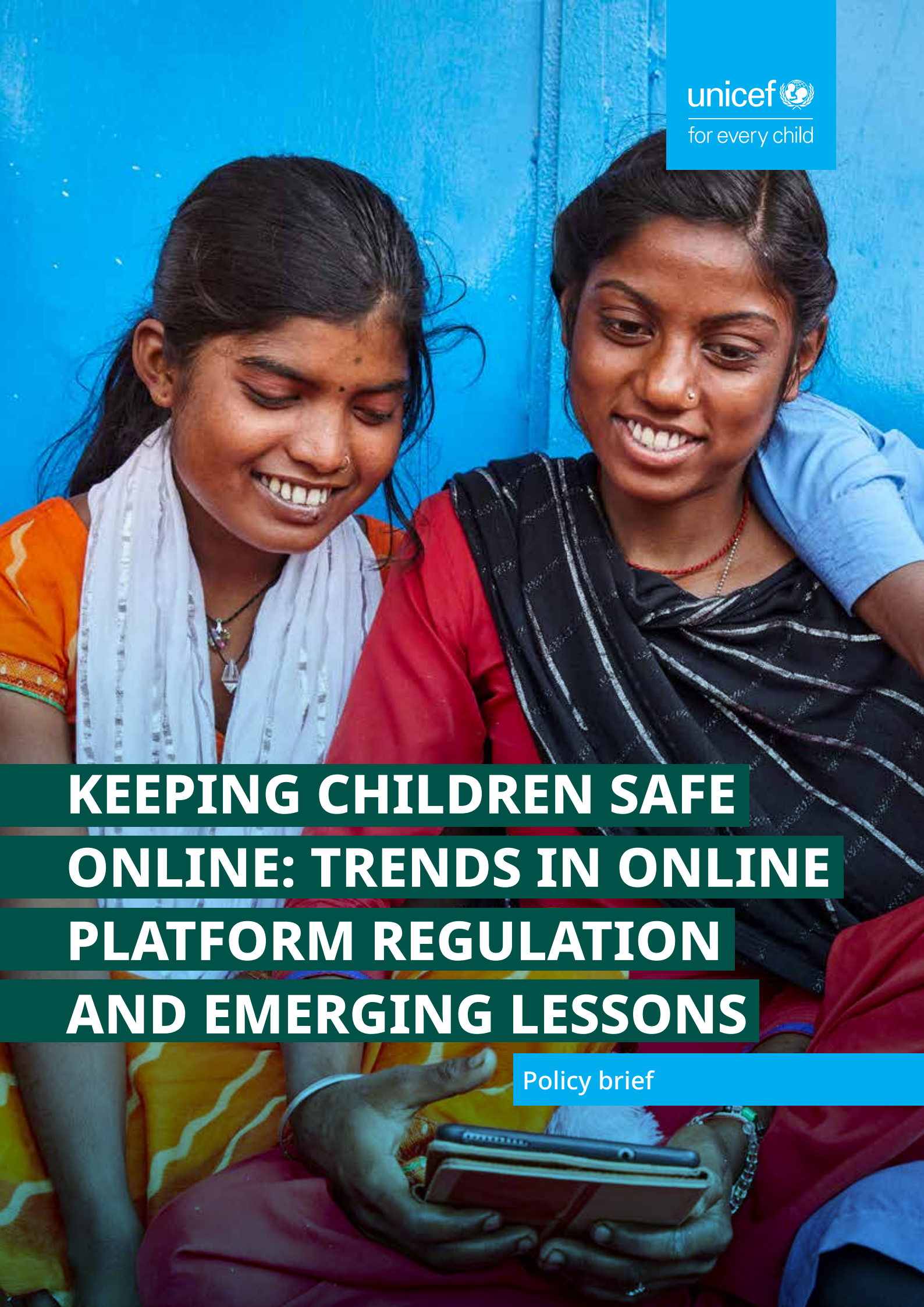


unicef
for every child

KEEPING CHILDREN SAFE ONLINE: TRENDS IN ONLINE PLATFORM REGULATION AND EMERGING LESSONS

Policy brief

ACKNOWLEDGEMENTS

This policy brief was developed by Afrooz Kaviani Johnson, Child Protection Specialist, and Josianne Galea Baron, Programme Specialist, under the supervision of Stephen Blight, Senior Adviser - Child Protection, and Ida Hyllested, Senior Adviser - Business Engagement and Child Rights. It is based on research and an initial draft by Sabine Witting and Emma Day, Tech Legality. We extend our thanks for feedback from UNICEF colleagues, Daniel Kardefelt Winther, Karla Correa, Ahustosh Sharma, and Aislu Bekmussa, and recognise the Expert Advisory Group, listed below, for their valuable insights and expertise through the process.

Esther Ruiz, Rachel Harvey and Steven Edwin Vosloo – UNICEF
Gayatri Khandhadai - Business and Human Rights Resource Centre
Gianclaudio Malgeiri - Leiden University
Jenny Jones - GSMA
Juan Carlos Lara Gálvez - Derechos Digitales
Marie-Ève Nadeu, Sami Jaber – 5Rights
Malavika Jayaram – Digital Asia Hub
Thobekile Matimbe – Paradigm Initiative

About this Policy brief

The document draws on a comparative analysis of online platform regulations in six jurisdictions, namely Australia, the European Union, India, Kazakhstan, South Africa, and the United Kingdom. The analysis aimed to identify and summarise key trends and best practices in online platform regulation addressing children's rights and safety.

This document reflects information available as of June 2025. While the authors have endeavoured to verify the accuracy and timeliness of the sources included, subsequent legislative and regulatory amendments may not be reflected.

Suggested citation: United Nations Children's Fund, 'Keeping children safe online: Trends in online platform regulation and emerging lessons, Policy brief', UNICEF, New York, December 2025.

TABLE OF CONTENTS

KEY MESSAGES

1. **Digital technology offers children opportunities but also exposes them to risks and harms.** Online 'platforms'[1] – such as social media, gaming, and video-sharing services – can help children learn, connect, and express themselves. But they also expose children to serious risks and harms, including exploitation and abuse. Governments around the world are responding by introducing laws and regulations to make digital spaces safer and more rights-respecting for all. Dedicated measures are required to ensure online platform regulation works for children.

2. **Businesses must respect children's rights in the digital environment.** Digital products, services, and business practices can negatively impact children's rights. Governments need to implement a 'smart mix' of measures – including laws, policies, incentives, and voluntary standards – to ensure all businesses operating in the digital environment respect human rights, including children's rights. This policy brief focuses on platform regulation as a key area of current regulatory development.

3. **Platform regulation must be grounded in international human rights and children's rights law.** Effective regulation should aim to prevent business-related abuses of children's rights, in line with international human rights law – including the UN Convention on the Rights of the Child.[2] These rights are universal, indivisible, and apply to all children, everywhere. While there is no 'one-size-fits-all' model and regulatory approaches should reflect national legal systems and capacities, contextualisation must not compromise human rights and children's rights standards.

4. **Regulation must be evidence-based and informed through meaningful engagement with children and young people.** Online platform regulation should be shaped by the lived experiences of users, including children, and supported by research or data to the greatest extent possible to make it relevant, effective, and responsive to context. Where an action may cause harm, but scientific evidence has yet to be established, governments may apply the precautionary principle to protect children's rights in the digital environment. However, its application must ensure that measures do not themselves create unintended negative consequences for human rights and children's rights.

5. **Eight common business obligations are emerging across jurisdictions.** While drivers and goals behind regulatory frameworks differ vastly, our comparative legal analysis identified a set of common business obligations present in at least half of the jurisdictions under review.[3] These include both preventative/systemic measures aimed at embedding safety and rights protections into platform design, and more reactive/content-focused measures that address illegal or harmful content and user complaints.

   The common business obligations include:

