Online gaming and risks to children

June 2026

Delivered in partnership

GLOBAL CYBERSECURITY FORUM | weprotect Global Alliance | TRILATERAL RESEARCH Ethical AI

Contributors

Srivatsan Rajagopalan, Trilateral Research
Dr Lorleen Farrugia, Trilateral Research

Table of Contents

Acronyms

4Cs	Content, Contact, Conduct, Commerce
ACAMS	Association of Certified Anti-Money Laundering Specialists
ADL	Anti-Defamation League
AI	artificial Intelligence
AML	anti-money-laundering
APAC	Asia-Pacific
AR	augmented reality
AVPA	Age Verification Providers Association
COPPA	Children's Online Privacy Protection Act
CSA	child sexual abuse
CSAM	child sexual abuse material
CSE	child sexual exploitation
CSO	civil society organisation
DM	direct message
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (5th edition)
FOMO	fear of missing out
ESRB	Entertainment Software Rating Board
EU	European Union
GCF	Global Cybersecurity Forum
GDPR	General Data Protection Regulation
HIC	high-income country
ICD-11	International Classification of Diseases (11th revision)
IGD	Internet Gaming Disorder
IP	internet protocol
IRB	Institutional Review Board
ISD	Institute for Strategic Dialogue
KYC	know-your-customer
LMIC	lower- and middle-income country
MAU	monthly active user
NGO	non-governmental organisation
NSPCC	National Society for the Prevention of Cruelty to Children
P2P	peer-to-peer
PEGI	Pan European Game Information
RAN	Radicalisation Awareness Network
RUSI	Royal United Services Institute
UGC	user-generated content
UK	United Kingdom
UNCRC	United Nations Convention on the Rights of the Child
UNICEF	United Nations Children's Fund
UNICRI	United Nations Interregional Crime and Justice Research Institute
US	United States
VPN	virtual private network
VR	virtual reality
WPGA	WeProtect Global Alliance

A. Executive summary

Online gaming is a core part of children's and young people's social infrastructure. It supports creativity, collaboration, identity exploration and belonging. Yet these same environments expose children to risks that can escalate into serious harms, especially where protections are fragmented, unevenly applied or poorly matched to how gaming ecosystems function.

This report distinguishes **risk** (features, conditions or behaviours that create vulnerability, such as unmoderated voice chat or gambling-like monetisation) from **harm** (negative outcomes, such as grooming leading to child sexual exploitation [CSE] or monetisation contributing to gambling-like behaviours). This distinction supports prioritisation of the highest-harm pathways rather than defaulting to broad debates about screen time. Stakeholders described a shifting threat landscape in which risks are increasingly financially motivated, more organised and networked across platforms, with rapid migration into adjacent services and encrypted channels. The central implication is that reliance on household supervision and reactive moderation is not sufficient, particularly in lower-resource contexts and for children using shared devices. Effective protection requires a layered safety infrastructure and safety-by-design approaches that match cross-platform threats and are rooted in knowledge of the evidence and lived experiences of gamers.

Evidence base and purpose

Commissioned by WeProtect Global Alliance (WPGA) and the Global Cybersecurity Forum (GCF) to strengthen global understanding of risks to children in online gaming and identify actionable improvements, this research synthesises a rapid narrative literature review, a survey of 41 WPGA members and 13 expert interviews across industry, academia, civil society and international organisations. It should be read as an expert synthesis of the understanding of risks and harm in the gaming ecosystem.

Core framing

Across the evidence, risks persist not primarily because tools are absent, but because incentives are misaligned. Commercial models often reward frictionless engagement and spending, while effective safety interventions introduce friction and delays. Responsibility is frequently pushed downstream onto families and schools, despite platforms controlling key design choices, affordances and safety data. Exposure is also not evenly distributed. Children who play more often and use social features more heavily are more likely to encounter risky interactions (for example, harassment, coercion or manipulation through chat and multiplayer play). This does not mean that time spent gaming causes harm, but it does mean higher-exposure settings are where safeguards and support need to work best. The report uses the 4Cs framework to organise risks in gaming: Content (what children see), Contact (who they interact with), Conduct (how people behave) and Commerce (how spending and monetisation operate). This helps make risk-to-harm pathways clearer across different games and services.

