Applying a “by design" approach to the digital sphere is likewise not new. Digital by design concepts aim to harness the influence of digital service providers, designers and policymakers to shape product and service development in ways that prioritise values that promote human well-being (Livingstone and Pothong, 2021[25]) (Livingstone and Pothong, 2023[26]). Two such concepts that are already well developed are security by design and privacy by design. In understanding digital safety by design for children, and any necessary actions to achieve it, it is helpful to examine these existing approaches and draw lessons from them.

Privacy by design refers to embedding privacy protections by default into the design, operation and management of organisations, as well as in products and services (OECD, 2020[27]). Cavoukian established seven foundational principles for privacy by design:

- proactive not reactive – preventive not remedial;
- privacy as the default setting;
- privacy embedded into design;
- full functionality – positive sum, not zero sum;
- end-to-end security – full lifecycle protection;
- visibility and transparency – keep it open;
- respect for user privacy – keep it user centric (Cavoukian, 2011[28]).

In practice, privacy by design may be used as a framing objective in laws and policies. For instance, the EU’s General Data Protection Regulation (GDPR) makes it clear that measures must be established to ensure data protection by design and by default (GDPR, 2018[29]) (GDPR, 2018[30]). The Australian Office of the Information Commissioner encourages privacy by design for organisations and government as an effective and efficient measure to ensure privacy is built into the design specifications and architecture of new processes and systems (Office of the Australian Information Commissioner, 2022[31]).

Security by design means building security into products and services from the start. It seeks to ensure that products are built in a way that reasonably protects against malicious cyber actors getting access to devices, connected infrastructure and data (Cybersecurity and Infrastructure Security Agency, 2023[32]). Security by design requires systematic implementation of rigorous standards with respect to supply chain, product design, and production environments (OECD, 2019[33]). Although products and services cannot be secured “once and for all” (OECD, 2021[34]), those that are less secure by design are more likely to become victims of digital security attacks (OECD, 2021[34]).

The OECD’s Recommendation on the Digital Security of Products and Services advocates taking a security by design and by default approach for products and services, which involves:

- setting out a need for baseline security measures,
- defining digital security requirements based on the assessment of the digital security risk of the product,
- pre-configuring and activating security features by default, and
- providing users information about preconfigured security settings and clear, simple instructions for security configuration (OECD, 2022[35]).

Whilst, of course, certain aspects of both privacy by design and security by design are unique to their respective subject matter, common themes in each can provide insight into necessary actions for digital safety by design for children. For instance:

- *Baseline safeguards* are features of both privacy by design and security by design. For instance, the United Kingdom's Product Security and Telecommunications Infrastructure Act mandates that manufacturers of consumer connectable products comply with baseline security requirements such as not using default passwords (UK Government, 2023[36]). Privacy by design requires limiting the collection of personal data whereby data controllers implement technical and organisational measures to ensure – by default – that only the personal data necessary for each specific purpose of the processing are processed (GDPR, 2018[29]).
- *Impact/risk assessments* are processes designed to identify and minimise risk as early as possible. The GDPR, for instance, requires a data protection impact assessment for processing of activities that might pose a significant risk to the rights and freedoms of individuals (GDPR, 2018[37]). Such assessments are recommended to occur both at initial stages of project design and development, but also throughout a product or services lifecycle to account for changes or any new risks (UK Government, 2024[38]).
- *Accountability* refers to a willingness and capacity to be responsible and answerable for business practices (OECD, 2023[39]). In the example of privacy by design it requires that organisations take responsibility for the personal data they handle. This can extend to a requirement to appoint a dedicated Privacy Officer to ensure that accountability requirements are met (European Data Protection Supervisor, 2023[40]) (GDPR, 2018[37]). It can also (i.e. in the security by design example) require appointing "risk owners" who are senior and experienced stakeholders accountable for managing risks throughout a products lifecycle (UK Government, 2024[41]).
- *User-centricity* refers to design processes that focus on the needs of the user. For privacy, this can manifest through requirements that give users control and ownership of their data such as rights to request disclosure or corrections of their personal data (Personal Information Protection Commission, 2003[42]) (Korea Legislation Research Institute, 2011[43]) (Parliamentary Counsel Office, 2020[44]). An example as part of security by design is relying on user experience research to implement security processes that are fit for purpose and easy to understand (UK Government, 2024[41]).
- *Transparency* is about open practices, and clear and accessible information provision. For instance, the UK Product Security and Telecommunications Infrastructure Act requires manufacturers to have a vulnerability disclosure policy that gives users information on how to report security issues and to provide transparent information on the amount of time the product will receive security updates (UK Government, 2023[36]) (UK Government, 2024[45]). Australia's Privacy Principles require that organisations manage personal information in an open and transparent manner, similarly the PIPEDA obliges organisations to provide clear and easily understandable information about their privacy policies and practices (Australian Government, 2014[46]) (Government of Canada, 2000[47]).

Applying a digital, by design approach to children

Outside of safety, advocates for children's rights have already turned their minds to applying by design approaches to meet children's unique needs in the digital environment. This includes frameworks for children's rights by design and playful by design. The former assumes a more holistic view (including provision, protection and participation rights). The latter sets out the importance of fostering enriching play opportunities for children. Both concepts spotlight the inherent responsibilities of digital service providers and underscore the role of intentional design in shaping safe and beneficial online experiences of children.

The Digital Futures Commission has developed resources on each of these concepts, summarised in Box 2.1.

Box 2.1. Children's rights by design and playful by design

Children's rights by design

This concept situates the protection of children within a holistic framework of children's rights encompassing protection of life, non-discrimination, participation, privacy, and information, amongst others. This guidance is for innovators of digital products and services that are likely to be used by, or likely to affect children, to support the practical implementation of rights by design.

The guidance is underpinned by eleven principles: i) equity and diversity; ii) best interests; iii) consultation; iv) age-appropriateness; v) responsibility; vi) participation; vii) privacy; viii) safety; ix) well-being; x) development; and xi) agency. The principles of equity and diversity, best interests, age appropriateness, and privacy are noted as key to enhancing children's safety in the digital environment. The resource recognises that collectively these principles promote safe, inclusive, and age-appropriate content whilst protecting children's personal data from exploitation. Additionally, it emphasises that responsibility and safety in design can help create safe digital spaces, actively mitigating potential risks.

Playful by design

Playful by design is a tool that aims to help designers enhance children's opportunities for play in the digital environment and address the challenge of developing digital products and services that respect children's rights.

It also seeks to support parents and game reviewers to evaluate the opportunities for free play in digital games. The tool includes resources to initiate discussion and provoke reflection and fresh ideas. It is directed at any stakeholder developing digital products and services used by children and can be used at any stage of the design process.

The playful by design tool is made up of cards that prompt child-focused thinking by designers and other stakeholders. They are grouped in the following categories: i) be welcoming; ii) enhance imagination; iii) support open-ended play; iv) adopt an ethical commercial model; v) ensure safety; vi) allow for experimentation; and vii) be age-appropriate.

Sources: (Livingstone and Pothong, 2023[26]) (Digital Futures Commission and 5Rights Foundation, 2023[48])

What lessons can be drawn from other by design approaches and applied to digital safety by design for children?

Several characteristics of the by design approaches outlined above are relevant to and could be applied in the context of digital safety by design for children, including:

- Proactive approaches that can identify and address vulnerabilities before they surface and harm children.
- Ongoing and dynamic risk management, such as embedding processes for finding and addressing safety problems that arise despite of any proactive approaches. This could involve regular risk assessment.
- User-centric design, for example child-friendly information provision and consideration as to age-appropriateness.

- Accountability, which is essential in all by design approaches, and for digital safety should require assigning responsibility within the design and development teams, but also fostering a company-wide mindset, including at the highest executive levels.
- Transparency, which again is essential in all by design approaches, and in the digital safety by design context involves transparent information on risk, actual harm that has already occurred on services, and actions being taken to address any risk or harm.

Safety by design may also need to be combined with other by design processes, and security by design and privacy by design have been advocated by the OECD and the United Nations Committee on the Rights of the Child as a means of enhancing protections for children's rights and interests from the digital products and services that they use (OECD, 2021[1]) (UN CRC, 2021[49]). For instance, limiting the collection and retention of personal data (i.e. privacy by design) to avoid privacy violations and data breaches that could undermine children's safety, is likely also necessary alongside any safety by design considerations.

3 International initiatives to understand and promote digital safety by design for children

Whilst there is no internationally accepted definition of digital safety by design for children, understanding, expanding upon and promoting the concept has become a prominent objective for a wide array of organisations, coalitions[3], and alliances[4]. A wealth of frameworks[5], guidance, codes of practice[6] and principles that have been developed (at least in part) to protect children in the digital environment explicitly call for digital safety by design.

At the OECD, safety by design is central to both the OECD Recommendation and the OECD Guidelines (OECD, 2021[1]) (OECD, 2021[2]). The OECD Recommendation encourages governments to promote safety by design through fostering the research, development, and adoption of privacy protective, interoperable and user-friendly technologies that can restrict contact and access to content that is inappropriate for children, taking into account their age, maturity, and circumstances (at III.5.a); and providing all stakeholders with clear information as to the trustworthiness, quality, user friendliness, and privacy by design of such technologies (at III.5.b) (OECD, 2021[1]).

The OECD Guidelines advise digital service providers to take a child safety by design approach in designing, delivering, and deploying services that are either directly intended for children or where it is reasonably foreseeable that they will be accessed or used by children (OECD, 2021[2]). In other words, this entails not just services whose intended audience is children, but those that children are actually using. Deciding whether it is reasonably foreseeable that a service will be accessed or used by children has to be decided on a case by case basis and depends on whether it has a particular appeal for children, ease of access, and the nature and content of that service (OECD, 2022[15]).

The OECD Guidelines set out that in taking a child safety by design approach, digital service providers should:

- pay due regard to providing a safe and beneficial digital environment for children through the design, development, deployment, and operation of their products and services, including through taking a safety by design approach to address risks;
- take necessary steps to prevent children from accessing services and content that should not be accessible to them, and that could be detrimental to their health and well-being or undermine any of their rights, and continue to review the efficacy of those measures and improve them where necessary;
- regularly review and update practices to take into account changes to technology, changes in use, and consequent changes in risks for children; and
- ensure that, where laws and policies require them, age-based restrictions are in place to prevent children below certain ages accessing a service, and that such restrictions are proportionate to risk, privacy-preserving and respected (OECD, 2021[2]).

The Companion Document to the OECD Recommendation suggests that a digital safety by design approach for children assumes its adoption as an embedded design objective, prior to the use of any system architecture, service or product. The Companion Document further notes that safety by design's main objective is to minimise risks to children by anticipating, assessing the impact of, detecting, and eliminating harms before they occur (OECD, 2022[15]).

The UN Committee on the Rights of the Child, in its General Comment 25, recognises that digital safety by design is necessary for fully protecting children's rights in the digital environment (UN CRC, 2021[49]) and recommends that safety by design be integrated into the services and products that children use, so as to minimise the risks they face in the digital environment (UN CRC, 2021[49]). An Explanatory Note to General Comment 25 states that safety by design is the practice of designing services with the goal of ensuring users' safety, for instance by default safe settings for accounts of underage users or by preventing adults from contacting children (5Rights Foundation, 2021[50]). General Comment 25 recommends that states require businesses whose activities affect children's rights in the digital environment implement regulatory frameworks, terms of service and industry codes that adhere to the highest standards of safety, ethics, and privacy in relation to the design, development, engineering, operation, distribution and marketing of their products and services. It further recommends that businesses maintain high standards of accountability and transparency and encourages them to take measures to innovate in the best interests of children. Lastly, it sets out that businesses should require the provision of age-appropriate explanations of their terms of service to children, or to parents and caregivers for very young children (UN CRC, 2021[49]).

The UN General Assembly Resolution on the Rights of the Child in the Digital Environment encourages states to urge companies to address negative effects on children's rights in the digital environment that are associated with their design, operations, products and services. It calls for the promotion of industry codes and terms of services that respect, protect and fulfil the rights of the child and which uphold ethics, privacy and safety standards in relation to the design, engineering, development, operation, distribution and marketing of technological products and services. It encourages private actors in the technology sector to take into account the particular needs of children and follow international standards and best practices for safety, privacy and security by design (United Nations, 2023[51]).

UNICEF research[7] notes that digital safety by design for children should include taking preventative measures to make sure that anticipated and known harms have been evaluated in the design and provision of a digital service. This research also underlines that user empowerment and autonomy should be secured as part of an in-service experience, that organisations should take ownership and responsibility for user safety, and that they should be transparent about the measures taken to address any concerns (UNICEF, 2020[52]).

The Council of Europe's (CoE) Guidelines to respect, protect and fulfil the rights of the child in the digital environment note the importance of safety by design for children. These guidelines call upon CoE member states to promote safety by design and to provide incentives to businesses to implement it as a guiding principle for products' and services' functionalities and features that might be used by children or addressed to them (Council of Europe, 2018[53]) (Council of Europe, 2020[54])[8].

The International Telecommunication Union (ITU) Guidelines for industry on Child Online Protection recommend that to create a safe and age-appropriate digital environment, companies should always consider safety by design in products and services that are addressed to, or commonly used by, children (International Telecommunication Union, 2020[55]). These guidelines stress that children's safety and the responsible use of technology should be carefully considered and not be an afterthought (International Telecommunication Union, 2020[55]).

The World Economic Forum (WEF) Global Principles on Digital Safety seeks to advance digital safety in accordance with human rights principles, drive multi-stakeholder engagement, and enable and inform industry, regulatory and societal efforts and innovations. The principles provide a framework for applying rights-respecting initiatives to online safety across any activity, from regulation to product development. In

particular, the principles call on digital service providers to "invest in and embed a multidisciplinary approach to safety by design throughout the business lifecycle of products and services, including empowering users by providing tools, methods and resources to tailor their experiences, help them safeguard themselves and report harm" (World Economic Forum, 2023[56]).

Table 3.1 provides an overview of the key international guidance documents, highlighting the main commonalities in their approaches towards digital safety by design for children and the actions that digital service providers are called to take. Some of the elements identified are incorporated in the overall guidance document but are not specifically linked with the concept of safety by design (these are denoted with *).

Table 3.1. Overview of selected international guidance documents on digital safety by design for children

Title of the guidance	Common Elements / Requirements						
	Accountability/ transparency	Age-appropriate service provision	Child-friendly information provision	Default safe settings	Integrate safety considerations at the outset	Privacy protections	Regularly review/ update practice
OECD Recommendation on Children in the Digital Environment; Guidelines for Digital Service Providers	*	●	*	●	●	●	●
UN CRC General Comments 25	●	*	●	●	●	●	*
UNGA Resolution on the Rights of the Child in the Digital Environment	*		*	●	●	●	
UNICEF Children Rights by Design	●	●	●		●	●	
COE Guidelines to respect, protect and fulfil the rights of the child in the digital environment	*	*	*	●	*	●	*
ITU Guidelines Industry on Child Online Protection	*	●	*	●	●	●	
WEF Global Principles on Digital Safety (*N.b. not child specific*)	●			●	●	*	

Source: OECD
Notes: "●" Denotes where the reference to safety by design calls for this element; "*" Denotes where this is called for in the overall guidance but is not specifically linked with safety by design.

Digital safety by design is called for in all these guidance documents, and whilst there are common themes as identified above, the level of detail in how the concept is defined varies. For the most part digital safety

by design for children is described as including a need to integrate safety considerations at the outset of product or service development, through it being made it clear that this must be a design imperative. However, the guidance documents rarely include a more in-depth description of what incorporating safety at the design or conceptual stage means in practice.

Protecting privacy is commonly included as an important element in safety by design, and in some documents (e.g. General Comment 25 and the UNGA Resolution) privacy by design is associated with safety by design as dual protective measures. Others, (such as the OECD's own guidance) do not necessarily make a combined call for privacy and safety by design but do call for any safety by design technologies that are employed to be privacy preserving. Default safe settings are another common feature, however specificity as to what this incorporates is not always expanded on. Nonetheless some of the guidance documents do call out specific measures such as age assurance (OECD) or technical measures to prevent unknown adults contacting children (General Comment no. 25, Explanatory note).