   a. Safety- and/or privacy-by design

   b. Impact and/or risk assessments

   c. Age assurance mechanisms/content age gating

   d. Proactive detection of illegal and/or harmful content

   e. Notice-and-action mechanisms

f. Mandatory reporting of illegal content

g. Complaints procedures

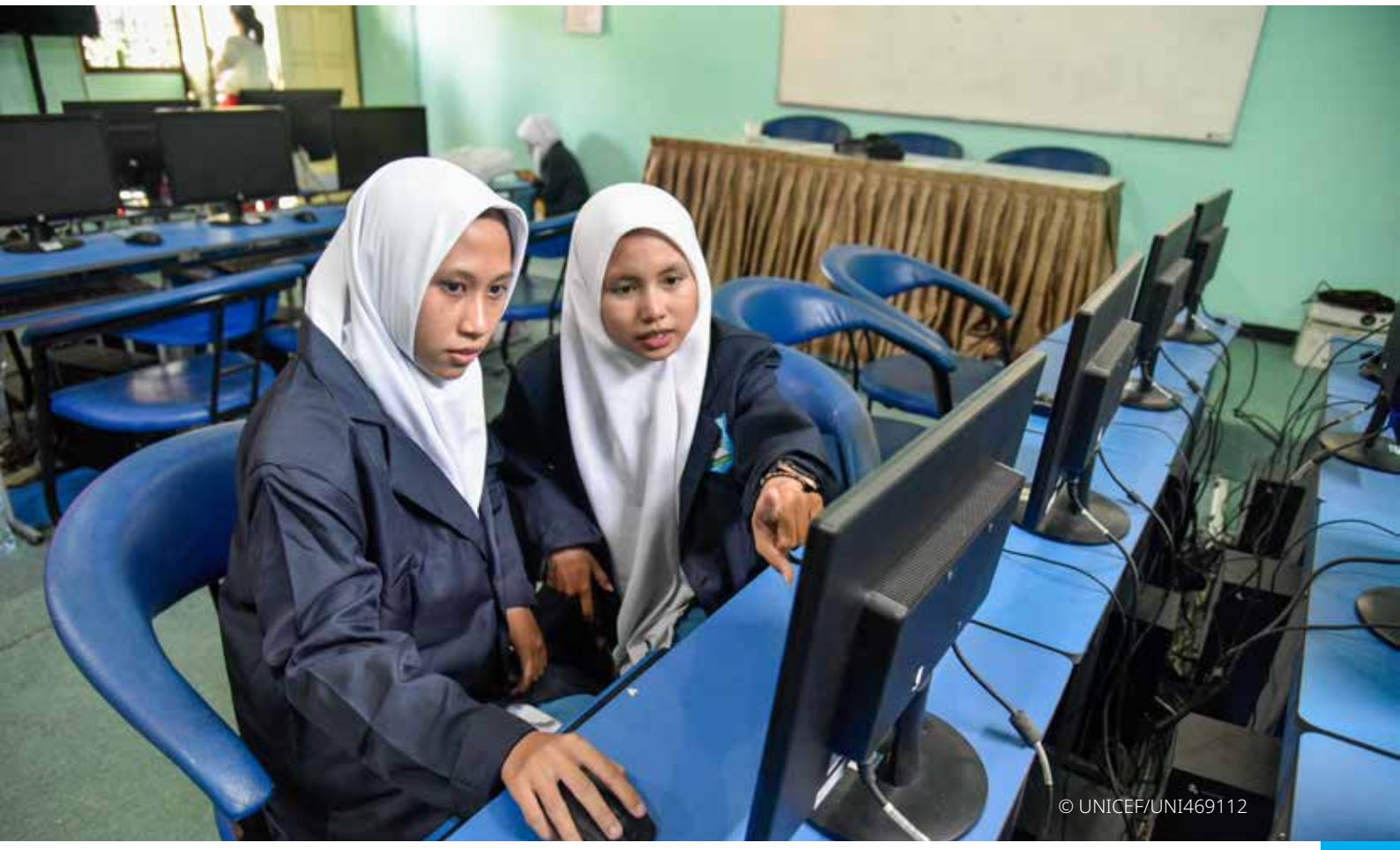
h. Transparency reporting.

This policy brief explores how each of these obligations relate to human rights, including children's rights, and how they may be adapted to different regulatory contexts. See Section 3.

6. **Regulation must ensure all affected rights reach their maximum potential.** Platform regulation affects a wide range of rights, such as privacy, freedom of expression, and protection from harm. Policymakers need to carefully assess the necessity and proportionality of each regulatory measure to ensure the maximum realisation of all affected rights. Regulations and accompanying enforcement mechanisms should also include safeguards to ensure that platforms implement measures in ways that respect the full range of human rights, including children's rights.

7. **Regulation should be developed through coordinated, cross-sectoral collaboration.** Effective regulation requires input from multiple ministries and agencies, including those responsible for children, human rights, technology, data protection, education, and justice. No single body holds all the expertise needed. Cross-sectoral coordination can ensure that regulation is both fit for purpose and practically enforceable, reflecting the complexity of digital environments and children's rights and needs.

8. **Strong enforcement is essential for effective regulation.** To ensure platforms comply with regulations, governments need enforcement mechanisms with legal authority and dedicated financial and human resources. In low-resource settings, regional cooperation, partnerships, and participation in global networks, offer opportunities to strengthen enforcement.



01. INTRODUCTION

The age of digital technology has brought tremendous benefits for children – opening up new possibilities for learning, connection, play, and self-expression. Online platforms are central to this experience, shaping how children interact in the digital environment. Yet these same technologies also expose children to serious risks and harms. While online platforms can support the realisation of children's rights, they may also cause, contribute, or be directly linked to child rights abuses.

In line with the UN Guiding Principles on Business and Human Rights[4] (UNGPs) and the Child Rights and Business Principles[5] (CRBPs), businesses have a responsibility to respect children's rights and prevent and remedy children's rights abuses. Governments around the world are now passing laws and regulations to make digital spaces – especially online platforms – safer and more rights-respecting for all.

1.1 METHODOLOGY AND SCOPE

Governments need to implement a 'smart mix' of measures – including laws, policies, incentives, and voluntary standards – to ensure all businesses operating in the digital environment respect human rights, including children's rights.[6] This policy brief focuses on platform regulation as a key area of current regulatory development. It builds on existing human rights-based guidance for policymakers and elaborates on the children's rights dimensions.[7]

This brief is based on a comparative legal analysis of platform regulation across six jurisdictions and is grounded in international human rights frameworks and authoritative guidance. In particular, it references **General comment No. 25 of the UN Committee on the Rights of the Child**, which outlines how States should apply the UN Convention on the Rights of the Child in relation to the digital environment. The Committee emphasises that States must ensure businesses operating in the digital environment respect children's rights, for example through the development, monitoring, implementation, and evaluation of legislation, regulations, and policies.[8]

This brief is structured in three sections. Section 1 explains what online platform regulation is and how it affects children's rights, including their right to protection from violence, abuse and exploitation. Section 2 highlights global regulatory trends as of June 2025. Section 3 offers key considerations and guidance for developing and implementing online platform regulation that works for children. Rather than prescribing a specific model, this brief provides a child rights-based analysis of global developments with the aim of supporting policymakers design locally relevant, globally aligned regulatory frameworks.

1.2 WHAT ARE ONLINE PLATFORMS?

This brief uses the term online 'platform' as shorthand to capture a broad range of digital environments where children interact with others and with content. As there is no agreed international definition of online platforms, the scope of services covered under regulations can differ quite considerably across jurisdictions.

The following definition of online platform is used in this brief:

> *'digitally enabled product that mediates relationships between two or more parties, usually featuring technical elements that allow third parties to build upon it or interact with it'.*[9]

The definition acknowledges that most online platforms aim to generate profit, although some are run as non-profits. Further, online platforms are not mere passive entities, but shape users' experience, how people use them and how they interact with one another – for example, by recommending content or profiles through algorithmic systems. See **Figure 1**.

In contrast, some digital products and services typically do not count as online platforms. Examples include:

- **Single-sided services:** These deliver services directly to a single user group without enabling interactions, for example a Weather app.
- **Internet infrastructure services:** These provide the technical foundation that makes digital services and products function, for example internet service providers (ISPs).

Figure 1 Broad platform categories and platform types within these categories[10]

Platform Category	Platform Types
Marketplaces	E-Commerce, App Stores, Online Labor Markets
Communication	Peer-to-Peer Messaging, Feed-Based Social Networks, Bulletin Boards
Entertainment	Copyrighted Content Streaming, User-Generated Content Streaming
Information Retrieval	Search, Wiki
Business-to-Consumer Software Services	Consumer Cloud, Payment
Business-to-Business Software Services	Internet Infrastructure, Enterprise Cloud, Enterprise Payment
'Locally Tethered' Services	Accommodation, On-Demand Transport, Food Delivery

AI regulation and online platform regulation

Artificial Intelligence (AI) technologies are becoming embedded in multiple domains, activities, and services that children rely on. Proactive regulation – anchored in children's rights law – is essential to ensure that AI technologies support, rather than undermine, children's safety, development, and well-being.

AI regulation and online platform regulation often differ in focus and scope. AI regulation typically focuses on the technology itself – its safe development, deployment and maintenance. In contrast, platform regulation tends to focus on how users interact with content and with each other within specified services. Because of these differences, many jurisdictions treat them as separate regulatory fields.

However, **overlaps do exist**, such as when AI is deployed on an online platform for tasks like content moderation. Whether or not AI use by platforms is regulated through separate AI regulations or as part of platform regulation depends on the regulatory context. If a country does not have any AI regulation, it might include guardrails for AI use by platforms within platform regulations instead.

Further reading: UNICEF's ***'Artificial Intelligence Governance in Motion: A rapid global review of AI regulation and its implications for children's rights'*** provides an overview of international, regional, and national AI frameworks as of April 2025. The review calls for ongoing research to monitor and assess the impact of AI regulation on children across different contexts. It warns against a 'wait-and-see' approach, arguing that the consequences for children may be more difficult to undo than to prevent.