Key findings

- **Beyond screen time**: Interviewed experts cautioned that screen time is a limited proxy for safety. Focusing on time alone can miss higher-harm pathways (including grooming and sexual exploitation) that escalate through social play, off-platform migration and commercial features.

- **Cheating-tier investment gap and privacy paradox**: Anti-cheat and fraud prevention often receive deeper investment and more intrusive technical measures than child safety. Stakeholders described a privacy paradox in which adult-centric interpretations of privacy and business risk can limit proactive detection, documentation and transparency. They also described how offenders exploit gaps between services, moving quickly from games to adjacent platforms (for example, social media or encrypted messaging) where oversight is weaker. They proposed proportionate, rights-governed signal sharing as a priority to address these issues, alongside capacity constraints for smaller studios. Age assurance is central for tailoring protections, but rigid or poorly designed mandates can increase displacement and inequity.

- **Inequality of safety and visibility**: Many tools assume a wealthy default, including personal devices, private accounts, stable connectivity and digitally skilled caregivers. In lower-resource contexts, shared devices, cybercafés, mobile-first access and modified gaming clients can reduce both protections and visibility. Children may be "digital heads of household", or the person in the household with most digital skills, but when they face risks, stigma and low trust inhibit reporting of sexual harms. The combined effect is a visibility gap where the highest-risk children are least visible to protection systems and least able to access remedy.

- **Parenting gap and trust erosion loop**: Policy expectations that parents and caregivers can manage safety do not match time poverty, gaming literacy gaps and adolescent development. Punitive responses from parents and caregivers can trigger a trust erosion loop, reducing disclosure and opportunities for early intervention.