Transparency and accountability, age-appropriate service provision, and child-friendly information provision, whilst specifically incorporated in the description of digital safety by design for children in only a few of the documents are all nonetheless a common feature across almost all of the guidance documents. This highlights the importance of child-centred and user-friendly actions, transparent business practices, as well as corporate responsibility. It is noted that where many of the documents call for child-friendly information provision, they also associate this with complaint procedures that are open and accessible to children.

For those guidance documents directed primarily at governments (e.g. CoE, OECD, UN) there is an emphasis on governmental responsibility to encourage digital service providers to adopt digital safety by design. Indeed, the CoE Guidelines go so far as to note that governments should offer incentives for businesses to adopt safety by design as a foundational principle for products and services that may be used by or targeted at children.

Despite the variations in how these common elements are defined or described in the different guidance documents, they nonetheless provide a good basis of the important factors to be applied or considered when implementing digital safety by design for children.

4 Existing and emerging laws and policies on digital safety encompassing children

Several governments have developed legislation and/or policies requiring that online safety considerations be at the centre of product and service development. At least one jurisdiction has clearly elaborated safety by design principles (alongside legislative requirements), whereas other jurisdictions have laws with safety by design requirements but not clearly elaborated principles. This section provides selected[9] examples of domestic and regional laws that include safety by design features, including those that feature safety by design aspects explicitly for children.

Australia's Online Safety Act, whilst not child specific, requires businesses to implement a number of different safety features and considerations (Australian Parliament, 2021[57]) (Australian eSafety Commissioner, 2021[58]). For example, mandatory industry codes create a set of minimum and enforceable commitments for digital service providers. Under the Act, industry bodies are required to develop codes (in concert with the eSafety Commissioner) to regulate illegal and restricted online material, with the codes then registered with and enforced by the eSafety Commissioner. For the codes to be registered, the eSafety Commissioner must agree that they meet the requirements set out in the Act. Should the codes not meet those statutory requirements, the eSafety Commissioner is empowered to impose industry-wide standards.

A number of industry codes came into effect in 2023, including for social media services, app distribution services, and hosting services (Australian eSafety Commissioner, 2024[59]). The code for social media services includes provisions requiring the services to integrate safety by design features and settings that can mitigate risks, including for children. Minimum requirements include, for example, making clear in terms and conditions the minimum age of users permitted to hold an account on the service, and taking reasonable steps to prevent or remove the accounts of children that are under the minimum age (Australian eSafety Commissioner, 2023[60]). The eSafety Commissioner is empowered to receive complaints from citizens to identify potential non-compliance with the codes, investigate potential breaches, direct online services to comply, and take enforcement action (Australian eSafety Commissioner, 2023[61]).

The Online Safety Act includes Basic Online Safety Expectations (BOSE), seeking to ensure online services are safe to use, and to encourage the technology industry to be transparent about their safety features, practices and policies. The BOSE apply to a variety of services and establish a benchmark for online service providers to be proactive in how they protect users, including children, from harmful content and abusive conduct. They require online service providers to minimise abuse, bullying and any other harmful content and activity. The eSafety Commissioner has the power to order online service providers to report how they are meeting any or all of the BOSE and to issue penalties for those that do not meet reporting obligations[10].

The Online Safety Act requires that online service providers have clear and easy-to-follow rules for users to submit complaints about unacceptable use, including cyberbullying targeted at children. If there is

reason to believe that a child was or is the target of cyberbullying, then either the child or an adult on behalf of the child may make a complaint to the eSafety Commissioner. The eSafety Commissioner can investigate the issue and require the online service provider to remove the content[11] from the full range of online services, for example social media platforms, games, websites and messaging services (Australian Parliament, 2021[57]). The eSafety Commissioner has also developed Safety by Design principles and a vision for young people (summarised in Box 4.1).

Box 4.1. The Australian eSafety Commissioner's Safety by Design principles and vision for young people

The eSafety Commissioner's Safety by Design framework provides digital service providers with a set of actionable measures to safeguard citizens online. The Safety by Design framework is a broad programme of resources and support that helps organisations to embed the rights of users and user safety into the functionality of services and products. The Safety by Design principles provide guidance in incorporating, assessing and enhancing user safety considerations throughout the development, design and deployment phases of a typical service lifecycle. The principles position user safety as a fundamental design principle to be embedded in the development of technological innovations from the beginning.

The principles include:

i) service provider responsibility (e.g. nominate experts and make them accountable for user safety policy creation, evaluation, implementation and operations; put in place infrastructure that supports internal and external reporting on all user safety concerns);

ii) user empowerment and autonomy (e.g. provide technical tools and measures that adequately allow users to manage their safety; evaluate all design and function features to ensure risks factors have been mitigated for users); and

iii) transparency and accountability (e.g. ensure that user safety policies, terms and conditions, community guidelines and processes about user safety are accessible, easy to find and understand, and regularly updated; commit to consistently innovate and invest in safety enhancing technologies).

Alongside the development of the Safety by Design principles, the eSafety Commissioner asked young people to prepare a vision statement that lays out their expectations on online safety and how they wish to be supported from the technology industry to ensure that they can navigate in the digital environment safely. The vision statement prioritises the following areas:

- empowering users by providing them greater control of their safety and experiences in the digital environment;
- providing clear guidance and rules that are easy to comprehend;
- providing users with safety features and tools;
- imposing sanctions and consequences in case there is a violation of the rules of the site; and
- using filtering technology and scanning to ensure user safety is upheld on the site and that users are not exposed to sensitive or inappropriate content.

Source: (Australian eSafety Commissioner, 2019[62])

In the EU, the Digital Services Act contains a number of digital safety by design elements for children (European Commission, 2022[63]). For instance, Article 14 requires providers of intermediary services primarily directed at children or used predominantly by children to make efforts to render the explanation

of their terms and conditions easily comprehensible to minors. Article 28 requires online platforms (such as social networks, online marketplaces, and app stores) to take appropriate and proportionate measures to protect children's safety, security and privacy.

Articles 34 and 35 set out a risk assessment and mitigation framework for very large online platforms and online search engines (these are search engines and online platforms with over 45 million users in the EU). Very large online platforms are required to identify and assess risks taking into account the severity and probability of harm, and consequently consider the protection of children. They are required to take targeted measures to protect children's rights, to assure age, and deploy tools that seek to support children by helping them to signal abuse or obtain support as well as parental control tools. The Act also requires that digital service providers do not present advertisements based on profiling when the user is a child. Additionally, the European Audiovisual Media Services Directive includes requirements for video-sharing platforms to establish and operate age verification systems and to take appropriate measures (inclusive of age assurance) to protect children from "programmes, user-generated videos, and audiovisual commercial communications which may impair their physical, mental or moral development" (European Parliament, 2018[64]).

At the policy level, the EU's Declaration on Digital Rights and Principles for the Digital Decade includes principles on the protection and empowerment of children in the digital environment (European Commission, 2022[65]). For instance, the Declaration calls for age-appropriate materials that improve the experiences, well-being and participation of children in the digital environment and notes that children should be empowered to make safe and informed choices and express their creativity online. The European Commission also encourages online platforms in the EU to consider best practices and guidance to ensure digital safety by design for children. These include the recommendations of the European Commission in "A Digital Decade for children and youth: the new European strategy for a better internet for kids" (the strategy is described in Box 4.2). Under this strategy, the EU established a multi-stakeholder group to develop a code of conduct on age-appropriate design (European Commission, 2023[66]).

Box 4.2. The EU's Better Internet for Kids (BIK+) Strategy

In 2022, the European Commission adopted the Better Internet for Kids (BIK+) Strategy to improve the age-appropriateness of digital services and ensure children are empowered, protected and respected in the digital environment. The Commission also developed a child-friendly version of the strategy.

The BIK+ Strategy proposes actions around three main pillars. First, it seeks to protect children from harmful and illegal content, conduct, contacts and consumer risks and enhance their well-being online through an age-appropriate, safe digital environment that respects their best interests. Second, it promotes the digital empowerment of children by supporting them in acquiring the necessary skills and competences to express themselves responsibly and safely in the digital environment and to make sound choices. Third, it encourages participation, including by promoting more child-led activities to foster creative and innovative digital experiences.

The BIK+ Strategy identifies digital safety by design for children as an essential component in ensuring a safe and beneficial digital environment for children, noting industry has a responsibility to ensure that the products it creates are safe for children by default and design.

The Strategy recommends all digital products and services likely to be used by children respect fair and basic design features that are in accordance with the Digital Services Act. For instance, the Strategy notes all products and services likely to be used by children should have age-appropriate, easily accessible and understandable information, such as instructions, warnings, and simple mechanisms to report harm. The Strategy recommends that all parties involved in digital design should understand the

potential harmful effects of design and development choices on children, and the possible risks and harms, for instance grooming[12], that may arise from children's use of various digital services.

Source: (European Commission, 2022[67]) (European Commission, 2022[68])

In 2023, the French Senate passed a law requiring social media platforms to implement age verification systems and seek explicit parental consent for users aged 15 and below (Sénat, 2023[69]). The law seeks to reduce the screen time of children, protect them from cyberbullying and safeguard them from harmful content. Accordingly, social media platforms are required to install functions that limit children's usage time and give parents the right to suspend accounts belonging to children under the age of 15. In 2022, France established the Children Online Protection Laboratory to enhance the safety of children in the digital environment (described in Box 4.3).

Box 4.3. The Children Online Protection Laboratory

France set up the Children Online Protection Laboratory to enhance the safety of children in the digital environment. The initiative involves researchers, online platforms and advocates who work together on issues such as digital literacy, harassment, privacy protection, transparency and moderation with a specific focus on gender-based risks.

Specific projects that the participants worked on during the Laboratory's first year include developing Artificial Intelligence technology to detect conversations in which sexual offenders pose as children; a trusted third party system to estimate the age of users of the digital environment; and a shared database amongst platforms to identify and remove non-consensual intimate images of children online.

Source: (Elysee, 2022[70]) (Kayali, 2022[71])

In Ireland, the Online Safety and Media Regulation Act empowers the Coimisiún na Meán to establish and enforce a regulatory framework for online safety, including for children (House of the Oireachtas, 2022[72]). In particular, it seeks to ensure that children are protected from age-inappropriate and harmful content online. In addition, the Commissioner is developing an Online Safety Code with a safety by design focus and a child-centred approach, in accordance with Article 24 of the EU Charter of Fundamental Rights (which sets out the rights of the child) and Article 3 of the UN Convention on the Rights of the Child (which states that the best interest of the child shall be a primary consideration in all actions concerning children) (European Union Agency for Fundamental Rights, 2000[73]) (United Nations, 1989[74]). Measures in the Online Safety Code are anticipated to include requirements that companies establish mechanisms for flagging and reporting harmful content, age verification systems, and parental control systems, and promote the development of media literacy measures and tools. The Code will be enforced by the Online Safety Commission, which will be able to impose fines on digital service providers of up to ten percent of relevant turnover or twenty million euros (whichever is the greater) and prosecute certain offences under the Act (House of the Oireachtas, 2022[72]).

Türkiye's Law on the Regulation of Internet Publications and Combatting Crimes Committed through Such Publications (Law No. 5651) (Government of Türkiye, 2015[75]) regulates the responsibilities and liabilities of content providers, hosting providers, access providers and public use providers (Articles 1, 2). The law sets out provisions aimed at combatting the spread of illegal content on the Internet. It includes measures for blocking access to and removing such content and establishes penalty should hosting or access providers fail to comply with a blocking or removal request (Articles 8, 9). One initiative in place to support the implementation of the law is Türkiye's "Safe Internet Service" is a filtering system that works in

conjunction with ISPs to block access to websites containing illegal content (Türkiye Information Technologies and Communications Authority, 2018[76]).

In the United Kingdom, the Online Safety Act (UK Parliament, 2023[77]) sets out several digital safety measures for children, including a requirement for user-to-user[13] and search services[14] likely to be accessed by children to carry out a suitable and sufficient children's risk assessment and take or use additional safety measures for child users to mitigate identified risks. User-to-user services likely to be accessed by children have a duty to use proportionate systems and processes *designed* to prevent children from encountering "primary priority" content. Primary priority content is content which children of all ages must be prevented from encountering, and covers pornography and content that encourages, promotes or provides instructions for suicide, self-harm, or eating disorders. User-to-user services must also protect children in age groups judged to be at risk from "priority" content that is harmful to children, including bullying and content that encourages or depicts serious violence.

Search services must take proportionate steps to minimise the risk of all children encountering search content that is harmful to children. When carrying out their children risk assessment, user-to-user and search services must assess the risk of harm to children stemming from the design of a service, for example its features, functionalities and algorithms, and must then mitigate and manage the risk to children they have identified under the child safety duties.

The UK Online Safety Act requires user-to-user services to use highly effective age verification or age estimation (or both) to prevent children of any age from encountering "primary priority" content if they allow, or do not prohibit, such content for all users of their service. Separately, providers of pornographic content[15] must also use highly effective age verification or estimation to prevent children from accessing such content. Alongside guidance for services on undertaking risk assessments, the UK's online safety regulator, Ofcom, will be required to produce codes of practice and guidance on the types of measures that are highly effective in determining whether a user is a child, taking into account clear principles for the use of age assurance measures in the Act. The principles include privacy, efficacy, accuracy, inclusiveness, proportionality and interoperability. The Act requires user-to-user services to put in place clear and accessible terms of services, explain how children of any age are prevented from encountering harmful content, and set out the policies and processes for user redress. The Act requires relevant online providers publish annual transparency reports to help users understand the steps digital service providers are taking to keep users, including children, safe. Services will need to ensure that parents are able to easily access reporting mechanisms and report instances of harmful content and behaviour where their child is either a user of the service or the subject of the content. Reporting mechanisms also need to be easy to navigate for child users.

Outside of the OECD, other jurisdictions have enacted or are proposing laws that focus on online safety and incorporate the needs of children on countries outside the OECD. For instance, a speaker at the Roundtable noted that in Brazil there is legislative debate regarding adopting a law to regulate digital service providers that would include provisions to protect the best interests of children (OECD, 2023[9]). In Fiji, the Online Safety Act protects children from harms such as cyberbullying, cyberstalking, and exposure to offensive or harmful content (Parliament of the Republic of Fiji, 2018[78]). In Romania, the Audiovisual Law (transposing the EU's Audio Visual Media Services Directive) includes provisions requiring video-sharing platforms to put in place age assurance systems, and transparent and user-friendly mechanisms allowing the reporting of content that may negatively affect children (CNA, 2009[79]). In Singapore, the Infocomm Media Development Authority issued a Code of Practice for Online Safety requiring designated social media services to put in place systems and processes to mitigate risks from harmful content to Singaporean users, especially children. It includes requirements for children's enhanced protection, such as differentiated accounts with more restrictive settings that are age-appropriate by default. (IMDA, 2023[80]).

Table 4.1 provides an overview of select elements from the laws in this section.

Table 4.1. Overview of selected laws from OECD countries and the European Union on digital safety encompassing children

Jurisdiction	Title of the law / Year	Specific digital safety law	Law that includes digital safety elements	Examples of child specific provisions	Other key provisions
Australia	Online Safety Act (2021)	√		- Requires clear mechanisms for reporting online harm and requires online platforms to be responsive to these reports, especially when they concern children. - Requires online service providers to prevent access by children to harmful material. - Complaints service for Australian children who experience serious cyberbullying. Under the scheme, eSafety can investigate complaints about serious cyberbullying material targeting an Australian child and require its removal. eSafety can require content be taken down from the full range of online services, for example social media platforms, games, websites and messaging services.	- The BOSE outline the Australian Government's expectations that social media, messaging and gaming service providers and other apps and websites will take reasonable steps to keep Australians safe. The Online Safety Act provides powers to require reporting on the implementation of the BOSE. - Industry codes and standards that provide a common set of objectives and outcomes, while granting the flexibility to implement measures to meet those objectives and outcomes that are most suited to business models and technologies.
European Union	Digital Services Act (2022)		√	- Requires providers of intermediary services directed towards or used by children to render the explanation of their terms easily comprehensible to children. - Requires online platforms to take measures to protect children's safety, security, and privacy. - Requires online service providers not to present advertisements based on profiling when the user is a child. - Requires very large online platforms to take into consideration the protection of children in risk assessment and mitigation. - Requires very large online platforms to take targeted measures to protect the rights of the child, including age verification.	Platforms must publish transparency reports on content moderation decisions and risk management.
France	Law for Securing and Regulating the Digital Space (2023)		√	Requires social media platforms to implement age verification systems; seek explicit parental consent for users aged 15 and below; give parents the right to suspend children's accounts; and install functions limiting children's screen time.	
Ireland	Online Safety and Media Regulation Act (2023)	√		Provisions to ensure that children are protected from age-inappropriate and harmful content online.	