1.3 HOW DO ONLINE PLATFORMS AFFECT CHILDREN'S RIGHTS AND SAFETY?

Online platforms pose a wide range of risks for children's rights and safety. The '5Cs framework'[11] provides a helpful categorisation of content, contact, conduct, consumer, and cross-cutting risks (see **Figure 2**). These encompass, among other things, violent content, cyberaggression and harassment, gambling, sexual exploitation and abuse, and the promotion of or incitement to suicide or life-threatening activities.[12] The UN Committee on the Rights of the Child affirms that States should take all appropriate measures to protect children from these risks.[13]

This brief primarily focuses on regulation addressing contact, conduct, content and cross-cutting risks – recognising that these categories often overlap. While commercial incentives play a significant role in shaping the design and operation of platforms, a detailed analysis of consumer risks falls outside the scope of this brief. Addressing these issues requires engagement with additional regulatory domains, such as data protection and consumer protection. For a deeper exploration of consumer risks, including those linked to digital marketing, UNICEF's complementary research and guidance may be consulted.[14]

Figure 2 **How online risks manifest in the digital environment**[15]

Risks for children in the digital environment				
Risk categories	**Content risks**	**Conduct risks**	**Contact risks**	**Consumer risks**
Risk manifestations	Hateful content	Hateful behaviour	Hateful encounters	Marketing risks
	Harmful content	Harmful behaviour	Harmful encounters	Commercial profiling risks
	Illegal content	Illegal behaviour	Illegal encounters	Financial risks
	Disinformation	User-generated problematic behaviour	Other problematic encounters	Security risks
Cross-cutting risks	Privacy risks (interpersonal, institutional & commercial)			
	Advanced technology risks (e.g. AI, IoT, predictive analytics, biometrics)			
	Risks on health & wellbeing			

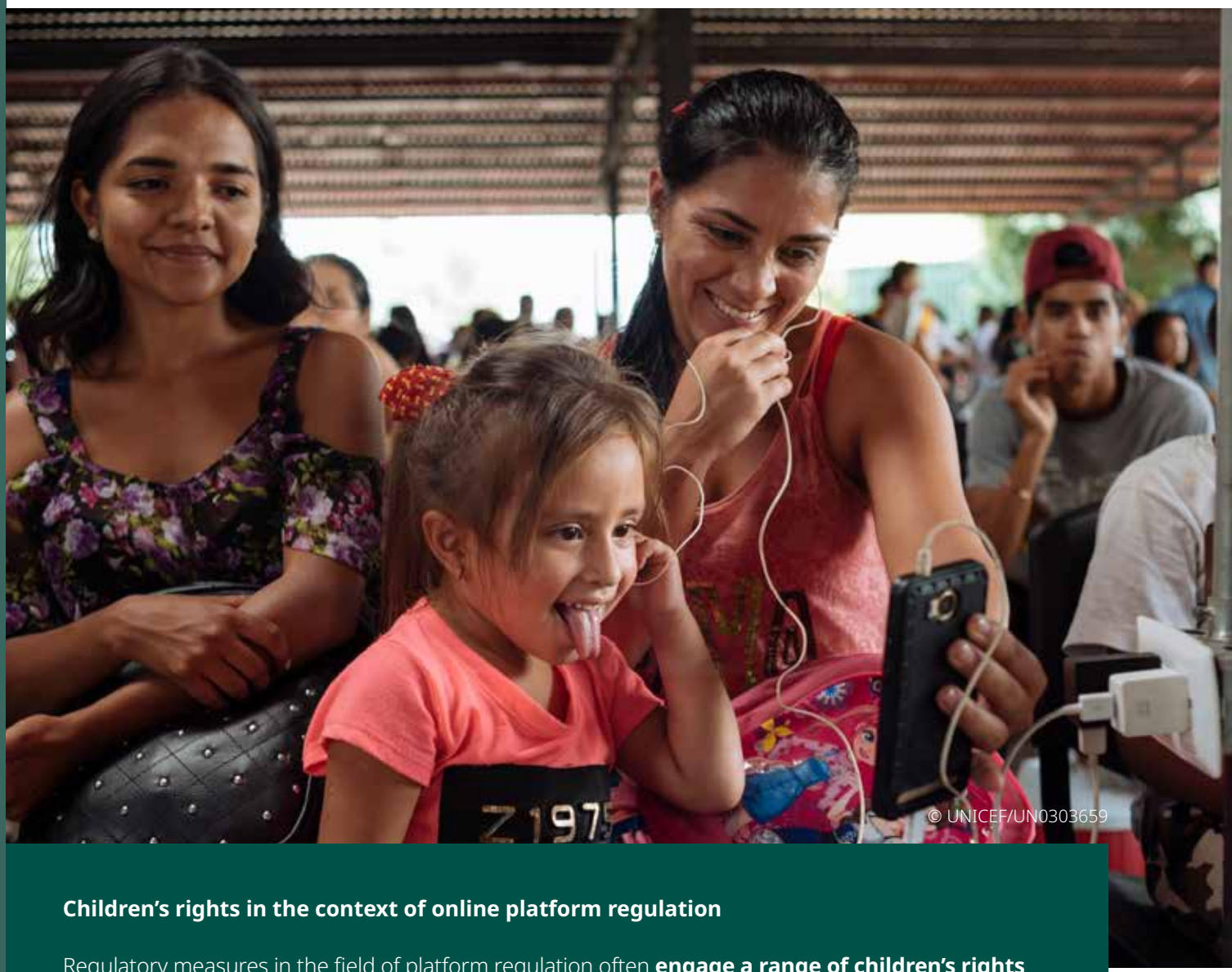
1.4 WHAT IS DRIVING ONLINE PLATFORM REGULATION AROUND THE WORLD?

Governments have historically taken different approaches to online platform regulation. These range from industry self-regulation through non-binding codes of conduct, to co-regulatory models where industry-developed standards are reviewed and adopted by regulators, to traditional 'top-down' regulation led by the state.[16] Certain forms of platform regulation, such as intermediary liability, have existed for some time. Safety-focused regulation is a relatively recent development.

Current trends reflect a growing role of the state as an active regulator, with national scrutiny of platforms intensifying and increasingly focused on the regulation of risks associated with platforms.[17] This trend is also fuelled by a broad perception of the failure of voluntary measures and accounts of numerous high-profile whistleblowers.[18]

The regulation of online platforms is not solely a technical or legal matter – it is also political. Online platforms are now central to public debate, the economy, and social life, including for children. They do more than just host information. Online platforms decide how content is created, shown, amplified, shared and accessed, through a combination of content moderation, algorithmic curation and recommendation systems. By enabling wide content sharing and interaction, they can advance democratic participation, including for children. But they can also influence and distort people's opinions through their design and/or misuse.

From a children's rights perspective, the goal of online platform regulation should be to prevent child rights abuses and bring the full range of children's rights to the maximum potential.[19] Regulatory frameworks should protect children from online harms, while still allowing them to benefit from the opportunities online platforms offer. Online platform regulation should carefully consider the impact of measures on human rights, including children's rights.



Children's rights in the context of online platform regulation

Regulatory measures in the field of platform regulation often **engage a range of children's rights** as set out in the UN Convention on the Rights of the Child (CRC). These include but are not limited to:

- Right to protection from all forms of violence, sexual abuse and exploitation (Articles 19 and 34 CRC)
- Right to freedom of expression (Article 13 CRC)
- Right to privacy (Article 16 CRC)
- Right to access to information (Article 17 CRC)
- Right to protection from economic exploitation (Article 32 CRC)
- Right to play and leisure (Article 31)
- Right to non-discrimination (Article 2 CRC)
- Best interests of the child (Article 3 CRC)
- Right to life, survival and development (Article 6 CRC)
- Right to be heard (Article 12 CRC)
- Respect the child's evolving capacities (Article 5 CRC).

Regulatory measures affect not only children's rights as set out in the CRC, but also the human rights of all users, including children. These include the rights to **freedom of expression, privacy and access to information** as set out in the International Covenant on Civil and Political Rights.

02. KEY TRENDS IN ONLINE PLATFORM REGULATION

This section draws on a comparative legal analysis of online platform regulation in six jurisdictions undertaken in June 2025, and insights from an international Expert Advisory Group, to identify emerging trends and potential implications for children's rights and safety online. The legal analysis was further informed by relevant international normative frameworks[20] and other authoritative guidance.[21]

The legal analysis identifies and compares key legal and policy approaches to online platform regulation that seek to address risks to children in the digital environment. Australia, the EU, India, Kazakhstan, South Africa, and the UK were selected for their diversity in geography, legal systems, and income levels,[22] as well as differences in how long regulations have been in place and their potential influence regionally and globally.

Table 1 Legislation reviewed

Jurisdiction	Legislation reviewed
Australia	Online Safety Act 2021
European Union	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)
India	Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules 2021)
Kazakhstan	2023 Law on Online Platforms and Online Advertising
South Africa	Film and Publications Act (Act No. 11 of 2019)
United Kingdom	Online Safety Act 2023

2.1 WHAT COMMON BUSINESS OBLIGATIONS ARE PRESENT IN EXISTING REGULATIONS?

While there are significant differences in regulatory scope and approach, the legal analysis identified a set of common business obligations present in at least half of the jurisdictions under review. These span both preventative/systemic measures aimed at embedding safety and rights protections into platform design, and more reactive/content-focused measures that address harmful content and user complaints. This selection of business obligations enabled a comparison between the jurisdictions.