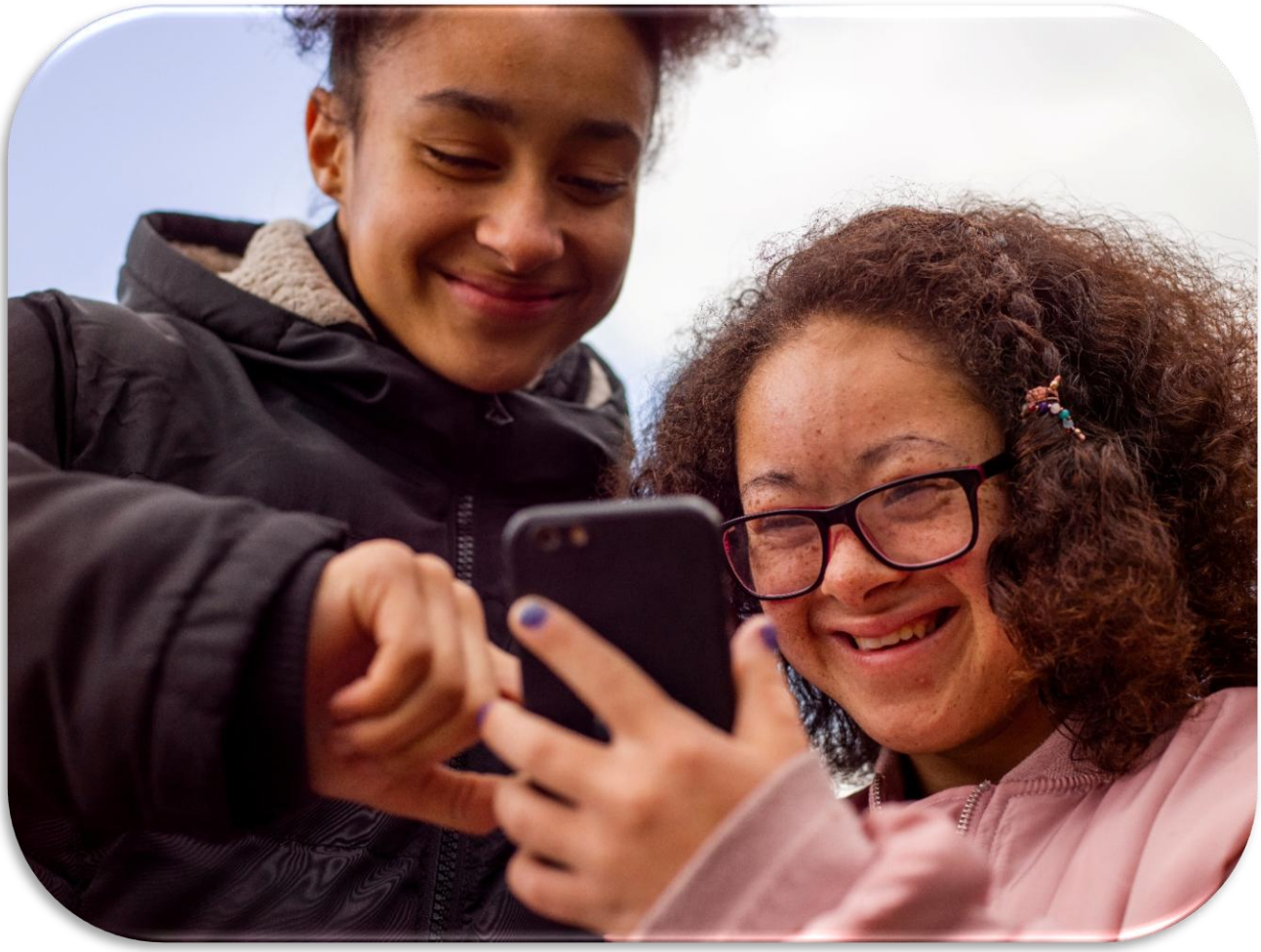
- **Displacement and contested AI frontier**: Poorly designed restrictions or age-related controls, especially when introduced without sufficient evidence or consultation, can displace children towards less governed, encrypted or offshore spaces, reducing visibility and access to remedy. Stakeholders favoured graduated, in-product interventions that keep children within moderated environments. The use of AI was seen as promising but contested, with concerns about legal complexity, entrapment risks and mission creep requiring strong governance and oversight.

Priorities for action

- **Industry**: Treat child safety as a matter of business integrity, not just compliance, and align investment with fraud and anti-cheat practices. Reform monetisation and engagement mechanics that exploit developmental vulnerabilities. Reduce offender mobility through proportionate cross-platform disruption with clear safeguards and feasible routes for smaller services.

- **Governments and regulators**: Mandate safety-by-design, including scrutiny of features, defaults and business models, not only content. Clarify when and how platforms can use and share limited safety signals for child protection purposes (for example, to detect grooming patterns or repeat offenders), with clear safeguards for necessity, proportionality, transparency and independent oversight. Prioritise measures that keep children within governed, moderated services through graduated protections, rather than measures that primarily force workarounds or migration to harder-to-reach spaces.

- **Civil society, research and cross-sector**: Close evidence and visibility gaps (shared devices, cybercafés, mobile-first and local-language contexts). Shift measurement towards child-centred outcomes and access to remedy. Move from transparency to verifiable accountability through independent scrutiny, auditing and robust testing, including for grooming and the financial sexual extortion of children.



1. Introduction

Online gaming is a core part of children's and young people's social interactions and leisure activities, offering opportunities for creativity, collaboration and community. However, these same environments expose children to a set of risks that, if unmonitored or unaddressed, can escalate into significant harms.