Jurisdiction	Title of the law / Year	Specific digital safety law	Law that includes digital safety elements	Examples of child specific provisions	Other key provisions
United Kingdom	Online Safety Act (2023)	√		Services (in scope of the Act) that are likely to be accessed by children must carry out children's risk assessments to assess the nature and level of risk of their service specifically for children. Under the child safety duties, these services must then: – protect children from online harms; – design proportionate systems that protect children from encountering harmful content; – mitigate and manage the risk of harm to children from features, functionalities or algorithms enabled by the design or operation of the service; – use highly effective age verification and/or age estimation to prevent children from encountering "primary priority" content, such as pornography; and – make terms of service easily accessible for children.	- Relevant online services are required to publish annual transparency reports.

Whilst the requirements of the laws as well as the mechanisms for enforcing them vary, a number of common themes are emerging. For instance, all the jurisdictions above (noting that France and Ireland are encompassed by the DSA) have transparency reporting requirements, although the mechanisms for them vary. Meanwhile, in Australia the Commissioner can request information from specific services, and other jurisdictions require regular reporting as standard procedure.

Codes of practice are a way to work with online providers to set minimum commitments, whilst making those commitments enforceable. These codes typically set out measures that companies can take to address illegal and harmful content, such as child exploitation and abuse material or age-inappropriate content. Whilst Australia and Ireland follow this model, other jurisdictions such as the EU and the UK have included requirements in their laws that safety mechanisms be implemented by providers of digital services and products. In the UK, services in scope of the Online Safety Act are required to implement measures in their design and operation to reduce and address the risks of harm to children and establish appropriate systems and processes to prevent children from accessing content deemed as "primary priority" like pornography. When determining such measures, user-to-user services must take into account functionalities that present higher levels of risk to children, including functionalities which allow adults to search for and contact other users of the service. Codes, however, also play a role in the UK framework, as under the Online Safety Act companies will need to follow the steps in Ofcom's risk-based codes of practice[16] or show their approach is equally effective. However, across the different schemes, the different commitments and requirements are enforceable and come with fines and penalties should they not be respected.

A common theme is requiring child-friendly mechanisms, such as providing information in child-friendly language or establishing accessible complaint mechanisms. Whilst most of the jurisdictions focus on ensuring products and services for children are age-appropriate, only two so far (France and the UK)

establish binding requirements for age assurance measures. Such requirements are, however, under consideration in other jurisdictions, including in Australia (Australian eSafety Commissioner, 2022[81])

5 Dedicated regulatory agencies focusing on digital safety, including for children

A growing number of regulators focus on digital safety, including for children. These regulators have the power to require digital service providers to remove content, issue transparency reports, and deploy technical solutions that protect users, including children, from harm. The duties of the regulators extend to awareness raising, community education, handling complaints and co-ordinating with national education bodies. Table 5.1 provides an overview of the key responsibilities and powers of selected regulatory bodies that focus on digital safety, including for children[17].

Table 5.1. Examples of selected regulatory agencies focusing on digital safety

Country and Regulatory Body	Key Digital Safety Responsibilities	Key Powers (Year power was granted)	Child-Specific Responsibilities
Australia **eSafety Commissioner**	Promote online safety for Australians. Provide information to Australians on the services the e-Safety Commissioner provides. Provide and distribute information to Australians and partners on critical online safety issues and trends. Coordinate the activities of government departments, authorities and agencies relating to online safety for Australians. Administer statutory schemes to respond to complaints and conduct regulatory investigations into child cyberbullying material, adult cyber abuse material, image-based abuse, and illegal and restricted online content. Ensure services (social media, messaging, gaming, file sharing, other app providers) and sites accessible from Australia take reasonable steps to keep Australians	Power to seek information from services, issue formal warnings, infringement notices, accepting enforceable undertakings, seeking court-ordered injunction and civil penalties **(2021)**.	Administer statutory schemes to respond to complaints and conduct regulatory investigations into child cyberbullying material, and illegal and restricted online content. Require social media services, messaging services, gaming services providers to report on how they are meeting any or all of the BOSE. Register mandatory industry codes requiring eight sectors of the digital industry to regulate harmful online content, including videos depicting the sexual abuse of children, and material which is inappropriate for children, such as online pornography.

Country and Regulatory Body	Key Digital Safety Responsibilities	Key Powers (Year power was granted)	Child-Specific Responsibilities
	users safe online through a set of Basic Online Safety Expectations. Register mandatory industry codes requiring eight sectors of the digital industry to regulate harmful online content, such as videos depicting the sexual abuse of children or terrorism, through to material which is inappropriate for children, such as online pornography. If these codes do not meet community expectations, eSafety can draft standards for the industry. Make financial grants on behalf of the government to foster online safety for Australians. Conduct and evaluate research about online safety for Australians.		
Ireland **Coimisiún na Meán**	Develop an Online Safety Code which sets out specific requirements for online platforms to ensure the online safety of users. Enforce the EU Digital Services Act setting out rules for online platforms, including requirements about how platforms handle complaints about harmful and illegal content.	Power to carry out investigations, impose sanctions on persons for failure to comply with obligations, impose administrative financial sanctions, impose levies **(2022)**.	Develop an Online Safety Code which, *inter alia*, sets out actions that online platforms must take to protect minors from harmful video content and from criminal offences, including those related to child sex abuse material. Enforce the EU Digital Services Act that sets out rules for online platforms about the protection of children online.
Korea **Korea Communications Standards Commission**	Monitor and regulate online content to ensure it adheres to Korean laws and standards. Handle public complaints about online content. Collaborate with other regulatory bodies to address online safety issues.	Imposition of punitive penalties, fines **(2010)**.	Ensure the safety and protection of children in the digital environment. Regulate and control the dissemination of online content that may be harmful for children.
UK **Ofcom**	Regulate online platforms to ensure they have proportionate measures in place to protect users from harm, including systems and processes to address illegal content and promote child safety online. Provide resources to companies to manage risks by i) analysing the causes and impacts of online harm, to support services carrying out risk assessments; ii) providing guidance on a recommended	Power to request information, impose fines, and seek court orders imposing business disruption measures **(2023)**.	Regulate online platforms likely to be accessed by children, to ensure they put in place proportionate measures to protect children from harmful and age-inappropriate content and activity, such as pornography, bullying and content that promotes eating disorders, suicide or self-harm. Produce: a register of risks to children; codes of practice setting out recommended measures to protect children

Country and Regulatory Body	Key Digital Safety Responsibilities	Key Powers (Year power was granted)	Child-Specific Responsibilities
	risk assessment process; iii) providing codes of practice, setting out what services can do to mitigate the risk of harm. Drive industry improvements by engaging with the largest and riskiest services through continuous regulatory supervision. Assure compliance by using investigative and enforcement powers.		online; and guidance to support online services to comply with the child safety duties (including: guidance on children's access assessments to determine if services are likely to be accessed by children, on content harmful to children, and to services on carrying out risk assessments.

Sources: (Australian eSafety Commissioner, 2022[82]) (Coimisiún na Meán, 2023[83]) (Ofcom, 2021[84]; Korea Communications Standards Commission, 2023[85]; Ofcom, 2023[86])

6 Key components of digital safety by design for children

Digital safety by design for children is unlikely to be achieved by just one action, instead requiring several complementary and overlapping actions that, when combined, can help to ensure a safe and beneficial digital environment for children. This section proposes key components of digital safety by design for children. They have been identified and elaborated on based on analytical work associated with the development and implementation of the OECD Recommendation (OECD, 2022[15]), key points made at the Roundtable (OECD, 2023[9]), consultations with the informal group of experts on children in the digital environment and feedback received from delegates to the Digital Policy Committee.

The key components themselves, whilst all ultimately aimed at advancing digital safety by design for children, have slightly different functions. They can be grouped into:

- practical tools and design features/decisions (i.e. age assurance, child-centred design, privacy protections, and technical tools to prevent harm);
- measures to create an environment that promotes and prioritises safety (i.e. child participation, child-friendly information provision, and promoting a culture of safety and well-being); and
- safety nets designed to mitigate or address harm should it nonetheless occur (i.e. complaint mechanisms and technical tools to detect harm).

Some of the components may overlap these different functions. For example, child-friendly information provision is likely to be central to a) creating an environment where children feel comfortable and safe, including through b) helping children to understand the different technical tools and mechanisms in place to protect them; but is also important in c) enhancing children's capacity to easily file a complaint should it become necessary.

Employing age assurance mechanisms

Establishing age is important for digital safety by design, as services are only going to be able to protect children from unsafe content and behaviours if they know which of their users are actually children (OECD, 2023[9]). Age assurance is an umbrella term used to describe the different processes to i) make sure that children (or certain age cohorts) are unable to access adult, harmful or otherwise inappropriate online content or digital services; and ii) estimate or verify the age of a user so that digital services can be tailored to their needs and age-appropriate safeguards put in place (United Kingdom Information Commissioners Office, 2024[87]).

At the Roundtable, experts made several observations regarding the underlying conceptual priorities for age assurance, including that:

- Age assurance should not be used only to bar children from adult services, but as far as possible should be used to deliver age-appropriate experiences.
- Age assurance mechanisms should be equitable, inclusive and reflect risk.

- Laws and regulations that require age assurance should be tech neutral, leave room for innovation, and not be overly prescriptive.
- Solutions should be principles-based, for example focusing on accuracy, usability, privacy and proportionality.

Age limits have already existed in various laws related to online content or activities (i.e. data protection, or age restricted goods and services) for some time. However, research has observed that barriers to children's access to age restricted online content, goods and services are usually ineffective and children will often find workaround strategies (Smirnova, Livingstone and Stoilova, 2021[88]). There is often a lack of public awareness regarding the purpose of age assurance. For instance, whilst age assurance is important for enforcing hard age limits (such as barring children's access to pornography or other products or services also prohibited offline), it is also used to deliver age-appropriate experiences. For example, on those services which have a mixed audience – such as social media – assigning the correct user age to accounts is necessary for putting in place safeguards like restricting adult content or direct messaging to child users accounts (Ofcom, 2022[89]). Research has shown that children and parents are often unaware that age assurance is used for this purpose, and indeed parents may help children to circumvent age assurance measures (Revealing Reality, 2022[90]). At the same time, age limits and age assurance for the purpose of enforcing these limits will only be truly effective if digital products and services (including their different age-tiered functionalities) are correctly labelled as appropriate/inappropriate for children or children of certain age ranges.

Age assurance takes several different forms but can be broadly divided into technologies or mechanisms that estimate a user's age or age range ("age estimation"); and those that verify the user's actual age ("age verification") (5Rights Foundation, 2021[91]) (Australian eSafety Commissioner, 2022[81]). Another process is verified parental consent, which is often used as part of privacy protection frameworks[18].

Various age assurance methods are already used across a wide range of services and platforms, such as social media, age restricted websites, video streaming services and online games (Revealing Reality, 2022[92]). The most basic age assurance mechanism is a simple self-declaration asking users to enter their date of birth or to tick a box indicating that they meet the minimum age. However, mere self-declaration of age is often not regarded as an effective age assurance technique as it can be easily misused (Comisiun na Mean, 2023[93]), and indeed the UK's Online Safety Act specifies that self-declaration alone is not to be regarded as age assurance (United Kingdom Government, 2023[94]). Age verification processes use verified sources of identification, such as a government-issued identity card or a passport, that display a date of birth, or rely on other verified sources of identification like a credit card. Biometric data, such as voice or facial features, can be used to estimate a user's age. Other approaches to age assurance, include profiling and inference models[19], capacity testing[20], and tokenised age checking using publicly held health or education data[21] (5Rights Foundation, 2021[91]).

Ongoing debates concern the accuracy and efficacy of the different age assurance methods, the ease of adoption for users (particularly children), the extent to which they can preserve privacy, as well as how best to balance children's developing need for autonomy with any necessary parental involvement in assuring age. Additionally, as legislation develops and different jurisdictions require different mechanisms, there may be ambiguity for online services and platforms as to the appropriate age assurance mechanisms that they are required to employ (Revealing Reality, 2022[92]).

In particular, protecting the privacy of users adds a layer of complexity to age assurance, as certain age assurance systems need to collect significant data, such as biometrics or ID documents, to be effective. To mitigate privacy risks, age assurance solutions should incorporate robust privacy protections, such as principles of data minimisation to collect and retain the minimal amount of data required. Another concern is the risk that certain age assurance methods may be biased against specific ethnic groups or be inaccessible to marginalised communities who lack digital access or access to official identification documents[22]. The combined use of a selection of age assurance methods (sometimes called "waterfall

techniques"[23]) may provide a higher level of confidence of effectiveness and privacy and reduce the risk of discrimination than methods used in isolation (ICO, 2024[95]).

Three different age assurance methods were presented at the Roundtable, and these are set out in Box 6.1.

Box 6.1. Examples of age assurance methods

To address the tension between effective age assurance solutions and minimising data collection and use, the French privacy enforcement authority (CNIL) is developing an age assurance technology that deploys a double-anonymity system. It can provide verified age information to a service or website without the age verification provider knowing what the service or website in question is.

Yoti provides age assurance services, such as facial age estimation, which determines someone's age from a live facial image. It converts the image into numbers and compares the patterns of numbers to patterns associated with known ages. It provides an age estimate and a confidence value. This privacy-preserving technology only estimates age, rather than identifying or authenticating the individual. As soon as someone's age is estimated, the facial image is deleted.

Another age assurance approach, developed by Privately, takes place directly on the user's device. It uses machine learning technology that checks voice and facial patterns to estimate age. This technology aims to ensure that no personally identifiable information leaves the device.

Sources: (CNIL, 2022[96]) (Information Commissioner's Office, 2021[97]) (Information Commissioner's Office, 2021[98]) (OECD, 2023[9])

Experts at the Roundtable agreed implementing age assurance requires cross-sectoral and international co-operation. Given the data collection implications of some age assurance technologies (for instance, scanning ID documents, biometrics, and the need to collect data from adults as well as children), safety and privacy regulators will need to work closely together on age assurance. There is a clear need for international collaboration on approaches to age assurance that incorporate different cultural, legal and language considerations. Such collaboration is necessary to avoid the onset of incompatible and/or inconsistent regimes across different jurisdictions, and to ensure that all children have the same levels of protection and access (OECD, 2023[9]).

Considerations for implementing child-centred design

Child-centred design is about enabling digital experiences that permit children to be creative, help them to regulate their emotions, build their confidence, give them a sense of purpose and empowerment, can facilitate social connection and can bring them joy (OECD, 2023[9]). Should digital spaces not be designed in a way that holistically considers children's rights and needs and rather focusses only on preventing harm, there is a risk that children could be excluded from digital spaces that advance their development and well-being. In other words, when digital spaces are focused only on preventing harms and not at all on positive opportunities, children may feel trapped and decide to experiment with digital spaces that may not be safe for them (OECD, 2023[9]).

Designing with children's safety in mind does not mean designing for an unrealistically homogenous group consisting of all children, but rather ensuring that digital spaces are age appropriate (IEEE Consumer Technology Society, 2021[99]) (OECD, 2023[9]). This requires recognising that what may be safe and

developmentally apt for a 16-year-old will likely be inappropriate for an 8-year-old, and indeed there is a clear correlation between the level of children's digital skills and their age (Cetic.br, n.d.[100]).

Children have unique and growing brains, and the different stages of development may be affected by some facets of different digital services they use, particularly services that were designed with adults in mind but have had significant uptake by children (e.g. some social media) (American Psychological Association, 2023[101]). Social media platforms have a variety of built-in features that trigger neural responses, giving not just the opportunity for peer reinforcement, but quantitative metrics that tell children how well they are doing in seizing those opportunities. As a result, between the ages of approximately 10 (when the brain makes adolescents hypersensitive to peer interaction and feedback) and 25 (when the brain's capacity to resist pursuing impulses for peer feedback is usually fully developed), social media may be providing opportunity but also creating risk. In light of this, in May 2023, the American Psychological Association, together with an expert advisory panel, published guidelines on adolescent use of social media. This includes recommendations regarding limiting adolescents' access to certain features that adversely affect them (e.g. the like button, recommended content, unrestricted time limits, endless scrolling) as well as recommendations on social media literacy skills for children (American Psychological Association, 2023[101]) (OECD, 2023[9]).