Figure 3 Summary and explanation of common business obligations

Safety- and/or privacy-by-design	Building platforms in ways that prevent harm by protecting users' safety and privacy from the outset
Impact and/or risk assessments	Evaluating how platforms may impact or create risks for users and taking steps to prevent or mitigate these risks
Age assurance mechanisms/ content age gating	Technologies to verify or estimate users' ages so that children are shielded from illegal or harmful content
Proactive detection of illegal and/or harmful content	Use of technological tools to detect and action illegal or harmful content
Notice-and-action mechanisms	Mechanisms that allows users to flag illegal or harmful content, which platforms must then review and act upon
Mandatory reporting of illegal content	Requirement to inform authorities when platforms detect certain types of illegal content
Complaints procedures	Mechanism for users to challenge a platform's decisions, e.g. regarding content removal or account suspension
Transparency reporting	Obligation to publish or share information with authorities on practices such as content moderation or recommender systems

2.2 WHAT REGULATORY APPROACHES ARE USED, AND WHAT CAN WE LEARN FROM THEM?

After identifying the common business obligations, the legal analysis examined which jurisdictions include these obligations and how they are structured. The following key points show where jurisdictions converge or diverge, and how this affects the scope and design of the common business obligations.

- **Underlying regulatory philosophy:** The regulatory philosophy in each country significantly influences the scope and detail of the common business obligations. For example, Australia emphasises child safety through co-regulatory mechanisms and safety-by-design principles, while India and Kazakhstan adopt frameworks that combine safety aims with more centralised oversight of online content.

- **Scope of businesses covered:** The scope of businesses can differ significantly across countries. For example, the impact/risk assessments under EU law only apply to 'Very Large Online Platforms' and 'Very Large Online Search Engines' (i.e. those with more than 45 million monthly users in the EU),[23] while the impact/risk assessments under UK law apply to all providers.

- **Richness in regulatory detail:** Some countries provide significant regulatory details for the common business obligations, while others leave the interpretation to the online platforms. For example, while the proactive detection obligation under Indian law provides detailed guardrails for the deployed technologies and for their impact on human rights, the South African and Kazakh law imply a detection obligation without providing further regulatory detail.

- **User-friendliness as priority:** Some countries require online platforms to design user-facing mechanisms in an accessible and user-friendly manner. For example, the Australian law provides for significant detail for the design of its notice-and-action mechanism, while the Kazakh law leaves this primarily to the online platforms.

- **Content-centric or systemic interventions:** Some countries consider online platform regulation mainly as a framework to manage content, while others aim to also influence the broader structures, policies and practices. For example, the Kazakh and South African law largely focus on preventing users' exposure to certain categories of content, while the EU law is aiming to influence the way online platforms function.

- **Holistic consideration for human rights, including children's rights:** Some countries place significant emphasis on the full scope of human rights, including children's rights, while others primarily focus on child protection. As an example, the EU embeds user safety within its broader fundamental rights framework, while the Australian law focuses on child protection, with limited attention to freedom of expression and the right to privacy.[24]

- **Role of secondary legislation, codes and guidelines:** Some countries set out the main regulatory parameters in the regulatory framework itself, while others rely heavily on accompanying secondary legislation, codes and guidelines. As an example, the main regulatory details for Indian online platform regulation are found in secondary legislation, while the UK covers most regulatory details in its actual law, complemented by extensive guidance for businesses to facilitate compliance.

- **Powers and resources of the enforcement agency:** Some countries provide extensive powers and resources to the enforcement agencies to ensure compliance with online platform regulations. For example, the EU has set up an elaborate enforcement structure with fines up to 6% of annual global turnover for non-compliance, while Indian law does not establish a dedicated monitoring and enforcement regime.

- **Stakeholder involvement:** Some countries actively seek stakeholder input for the creation and implementation of the law. For example, the UK law has undergone extensive public consultation, with additional avenues for stakeholder input for the development of accompanying guidance documents. For the Indian law, the consultation process appears to have been more limited.

The following table provides a simplified overview of the common business obligations (left column) across the reviewed jurisdictions (right columns).

Figure 4 Common business obligations

Common business obligations	Australia	EU	India	Kazakhstan	South Africa	UK
1. Safety- and/or privacy-by-design	● [25]	●	○	○	○	●
2. Impact and/or risk assessments	● [26]	● [27]	○	● [28]	○	●
3. Age assurance/ content age gating	●	● [29]	○	○	● [30]	●
4. Proactive detection of illegal and/or harmful content	● [31]	○ [32]	● [33]	● [34]	● [35]	●
5. Notice-and-action mechanism	●	●	●	●	●	●
6. Mandatory reporting of illegal content	● [36]	● [37]	○	● [38]	●	● [39]
7. Complaints procedures	● [40]	●	●	● [41]	○	●
8. Transparency reporting	●	●	●	●	○	● [42]

Legend
● Expressly included ● Partially included or implied ○ Not included

03. DEVELOPING ONLINE PLATFORM REGULATION THAT WORKS FOR CHILDREN

3.1 WHAT 'BASELINE' EXPECTATIONS SHOULD BE ESTABLISHED?

Online platform regulation – like many other pieces of regulation – needs to meet certain baseline expectations to ensure it indeed addresses the problem it aims to solve. Without these foundational elements, regulation is unlikely to be effective in addressing online risks and harms experienced by children. The following baseline expectations provide a foundation for ensuring that business obligations are embedded in a rights-based, enforceable, context-specific, and evidence-based regulatory framework.

1. **Online platform regulation must explicitly recognise the needs and rights of children.** While online risks and harms are a concern for all, dedicated measures are required to ensure that the safety and protection of children.

2. **Online platform regulation must be grounded in international human rights and children's rights law.** Effective regulation should aim to prevent business-related abuses of children's rights, in line with international human rights law. These rights are universal, indivisible, and apply to all children, everywhere. While there is no 'one-size-fits-all' model and regulatory approaches should reflect national legal systems and capacities, contextualisation must not compromise child rights standards.

3. **Regulation must be evidence-based and informed by meaningful engagement with children and young people:** Online platform regulation should be grounded in the lived experiences of users, especially children, and supported by research or data to the greatest extent possible. This evidence is crucial to ensure that the regulation is relevant and effective in the specific context. Where evidence is limited, governments may act under the *precautionary principle*. This principle states that where an action may cause harm to the public or the environment, but scientific evidence has yet to be established, steps should be taken to prevent or mitigate the harm. However, its application must ensure that measures do not themselves create unintended negative consequences for children's rights.

4. **Regulation should directly address the issues it aims to resolve:** Online platform regulation should clearly demonstrate how it responds to the specific challenges children face. As outlined in the 5Cs framework (see **Figure 2** above), these include contact, conduct, content, consumer, and cross-cutting risks. Effective regulation should be informed by the best available research and children's lived experiences, so that it is relevant and responsive to real needs. This helps ensure that measures are meaningful and not shaped by unrelated political priorities. Measures to ensure ongoing monitoring of effectiveness are also critical such as allowing access to data systems by independent auditors of researchers to determine efficacy of solutions.

5. **Regulation must clearly differentiate and define 'harmful' and 'illegal' content:** While platform design and commercial incentives are critical to online safety, regulation often deals specifically with how online platforms handle harmful and illegal content. These terms must be clearly defined. Not all harmful content is illegal under international human rights law. Freedom of expression protects even speech that may offend or upset people, and only very extreme types of expression may be lawfully restricted.[43] As such, governments must ensure any restrictions are clearly defined, necessary, and proportionate. Vague or overly broad rules risk limiting human rights, including children's rights, such as the rights to freedom of expression and access to information.[44]

6. **Regulation must cover the businesses responsible for creating safer online platforms:** Online platform regulation should apply to all online platforms that children use or which are likely to impact their rights. Further, regulation should cover all online platforms offering services in the relevant jurisdiction, including multi-national companies headquartered elsewhere.

7. **Regulations should contain measures for ensuring their enforceability:** Establishing a regulation enforcement mechanism with sufficient legal powers, expertise, and dedicated financial and human resources is crucial to ensure online platforms' compliance. Simply enacting a law or regulation does not bring the changes envisaged by the legislator. Only if laws and regulations are also enforced do such changes become tangible.

8. **Regulation should be developed through coordinated, cross-sectoral collaboration.** Effective regulation requires input from multiple ministries and agencies, including those responsible for children, human rights, technology, data protection, education, and justice. No single body holds all the knowledge needed, so coordination ensures the law is fit for purpose and practically enforceable.

3.2 COMMON BUSINESS OBLIGATIONS: IMPACTS ON CHILDREN'S RIGHTS AND SAFETY, AND KEY POLICY CONSIDERATIONS

3.2.1 Safety- and/or privacy-by-design

What is it?

Design approaches, such as safety-by-design or privacy-by-design, seek to ensure that digital products and services respect human and children's rights by-design. The UN Committee on the Rights of the Child explains –

> *'States parties should require all businesses that affect children's rights in relation to the digital environment to implement regulatory frameworks, industry codes and terms of service that adhere to the highest standards of ethics, privacy and safety in relation to design, engineering, development, operation, distribution and marketing of their products and services.'*[45]

This affirms the expectation that privacy and safety considerations should be embedded throughout the entire product cycle – including design. These design approaches are not mutually exclusive but

rather reinforcing. Contemporary regulatory trends increasingly demand their simultaneous integration to ensure comprehensive and rights-respecting user protection.