A practical distinction is needed:

- **Risk** refers to features, conditions or behaviours that create vulnerability, such as unmoderated voice chat or gambling-like monetisation mechanics.
- **Harm** refers to negative outcomes that occur when risks materialise, such as grooming that leads to sexual exploitation or monetisation that contributes to problematic gambling-like behaviours (NSPCC, 2024; ACAMS Today, 2023).

The scale of exposure to online gaming and associated risks is significant. This is not a marginal issue but an underlying feature of many children's lives. In the US, 85% of teens report playing video games, and 41% play at least once a day (Gottfried and Sidoti, 2024). There are direct risks associated with online gaming. For example, a recent meta-analysis found that about 8.6% of adolescents worldwide meet criteria for gaming disorder[1] (Satapathy et al., 2025). There are many other serious harms that children face in the online world, and they can overlap with gaming. The global systematic review and meta-analysis estimated that around 1 in 12 children worldwide (about 8%) experienced online CSE or abuse in the past year, while noting that rates vary across studies because definitions and measurement approaches differ (Fry et al., 2025).

Risk exposure is also patterned by the intensity of children's engagement. Put simply, as time spent gaming increases, so do the number of interactions, communication opportunities and points of contact with wider networks, which in turn increases the probability of encountering harmful behaviour (eSafety Commissioner, 2024). Empirical evidence aligns with this exposure dynamic, including findings that link online gaming to cybervictimisation, with time spent online accounting for part of this relationship, consistent with a dose-like association between heavier engagement and greater exposure to negative experiences (Marinoni et al., 2024).

The core concern is not any single platform, but the recurring behavioural patterns and design choices that enable harm across changing ecosystems. While particular games and community platforms may be identified as hotspots for grooming, coercion or exposure to harmful content, the underlying risks are not platform-specific. Behaviours and tactics shift as children move across services and technologies evolve. For this reason, the report focuses on risk-to-harm pathways and the enabling features that make those pathways viable, rather than transient platform trends.

Maintaining the distinction between risk and harm is critical for policy-oriented analysis because it clarifies where interventions should operate upstream to reduce exposure to risky features and conditions and downstream to respond when harms occur and prevent recurrence (ISD, 2024; RUSI, 2025). This aligns with the scope of the project, which prioritises the highest-harm risks, particularly CSE, child sexual abuse (CSA), grooming and the sexual extortion of children as enabled and amplified within gaming ecosystems (NSPCC, 2024; Kilmer and Kowert, 2024). For clarity, CSE refers to exploitative situations where a child is coerced or manipulated into sexual activity in exchange for something of value (for example, grooming and financial sexual extortion), while CSA refers to sexual offences against children that may involve direct abuse or the creation or distribution of child sexual abuse material.

[1] Persistent or recurrent gaming behaviour, which results in marked distress or significant impairment in personal, family, social, educational, occupational or other important areas of functioning (Király et al., 2023).

2. Methodology

2.1 Approach and aims

This research used a variety of approaches and an exploratory design combining a rapid literature review, survey data from members of the WPGA and semi-structured stakeholder interviews. The aim was to understand extreme risks that children and young people may encounter in online gaming environments — including social, behavioural, psychological and platform-related dynamics — and to identify best practice and knowledge gaps.

The approach intentionally integrated academic findings with professional expertise from civil society organisations (CSOs), nongovernmental organisations (NGOs), researchers and trust and safety practitioners within the gaming industry. It aligns with WPGA's child-centred, rights-driven and technology-informed mandate.

2.2 Literature review

A rapid narrative review of 40 sources was conducted, encompassing peer-reviewed publications and grey literature published between 2019 and 2025. Sources were identified through academic database searches, industry and NGO reports and citation tracking.

Inclusion criteria prioritised sources related to:

- child and youth engagement with gaming platforms
- risks and harms in gaming communities
- exposure to violent, abusive or sexual content in gaming
- harassment, identity-based abuse and community norms
- moderation, reporting and platform governance
- in-game social dynamics and communication channels
- regulation and risk-mitigation frameworks.