Designing with children's safety in mind also means considering the socio-economic differences between children, how these differences can exacerbate vulnerabilities and what (if any) particular protections may need to be put in place for children from different regions or with different characteristics. For instance, systemic concerns such as inequalities and poverty as well as discrimination against children with ethnic or cultural minority backgrounds make children more vulnerable to negative online experiences such as grooming and cyberbullying (OECD, 2019[102]). Children with disabilities are also more likely to encounter more risks in the digital environment (Livingstone and Palmer, 2012[103]) (OECD, 2023[104]).

Child centred design should be underpinned by a comprehensive evidence base. Experts at the Roundtable highlighted that whilst seemingly there is a lack of data on outcomes for children in the digital environment, in reality many of the key actors hold their own data sets. International collaboration on data sharing is needed, as well as access to that data as an evidence base for regulators and policy makers, to enable better policies. Different stakeholders need to learn from each other to find the best possible responses and comprehensive research that incorporates children from all regions is essential (OECD, 2023[9]).

An example of a framework for child-centred design is provided by the Responsible Innovation in Technology for Children (RITEC) initiative – a collaboration between the private sector, academia and UNICEF (see Box 6.2).

Box 6.2. The RITEC framework

RITEC provides an interim framework mapping how the design of children's digital play experiences affects their well-being, and offering guidance on how well-informed design choices can advance positive well-being outcomes.

The framework was developed with participatory workshops involving over 300 children from 13 countries, combined with the analysis of international datasets on over 34 000 children. This analysis yielded information about digital play and children's well-being, and the potential relationship between them, providing the basis of the framework and accompanying indicators.

The framework and the indicators are intended for companies that produce digital services and products that are likely accessed by children, and for governments that seek to promote the well-being of citizens. It sets out aspects that should be prioritised when designing digital experiences for children.

The key elements of the well-being framework to which digital experiences should contribute are:

i) Competence (digital play can support children in their perception of their own competence, ability and knowledge);

ii) Emotional regulation (children's participation in the digital environment can help them relax, adjust their mood, and support them in regaining energy to connect with peers);

iii) Empowerment (children who engage in digital play can have a reinforced sense of autonomy, choice and can feel a sense of control and achievement);

iv) Social connection (children engaging with the digital environment, including during play, can help them form social connections with peers, family or other people);

v) Creativity (children's participation in the digital environment can make them more curious and enhance their creative ability);

vi) Safety and security (when children play online, they should both feel safe and be safe);

vii) Diversity, equity and inclusion (digital play experiences should be diverse, equitable and inclusive in order to include children from all circumstances and backgrounds);

viii) Self-actualisation (children's digital play equips them with a sense of purpose and enhancements in their social connections).

Source: (UNICEF Office of Research - Innocenti, 2022[105])

Preventing and detecting harm

Preventing harm

Different prevention tools and features can be built into digital products or services to minimise the risk of harm occurring. These include default settings, platform level filters, and tools designed to identify risky behaviour and provide diversion pathways.

Default settings are safety parameters that are automatically turned on should a user account be associated with a child. These features can turn off the capacity for adult or unknown users to contact children, limit the visibility of posts, limit what can be seen (such as content designated as adult), and turn off personalised content. It is often the case today, however, that whilst such default settings may be

automatic if a user account is associated with a child, they can be turned off either by the child or a parent (Integrity Institute, 2024[106]).

Platform level filters act to identify harmful content or contacts, and then hide the content or prevent the contact so that the user does not see it. For instance, Instagram's anti-bullying comment filter proactively detects and hides bullying comments. Its direct message filter automatically filters message requests that contain a predefined set of offensive words, phrases and emojis (Instagram, 2021[107]) (Australian Government eSafety Commissioner, 2019[108]). Discord's sensitive content filters automatically blur media that may be sensitive, which is in direct or group messages to teenagers to encourage caution when viewing the media (Discord, 2023[109]). As with the default settings mentioned above, there is capacity for children to opt out off or turn off of these filters.

Other prevention tools include those which seek to divert users from engaging in harmful actions. For example, Koko is an automated AI-powered conversational tool that can be deployed on messaging platforms, uses a hybrid human-machine moderation system to identify users displaying suicidal intent or acute distress, and directs the user to crisis lifelines or peer support networks (Koko, 2022[110]) (Australian Government eSafety Commissioner, 2019[108]). Some diversion measures are not necessarily aimed directly at children, but rather at those who would harm them. For instance, one adult website in partnership with civil society organisations working to combat child sexual exploitation and abuse (CSEA) deployed a warning message when a user searched a term associated with CSEA notifying the user that the kind of content searched for is illegal. A chatbot was also deployed pointing people to a helpline. The analysis of this pilot program identified that the chatbot and warning message played a role in a meaningful decrease in searches for CSEA on the website (Internet Watch Foundation, 2023[111]).

Other safety tech initiatives are emerging that seek to prevent harm. For instance, PlaySafeID for Kids is a tool which would generate an anonymous and unique ID for gamers, which could then be associated with that users ID on different gaming platforms. Should a user with a PlaySafeID break the terms and conditions on any platforms using the tool (including being inappropriate to children online) their ID would then be blacklisted, preventing them from ever obtaining another PlaySafeID or interacting with another PlaySafeID user (PlaySafeID, 2024[112]).

Detecting harm

Other tools and measures are more reactive, serving to detect existing harmful or illegal content. For example, trusted flaggers are specific networks of people who, when they flag violative content, should have their report prioritised. The EU's Digital Services Act establishes a trusted flaggers framework and authorises Digital Services Coordinators to appoint flaggers who demonstrate expertise and independence (European Union, 2022[113]). EU co-funded safer Internet helplines and hotlines assist the public, in particular children, when confronted with harmful and illegal content. If granted the status of "trusted flagger" under the Digital Services Act, the helpline/hotline staff will be able to contribute to a swifter assessment of and action upon notifications of illegal content online (European Commission, 2022[114]). Certain digital service providers also have their own trusted flagger programme, such as YouTube's trusted flagger programme providing specific tools and pathways for approved government and non-government organisations to flag violative content. This programme establishes a specific reporting form and pathway for these partners, prioritises their reports, and facilitates training and discussion between the organisations and the company (YouTube, 2023[115]).

Most prominently, digital service providers use technical tools to detect and report CSEA on their platforms. For example, a 2021 survey by WeProtect Global Alliance of 32 technology companies found that 87% of those companies use image hash-based tools[24] and 76% used video hash-based detection tools to detect CSEA. Companies may also use advanced detection measures, such as artificial intelligence classifiers (WeProtect Global Alliance, 2021[116]). Another example is Lantern, a child safety cross-platform signal sharing programme that brings together technology companies to securely and responsibly share signals

about activity and accounts that violate their CSEA policies. Signals can be, for example, information tied to policy-violating accounts like email addresses, usernames, CSEA hashes, or keywords used to groom as well as buy and sell child sexual abuse material. Signals are not definitive proof of abuse, but rather offer clues for further investigation and can enable digital service providers to uncover a real-time threat to a child's safety (Litton, 2023[117]).

In employing the different tools, digital service providers should promote transparency and accountability to make sure they do not create new risks for children and other users of the service, and should pay special attention to compliance with existing regulations.

Protecting children's privacy and personal data

Privacy in the context of safety by design for children involves considering a) how privacy infringements affect children's safety; b) how digital safety by design can be leveraged to safeguard against privacy infringements; and c) how to balance risks to children's privacy that may arise from measures that themselves seek to ensure children's safety online (e.g. age assurance mechanisms that use sensitive or biometric data, or from tools designed to detect or prevent harm).

The OECD's Revised Typology of Risks recognises privacy[25] infringement as a cross cutting risk that can have a wide-ranging impact on children's lives (OECD, 2021[14]). Breaches of children's privacy and the misuse of their data can have a direct impact on a child's safety. For example, algorithms that operate on a child's data traces and digital footprint could promote harmful content (e.g. violent or self-harm material). Unauthorised access to personally identifiable information about children may present overall safety risks. A child's location data or personal information included in their social media user profiles can be used to suggest random contacts and potentially lead to grooming, stalking or harassment (Livingstone and Pothong, 2021[118]) (Digital Futures Commission, 2023[119]).

It is essential that safety tools are privacy preserving and comply with privacy and data protection obligations. For instance, safety tools that monitor the conduct of children online and collect sensitive information about them can pose risks of privacy violations. Careful consideration needs to be given to the balance between measures to promote safety and children's ability to use and enjoy the digital environment without their privacy and autonomy being infringed. For age assurance, several Privacy Enforcement Authorities are turning their attention to safeguards for ensuring that age assurance can also be privacy preserving, for example in France (Commission Nationale de l'Informatique et des Libertés, 2022[120]), Spain (Agencia Española de Protección de Datos, 2023[121]), Singapore (Personal Data Protection Commission Singapore, 2023[122]) and the UK (United Kingdom Information Commissioners Office, 2024[87]).

At the national level, several laws protect children's privacy and an increasing number specifically call for privacy by design for children, requiring digital service providers to take privacy and data protection into account during the first stages of the development process and throughout the entire design process (including with regard to the collection, use and storage of children's personal data) (5Rights Foundation, 2022[123]) (Australian Government eSafety Commissioner, 2019[108]). For instance, the Netherlands' Code for Children's Rights, which consists of ten principles for designers and developers for safeguarding the fundamental rights of children in digital services, sets out that designers and developers of digital services should provide a child-friendly privacy design, should not process more personal data than is strictly necessary, and should make any default settings as privacy-friendly as possible (Ministry of the Interior and Kingdom Relations, 2021[124]). Laws often require that the mechanisms of collecting, using and sharing children's data be transparent and children be provided with clear, simple and easily accessible information regarding the use of their data (Hartung, 2020[125]). In many of these laws, requiring privacy by design is inherently linked with safety by design, two issues which often overlap in the context of children's digital experiences.

Existing or emerging regulatory and policy measures that protect children's privacy include the UK's Children's Code (UK Information Commissioners Office, 2020[126]), France's 'Eight Recommendations to Enhance the Protection of Children Online' (developed by the CNIL) (CNIL, 2021[127]), and the U.S. Children's Online Privacy Protection Act (COPPA) (FTC, 1998[128]). For instance, the UK's Children Code emphasises both privacy by design and safety, ensuring that children's personal data is treated with special care and that online services are designed to prevent potential harms. These regulatory and policy examples are highlighted in Box 6.3.

Box 6.3. Privacy by Design Initiatives

The UK's Children's Code focusses on UK companies' obligations under data protection law to protect children's data. It contains 15 standards that online services need to meet if children are likely to access their service. The standards take a risk-based approach and focus on providing default settings ensuring children have the best possible access to online services, whilst minimising their data collection and use. The standards in the code include:

- ensuring the best interests of the child are a primary consideration in the design and development of services;
- undertaking data protection impact assessments to mitigate the risks to the rights and freedoms of children;
- using an age-appropriate application and ensuring that the standards in the code are applied to child users. When it is not possible to verify if the user is a child, the standards should be applied to all users;
- ensuring that privacy information is transparent and provided in an age-appropriate manner; and
- using "high privacy" settings by default, not disclosing children's data, and switching off both geolocation and profiling options by default, unless there is a compelling reason not to use these default settings, or to disclose children's data.

France's 'Eight Recommendations to Enhance the Protection of Children Online' (developed by the CNIL) include a recommendation on safeguards to protect the interests of the child, encouraging digital service providers to adopt good practices or establish a code of conduct, and to put in place specific safeguards to meet the child's best interests. These safeguards include:

- Strict default privacy settings;
- Deactivation by default of any profiling system for children, particularly when the profiling is for the purposes of targeted advertising; and
- Prevention of the reuse and sharing of children's data for commercial or advertising purposes, unless it can be demonstrated that it is reused or shared for overriding reasons in the best interests of the child.

In the United States, the COPPA Rule requires that data only be retained for the period necessary, and that child's participation in an online activity is not conditioned on them providing more information than is necessary for the activity. In addition, the US' Federal Trade Commission Act, enables the Federal Trade Commission (FTC) to challenge a wide array of deceptive or unfair trade practices. The FTC has used this authority to address unsafe video game features (e.g. on-by-default voice and text chat

features that expose children to bullying, harassment and risk of self-harm and suicide), unnecessary collection and use of sensitive location data, and manipulative website designs, amongst other things. For instance, in December 2022, the FTC brought a complaint against Epic Games (Epic), developer of the popular video game Fortnite. The complaint alleged that Epic violated the COPPA Rule by, amongst other things, failing to provide notice to parents of its practices for collecting and using and disclosing the information it collects from children in Fortnite and failing to obtain parental consent before collecting or using children's personal information.

Source: (UK Information Commissioners Office, 2020[126]) (CNIL, 2021[129]) (FTC, 2022[130]; FTC, 1998[128])

Ensuring child-friendly information provision

Child-friendly information provision helps children to understand the terms of service associated with the various digital products and services they engage with, any risks associated with these products and services, and how to report harms and flag unsafe or illegal content. The OECD Recommendation and the OECD Guidelines call for information (i.e. on risks, privacy rights and complaint mechanisms) to be provided to children in a concise, intelligible, easily accessible manner using clear and age-appropriate language (OECD, 2021[1]) (OECD, 2021[2]). The Committee on the Rights of the Child calls for child-friendly information to be given on complaint mechanisms and children's rights (UN CRC, 2021[49])[26]. Children themselves call for easily understandable information, clarity around rules and procedures, and clear and easily accessible mechanisms for making complaints or flagging concerning content or contacts (Third and Moody, 2021[131]; Meurens et al., 2022[132]).

The Australian eSafety Commissioner's Safety by Design initiative calls for safety policies, terms of service, community standards and processes related to user safety to be in plain language that is comprehensible to younger users (see in Box 4.1) (Australian Government eSafety Commissioner, 2019[108]). Both the UK's Children's Code and France's "Eight Recommendations to Enhance the Protection of Children Online" (both of which centre on privacy by design for children) call for child-friendly language and techniques when providing information to children. For example, the UK's Children's Code requires that information regarding privacy, terms of use, policies, and community standards be "concise, prominent and in clear language suited to the age of the child" (UK Information Commissioners Office, 2020[126]). France's Eight Recommendations specifies that children need to be spoken to in their own language and in a manner that makes them want to pay attention. For example, through using clear, simple and short sentences, providing information only when it is necessary to make a decision, and using interactive tools (i.e. icons, videos or images) (CNIL, 2021[127]).

In addition to ensuring that information regarding actual services is delivered in a clear and age-appropriate way, child-friendly language and child-directed communication methods can also be used in awareness raising campaigns. For example, 5Rights twisted toys campaign (aimed at parents and children) uses mock toys and plain language to highlight risks associated with digital technologies and connected toys (e.g. algorithm facilitated contact with strangers, bullying and harassment, and misuse of data) (5Rights Foundation, n.d.[133]).

Facilitating complaints and redress

Helping children understand how to report incidents and harms, and flag unsafe or illegal content is essential. If complaints mechanisms are hard for children to find, access or use, they will have little utility. Although, ideally, services would be designed to be completely safe for children, that is not always (if ever) achievable, so child-friendly complaint mechanisms are essential.

Box 6.4 briefly explains accessible complaint mechanisms.

Box 6.4. Child-friendly and accessible complaint mechanisms

Providing effective and child-sensitive means for children to make complaints is key to ensure that they can enjoy the benefits of the digital environment. A complaint mechanism is a means through which children can request redress and make the assumed violation of their rights stop. It is therefore an essential tool for correcting mistakes and ensuring that children are protected from harm online. Having accessible complaint mechanisms for children provides a clear recognition that they are full citizens. Such a mechanism can be particularly important for marginalised children who can be at increased risk of having their rights violated.

Complaint mechanisms may seem formal and complex to children, which makes it even more important that digital service providers regularly review the accessibility of their complaint mechanism and make children aware of it. Several elements can make a complaint mechanism accessible and effective for children. To begin with, it is important to provide necessary information and answers to children's questions, to make sure the child is effectively heard and that his/her views are not undermined. Children should be informed about who they are sharing their views with, how the complaint will be used and what privacy protections are in place for the child during the complaint process – all without the process being overly formal. Children should be involved in the development of complaint mechanisms to ensure that their views and experiences are reflected in the complaint mechanism, and that they are user tested. The child's privacy and dignity should be respected throughout the complaint process.