- **Privacy-by-design:** Privacy-by-design aims to prevent privacy abuses before they occur. It integrates privacy protections into the design and operation of digital products and services from the outset. It is codified in the EU General Data Protection Regulation (GDPR)[46] and emphasises data minimisation, purpose limitation, access controls, and default settings that safeguard user data.
- **Safety-by-design:** Safety-by-design aims to build digital products and services that prevent online harms from the outset, particularly for children and vulnerable users. This approach gained global attention through Australia's eSafety Commissioner. Safety-by-design includes integrating service provider responsibility, user empowerment and autonomy, and transparency and accountability as core principles of this approach.[47]
- **Child-rights-by-design** is another by-design approach that has emerged to uphold children's rights in the digital environment. A child-rights-by-design approach aims to apply the range of rights set out in the CRC within the design, development, and execution of online services or products used by children.[48]

Key considerations

Design approaches can be considered essential features to safeguard children's rights and safety on online platforms. When effectively applied, they shift the regulatory emphasis from after-the-fact accountability to before-the-fact responsibility – embedding rights, including safety and privacy, into the design and operation of online platforms from the outset.

However, poorly *implemented* design measures may have the opposite effect. It is therefore important to examine not only whether platforms adopt such approaches, but how they are put into practice. Design approaches should never serve as a pretext for measures that conflict with international human rights and children's rights law. As an example, a platform automatically sharing all users' location data with law enforcement without consent cannot be justified as a 'safety-by-design' feature. While intended to enhance the safety of users, such a measure would clearly violate users' right to privacy and personal data protection.

To address these implementation risks, legal or regulatory obligations could require online platforms to assess potential adverse impacts on children's rights and human rights, and enforcement agencies could issue guidance on the appropriate implementation of design approaches. Independent monitoring, including by national human rights institutions or independent auditors, can also play an important role in identifying and addressing poorly implemented measures.

3.2.2 Impact and/or risk assessments

What is it?

Impact and/or risk (i.e. potential impact) assessments are key tools for **identifying, preventing, and mitigating potential adverse impacts** of business activities on human rights and children's rights. Under the UNGPs and the CRBPs, all companies, including those developing, deploying, and using digital technologies, have a responsibility to identify and address the adverse human rights impacts with which they are involved – both online and offline.[49]

Governments are increasingly including impact or risk assessments in platform regulation. The purpose and scope of such assessments differ considerably, with some assessing whether and how children use a specific platform, and which risks they might encounter, and others more explicitly looking at how a platform impacts children's rights or human rights more broadly.

Human rights impact assessments (HRIAs) and/or Child Rights Impact Assessments (CRIAs) in line with the UN Guiding Principles on Business and Human Rights constitute best practice in this space. The UN Committee on the Rights of the Child highlights the importance of legally mandating CRIAs for businesses –

> *'States parties should require the business sector to undertake child rights due diligence, in particular to carry out child rights impact assessments and disclose them to the public, with special consideration given to the differentiated and, at times, severe impacts of the digital environment on children.'*[50]

HRIAs and CRIAs should form part of businesses' broader, continuous human rights due diligence (HRDD) efforts. HRDD requires companies not only to carry out HRIAs or CRIAs but also to act on findings, track responses, and communicate externally.

- A **HRIA** is a process for identifying, understanding, assessing and addressing the actual or potential adverse effects of a business project or business activities on the human rights enjoyment of impacted rightsholders.[51] If not conducted in conjunction with a CRIA, the HRIA should include a dedicated and comprehensive assessment of child rights dimensions.

- A **CRIA** specifically analyses the impacts of business operations, products or services on children's rights. CRIAs can complement HRIAs processes, or be conducted separately.[52] Ideally, CRIAs are conducted in conjunction with or built on a HRIA, to ensure that the CRIA assesses the impact of the business activity on human rights more broadly. Reference may be made to UNICEF's D-CRIA Toolbox, a detailed guide on conducting CRIAs in relation to the digital environment.

Key considerations

HRIAs/CRIAs are a key regulatory intervention which can help online platforms, enforcement agencies and the general public to better understand the impact of a specific business on the full range of human rights and children's rights and mitigate adverse impacts.

At the same time, there are some potential risks – not inherent to the requirement to conduct a HRIA or CRIA itself, but rather related to the quality and implementation of such assessments:

- HRIAs/CRIAs as 'tick-box'-exercise: Companies might see HRIAs/CRIAs as something they do simply to meet a requirement, without engaging with the process or its purpose meaningfully. To address this risk, enforcement agencies should set quality standards for these assessments, including guidance on methodology and structure. Robust oversight and consistent enforcement action are essential to hold companies accountable when they fail to meet these standards.

- Risk mitigation strategies not in line with international human rights and children's rights: Companies might develop risk mitigation strategies which are not in line with international human rights and children's rights. Risk mitigation measures need to pursue a legitimate aim, be necessary, and proportionate. To address this risk, enforcement agencies should thoroughly review HRIAs/CRIAs and request companies to take corrective action when quality issues are detected.

3.2.3 Age assurance / content age-gating

What is it?

Age assurance obligations aim to identify the age or age range of users online to help platforms tailor experiences and protection. For children, this may involve restricting access to specific content, features, or online platforms entirely, with the aim of directing them to age-appropriate experiences and protecting them from content, conduct, contact, consumer and cross-cutting risks.

There are three main approaches to age assurance, each with different levels of reliability, intrusiveness, and data protection implications:

- **Self-declaration:** Users provide their age or confirm their age range, either by voluntarily providing their date of birth or age, or by declaring themselves to be above a certain age, typically by clicking on a button online.
- **Age estimation:** Methods which allow a provider to establish that a user is likely to be of a certain age, to fall within a certain age range, or to be over or under a certain age.
- **Age verification:** Systems that rely on physical identifiers or verified sources of identification that provide a high degree of certainty in determining the age of a user.[53]

The main difference between age estimation and age verification lies in their level of accuracy and the degree of privacy risks involved. Age estimation provides an approximate age, while age verification provides the exact age of the user.[54]

Key considerations

Age assurance can offer significant benefits for children's experience in the digital environment. However, age assurance tools can also carry considerable risks for human rights, including children's rights. Whether age assurance is an appropriate solution depends on the identified use case, the benefits and risks associated with age assurance in a specific context, the nature of the age assurance tool, and how age assurance is implemented.

Once the use case, benefits and risks have been determined, a proportionate response is chosen. This means weighing the benefits against the risks, determining whether age assurance is necessary and proportionate. If the age assurance tool is deemed necessary, it must be proven to be effective and the least intrusive measure. Platforms should be able to demonstrate how knowing a user's age or age range will result in meaningful safeguards.

Any age assurance solution must be rights-respecting and proportionate to the risk and/or harm it is deployed to address. Age assurance carries considerable risks for human rights, including children's rights:

- Processing of (sensitive) personal data: Age assurance often requires the processing of (sensitive) personal data, such as biometric information, official identity documents, or behavioural patterns. This raises concerns about privacy and data protection. To mitigate these risks, age assurance providers should conduct a data protection impact assessment (DPIA) to better understand their data processing practices and to come up with mitigation measures for any identified risks.[55] Such a DPIA should consider the specific vulnerabilities of children and other groups in the respective context wherever the age assurance tool will be deployed.

- Risk of discrimination for vulnerable groups: Age assurance measures should be consistently effective across a diverse range of users – that is, they should correctly assess a user's age regardless of differences in age group, gender, race, or other characteristics. Age assurance technologies that are AI-driven rely on training data to develop their algorithms. If this data is biased, for example by over-representing certain age groups, genders or racial backgrounds, the resulting tools might suffer from biases. To mitigate this risk, regulators should require age assurance providers to use diverse and representative training datasets, to conduct regular bias testing and independent audits, and to adopt privacy-respecting alternatives where possible.

- Chilling effect on freedom of expression: When users are compelled to disclose their identity or undergo checks which involve personal data collection, they may refrain from engaging in lawful and socially valuable expression, especially in contexts where engagement with certain types of content (for example, on sexual orientation) can lead to persecution. Ensuring that the age assurance requirement is necessary and proportionate is essential to mitigate this risk.

Importantly, age assurance is a rapidly evolving area, with new technologies entering the market every year. Assessing the necessity and proportionality of different kinds of age assurance tools requires expertise from technologists, child protection specialists, and data protection experts. National data protection authorities should be consulted at an early stage where age assurance is considered.

3.2.4 Proactive detection of illegal and/or harmful content

What is it?

Proactive detection obligations require online platforms to **actively monitor and identify certain types of content.** These obligations are particularly common in the context of illegal content, such as child sexual abuse material (CSAM) or content designated as 'terrorist' but are sometimes also deployed for harmful content.

Proactive detection obligations typically include automated content scanning of publicly or privately shared content, or the use of so-called 'upload filters', i.e. scanning of content before it is uploaded to the online platform. Such obligations are usually implemented with the support of detection technologies, which are designed to detect illegal or harmful content in photos, videos, or live-streamed content, as well as in text.