Research exclusively addressing adult-only gaming spaces, as well as technical game-engineering literature without behavioural context, was excluded. The review aimed to provide a rigorous and policy-relevant synthesis of available knowledge, identifying thematic patterns, divergences and conceptual contributions relevant to child safety in online gaming.

2.3 Survey with WPGA membership

A structured survey was distributed to WPGA members, yielding 41 responses from organisations engaged in child protection, online safety, trust and safety, regulatory policy, youth support services and research. The survey gathered both quantitative and qualitative data to understand organisations' engagement with gaming-related harms, roles and expertise and any relevant projects underway. It also aimed to identify perceived challenges and unanswered questions in online gaming environments. All responses were anonymised and analysed in aggregate. Charts summarising the survey results are provided in Annex 3, while thematic analysis of open-ended responses is integrated into the Findings chapter.

2.4 Stakeholder interviews

Thirteen semi-structured interviews were carried out with stakeholders across CSOs, NGOs, academic researchers and trust and safety teams within gaming and platform companies (Annex 1). Participants were purposively selected

to represent different roles, sectors and geopolitical regions, including contributors from South America and the Asia-Pacific region, ensuring diverse perspectives from both high-income countries (HICs) and lower- and middle-income countries (LMICs). Snowball sampling was applied where relevant to reach additional specialists.

A semi-structured interview guide designed specifically for the gaming context was used across all sessions (Annex 2). Questions focused on participants' perspectives on risks in online gaming, current responses and interventions, opportunities for improvement and the gaps and challenges that remain in addressing these risks. Participants were also asked stakeholder-specific questions to explore evidence gaps in the research, practitioners' experiences of prevention and awareness-raising and challenges faced by industry. All participants received an information sheet and provided informed consent prior to participation (Annex 2).
Interviews were conducted online between October and November 2025, each lasting roughly 60 to 90 minutes. Conversations were audio-recorded with permission, and interviewees were given the option to remain anonymous.

2.5 Data analysis

Survey results were analysed using descriptive statistics and qualitative thematic coding of open-text responses. Interview audio recordings were transcribed verbatim and coded through an iterative qualitative coding framework.

2.6 Ethical considerations

The analysis of the findings centres around the following themes:

- Types and mechanisms of risk within gaming platforms
- Behavioural interaction and peer dynamics
- Moderation systems, reporting practices and governance responsibilities
- Game design features influencing safety and vulnerability
- Regulatory and structural constraints
- Opportunities for safety-by-design, education and cross-sector collaboration.

Rigorous ethical standards were considered and upheld throughout the research process. Full compliance with General Data Protection Regulation (GDPR) requirements was maintained at every stage of the research process. Participation in both the survey and interviews was voluntary, with the right to withdraw at any point. Data confidentiality was maintained through anonymisation procedures and secure encrypted data storage.

Participants were informed of potentially sensitive topics in advance and could pause or terminate interviews at any time. Researcher well-being protocols were implemented given the nature of the discussions.

2.7 Limitations

The literature review was rapid, not systematic, and may not fully capture emerging sources or non-English language research. Survey respondents reflect perspectives within the WPGA network and may be influenced by organisational position and mandate. The interview sample, while covering diverse regions, stakeholder types and professional roles, is not representative of the entire gaming ecosystem. Notably, the study did not include children or individuals with lived experience of gaming-related harms, limiting experiential perspectives. Findings should therefore be interpreted as synthesised expert insight anchored in professional practice.