Sources: (UNICEF, 2019[134]) (McIlraith and Harrow, 2018[135])

Beyond accessible mechanisms such as those described above, it may be necessary to put in place systems to accompany children throughout the complaint process. The process may be complex or daunting for children, or the subject of the complaint (e.g. material which is bullying or harassing, or sexual abuse imagery) may be distressing, and the child could need emotional and practical support during the complaint process. One example of a service that aims to support children in reporting digital safety issues is New Zealand's Netsafe service, which helps children to navigate concerns related to their digital safety, such as grooming, bullying, illegal content and scams (Netsafe, 2023[136]). Through Netsafe, children can complete an online form (or call, email, or text Netsafe) which is then reviewed by helpline staff. After any requests for further information or evidence, Netsafe staff provide direct advice to the child, including options for taking the content down, mediation or negotiation with the other party, and/or referrals to other agencies.

The Global Privacy Assembly's Digital Education Working Group has identified several complaint mechanisms which combine the work of data protection authorities and law enforcement. Through co-operation mechanisms, data protection authorities assist children and/or their parents and provide advice on the steps necessary to lodge a complaint. Trained counsellors then guide children towards psychological support and as appropriate, to other competent institutions (Global Privacy Assembly, 2023[137]).

Encouraging child participation and putting children at the centre of decision making

Children have a right to be heard in matters that affect them (OECD, 2022[15]; UN CRC, 2021[49]), and when children (and young people) play a part in helping to develop programs and policies that will affect them, it is more likely that they will be better aligned with children's needs, interests and backgrounds (Cortesi, Hasse and Gasser, 2021[138]). Children need and want digital technologies, and the digital environment should be designed with their rights in mind – including freedom of expression, right to participation, access to information and to play (OECD, 2023[9]). The digital environment can provide community and a sense of belonging to disadvantaged or marginalised children who may lack support and community in the offline world.

The importance of child participation is reflected in the OECD Recommendation, which recognises that children themselves play an essential role in ensuring that the digital environment is both safe and beneficial for them, and sets out a number of recommendations that seek to foster their participation (OECD, 2021[1]). Notably, the OECD Recommendation calls on Actors:

- to uphold and respect the child's right to freely express their views and their ability, as appropriate considering their age and maturity, to participate in matters that affect them in the digital environment (at II.2.d); and
- to engage in multi-stakeholder dialogue, including with children themselves (at II.5.a) (OECD, 2021[1]).

Children are undoubtedly concerned about their safety online. They want to be able to employ strategies to protect themselves, and importantly, have definitive views regarding the manner in which the adults in their lives, policy makers and digital service providers should support them to be safe. One consultation with children[27] highlighted their concerns that digital technologies have significantly amplified the possibility that they will encounter serious forms of violence. The children in the study identified their key concerns to include that they may encounter inappropriate content, be bullied or harassed, suffer sexual exploitation or even that they may be kidnapped or murdered. The latter concerns arising over the possibility that location sharing could mean that they could be tracked, or that they may be tricked into meeting a bad actor. They expressed concerns regarding how children themselves might pose risks to others, the risk that online violence could lead to actual violence (and vice versa), as well as to the mental health impacts of having their safety compromised via the digital environment (Third and Moody, 2021[131]).

Children's right to be heard should be respected and listening to children's views (and acting upon them) is key to designing safe and beneficial digital spaces. Engaging with children regarding their safety online allows for an understanding of what concerns are pertinent for children, how they would like to be supported to address these concerns, and how they currently act themselves to mitigate risk. For example, in the same consultation, children reported that whilst they do take some steps to protect themselves from harms online (e.g. blocking people, not accepting friend requests from strangers, and being careful regarding what information they share) they want – and see as vital – the support of trusted adults (e.g. parents, carers and teachers) in keeping them safe. At the same time, children do not want to be reliant on their parents solely to protect them and want the digital environment itself to be safe and protective, advocating for governments, digital service providers and other relevant stakeholders urgently to put in place greater protections to prevent them from encountering serious harm (Third and Moody, 2021[131]).

At the Roundtable, an expert noted lessons learnt from consulting with children, including that children:

- care about safety in all aspects of their digital engagement, be it for learning, for play or even when they are testing boundaries;
- want to be offered age-appropriate experiences, including those that enhance their creativity, support their development and promote their well-being;

- care about their privacy and data protection, and expect those that handle their data to take only what is necessary, ask their consent for data collection, and to be transparent about how their data is used;
- expect to be treated fairly and inclusively;
- expect to be able to exercise their agency, and do not want to fall victim to manipulative design; and
- expect the digital environment to put their best interests at least on par with the interests of businesses (OECD, 2023[9])

UNICEF provides guidance for businesses when consulting with children setting out that:

- participation should be voluntary, with informed consent from the child and their parents;
- consultations should have a clear purpose and focus on specific issues relevant to children's lives and concerns;
- children's time is precious, and engagements should fit in with their routines;
- child rights stakeholders should also be consulted throughout the process;
- companies should rely on expert third-party facilitation;
- consultations should be part of a wider long-term approach to stakeholder engagement;
- child safeguards and confidentiality should be ensured throughout the process; and
- consultations should be carried out with respect for the cultural practices, beliefs and norms of each community or group (UNICEF, 2014[139]).

Actions of governments provide examples of mechanisms for consulting with children and incorporating their views in decision making. For instance, In Ireland, the Coimisiún na Meán is establishing a Youth Advisory Committee that it intends to consult regularly. France seeks to directly involve children in policy making through the Children's Parliament project, which allows school children to discuss societal issues through democratic debate, including discussing children's use of the Internet (Ministère de l'Éducation Nationale, 2023[140]). The Netherlands' Code for Children's Rights states that children (in a manner reflecting their age and circumstances) must be able to participate in the design and development of digital services that affect them (Ministry of the Interior and Kingdom Relations, 2021[124]). In the UK, the view of children and parents were surveyed during the development of the Children's Code (ICO, 2019[141]). Youth councils may also involve children in a local level (Better Internet for Kids, 2021[142])

The European Commission's Better Internet for Kids Strategy expressly recognises that the active participation of children is necessary to respect them and give them a voice in the digital environment, including through child-led activities to foster innovative and creative safe digital experiences (European Commission, 2022[114]). In Australia, the eSafety Commissioner's Youth Council also provides an example of child participation in practice. Box 6.5 provides an overview of this initiative.

Box 6.5. Australia's eSafety Youth Council

To guide the Australian eSafety Commissioner's Engagement Strategy for Young People, the Commissioner commissioned the Young and Resilient Research Centre to run a Living Lab process with children.

The process used youth-centred, co-design and participatory co-research methods to gather children's insights on digital safety and develop recommendations for eSafety's digital safety messaging and resources, and their outreach to children. The work included designing the process for establishing the

Youth Advisory group, and the development of a draft Aspirational Statement to support the Engagement Strategy for Young People.

Some of the key findings of the Living Lab on digital safety include that:

- Children have high expectations for the Internet and digital safety. Ideally, they expect the digital environment to be inclusive, enabling and youth-centred. They would like to be involved in decision-making in digital safety and to design digital resources that can support children's diverse needs online. Children would like to have independence online, but with an appropriate level of guidance.
- Children's main digital safety concerns are related to the interactions with others, with privacy, and security issues. Cyberbullying is also a key concern. Other risks that children worry about include accessing or being exposed to inappropriate content, sexual exploitation, misinformation and fake news, commercial exploitation and being judged by their peers for their opinions.

Source: (Moody et al., 2021[143])

Promoting a culture of safety and well-being

The role of corporate responsibility, whereby digital service providers can address safety through business culture, standards, codes of conduct and impact assessment, is an important consideration (5Rights Foundation, 2022[144]). Digital safety by design for children should be a key aspect of corporate responsibility, in product system design but also as a company-wide priority – incorporating not only proactive preventative measures, but also remedial measures to address any harm (OECD, 2023[9]). For instance, terms and conditions can reflect the child's best interests (5Rights Foundation, 2022[144]). In addition, standards and codes of practice can include provisions that seek to prevent children from being offered harmful or inappropriate content or contact, to protect children's privacy on the device or system-level, and to address security risks raised by the Internet of Things (5Rights Foundation, 2022[144]).

Awareness raising activities, directed at both the public at large as well as part of corporate responsibility can serve to further safety. For example, user safety considerations can be incorporated in staff training, or steps can be taken to embed user safety considerations into roles, functions, and working practices. Targeted training can be directed at all actors engaged in protecting children (i.e. law enforcement, education and therapeutic professionals) on the specific risks and factors relevant to the digital environment and effective dissemination can help to ensure that awareness-raising activities and training programmes achieve their intended impact (Australian Government eSafety Commissioner, 2019[108]) (5Rights Foundation, 2022[144]).

A child rights impact assessment is a risk mitigation tool that can support positive cultural change, helping businesses to identify how their products affect children including through helping them to offer an age-appropriate service (5Rights Foundation, 2022[144]). It can help identify the direct or indirect effect of decisions on children and mainstream children right's principles into strategic planning and commissioning, problem solving, policy development, programme prioritisation, budget setting and service design, delivery and evaluation (UNICEF, 2021[145]). Child rights impact assessments usually involve meaningful participation and engagement with children, include feedback and review mechanisms; promote cross-departmental and cross-agency collaboration and are evidence-based (UNICEF, 2021[145]).

In the private sector, there are examples of integrating child rights impact assessments into business operations. For instance, Telia (a telecommunications and mobile network operator) used a child rights impact assessment to develop actions for integrating child rights aspects across relevant processes within the organisation and in the supply chain (Ring, 2023[146]). As part of this process, the company also

engaged with children directly through Children's Advisory Panels and with child rights organisations (Ring, 2023[146]).

The actions of governments also provide an example, and across the OECD, several countries have developed policies or require child rights impact assessment. In the EU, several countries have specific provisions that require that a child rights impact assessment takes place when developing laws and policies, and taking administrative decisions that impact children (European Union Agency for Fundamental Rights, 2014[147]) In New Zealand, the government released a Child Impact Assessment Tool that includes templates that agencies can use to test and assess any proposed policy or law for consistency with the UN Convention on the Rights of the Child (Ministry of Social Development, 2018[148]).

7 Case Studies

This section contains case studies of three services. One is directed at children (LEGO Life App), another serves both adults and children (Roblox), and the third is directed at adults (Omegle). The boxes below describe the services and how their actions align with the digital safety by design components described above. An analysis of their risk profiles follows. Based on the information outlined in the boxes and the risk profiles. Table 7.1 analyses the key components in section 6 against these case studies.

Box 7.1. Digital Service Directed at Children

LEGO's Life App

LEGO Life is a social media application designed for children to share stories and pictures of their LEGO creations, where children can access interactive building ideas and instructions and comment on other user's photos.

Digital Safety by Design Components:

- *Employing age assurance mechanisms*: When children attempt to create an account on LEGO Life, they are asked to provide a parent's email address. LEGO then sends an email to the parent to inform them of the account creation attempt and asks for their consent. This is a form of age verification and parental supervision. In addition, an age-gate prompts users to declare age. To reduce the incentive to lie, children can still enjoy a "boxed" version of the app if they do not receive parental consent - with fuller social features being unlocked once verifiable parental consent is achieved.
- *Implementing child-centred design:* Given its audience, the LEGO Life App's features are child-centred, allowing them to be creative and engage in digital play.
- *Detecting and preventing harm*: Moderators review all posts to ensure that they do not include any personal information and that they are age-appropriate.
- *Protecting children's privacy and personal data*: Children can choose from pre-approved, auto-generated usernames and create their LEGO avatar to remain anonymous, without personal information being shared. The company states that it follows standards of data privacy and security in accordance with the GDPR anywhere in the world where the company collects, stores, uses or shares users' personal data. Where local rules require more than that, the company states that it will adjust its practice to make sure users' data is safe.

Box 7.1. Digital Service Directed at Children (continued)

- *Ensuring child-friendly information provision*: A tool within the Life App called "Captain Safety" provides guidance for children on behaving like responsible digital citizens. All users learn how they should behave appropriately through a safety pledge and are provided periodic reminders as they engage with the app. Captain Safety also explains moderation and safety decisions to children (e.g. on why they need to be thoughtful when they upload images) and helps them understand LEGO's privacy policies.
- *Facilitating complaints and redress:* No specific complaint and redress information is specified.
- *Encouraging child participation and putting children at the centre of decision making*: No information is available on child participation and putting children at the centre of decision making.
- *Promoting a culture of safety and well-being*: LEGO launched Digital Design Principles that seek to prompt designers in the LEGO Group to consider children's perspectives and needs through strategic, design and engagement lens. Children need to sign a code of conduct that includes topics such as "I will protect my privacy" and "I will be kind to others". Tools and articles address topics such as safety and privacy and promote digital literacy.
- *Additional Features*:
    - Parental tools are included in the app. For example, Verified Parental Consent enables parents to verify their identity and give consent for their children to play and share within the application.
    - Through certain exercises, the Life App guides children to make sure they are aware of their own and others' feelings and to help them develop empathy.

Source: (LEGO, 2023[149]) (UNICEF Office of Research - Innocenti, 2022[105])

Risk profile

LEGO's Life App incorporates child-centric design, strong moderation features, and a capacity for parents, carers and guardians to monitor their child's usage. As zero risk is unattainable, there remains potential exposure to risks such as:

- *Content risks*: even with moderation in place, there is always a chance that inappropriate content might affect children. Whilst LEGO employs moderators to review uploads, human and algorithmic methods can occasionally miss harmful content.
- *Privacy risks*: As with any app that collects data, there remains concern that that data may be accidentally shared or subject to a data breach (Valentino-DeVries and Singer, 2018[150]).

Although LEGO's Life App is safer than many other platforms due to its child-centric design and strong moderation features, parents, carers and guardians should still monitor their child's usage, ensuring they maintain a balanced approach to both digital and physical experiences.

Box 7.2. Digital Service Directed at Both Children and Adults

Roblox

Roblox is a game-creation platform that allows users to design their own games and play a range of games created by other users. Roblox aims to provide an intersection for immersive social spaces and experiences for friends. Roblox has no minimum or maximum age range, but the recommended age on the platform is above eight years old. The platform hosts millions of virtual worlds and user-created games that cover a wide range of genres, such as role-playing games, racing, obstacle courses or stimulations. Roblox can play an educational role and help children learn coding and explore aerospace and biomedical engineering and 3D design amongst others. Roblox also provides an opportunity for its users to create, buy and sell virtual items.

Digital Safety by Design Components:

- *Employing age assurance mechanisms*: Currently there is no age verification on Roblox. However, Roblox is testing age verification as a new feature. Users will need to prove with a government-issued photo ID or with another government-issued identification document that they are at least 13 years old.
- *Implementing child-centred design*: No information on child-centred design is provided.
- *Detecting and preventing harm*: Roblox uses human and automated image review to ensure that images, videos and audio files are reviewed at the time of upload. Safety features include an automated chat filter and rules (with the ability to turn off chat), and community reporting provides an additional layer of protection for users, including children.
- *Protecting children's privacy and personal data:* Roblox provides customised privacy settings and the opportunity to enable privacy features. Moderators in the app review all posts to ensure that they do not include any personal information.
- *Ensuring child-friendly information provision*: No information on child-friendly information provision is given.
- *Facilitating complaints and redress*: Users can report inappropriate behaviour, content, or any violations of the platform's terms of service directly within the game.
- *Encouraging child participation and putting children at the centre of decision making*: No information on child participation and on any initiative to put children at the centre of decision making is provided.
- *Promoting a culture of safety and well-being*: Roblox uses a "Trust by Design" process to create risk assessments and mitigations for new functionality early in product development. During this process, designers and product managers bring concepts to a cross-functional team that includes members from safety, operations, moderation, public policy, civility, legal and other teams to provide feedback. Based on the safety assessment, the concept can move to the next phase or undergo further iteration.
- *Additional Features*:
    - Roblox provides support for parents and a range of online safety tools to protect their children. This includes guidance for parents explaining parental controls that can be used for very young users.