Key considerations

The term 'proactive detection' is an umbrella term which covers different detection approaches. These rely on a variety of technologies to be implemented. Proactive detection obligations in online platform regulation laws should be carefully considered on an individual basis, taking into account their scope and scale and the severity of adverse impacts on human rights and children's rights.

Regulators need to ensure such measures are necessary and proportionate and bring all affected rights to their maximum potential, considering the following parameters:

- Is the detection obligation targeted or general and indiscriminate? Detection obligations can be targeted, applying only in cases of concrete suspicion, or general and indiscriminate, applying to all users. If the detection obligation is general and indiscriminate, concerns around mass

surveillance have been raised which is not considered in line with international human rights law.[56] This might also have a so-called 'chilling effect' on the right to freedom of expression as users may self-censor for fear of being constantly tracked.[57]

- Does the detection obligation focus on publicly available content or content in private communications? Detection obligations can target publicly available content and/or content shared in private communications. Proactive detection obligations targeting private communications are considered a particularly severe interference with the right to privacy. This level of scrutiny is further heightened when the private communication takes place in an end-to-end encrypted (E2EE) environment. There is an ongoing debate amongst stakeholders whether detection of content in E2EE communications is technically feasible, and whether this can be done in a privacy-respecting manner.[58]

- Does the detection obligation focus on known illegal content and/or potentially illegal content? Detection obligations can focus on known illegal content and/or potentially illegal content. The technologies used for detecting these types of content considerably differ in terms of their accuracy rate. If technologies with a low accuracy rate are deployed, legal content might wrongfully be categorised as illegal. This can have an adverse impact on freedom of expression and access to information.

3.2.5 Notice-and-action mechanisms

What is it?

A notice-and-action mechanism is a formal process that **allows individuals or entities to notify an online platform about potentially illegal content or conduct that violates the platform's terms of service.** Upon receiving this notice, the online platform is expected to assess its validity and respond in a timely, proportionate and legally justified manner. This may involve removing the content, disabling access, or determining that no action is warranted if the content does not breach legal or policy standards.

Importantly, content moderation is not limited to a binary choice between removing or retaining content. Platforms have a range of options at their disposal, such as restricting access to users above a certain age (e.g. over 18), demoting content in recommendation systems, or disabling sharing functions.

Key considerations

Notice-and-action mechanisms are key safety tools. They offer children – and others acting on their behalf – a channel to report negative experiences. As emphasised by the UN Committee on the Rights of the Child, such mechanisms 'should be free of charge, safe, confidential, responsive, child-friendly and available in accessible formats.'[59]

Regulation should set out parameters for their design and operation, including:

- Interfaces that are user-friendly and accessible, with clear and easy to understand instructions on the process of submitting and receiving responses to notices;

- Multiple ways to submit notices on the platform; and

- Mechanisms to allow users to easily follow up on reports, such as through case tracking numbers.

While notice-and-action mechanisms are essential tools for online safety and accountability, it is important to be aware of potential risks – particularly when these mechanisms are used to target content that constitutes protected speech under international human rights and children's rights law:

- National criminal law that violates freedom of expression: In some jurisdictions, national criminal law classifying content as 'illegal' may not align with international human rights and children's rights law, for example when laws refer to vague and overly-broad language.[60] In such cases, notice-and-action mechanisms risk becoming tools for the removal of vast amounts of protected speech under the pretext of 'illegal content'. This risk can only be mitigated by ensuring that national laws are aligned with international human rights law on freedom of expression.

- Difficulties in determining illegality or harmfulness: Online platforms can struggle to accurately assess whether specific content is illegal or harmful, especially at scale. Automated content moderation tools lack contextual understanding,[61] while human moderators are frequently under pressure to make rapid decisions.[62] In ambiguous cases, platforms might err on the side of caution and remove content to avoid liability, leading to over-removal and potential infringements on freedom of expression. This risk emphasises the need for strong regulatory oversight of content moderation practices.

3.2.6 Mandatory reporting of illegal content

What is it?

Mandatory reporting mechanisms require online platforms to **report (potentially) illegal content to law enforcement for further investigation.** Such reporting mechanisms are often limited to certain categories of illegal content, such as serious criminal offences, but can also apply to any category of criminal content.

Key considerations

Mandatory reporting obligations can be a powerful tool to ensure that online platforms not only remove illegal content but also refer it to law enforcement for further investigation and intervention. This is especially critical in cases of technology-facilitated child sexual abuse and exploitation, where the child may still be at risk and require urgent identification and safeguarding.

Mandatory reporting obligations can play a critical role in keeping children safe, but its impact depends on how it is implemented. Risks to children's rights arise not from the existence of mandatory reporting, but from how platforms operate and manage these obligations. For example:

- Failure to report incidents: If an online platform fails to report illegal content that falls within the scope of mandatory reporting, children who are still in danger may not receive the urgent protection and assistance they need. To prevent this, enforcement agencies must ensure that online platforms consistently report all relevant content to law enforcement, enabling timely victim identification, assistance, and safeguarding.

- Wrongful reporting of legal content: Risks to freedom of expression may arise if platforms mistakenly report legal content to law enforcement. This can happen when content assessment benchmarks are applied inconsistently or arbitrarily – or are not aligned with international human rights law. Content assessment benchmarks, such as national laws, community guidelines or terms of service, should align with international human rights and children's rights law.

3.2.7 Complaints procedure against the online platform

What is it?

A complaints procedure against the online platform is a **grievance mechanism which allows users of an online platform to lodge complaints** when they feel wronged by the online platform itself, typically about unfair decisions, such as account suspension, or content removal.

Key considerations

Effective and user-friendly complaints procedures and grievance mechanisms are essential for online platforms to uphold user rights and promote accountability. As discussed in relation to notice-and-action mechanisms (section 3.2.5), the UN Committee on the Rights of the Child explains that such mechanisms should be 'free of charge, safe, confidential, responsive, child-friendly and available in accessible formats.[63]

Regulation should set out specific parameters on the design and operation of such complaints procedures as set out in section 3.2.5 on notice-and-action mechanisms. In addition, complaints procedures should notify users and explain the appeal process when their content is removed or restricted and provide access to appeals and escalation pathways.

While such mechanisms are essential, it is important to be aware of how their implementation can affect rights. Risks do not stem from complaints procedures themselves, but from how platforms operate and manage these mechanisms. Key concerns include:

- Unsafe storage or processing of complaints and related data: Children's right to privacy might be compromised if platforms fail to safely store and process the complaints and related data. To mitigate this risk, platforms should implement robust data handling policies, limit data retention periods, and conduct regular security audits.
- Improper content decisions: The right to freedom of expression might be undermined if platforms inadequately assess complaints, apply benchmarks for content assessment in an inconsistent and arbitrary manner, or use benchmarks that are not aligned with international human rights and children's rights law. To address this, both regulators and platforms must ensure that benchmarks used to assess complaints are rights-based, with oversight provided by the enforcement authority.

3.2.8 Transparency reporting

What is it?

Transparency reporting obligations require online platforms to **publicly disclose how they govern content, enforce community standards, and mitigate systemic risks** to human rights and children's rights. The primary purpose is to enhance accountability, public trust, and regulatory oversight.[64]

Transparency reporting obligations serve as a check against arbitrary, discriminatory, unlawful or rights-abusing practices by allowing external scrutiny. They provide a means for regulators, researchers, civil society, and the general public to assess whether online platforms are respecting human rights.[65]