3. Literature review

3.1 The structural context: incentives and design

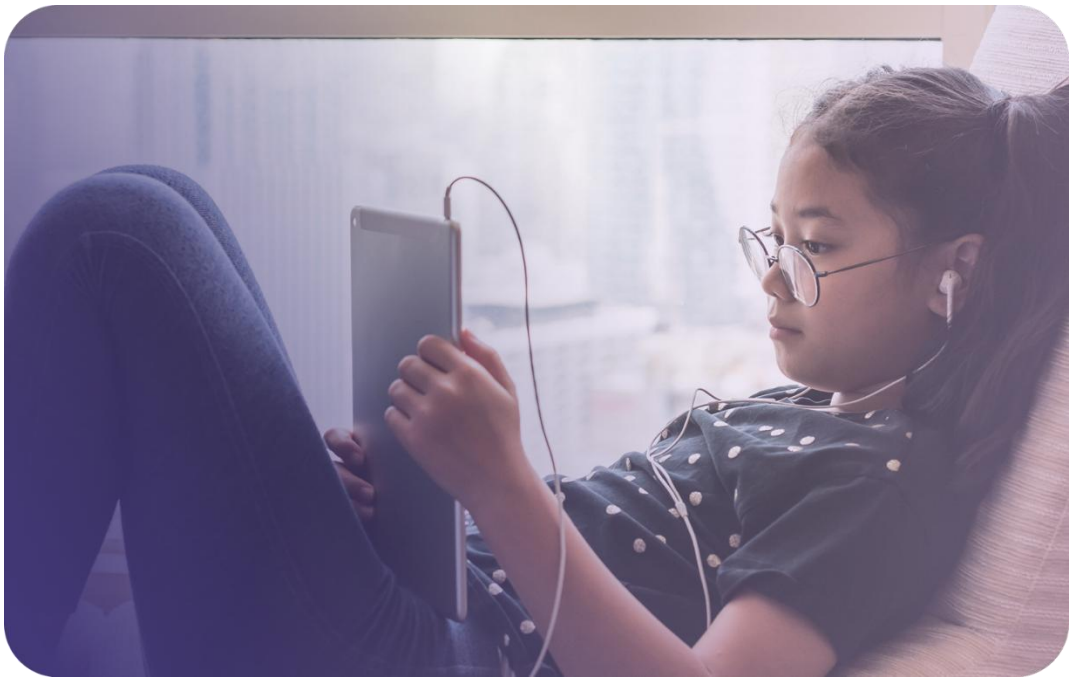
The literature suggests that the persistence of risks in gaming environments is a direct consequence of a fundamental misalignment of incentives. Commercial incentives that reward long play sessions can discourage safety and stricter moderation because they interrupt the velocity of play or spending, meaning the speed at which users can move, message and transact without interruption (ISD, 2024; RUSI, 2025). A further dimension is the transfer of responsibility. Families, schools and civil society are often expected to carry the burden of prevention and response, while platforms retain control over key design choices and safety data.

3.2 The landscape of risks in online gaming

Across the evidence base, a recurring pattern is that exposure is not evenly distributed. Children who spend more time in games and who participate more heavily in social play are more likely to report negative experiences (eSafety Commissioner, 2024). This is best understood as an exposure mechanism rather than a deterministic pathway: higher-frequency play and frictionless interaction expand the surface area for contact- and conduct-related risks (for example, harassment, coercion or manipulation), particularly in multiplayer contexts (eSafety Commissioner, 2024). While this does not imply that time spent gaming causes harm, it supports the view that prevention efforts should focus on the highest-exposure contexts (high-frequency use, socially networked play and high-interaction features), rather than treating risk as uniformly distributed across all players (Marinoni et al., 2024). This matters because it shifts attention from generic debates about time spent towards the specific features and interactions through which risks concentrate and escalate.

To map how risks arise, cluster and escalate across gameplay and adjacent services, the 4Cs framework (Content, Contact, Conduct, Commerce) is used and adapted to gaming by specifying the platform features that enable each category, including user-generated content pipelines, real-time voice and text, discovery systems and virtual economies (Livingstone and Stoilova, 2021).