- Roblox encourages parents to help their children make responsible decisions without relying on the extensive use of parental controls.
- Other specific safety measures include asset pre-moderation that includes 3D content and holistic evaluation of a complete immersive content experience including 3D and 2D content plus interactivity.

Sources: (Roblox, 2023[151]) (Australian eSafety Commissioner, 2023[152]) (Kochan and Magid, 2018[153]) (Fore, 2023[154])

Risk profile

Whilst Roblox offers many opportunities for creativity and social interaction and has several digital safety features in place, children still face risks and potential harm:

- *Content Risks*: Since Roblox allows user-generated content, children might encounter games or other content that are inappropriate or not designed for their age group, such as naked avatars (Zapal, 2023[155]). Although Roblox has moderation practices, with the vast amount of content generated daily, some unsuitable materials might escape moderation (Robertson, 2022[156]).
- *Contact Risks*: Roblox is a multiplayer platform, meaning children can interact with other players, who are often anonymous. This opens potential risks of engaging with unknown adults, grooming, bullying, or exposure to inappropriate conversations. For instance, one report found a 7-year old girl's avatar was sexually assaulted in the game (Shamsian, 2018[157]). In another case, a child using the game was groomed to send sexually explicit pictures of himself (Stonehouse, 2019[158]). Roblox does offer customisable privacy settings to restrict interactions, but younger players might not always use such settings effectively.
- *Consumer Risks*: Roblox's virtual currency, Robux, can be used to buy in-game items. Children might be tempted to spend large amounts of real money, and there have been reports of unauthorised purchases or scams related to virtual goods (Latham, 2022[159]) (Bird, 2023[160]). The use of Robux can also raise concerns of exploitation, and in at least one instance an offender has been found by a court to have used Robux to persuade child victims to share explicit images (BBC, 2019[161]).
- *Health and Well-being-Risks*: The immersive nature of many Roblox games and the social interactions has raised concerns that children may spend excessive amounts of time on the platform, potentially affecting their well-being and other offline activities (Brill, 2021[162]).

Box 7.3. Digital Service Directed at Adults

Omegle

Omegle was a free online chat website (it shut down in November 2023) that paired two random users together in a text or video chat. Users could add topics that they would like to discuss to find users with common interests. The user had the possibility, once finished chatting, to select all the chat text to copy to the clipboard or download a PNG image of the text making it easy to share it further online.

Digital Safety by Design Components:

- *Employing age assurance mechanisms*: Omegle's Terms of Service noted: "The Services are not available to, and shall not be accessed or used by, persons under the age of 18. By accessing or using the services, you represent and warrant that you are at least 18 years of age."
- *Implementing child-centred design:* No information was provided.
- *Detecting and preventing harm*: Omegle did not allow users to flag, report or block chats that may be inappropriate or even illegal.
- *Protecting children's privacy and personal data:* Omegle's privacy notice stated that the website does not knowingly collect personal information from children under the age of 18 on the site or through the chat services. If the user was under the age of 18, they were encouraged not to give Omegle any personal information. Omegle encouraged parents and/or legal guardians to monitor their children's Internet usage and help protect their privacy by instructing them never to provide personal information via the Internet without parental or legal guardian permission.
- *Ensuring child-friendly information provision*: No information on available child-friendly information provision was given.
- *Facilitating complaints and redress:* No information was provided.
- *Child participation and putting children at the centre of decision making*: No information on child participation or on any initiative to put children at the centre of decision making was provided.
- *Promoting a culture of safety and well-being:* No information was provided.
- *Additional Features*: No other applicable safety features.

Sources: (Omegle, 2023[163]) (Tidy, 2023[164]) (Australian eSafety Commissioner, 2023[165]) (Lockwood, 2021[166]) (Omegle, 2022[167])

Risk profile

Omegle was a platform that connected users for one-on-one chat sessions with strangers. An investigation found that whilst the website was intended for adults only, during a ten-hour period the investigator was paired with dozens of children (Tidy, 2021[168]). Due to its nature, the website presented significant risks for children.

- *Content Risks*: Omegle chats were unpredictable. Users, especially children, could encounter explicit or harmful content, as conversations were not moderated. Whilst there was a moderated video section, as always, there was no guarantee of complete safety.
- *Contact Risks*: Omegle's tagline, "Talk to strangers!" itself suggested a fundamental risk for children. Sharing personal information led to potential real-world threats, stalking, or cyberbullying.

An investigation uncovered that the site had children engaging with adults in sexual activity and self-creating child sexual exploitation and abuse material (Tidy, 2021[168]). The Internet Watch Foundation (IWF), an organisation that is responsible for finding and removing child sexual abuse online, also found such abuse material that was created by sexual predators who have captured and distributed such images from the Omegle chat website (Tidy, 2021[168]).

- *Privacy Risks*: At Omegle, there was a risk of chat sessions, especially video chats, being recorded and shared without consent, leading to potential embarrassment or more severe consequences for the child involved.

Several lawsuits had been launched due to harm occurring to children on the website. For instance, in Oregon (US) a victim who encountered child sexual exploitation and abuse through Omegle filed a product liability lawsuit against the company (District Court of Oregon, 2023[169]). In another case, an individual was arrested and later convicted after broadcasting CSEA material through Omegle (McNaughton, 2023[170]).

Digital safety measures in place for digital services based on their target audience

Table 7.1 provides a synthesis of the key components set out in section 6 and considers how they may be necessary for the types of services outlined as case studies.

Table 7.1. Safety measures necessary for different kinds of digital services

Type of digital service	Type of measure							
	Age assurance	Detecting and preventing harm	Child-centred design	Protecting children's privacy and personal data	Child-friendly information provision	Facilitating complaints and redress	Child participation and putting children at the centre of decision making	Promoting a culture of safety and well-being
Child Directed Digital Service (LEGO Life App)	Service directed at children with moderated content and contact. Provided appropriate age assurance measures are in place the highest level of this component should not be necessary.	Whilst risk of harmful contact and content remains low, moderation remains necessary.	As the service is directed at children, child-centred design is necessary.	Service processes children's data, strong protection of children's privacy is necessary	Service is directed at children, so it is necessary that all information is provided in a child-friendly manner.	Low risk profile, but complaints and redress should be made available.	Given the audience of the service, engaging with children is essential to ensure that their concerns are addressed and that they can fully realise the benefits of the service.	All businesses have responsibility to promote a culture of safety and well-being and should undertake due diligence. Given that children are the intended audience of their platforms, the first two services (LEGO and Roblox) should undertake child rights impact assessment.
Digital Service Directed at Both Children and Adults (Roblox)	Service directed at both children and adults. Whilst the service has several safety features, given the potential risks, age assurance	Given that the service is directed at both adults and children, moderation and other technical measures to detect and prevent harmful	Service is directed at both adults and children, child-centred design would be necessary.	Strong privacy protection are necessary, as service processes children's data.	Service has child users, therefore child-friendly information is necessary.	Medium risk profile, complaints and redress mechanisms should be made available for children in a child-friendly manner.	As the service is also directed at children, child participation is necessary to address their concerns, shield them from harmful contact and content and to	

	would be required to protect children from harmful content and contacts.	contact and content are necessary.					ensure a safe experience on the platform.	
Digital Service Directed at Adults (Omegle)	Service was directed at adults, and the highest level of age assurance would have been required to ensure children cannot access the site.	Whilst ostensibly children should not have accessed the service, strong moderation processes and measures to detect and prevent harm to children should have been implemented, given the severity of risk to children.	Child-centred design in this case would have involved measures to effectively prevent access by children, and to protect / remove them from site if they do gain access.	Privacy harms for children could be significant, strong privacy protective measures would have been required.	Service should have clearly set out in a child-friendly manner that it was not intended for children.	Service was directed at adults. Complaints and redress are nevertheless necessary to protect children if they do gain access.	Engaging with children to understand the risks of service for them could have helped in mitigating risks, in designing policies and technical initiatives to effectively prevent children from accessing services, and moderation / complaint mechanisms should they nonetheless have been able to access the service.	

Source: OECD

Digital safety by design components vary depending on the service's risk profile, but all are necessary in all services. In those used by adults or both adults and children, certain protections are vital given the significant harm that can flow from children's access or use of the service.

In the case of LEGO, the Life App is primarily tailored for children, with moderation to ensure content and contact appropriateness. Although the highest level of age assurance is not vital some form of age assurance is important, that should be principles-based with an emphasis on usability, accuracy, privacy and proportionality. The design of the Life App is inherently child-centred, prioritising child safety and privacy. It is essential that all information provided is easily understandable by children. Despite its low-risk profile, a mechanism for complaints and redress is necessary. Moreover, promoting a culture of safety, well-being, and involving children in decision-making is of utmost importance.

Roblox, catering to both children and adults, necessitates age assurance due to potential risks. It is essential to have moderation and other preventative measures to curb harmful content and contacts. Whilst its design caters to both audiences, there is an emphasis on child safety and privacy. The platform should provide child-friendly information and ensure children have a mechanism to address complaints in an accessible manner. Encouraging a culture of safety and well-being, as well as child participation is crucial due to the mixed audience.

Robust safeguards are essential to ensure that children do not access platforms like Omegle, designed primarily for adults. The foremost protection is stringent age assurance; if effectively implemented, many subsequent child-centric measures become redundant. However, if a child does manage to access the platform, immediate moderation is crucial to remove or protect them. Additionally, easily accessible complaint mechanisms are vital to address any issues that arise. Omegle lacked these essential protective measures, and as a result has been proven to be a platform where significant harm has come to children.

8 Conclusion and Next Steps

As the digital frontier continues to expand, the challenges and opportunities it presents evolve in tandem. This report underscores the multifaceted nature of incorporating digital safety for children into design, highlighting the interplay of technical, societal, and corporate initiatives. Whilst the collaborative efforts of various stakeholders, from tech companies to policymakers, educators to parents, and children themselves, are crucial in shaping a digital environment that is both enriching and safe, policy makers are clearly signposting the responsibility of digital service providers to embed safety for children in their products and services by design.

Various international and regional initiatives are emerging, underscoring the emergence of digital safety, particularly for children, as a clear policy priority. Digital safety for children is now a global issue, but if each jurisdiction acts in isolation, their efforts will not serve children as well as they could. For instance, if each jurisdiction develops its own, distinct version of digital safety by design for children, international norms will be slow to develop and children from countries with the least institutional capacity might be disadvantaged. To facilitate international dialogue on digital safety and to effectively tackle online harms, it is necessary to build a comprehensive evidence base to support principles and standards. Developing this evidence base requires multi-stakeholder co-operation and at the Roundtable experts noted that the OECD is well placed to help.

The key components of digital safety for children, whilst diverse in nature, are intricately interlinked, forming a cohesive framework that prioritises children's safety and well-being. These components span a spectrum, from those embedded at the design stage to broader framing elements that support digital safety by design on a global scale, such as promoting a culture of safety and well-being for children.

The case studies presented in this report further illuminate the complexities of implementing digital safety measures. They underscore that there is no "one-size-fits-all" solution. Different components may need nuanced applications based on the risk profile of a particular service or platform. Furthermore, the mere declaration that a service is not intended for children does not absolve providers from responsibility. Proactive measures are essential to ensure that children, whether intentional users or inadvertent participants, are shielded from potential digital harms.

Looking ahead, there are several avenues for future work. Continuous horizon scanning, mapping and analysis are required to stay abreast of emerging technologies and the new challenges they pose, as well as the opportunities they present, for children’s safety in the digital environment. The OECD could facilitate international dialogues, fostering the exchange of best practices and innovative solutions across countries to further advance digital safety for children. Furthermore, the development of standardised metrics for assessing digital safety initiatives could provide clearer benchmarks for progress. As the digital landscape is ever-changing, a proactive and adaptive approach will be key to ensuring that children's online experiences are both fulfilling and secure.

References

5Rights Foundation (2022), *Approaches to Children's Data Protection*, https://5rightsfoundation.com/Approaches-to-Childrens-Data-Protection---.pdf. [123]

5Rights Foundation (2022), *Child Online Safety Toolkit*, https://childonlinesafetytoolkit.org/toolkit/. [144]

5Rights Foundation (2021), *But how do they know it is a child?*, https://5rightsfoundation.com/uploads/But_How_Do_They_Know_It_is_a_Child.pdf. [91]

5Rights Foundation (2021), *Explanatory Notes: General comment no. 25 (2021) on children's rights in relation to the digital environment*, https://5rightsfoundation.com/uploads/ExplanatoryNotes_UNCRCGC25.pdf. [50]

5Rights Foundation (n.d.), *Twisted Toys*, https://twisted-toys.com/. [133]

Agencia Española de Protección de Datos (2023), *Decalogue of Principles: Age verification and protection of minors from inappropriate content*, https://www.aepd.es/guides/decalogue-principles-age-verification-minors-protection.pdf. [121]

American Psychological Association (2023), *Health Advisory on Social Media Use in Adolescence*, https://www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use.pdf. [101]

Australian eSafety Commissioner (2024), *Industry codes and standards*, https://www.esafety.gov.au/industry/codes. [59]

Australian eSafety Commissioner (2023), *Industry codes complaints*, https://www.esafety.gov.au/industry/codes/complaints. [61]

Australian eSafety Commissioner (2023), *Omegle*, https://www.esafety.gov.au/key-issues/esafety-guide/omegle. [165]

Australian eSafety Commissioner (2023), *Our legislative functions*, https://www.esafety.gov.au/about-us/who-we-are/our-legislative-functions. [173]

Australian eSafety Commissioner (2023), *Register of industry codes and industry standards for online safety*, https://www.esafety.gov.au/industry/codes/register-online-industry-codes-standards. [60]

Australian eSafety Commissioner (2023), *Second set of tech giants falling short in tackling child sexual exploitation material, sexual extortion, livestreaming of abuse*, https://www.esafety.gov.au/newsroom/media-releases/second-set-of-tech-giants-falling-short-in-tackling-child-sexual-exploitation-material-sexual-extortion-livestreaming-of-abuse. [175]

Australian eSafety Commissioner (2023), *What is Roblox?*, https://www.esafety.gov.au/key-issues/esafety-guide/roblox. [152]

Australian eSafety Commissioner (2022), *Age verification*, https://www.esafety.gov.au/about-us/consultation-cooperation/age-verification. [81]

Australian eSafety Commissioner (2022), *eSafety Strategy 2022-2025*, https://www.esafety.gov.au/about-us/who-we-are/strategy. [82]

Australian eSafety Commissioner (2021), *Learn About the Online Safety Act*, https://www.esafety.gov.au/newsroom/whats-on/online-safety-act. [58]

Australian eSafety Commissioner (2019), *Safety by design: principles and background*, https://www.esafety.gov.au/industry/safety-by-design/principles-and-background. [62]

Australian Government (2014), *Australian Privacy Principles*, https://www.oaic.gov.au/privacy/australian-privacy-principles/read-the-australian-privacy-principles. [46]

Australian Government eSafety Commissioner (2019), *Safety by Design Overview*, https://www.esafety.gov.au/sites/default/files/2019-10/SBD%20-%20Overview%20May19.pdf. [108]

Australian Parliament (2021), *Online Safety Bill*, https://www.aph.gov.au/Parliamentary_Business/Bills_LEGislation/Bills_Search_Results/Result?bId=r6680. [57]

BBC (2019), *How Pontypridd paedophile Owain Thomas was caught*, https://www.bbc.com/news/uk-wales-48880997. [161]

Better Internet for Kids (2021), *BIK Policy Map Country Profiles: France*, https://www.saferinternetday.org/documents/167024/6823249/France+-+BIK+Policy+Map+Infosheet+-+FINAL.pdf/ca55892f-7294-e7f4-a474-a9f42196d076?t=1622798008960. [142]

Bird, N. (2023), *Roblox: Ten-year-old spent £2,500 of mum's money without her knowing*, https://www.bbc.com/news/uk-wales-65659896. [160]

Brill, A. (2021), *Is my child addicted to Roblox?*, https://www.washingtonpost.com/lifestyle/2021/05/25/roblox-addiction-advice/. [162]

Cavoukian, A. (2011), *Privacy by Design: The 7 Foundational Principles*, https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf. [28]

Cetic.br (n.d.), *ICT Kids Online Brazil*, https://cetic.br/en/pesquisa/kids-online/. [100]

CNA (2009), *The Audio-visual Law*, https://www.cna.ro/The-Audio-visual-Law,1655.html. [79]