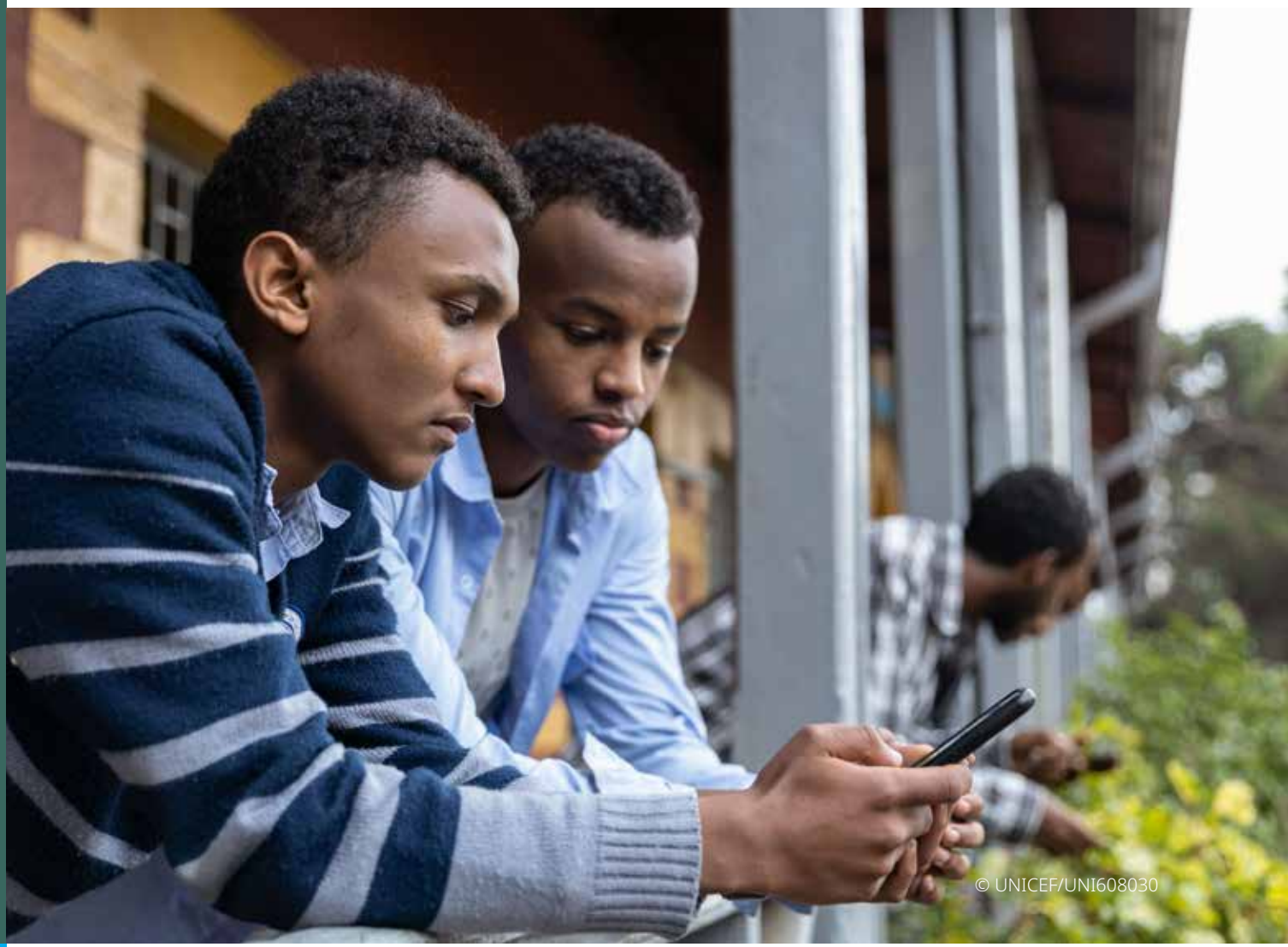
Key considerations

Transparency obligations can be a critical tool for assessing how platforms are respecting children's rights. However, they must be carefully designed to avoid unintended consequences. To be effective, transparency reports should go beyond 'data dumps' and provide clear, accessible, and actionable information that enables regulators and other stakeholders to make informed decisions.[66]

Regulations should prescribe clear timeframes and reporting parameters for transparency reports.

- Establish clear reporting timeframes: Requiring online platforms to submit transparency reports regularly enables regulators, researchers, civil society and the general public to monitor an online platform's behaviour over time and compare practices across companies.[67] In addition to periodic reporting, provisions for *ad hoc* reports (triggered by significant events such as changes in content moderation policies or practices) can enhance accountability and transparency.

- Define reporting parameters: Clear reporting parameters help ensure that transparency reports include relevant, comprehensive and comparable information.[68] This should include data on user reports, enforcement actions, appeals procedures, content moderation practices, and the functioning of recommender systems. UNICEF's disclosure recommendations on child rights in relation to the digital environment lists child rights-based disclosures for incorporation in company reporting, and outlines how these link with existing mandatory and voluntary reporting standards and frameworks.[69]

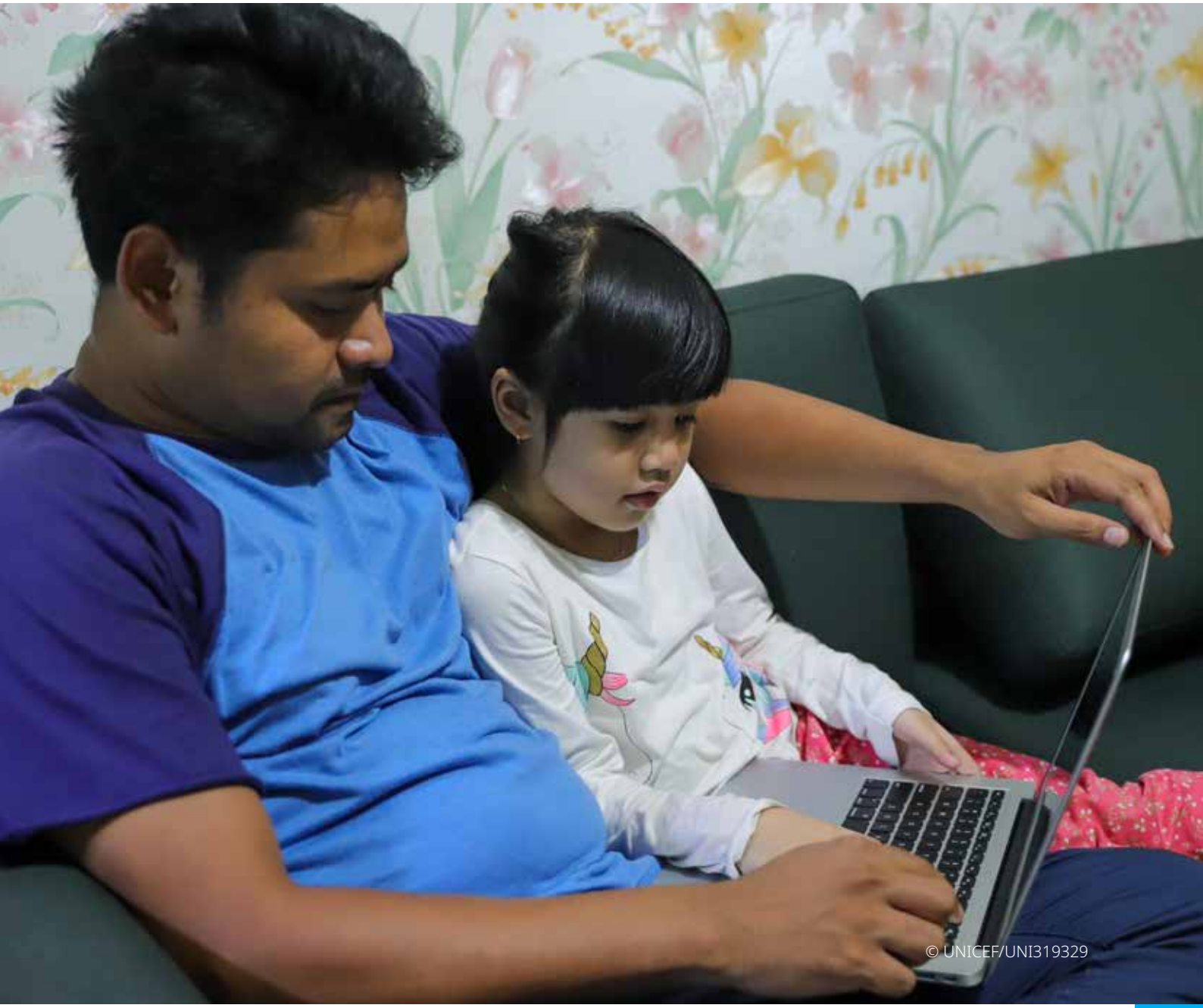


© UNICEF/UNI608030

04. WAY FORWARD

Online platform regulation is a fast-moving and oftentimes complex area, with significant implications for children's rights. While many of the frameworks reviewed here are newly adopted and still developing in practice, waiting for a 'perfect' regulatory model is not an option. Policymakers must act now to protect children and their rights, drawing on the recommendations in this brief and lessons from other jurisdictions to design regulations that are effective, enforceable, and rights based.

As these frameworks take shape, ongoing monitoring is crucial to assess whether they are achieving their intended goals, protecting human rights and children's rights, adapting to new risks, and ensuring compliance by online platforms. Importantly, children's voices must be central to these efforts – not only in identifying problems, but in co-creating solutions. A comprehensive, child-centred approach will be a crucial to identify what really 'works' to protect children's rights and safety in a digital world.



© UNICEF/UNI319329

Endnotes

1 This brief uses the term online 'platform' as shorthand to capture a broad range of digital environments where children interact with others and with content (see further, section 1.1). The use of this term does not imply any limitation or exemption of such services from applicable legal responsibilities.

2 See also UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), which explains how States parties should implement the Convention on the Rights of the Child in relation to the digital environment and provides guidance on relevant legislative, policy and other measures to ensure full compliance with their obligations under the Convention and the Optional Protocols.

3 Australia, the EU, India, Kazakhstan, South Africa, and the UK.

4 UN, Guiding Principles on Business and Human Rights, 2011.

5 UNICEF/UN Global Compact/Save the Children, Children's Rights and Business Principles, 2010.

6 UNICEF/OHCHR, Taking a child rights-based approach to implementing the UNGPs in the digital environment, 2024, p. 5.

7 For example: Office of the High Commissioner for Human Rights, Online Platform Governance & Human Rights, 2025; UNESCO, Guidelines for the Governance of Digital Platforms: Safeguarding freedom of expression and access to information through a multistakeholder approach, 2023.

8 UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), paras. 35-36.

9 Gorwa, The Politics of Platform Regulation. How Governments Shape Online Content Moderation, 2024, p. 16.

10 Based on Gorwa, The Politics of Platform Regulation. How Governments Shape Online Content Moderation, 2024, p. 18.

11 OECD, Children in the digital environment. Revised typology of risks, 2021.

12 UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), para. 14.

13 Ibid.

14 Please refer to UNICEF's work on 'Children and Digital Marketing: promoting responsible commercial practices in a hyperconnected world'.