The analytic value of the framework in gaming is in making pathways explicit. In practice, risks in one domain often create routes into others, for example, contact risks escalating through off-platform movement (such as moving from in-game chat to social media or encrypted messaging, where oversight is weaker) or conduct risks interacting with recommendation and community dynamics. Table 1 summarises these categories as risk-to-harm pathways and proposes leading indicators that, while not yet standardised, reflect best-practice safety measurement.



C	Risk definition (gaming-specific)	Typical enablers (platform features)	Risk → Harm pathway (illustrative)	Leading indicators to monitor*
Content	Exposure to illegal/inappropriate material via user-generated content (UGC), streams, mods, in-game assets.	UGC upload at scale; weak pre- and post-moderation; permissive asset marketplaces; algorithmic discovery/surfacing.	**Risk**: repeated exposure to sexual, violent, or extremist content → **Harm**: desensitisation, psychological distress, normalisation of hate, potential grooming priming.	Rate of UGC takedowns by type; time-to-removal; age verification efficacy; repeat violator recidivism; prevalence of sensitive asset tags.
Contact	Harmful interactions with others (grooming, coercion, financial sexual extortion) via voice/text chat, direct messages (DMs), group servers.	Real-time chat; anonymity/pseudonymity; frictionless friending; off-platform "hops" to private/encrypted apps.	**Risk**: rapport-building in-game → hop to semi-private server → move to encrypted DMs → **Harm**: sexual exploitation, child sexual abuse material (CSAM) production, financial sexual extortion.	Grooming signal detections; share of new-contact DMs to minors; cross-platform migration events; report-to-action latency; law enforcement referral volumes.
Conduct	Harmful behaviours by or against the child (toxicity, hate, bullying, doxxing, brigading).	Competitive ranking/leaderboards; weak civility frictions; poor repeat-abuse controls; permissive clan/guild tools.	**Risk**: persistent harassment/misogyny → **Harm**: mental health impacts, exclusion, gateway into extremist subcultures and male-supremacist spaces.	Abuse report volumes per monthly active users (MAUs); repeat offender rates; voice/text toxicity scores; suspension efficacy (time-to-reoffence); targeted-harassment clustering.
Commerce	Financial risks: scams, loot boxes/chance-based rewards, "pay-to-win" pressure, laundering via virtual economies.	Opaque pricing via premium currencies; time-limited offers/fear of missing out (FOMO); P2P trading; weak know-your-customer (KYC) on marketplaces.	**Risk**: compulsive spend & scam exposure → **Harm**: debt, family conflict, Internet Gaming Disorder (IGD) comorbidity; systemic: laundering/illicit flows leveraging tradable skins/currency.	Youth spend distributions; chargebacks/refunds; scam/phishing reports; loot box engagement vs. age; anti-money-laundering (AML) red flags on asset flows.

*__Note__: Many of the "Leading indicators to monitor" represent best-practice safety metrics not yet standard in public transparency reporting.

3.2.1 Contact risks: child sexual exploitation, grooming and sexual extortion

Grooming is a primary contact risk in gaming, enabled by anonymity and real-time interaction channels that support rapid rapport-building and escalation (NSPCC, 2024). One study reported that grooming interactions can escalate extremely quickly in some gaming contexts, including cases where high-risk sexual conversation begins within seconds and progresses to abuse within an hour (WeProtect Global Alliance, 2023). Offenders also commonly move children off-platform to reduce visibility and oversight, migrating from in-game contact to semi-private community spaces and then to encrypted messaging (ADL, 2021; NSPCC, 2024). Where safeguards are weak, investigations have documented systemic vulnerabilities that facilitated grooming and sexual exploitation (Hindenburg Research, 2024).

Literature also highlighted overlaps in the persuasive tactics used across grooming and extremist exploitation pathways (ISD, 2021; Kilmer and Kowert, 2024).