CNIL (2022), *Vérification de l'âge en ligne : trouver l'équilibre entre protection des mineurs et respect de la vie privée*, https://www.cnil.fr/fr/verification-de-lage-en-ligne-trouver-lequilibre-entre-protection-des-mineurs-et-respect-de-la-vie. [96]

CNIL (2021), *CNIL publishes 8 recommendations to enhance the protection of children online*, https://www.cnil.fr/en/cnil-publishes-8-recommendations-enhance-protection-children-online. [127]

CNIL (2021), *Recommandation 8 : prévoir des garanties spécifiques pour protéger l'intérêt de l'enfant*, https://www.cnil.fr/fr/recommandation-8-prevoir-des-garanties-specifiques-pour-proteger-linteret-de-lenfant. [129]

Coimisiún na Meán (2023), *Work Programme*, https://www.cnam.ie/wp-content/uploads/2023/06/Coimisiun-na-Mean-Work-Programme-_Web.pdf. [83]

Comisiun na Mean (2023), *Responses to Coimisiún na Meán Call for Inputs: Online Safety Code*, https://www.cnam.ie/wp-content/uploads/2023/12/CallForInputs_ResponsesReceived.pdf. [93]

Commission Nationale de l'Informatique et des Libertés (2022), *Demonstration of a privacy-preserving age verification process*, https://linc.cnil.fr/demonstration-privacy-preserving-age-verification-process. [120]

Cortesi, S., A. Hasse and U. Gasser (2021), "Youth Participation in a digital world: Designing and implementing spaces, programs and methodologies", https://cyber.harvard.edu/publication/2021/youth-participation-in-a-digital-world. [138]

Cortesi, S. et al. (2019), *Youth and Artificial Intelligence: Where We Stand*, https://dash.harvard.edu/bitstream/handle/1/40268058/2019-05_YouthAndAI.pdf?sequence=5&isAllowed=y. [5]

Council of Europe (2020), *Handbook for policy makers on the rights of the child in the digital environment*, https://rm.coe.int/publication-it-handbook-for-policy-makers-final-eng/1680a069f8. [54]

Council of Europe (2018), *Guidelines to respect, protect and fulfil the rights of the child in the digital environment*, https://edoc.coe.int/en/children-and-the-internet/7921-guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-the-digital-environment-recommendation-cmrec20187-of-the-committee-of-ministers.html. [53]

Cybersecurity and Infrastructure Security Agency (2023), *Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Security-by-Design and -Default*, https://www.cisa.gov/sites/default/files/2023-04/principles_approaches_for_security-by-design-default_508_0.pdf. [32]

Digital Futures Commission (2023), *Privacy matters when enabling a safer online experience for children*, https://digitalfuturescommission.org.uk/blog/privacy-matters-when-enabling-a-safer-online-experience-for-children/. [119]

Digital Futures Commission and 5Rights Foundation (2023), *Playful by Design Toolkit*, https://digitalfuturescommission.org.uk/playful-by-design-toolkit/. [48]

Discord (2023), *Building a safer place for teens to hang out*, https://discord.com/safety/safer-place-for-teens. [109]

District Court of Oregon (2023), *A.M. v. Omegle.com*, https://casetext.com/case/am-v-omeglecom-1. [169]

Elysee (2022), *Laboratory for Childhood Protection Online Charter*, https://www.elysee.fr/en/emmanuel-macron/2022/11/10/laboratory-for-childhood-protection-online-charter. [70]

European Commission (2023), *Special group on the EU Code of conduct on age-appropriate design*, https://digital-strategy.ec.europa.eu/en/policies/group-age-appropriate-design. [66]

European Commission (2022), *A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+) (COM(2022) 212 final)*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022DC0212. [114]

European Commission (2022), *A European strategy for a better internet for kids (BIK+)*, https://digital-strategy.ec.europa.eu/en/policies/strategy-better-internet-kids. [67]

European Commission (2022), *Child-friendly version of European strategy for a better Internet for kids*, https://digital-strategy.ec.europa.eu/en/library/child-friendly-version-european-strategy-better-internet-kids-bik. [68]

European Commission (2022), *European Declaration on Digital Rights and Principles for the Digital Decade*, https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles. [65]

European Commission (2022), *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065. [63]

European Commission (2021), *The European Union's Plan for Children's Rights*, https://op.europa.eu/webpub/just/eu-plan-for-children-rights/en/. [174]

European Data Protection Supervisor (2023), *Accountability*, https://edps.europa.eu/data-protection/our-work/subjects/accountability_en#:~:text=The%20General%20Data%20Protection%20Regulation,and%20its%20effectiveness%20when%20requested. [40]

European Medicines Agency (2023), *Marketing authorisation*, https://www.ema.europa.eu/en/human-regulatory-overview/marketing-authorisation. [21]

European Parliament (2018), *Audiovisual Media Services Directive*, https://eur-lex.europa.eu/eli/dir/2018/1808/oj. [64]

European Union (2022), *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance)*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065. [113]

European Union (2009), *Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02009L0048-20191118. [24]

European Union Agency for Fundamental Rights (2014), *Child rights impact assessment*, https://fra.europa.eu/en/content/child-rights-impact-assessment. [147]

European Union Agency for Fundamental Rights (2000), *Article 24 - The rights of the child*, https://fra.europa.eu/en/eu-charter/article/24-rights-child. [73]

European Union Aviation Safety Agency (2023), *Regulations*, https://www.easa.europa.eu/en/regulations. [19]

Federal Aviation Administration (2023), *Regulations & Policies*, https://www.faa.gov/regulations_policies. [18]

Fore, P. (2023), *Ever wanted to learn how to code, build robotics, or explore life on Mars? Roblox's $25 million investment is starting to make that a reality*, https://fortune.com/education/articles/roblox-community-fund-education-investments-david-baszucki/. [154]

FTC (2022), *Fortnite Video Game Maker Epic Games to Pay More Than Half a Billion Dollars over FTC Allegations of Privacy Violations and Unwanted Charges*, https://www.ftc.gov/news-events/news/press-releases/2022/12/fortnite-video-game-maker-epic-games-pay-more-half-billion-dollars-over-ftc-allegations. [130]

FTC (1998), *Children's Online Privacy Protection Rule ("COPPA")*, https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa. [128]

GDPR (2018), *Article 25 GDPR. Data protection by design and by default*, https://gdpr-text.com/read/article-25/. [29]

GDPR (2018), *Recital 78*, https://gdpr-text.com/read/recital-78/?col=2&lang1=fr&lang2=en&lang3=sv. [30]

GDPR (2018), *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*, https://eur-lex.europa.eu/eli/reg/2016/679/oj. [37]

Global Privacy Assembly (2023), *Digital education Working group - Report*, https://globalprivacyassembly.org/wp-content/uploads/2023/10/5.-DEWG-Annual-activity-Report-2022-2023-V2-28-September-2023.pdf. [137]

Government of Canada (2000), *Personal Information Protection and Electronic Documents Act*, https://laws-lois.justice.gc.ca/eng/acts/p-8.6/FullText.html. [47]

Government of Türkiye (2015), *Law on the Regulation of Internet Publications and Combating Crimes Committed through Such Publications*, https://www.venice.coe.int/webforms/documents/default.aspx?pdffile=CDL-REF(2016)026-e. [75]

Hartung, P. (2020), *The children's rights-by-design standard for data use by tech companies*, https://www.unicef.org/globalinsight/media/1286/file/%20UNICEF-Global-Insight-DataGov-data-use-brief-2020.pdf. [125]

House of the Oireachtas (2022), *Online Safety and Media Regulation Act 2022*, https://www.oireachtas.ie/en/bills/bill/2022/6/. [72]

ICO (2024), *Age assurance for the Children's code*, https://ico.org.uk/about-the-ico/what-we-do/information-commissioners-opinions/age-assurance-for-the-children-s-code/. [95]

ICO (2019), *Towards a better digital future: Informing the Age Appropriate Design Code*, https://ico.org.uk/media/about-the-ico/consultations/2614763/ico-rr-report-0703.pdf. [141]

IEEE Consumer Technology Society (2021), *Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children*, https://5rightsfoundation.com/static/ieee-2089-2021.pdf. [99]

IMDA (2023), *IMDA's Online Safety Code comes into effect*, [80]

https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/imdas-online-safety-code-comes-into-effect.

Information Commissioner's Office (2021), *Privately: Age Appropriate Design Code Engagement Report - Executive Summary*, https://ico.org.uk/media/action-weve-taken/audits-and-advisory-visits/4019196/privately-aadc-exec-summary-202112.pdf. [98]

Information Commissioner's Office (2021), *Regulatory Sandbox Final Report: Yoti*, https://ico.org.uk/media/for-organisations/documents/4020427/yoti-sandbox-exit_report_20220522.pdf. [97]

INHOPE (2021), *What is Safety by Design*, https://inhope.org/EN/articles/what-is-safety-by-design?locale=en#:~:text=Safety%20by%20Design%20is%20an,after%20the%20harm%20has%20occurred. [17]

Instagram (2021), *Introducing new tools to protect our community from abuse*, https://about.instagram.com/blog/announcements/introducing-new-tools-to-protect-our-community-from-abuse. [107]

Integrity Institute (2024), *Child Safety Online*, https://integrityinstitute.org/blog/child-safety-online. [106]

International Telecommunication Union (2020), *Guidelines for industry on Child Online Protection*, https://www.unicef.org/media/90796/file/ITU-COP-guidelines%20for%20industry-2020.pdf. [55]

Internet Watch Foundation (2023), *reThink Chatbot evaluation*, https://www.iwf.org.uk/about-us/why-we-exist/our-research/rethink-chatbot-evaluation/. [111]

Kayali, L. (2022), *France's global push for online child safety*, https://www.politico.eu/article/emmanuel-macron-france-child-protection-online-harmful-internet-content/. [71]

Kochan, M. and L. Magid (2018), *The Parent's Guide to Roblox*, https://corp.roblox.com/wp-content/uploads/2018/06/Roblox-ConnectSafely-Parents-Guide-v2.pdf. [153]

Koko (2022), *Koko*, https://www.kokocares.org/. [110]

Korea Communications Standards Commission (2023), , https://www.kocsc.or.kr/eng/PageLink.do. [85]

Korea Legislation Research Institute (2011), *Personal Information Protection Act*, https://elaw.klri.re.kr/eng_service/lawView.do?hseq=62389&lang=ENG. [43]

Latham, K. (2022), *How computer games encourage kids to spend cash*, https://www.bbc.com/news/business-65372710. [159]

LEGO (2023), *Digital Safety*, https://www.lego.com/en-us/life/digital-safety. [149]

Litton, S. (2023), *Announcing Lantern: The First Child Safety Cross-Platform Signal Sharing Program*, https://www.technologycoalition.org/newsroom/announcing-lantern. [117]

Livingstone, S. and T. Palmer (2012), *Identifying vulnerable children online and what strategies can help them*, https://eprints.lse.ac.uk/44222/. [103]

Livingstone, S. and K. Pothong (2023), *Child Rights by Design*. [26]

Livingstone, S. and K. Pothong (2021), *UK "Secure by Design" vs Australian "Safety by Design"*, https://blogs.lse.ac.uk/parenting4digitalfuture/2021/09/29/secure-by-design/. [118]

Livingstone, S. and K. Pothong (2021), *What is meant by "by design"?*, https://digitalfuturescommission.org.uk/blog/what-is-meant-by-by-design/. [25]

Lockwood, A. (2021), *What is Omegle? Key things parents and carers need to know*, https://saferinternet.org.uk/blog/what-is-omegle-key-things-parents-and-carers-need-to-know. [166]

McIlraith, J. and C. Harrow (2018), *Feedback and Complaints Systems: A Rapid Review*, https://orangatamariki.govt.nz/assets/Uploads/About-us/Research/Latest-research/Feedback-and-complaints-report/Feedback-and-complaints-systems-A-rapid-review.pdf. [135]

McNaughton, G. (2023), *Guelph man can no longer be teacher after child porn conviction*, https://www.guelphmercury.com/news/crime/guelph-man-can-no-longer-be-teacher-after-child-porn-conviction/article_7b1fca76-cef1-56e5-a9e7-cb9091ac43bb.html. [170]

Meurens, N. et al. (2022), *Child safety by design that works against online sexual exploitation of children*, https://www.datocms-assets.com/22233/1652864615-child-safety-by-design-report-final-1.pdf. [132]

Ministère de l'Éducation Nationale (2023), *Le Parlement des enfants*, https://eduscol.education.fr/3310/le-parlement-des-enfants. [140]

Ministry of Social Development (2018), *Child Impact Assessment Tool*, https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/resources/child-impact-assessment.html. [148]

Ministry of the Interior and Kingdom Relations (2021), *The Netherlands Code for Children's Rights*, https://codevoorkinderrechten.nl/wp-content/uploads/2021/07/Code-voor-Kinderrechten-Wordversie_EN.pdf. [124]

Moody, L. et al. (2021), *Consultations with young people to inform the eSafety Commissioner's Engagement Strategy for Young People: A report on the findings*, https://www.esafety.gov.au/sites/default/files/2022-01/YRRC%20Research%20Report%20eSafety%202021_web%20V06%20-%20publishing_1.pdf. [143]

Netsafe (2023), *Netsafe*, https://netsafe.org.nz/aboutnetsafe/our-service/. [136]

OECD (2023), *Digital equity and inclusion in education: An overview of practice and policy in OECD countries*, https://www.oecd-ilibrary.org/docserver/7cb15030-en.pdf?expires=1715676646&id=id&accname=ocid84004878&checksum=342ADE68F29E8616BB0756F8F6952AE8. [104]

OECD (2023), "Emerging privacy-enhancing technologies: Current regulatory and policy approaches", *OECD Digital Economy Papers*, No. 351, OECD Publishing, Paris, https://doi.org/10.1787/bf121be4-en. [16]

OECD (2023), *Expert Roundtable on Digital Safety by Design for Children: Summary of Key* [9]

Points (Internal Document), https://one.oecd.org/document/DSTI/CDEP(2023)28/en/pdf.