15 OECD, Children in the digital environment. Revised typology of risks, 2021.

16 Finck, Digital Regulation: Designing a Supranational Legal Framework for the Platform Economy, LSE Law, Society and Economy Working Papers 15/2017, pp. 5 et seq.

17 Gorwa, The Politics of Platform Regulation. How Governments Shape Online Content Moderation, 2024, p. 50.

18 See for example: Sheera Frenkel, 'Key Takeaways from Facebook's Whistle-Blower Hearing,' New York Times, October 5, 2021.

19 United Nations Committee on the Rights of the Child, General comment General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25, para. 24.

20 Including the UN Convention on the Rights of the Child (CRC), the CRC Optional Protocol on the Sale of Children, Child Prostitution and Child Pornography (OPSC), and the International Covenant on Civil and Political Rights (ICCPR), amongst others.

21 Including General Comments from relevant UN treaty bodies and reports from relevant UN Special Rapporteurs, amongst others.

22 Based on World Bank Classification 2024.

23 In its 2025 Guidelines on the protection of minors under the DSA, the European Commission proposes Child Rights Impact Assessments (CRIAs) as a tool for online platforms to assess the risk they pose to children.

24 For recent developments see LexisNexis, Australia's bold leap into a new area of privacy law, 9 June 2025.

25 These provisions are not set out directly in Australia's Online Safety Act (2021), but in the secondary legislation called the Basic Online Safety Expectations (BOSE) and in the accompanying Industry Codes.

26 As above.

27 Only mandatory for Very Large Online Platforms and Very Large Online Search Engines (i.e. those with more than 45 million monthly users in the EU), see Art 34 EU Digital Services Act (2022). In its 2025 Guidelines on the protection of minors under the DSA, the European Commission proposes Child Rights Impact Assessments (CRIAs) as a tool for online platforms to assess the risks they pose to children.

28 Article 13 (2) of Kazakhstan's 2023 Law on Online Platforms and Online Advertising states that online platforms need to publish annual reports about identified systemic risks and the measures they have taken to reduce these risks. Even though not explicitly stated, this implies that online platforms need to conduct systemic risk assessments and take actions to respond to identified risks.

29 Age assurance is not a legal obligation under the EU Digital Services Act (2022), but considered a potential measure to ensure a high level of privacy, safety and security of minors on online platforms under the 2025 Guidelines on the protection of minors under the DSA.

30 Age assurance is not a legal obligation under South Africa's Film and Publications Act (Act No. 11 of 2019), but a potential avenue to create safer and age appropriate online environments under the accompanying Industry Code on the Prevention of Online Harm (2023).

31 These provisions are not set out directly in Australia's Online Safety Act (2021), but in the secondary legislation called the Basic Online Safety Expectations (BOSE) and in the accompanying Industry Codes.

32 Article 8 EU Digital Services Act (2022) prohibits a general monitoring obligation, meaning a process whereby an intermediary is obliged to introduce technological measures which monitor all user activity on its services. However, it allows for voluntary own-initiative investigations under specific circumstances, see Art 7 EU Digital Services Act (2022).

33 Section 3 (1) (b) of India's IT Rules 2021 state that intermediaries shall make reasonable efforts to not host, display, upload, modify, publish, transmit, store, update or share any information falling under specific content categories, such as content which is 'obscene' or 'harmful to child'. The Rules do not specify what 'reasonable efforts' means in this context. Even though this does not explicitly provide for a proactive detection obligation, intermediaries can only comply with this order if they put in place some form of monitoring system, either at upload level or post-facto.

34 Article 9 (4) (1) of Kazakhstan's 2023 Law on Online Platforms and Online Advertising requires platforms to take measures to counter the spread of illegal content on the territory of the Republic of Kazakhstan. While it is not clearly defined which measures should be taken, it presumably includes active scanning and detection, similar to the monitoring obligation for Internet Service Providers established under the 2004 Communications Law.

35 Section 24C of South Africa's Film and Publications Act (Act No. 11 of 2019) targets child-centred content and contact services. Such service providers must moderate content and take reasonable steps to ensure that such services are not being used to commit an offence against children. To comply with the latter obligation, in-scope businesses will have to be able to detect any suspicious content or conduct before an offence against a child is committed, which implies the need for some form of proactive detection.

36 These provisions are not set out directly in Australia's Online Safety Act (2021), but in the accompanying Industry Codes. Mandatory reporting obligations are only applicable for child sexual abuse material and so-called 'pro-terror' materials, see Section 7, Compliance Measure 1, Social Media Code.

37 Only applicable to criminal offence involving a threat to the life or safety of a person, see Article 18 EU Digital Services Act (2022).

38 Only applicable to illegal content that entails a threat to the life or security of a person and citizen, see Article 9 (4) (3) of Kazakhstan's 2023 Law on Online Platforms and Online Advertising.

39 Only applicable for UK linked child sexual abuse and exploitation content, see section 66 UK Online Safety Act (2023).

40 These provisions are not set out directly in Australia's Online Safety Act (2021), but in the in the secondary legislation called the Basic Online Safety Expectations (BOSE) and in the accompanying Industry Codes.

41 Refer Articles 11.1-11.3 of Kazakhstan's 2023 Law on Online Platforms and Online Advertising.

42 Applicable only to so-called Categorised Services, see section 77 UK Online Safety Act (2023).

43 UN Human Rights Committee, General comment No. 34 Article 19: Freedoms of opinion and expression, UN Doc CCPR/G/GC/34, para. 11.

44 See also: Office of the High Commissioner for Human Rights, Online Platform Governance & Human Rights, 2025.

45 UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), para 39.

46 The General Data Protection Regulation (GDPR) is the EU's comprehensive data protection law, applicable to organisations processing personal data of individuals in the EU, regardless of where the organisation is based. Its extraterritorial scope and strict standards have influenced global tech companies, many of which have adapted their practices worldwide to align with its requirements.

47 eSafety Commissioner, Safety by Design principles.

48 See for example: Livingstone, S. & Pothong, K. (2023). Child Rights by Design: Guidance for Innovators of Digital Products and Services Used by Children. Digital Futures Commission, 5Rights Foundation.

49 UNICEF, Developing Global Guidance for Child Rights Impact Assessments in Relation to the Digital Environment, 2024, p. 2.

50 UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), para 38.

51 Danish Institute for Human Rights, Guidance on human rights impact assessment of digital activities, 2020, p. 15.

52 UNICEF, Developing Global Guidance for Child Rights Impact Assessments in Relation to the Digital Environment, 2024, p. 3.

53 European Commission, Guidelines on measures to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of Regulation (EU) 2022/2065, 13 May 2025, para 29.

54 Ibid.

55 European Data Protection Board, Statement 1/2025 on Age Assurance, adopted on 11 February 2025.

56 UN Human Rights Council, The right to privacy in the digital age, UN Doc A/HRC/39/29, 3 August 2018, para. 17.

57 UN Human Rights Council, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc A/HRC/32/38, 11 May 2016, para. 57.

58 Arguing for privacy-respecting solutions, see for example WeProtect Global Alliance, Global Threat Assessment 2023, 2023; arguing that such solutions do not exist, see for example ISOC, Client-side scanning. What It Is and Why It Threatens Trustworthy, Private Communications, 2022. See also UN Human Rights Council, The right to privacy in the digital age, UN Doc A/HRC/39/29, 3 August 2018, para. 20.

59 UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), para 44.

60 In such a situation, platforms might be caught between complying with local laws and upholding human rights and children's rights. The Global Network Initiative's Principles on Freedom of Expression and Privacy provide guidance to companies in these situation: 'ICT companies should comply with all applicable laws and respect internationally recognized human rights, wherever they operate. If national laws, regulations and policies do not conform to international standards, ICT companies should avoid, minimize, or otherwise address the adverse impact of government demands, laws, or regulations, and seek ways to honour the principles of internationally recognized human rights to the greatest extent possible.'

61 UN Human Rights Council, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc A/HRC/38/35, 6 April 2018, para. 56

62 Ibid., para. 35.

63 CRC Committee, General comment No. 25 (2021) on children's rights in relation to the digital environment, UN Doc CRC/C/GC/25 (2 March 2021), para. 44.

64 Brookings, Transparency is essential for effective social media regulation, 2022.

65 Santa Clara Principles on transparency and accountability in content moderation, 2021.

66 UNICEF, Corporate Reporting on Child Rights Impacts in Relation to the Digital Environment: Guidance for Business, 2025, p. 11.

67 Tech Coalition, Trust: Transparency reporting implementation guide, 2024, p. 8.

68 UNICEF/OHCHR, Taking a child rights-based approach to implementing the UNGPs in the digital environment, 2024, p. 18.

69 UNICEF, Corporate reporting on child rights impacts in relation to the digital environment. Disclosure recommendations, 2025, pp. 5-31.

Published by UNICEF *Child Protection* and *Business Engagement and Child Rights* Teams, Programme Group, 3 United Nations Plaza, New York, NY 10017.
Email: childprotection@unicef.org.
Website: www.unicef.org.