OECD (2023), *Global Recalls Portal*, https://globalrecalls.oecd.org/#/. [7]

OECD (2023), *OECD Privacy Guidelines Implementation Guidance: Foreword and Chapter on Accountabilty*, https://one.oecd.org/document/DSTI/CDEP/DGP(2022)8/FINAL/en/pdf. [39]

OECD (2023), *Roundtable on Digital Safety by Design for Children: Revised Agenda (internal document)*, https://one.oecd.org/document/DSTI/CDEP(2023)19/REV1/en/pdf. [8]

OECD (2023), *Transparency Reporting on Child Sexual Exploitation and Abuse Online*, https://www.oecd-ilibrary.org/docserver/554ad91f-en.pdf?expires=1693988793&id=id&accname=ocid84004878&checksum=0C2DA4AD664AF2EC39CE58CF8E1F0AAE. [3]

OECD (2022), *Companion Document to the OECD Recommendation on Children in the Digital Environment*, https://www.oecd-ilibrary.org/science-and-technology/companion-document-to-the-oecd-recommendation-on-children-in-the-digital-environment_a2ebec7c-en. [15]

OECD (2022), *Recommendation of the Council on the Digital Security of Products and Services*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0481. [35]

OECD (2021), *Children in the Digital Environment: Revised Typology of Risks*, https://www.oecd-ilibrary.org/docserver/a2ebec7c-en.pdf?expires=1653053778&id=id&accname=guest&checksum=AD7D1497BD71DD3BA15ABD5B1557BF8F. [14]

OECD (2021), *Guidelines for Digital Service Providers*, https://one.oecd.org/document/C/MIN(2021)7/ADD1/en/pdf. [2]

OECD (2021), *Recommendation of the Council on Children in the Digital Environment*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389%20. [1]

OECD (2021), *Understanding the Digital Security of Products: An In-Depth Analysis*, https://www.oecd-ilibrary.org/docserver/abea0b69-en.pdf?expires=1655130260&id=id&accname=ocid84004878&checksum=3A99D35A2AA87DCA7C675B7D5CCA1718. [34]

OECD (2020), *Digital Economy Outlook 2020*, https://www.oecd-ilibrary.org/docserver/bb167041-en.pdf?expires=1654854924&id=id&accname=ocid84004878&checksum=856963E884B92AEB2496B413B4175AA1. [27]

OECD (2020), *Protection of Children Online: An Overview of Recent Developments in Legal Frameworks and Policies*, https://www.oecd-ilibrary.org/docserver/9e0e49a9-en.pdf?expires=1653410836&id=id&accname=ocid84004878&checksum=4A109A32D62EB1DA07305E7CEE85747D. [13]

OECD (2019), *Educating 21st Century Children : Emotional Well-being in the Digital Age*, https://www.oecd-ilibrary.org/education/educating-21st-century-children_b7f33425-en. [102]

OECD (2019), *Roles and Responsibilities of Actors: Governance of Digital Security in Organisations and Security of Digital Technologies*, https://one.oecd.org/document/DSTI/CDEP/SPDE(2019)4/FINAL/en/pdf. [33]

OECD (2017), *Protection of Children Online: Preliminary Country Survey Findings and Proposal For Next Steps*, https://one.oecd.org/document/DSTI/CDEP/SPDE(2017)3/en/pdf. [12]

OECD (2011), *The Protection of Children Online: Risks Faced by Children Online and Policies to Protect Them*, https://www.oecd-ilibrary.org/science-and-technology/the-protection-of-children-online_5kgcjf71pl28-en. [11]

OECD (2008), *Declaration for the Future of the Internet Economy (The Seoul Declaration)*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0366. [10]

Ofcom (2023), *Ofcom's approach to implementing the Online Safety Act*, https://www.ofcom.org.uk/__data/assets/pdf_file/0017/270215/10-23-approach-os-implementation.pdf. [86]

Ofcom (2022), *Children's Online User Ages Quantitative Research Study*, https://www.ofcom.org.uk/__data/assets/pdf_file/0015/245004/children-user-ages-chart-pack.pdf. [89]

Ofcom (2021), *Regulating video-sharing platforms: what you need to know*, https://www.ofcom.org.uk/online-safety/advice-for-consumers/video-sharing-platforms. [84]

Office of the Australian Information Commissioner (2022), *Privacy by design*, https://www.oaic.gov.au/privacy/privacy-for-organisations/privacy-by-design. [31]

Omegle (2023), *Omegle: Talk to Strangers*, https://www.omegle.com/. [163]

Omegle (2022), *Omegle Privacy Notice*, https://www.omegle.com/static/privacy.html. [167]

Parliament of the Republic of Fiji (2018), *Online Safety Act 2018*, https://www.parliament.gov.fj/wp-content/uploads/2018/05/Act-8-Online-Safety.pdf. [78]

Parliamentary Counsel Office (2020), *Privacy Act 2020*, https://www.legislation.govt.nz/act/public/2020/0031/latest/LMS23223.html. [44]

Personal Data Protection Commission Singapore (2023), *Launch of the PDPC Innovation Challenge*, https://www.pdpc.gov.sg/news-and-events/announcements/2023/06/launch-of-the-pdpc-innovation-challenge. [122]

Personal Information Protection Commission (2003), *Act on the Protection of Personal Information*, https://www.ppc.go.jp/en/legal/. [42]

PlaySafeID (2024), *PlaySafe ID*, https://www.playsafeid.com/. [112]

Revealing Reality (2022), *Families' attitudes towards age assurance*, https://assets.publishing.service.gov.uk/media/6343dd3f8fa8f52a5803e669/Ofcom_ICO_joint_research_-_age_assurance_report.pdf. [90]

Revealing Reality (2022), *Families'attitudes toward age assurance (Research commissioned by the ICO and Ofcom)*, https://www.ofcom.org.uk/__data/assets/pdf_file/0026/245195/DRCF-Ofcom-ICO-age-assurance.pdf. [92]

Ring, H. (2023), *Children's Rights Impact Assessment helps identify and mitigate impacts on children in a digital environment*, https://www.oecd-forum.org/posts/children-s-rights-impact-assessment-helps-identify-and-mitigate-impacts-on-children-in-a-digital-environment. [146]

Robertson, A. (2022), *Parents guide to Roblox and how your kids can play it safely*, https://www.internetmatters.org/hub/esafety-news/parents-guide-to-roblox-and-how-your-kids-can-play-it-safely/. [156]

Roblox (2023), *Age ID Verification*. [151]

Sénat (2023), *Projet de loi visant à sécuriser et réguler l'espace numérique*, https://www.senat.fr/basile/visio.do?id=d0150842&idtable=d172203-114375_20%7Cd0150842&_c=m%C3%A9dias&rch=ds&de=20230615&au=20230630&dp=15+jours&radio=dp&aff=72203&tri=p&off=0&afd=ppr&afd=ppl&afd=pjl&afd=cvn. [69]

Shamsian, J. (2018), *A mother says her 7-year-old daughter's character was raped in the kid's game 'Roblox'*, https://www.insider.com/roblox-girl-avatar-rape-toxic-trolling-community-2018-7. [157]

Smirnova, S., S. Livingstone and M. Stoilova (2021), *Understanding of user needs and problems: A rapid evidence review of age assurance and parental controls*, https://euconsent.eu/download/understanding-of-user-needs-and-problems-a-rapid-evidence-review-of-age-assurance-and-parental-controls/. [88]

Stonehouse, R. (2019), *Roblox: 'I thought he was playing an innocent game'*, https://www.bbc.com/news/technology-48450604. [158]

Taylor, J. (2023), *X fined $610,500 in Australia first for failing to crack down on child sexual abuse material*, https://www.theguardian.com/technology/2023/oct/16/x-fined-610500-in-australia-first-for-failing-to-crack-down-on-child-sexual-abuse-material. [171]

Third, A. and L. Moody (2021), *Our rights in the digital world: A report on the children's consultations to inform UNCRC General Comment 25*, https://5rightsfoundation.com/uploads/OurRIghtsinaDigitalWorld-FullReport.pdf. [131]

Thorn (2023), *Online grooming: What it is, how it happens, and how to defend children*, https://www.thorn.org/blog/online-grooming-what-it-is-how-it-happens-and-how-to-defend-children/. [172]

Tidy, J. (2023), *Omegle: Suing the website that matched me with my abuser*, https://www.bbc.com/news/technology-64618791. [164]

Tidy, J. (2021), *Omegle: Children expose themselves on video chat site*, https://www.bbc.com/news/technology-56085499. [168]

Türkiye Information Technologies and Communications Authority (2018), *Güvenli internet merkezi*, https://www.gim.org.tr/. [76]

UK Government (2024), *Performing a Security Risk Assessment*, https://www.security.gov.uk/guidance/secure-by-design/activities/performing-a-security-risk-assessment. [38]

UK Government (2024), *Product Security and Telecommunications Infrastructure Act*, https://www.gov.uk/government/publications/the-uk-product-security-and-telecommunications-infrastructure-product-security-regime. [45]

UK Government (2024), *Secure by design principles*, https://www.security.gov.uk/guidance/secure-by-design/principles/. [41]

UK Government (2023), *Secure by design*, https://www.gov.uk/government/collections/secure-by-design. [36]

UK Government (2020), *Best practice guidance on the labelling and packaging*, https://assets.publishing.service.gov.uk/media/5fe09fa48fa8f51489f14374/Best_practice_guidance_labelling_and_packaging_of_medicines.pdf. [22]

UK Information Commissioners Office (2020), *Age Appropriate Design Code*. [126]

UK Parliament (2023), *Online Safety Act 2023*, https://bills.parliament.uk/bills/3137. [77]

UN CRC (2021), *General comment No. 25 (2021) on children's rights in relation to the digital environment*, https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/053/43/PDF/G2105343.pdf?OpenElement. [49]

UNICEF (2021), *Child Rights Impact Assessment: Template and Guidance for Local Authorities*, https://www.unicef.org.uk/child-friendly-cities/wp-content/uploads/sites/3/2022/06/CRIA_June-2022.pdf. [145]

UNICEF (2020), *The children's rights-by-design standard for data use by tech companies*, https://www.unicef.org/globalinsight/media/1286/file/%20UNICEF-Global-Insight-DataGov-data-use-brief-2020.pdf. [52]

UNICEF (2019), *Child-Friendly Complaint Mechanisms*, https://www.unicef.org/eca/sites/unicef.org.eca/files/2019-02/NHRI_ComplaintMechanisms.pdf. [134]

UNICEF (2014), *Engaging stakeholders on children's rights*, https://www.unicef.ch/en/media/1050/download?attachment=. [139]

UNICEF Office of Research - Innocenti (2022), *Responsible Innovation in Technology for Children*, https://www.unicef-irc.org/publications/pdf/RITEC_Responsible-Innovation-in-Technology-for-Children-Digital-technology-play-and-child-well-being.pdf. [105]

United Kingdom Government (2023), *Online Safety Act*, https://www.legislation.gov.uk/ukpga/2023/50/enacted. [94]

United Kingdom Information Commissioners Office (2024), *Age assurance for the Children's code*, https://ico.org.uk/about-the-ico/what-we-do/information-commissioners-opinions/age-assurance-for-the-children-s-code/. [87]

United Nations (2023), *Resolution on the Rights of the Child in the digital environment*, https://childrightsconnect.org/un-general-assembly-groundbreaking-resolution-with-193-states-committed-to-implementing-childrens-rights-in-the-digital-environment/#:~:text=The%20UNGA%20resolution%20notes%20that,the%20exercise%20of%20their%20rights. [51]

United Nations (1989), *Convention on the Rights of the Child*, https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child#:~:text=or%20family%20members.-,Article%203,2. [74]

US Food & Drug Administration (2023), *The Drug Development Process*, https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process. [20]

US Surgeon General (2023), *Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory*, https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf. [6]

Valentino-DeVries, J. and N. Singer (2018), *Senators Call for Federal Investigation of Children's Apps*, https://www.nytimes.com/2018/10/03/technology/kids-apps-federal-investigation.html. [150]

Vosloo, S. (2023), *Generative AI: Risks and Opportunities for Children*, https://www.unicef.org/globalinsight/media/3061/file. [4]

WeProtect Global Alliance (2021), *Findings from WeProtect Global Alliance / Technology Coalition survey of technology companies: Summary of Findings*, https://www.weprotect.org/wp-content/uploads/Survey-of-technology-companies-2021.pdf. [116]

White, N. and W. Kibalama (2018), *Prevention of Pediatric Pharmaceutical Poisonings*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6124998/. [23]

World Economic Forum (2023), *Global Principles on Digital Safety: Translating International Human Rights for the Digital Context*, https://www3.weforum.org/docs/WEF_Global_Charter_of_Principles_for_Digital_Safety_2023.pdf. [56]

YouTube (2023), *About the YouTube Trusted Flagger program*, https://support.google.com/youtube/answer/7554338?hl=en#:~:text=The%20YouTube%20Trusted%20Flagger%20program%20helps%20provide%20robust%20tools%20to,that%20violates%20our%20Community%20Guidelines. [115]

Zapal, H. (2023), *The Top 5 Hidden Dangers of Roblox*, https://www.bark.us/blog/hidden-dangers-roblox/. [155]

Notes

[1] As defined in the OECD Recommendation, “digital service providers” refers to any natural or legal person that provides products and services, electronically and at a distance (OECD, 2021[1]).

[2] Throughout this paper, where “parent” is used it should be read to cover also carers and guardians.

[3] For instance, the ICT Coalition for Children Online; Better Internet for Kids: CEO Coalition; Coalition for Digital Intelligence amongst others.

[4] For instance, Alliance to Better Protect Minors Online, WeProtect Global Alliance amongst others.

[5] For instance, the European Framework for Safer Mobile Use by Younger Teenagers and Children.

[6] For instance, the EU Code of Practice for Disinformation.

[7] This research was a contribution to UNICEF’s Data Governance Manisfesto, but is not an adopted standard.

[8] The European Strategy for a Better Internet for Kids, adopted in May 2022, also seeks to ensure that children are protected, empowered and respected in the digital environment, in accordance with the European Digital Principles (European Commission, 2022[67]).

[9] The examples were selected based on the presentations at the Roundtable and desk research.

[10] For instance, in October 2023 the eSafety Commissioner issued an AUD 610,500 fine to an online platform for its failure to comply with the non-periodic reporting notice, in relation to the steps it is taking to tackle child sexual exploitation and abuse. From the date of issue, the platform had 28 days to pay the fine. (Taylor, 2023[171]) (Australian eSafety Commissioner, 2023[175]).

[11] Cyberbullying content is defined as anything posted on an online service which is intended to target an Australian child, and which has the effect of seriously harassing, humiliating, intimidating, or threatening the child (Australian eSafety Commissioner, 2023[173]).

[12] Online grooming refers to the range of strategies that predators use over the Internet to sexually abuse children (Thorn, 2023[172]).

[13] The Online Safety Act (at Part 2) defines “user-to-user service” as “an internet service by means of which content that is generated directly on the service by a user of the service, or uploaded to or shared on the

service by a user of the service, may be encountered by another user, or other users, of the service" (UK Parliament, 2023[77]).

[14] The Online Safety Act (at Part 2) defines "search service" as "an internet service that is, or includes, a search engine" (UK Parliament, 2023[77]).

[15] The Online Safety Act (at s79) defines "provider pornographic content" as pornographic content that is published or displayed on a service by the provider of the service or by a person acting on behalf of the provider. It includes pornographic content published or displayed on the service by means of software, an automated tool, or an algorithm that is either applied or made available on the service by the provider or a person acting on their behalf. It does not include user-generated content (see s55 of the Act for a definition of "user-generated") (UK Parliament, 2023[77]).

[16] Further information on the codes of practices can be found at https://www.ofcom.org.uk/online-safety/information-for-industry/guide-for-services/codes-of-practice.

[17] The examples were selected based on the presentations at the Roundtable and desk research.

[18] For example, the US' Children's Online Privacy Protection Act (COPPA) and the GDPR require consent from a parent or caregiver when a child has not yet reached the specified age. Often such requirements relate to consenting to the underlying data processing needed for a service to operate. COPPA requirements apply when services are "directed to children" – when it is effectively assumed that the audience is going to be made up of children – and when there is "actual knowledge" of a child or that the operator is collecting information from a service directed to children.

[19] Profiling for the purposes of age assurance is referred to as using Artificial Intelligence or 'inference models' to estimate age (5Rights Foundation, 2021[91]).

[20] In capacity testing, a service seeks to estimate the age of the user based on his or her capacity. For instance, children may be required to solve a puzzle, complete a language test, or undertake a task that gives an indication of their age or age range (5Rights Foundation, 2021[91]).

[21] An age token exclusively holds data pertaining to the exact age or age bracket of a user. This enables the service to determine whether a user satisfies their age criteria without gathering additional personal details (5Rights Foundation, 2021[91]).

[22] Noting that not all countries have national identification that is available to children.

[23] The waterfall technique combines different age assurance approaches. Waterfall techniques build on the output of successive age assurance approaches to provide a cumulative result with a greater level of confidence than any of these approaches in isolation. When used correctly, waterfall techniques have the potential to offer high levels of confidence, whilst providing a more privacy respecting approach with less user friction. A common example is where an age estimation method is combined with a secondary age verification method, when a high level of assurance is required.

[24] Hash matching or digital fingerprinting are the most widely used and longest established forms of automated content moderation. These technologies are used to tag, remove and prevent re-upload of known images and videos of known child sexual abuse material (OECD, 2023[3]).

[25] The OECD Recommendation reflects the prominent need to protect children's privacy in several places, acknowledging in the preamble that "safeguarding children's privacy and protecting children's personal data is vital for children's well-being and autonomy, and for meeting their needs in the digital environment". It calls on all actors to support children in understanding their rights as data subjects, as well as the ways in which their personal data is collected, processed, shared, and used; and calls on governments (as part of digital literacy measures) to support children in understanding how their personal data is collected, disclosed, made available or otherwise used. (OECD, 2021[1]) Additionally, the OECD Guidelines give guidance regarding the responsibility of digital service providers to safeguard children's privacy (OECD, 2021[2]).

[26] Certain rights documents and standards can also be found in child-friendly versions. For example, both the European Union's Strategy on the Rights of the Child (European Commission, 2021[174]) and its Better Internet for Kids Strategy (European Commission, 2022[68]) had child-friendly versions.

[27] This consultation underpinned the Committee on the Rights of the Child General Comment No. 25 on Children's Rights in relation to the Digital Environment (UN CRC, 2021[49]).