3.2.2 Conduct risks: toxicity, harassment and hate

The principal conduct risks are toxicity, harassment and hate speech, which are pervasive in online games and disproportionately target girls and women (ADL, 2023; Unity, 2023). These risks can escalate into harms such as psychological distress, exclusion from gaming spaces and normalisation of hostility. Toxic environments can normalise hostility and, in some cases, make it easier for misogynistic and hate-based communities to recruit or socialise young people into more extreme subcultures (Miller-Idriss, 2025).

3.2.3 Extremist exploitation and radicalisation risks

Extremist exploitation cuts across content, contact and conduct dynamics, including propaganda circulation, recruitment attempts and hate-based harassment. Extremist actors refer to individuals or organised groups promoting ideologies that justify violence, hate or exclusion, including efforts to recruit or radicalise young people (ISD, 2021; Lamphere-Englund, 2025). The literature documents the use of gaming-adjacent platforms and community infrastructures to create insular spaces for socialisation and recruitment, including on services commonly used by gaming communities (Life After Hate, 2024). High-profile cases illustrate how violent extremist narratives can draw on gaming imagery and community cues to appear familiar and appealing (Lakhani, 2021). Risks escalate into harm when children adopt extremist worldviews, enter extremist networks or are encouraged towards violence (RAN, 2022).

3.2.4 Commerce risks: financial exploitation and illicit finance

Commerce risks include scams, gambling-style loot boxes[2] and financial coercion (including the financial sexual extortion of children) as well as system-level misuse of virtual economies for illicit finance (ACAMS Today, 2023). Loot box mechanics, in particular, mimic gambling, creating significant risks of normalising addictive behaviours among children (ACAMS Today, 2023). At the systemic level, gaming platforms have been exploited for money laundering and terrorist financing, using tradable assets and virtual currencies to move illicit funds across borders (Hindenburg Research, 2024). If unchecked, these risks escalate into harms affecting both children and wider security systems.

3.2.5 Problematic use/gaming disorder (cross-cutting risk)

Excessive or problematic gaming represents another significant risk. Design features that encourage continuous play, such as reward loops and in-game incentives, create the conditions for excessive use. This risk can escalate into harms captured by IGD, including disrupted sleep, reduced school performance and strained family relationships (Király et al., 2023; Technavio, 2025). It is important to note that while IGD is recognised by the World Health Organization in the International Classification of Diseases, 11th Revision (ICD-11), it remains listed in DSM-5 Section III (Emerging Measures and Models), which is reserved for conditions requiring further research and clinical experience before formal inclusion (American Psychiatric Association, 2023). This is one reason prevalence estimates and severity claims should be interpreted cautiously. While most children play games without severe consequences, vulnerable groups face disproportionate risks of escalation, and youth voices highlight the personal toll of problematic use (Thorn, 2024).

[2] Loot boxes are in-game purchases (or earned rewards) where the player receives a randomised item or set of items, meaning the contents are unknown at the time of opening and vary in value or usefulness.

3.3 Enablers of risk

These risks are amplified by low-friction social design, including easy contact, rapid movement across services and monetisation systems that encourage frequent spending, while safeguards remain uneven and hard to evaluate independently. Although platforms are deploying measures such as automated moderation, user reporting tools and transparency reporting, evidence indicates that implementation is inconsistent across regions and languages and often emphasises enforcement outputs rather than child safety outcomes (Roblox, 2024; ISD, 2024). While regulatory frameworks such as the EU Digital Services Act and the UK Online Safety Act are beginning to formalise expectations around risk assessment, implementation remains fragmented (ISD, 2024; RUSI, 2025). As a result, the current response is frequently characterised as reactive. This disconnect between the availability of safety tools and the persistence of harm is a core rationale for the primary research presented in the subsequent chapter. In practice, these enabling features also shape the distribution of exposure: higher-frequency play and more socially networked interaction increase the surface area for negative experiences, particularly for contact and conduct risks (eSafety Commissioner, 2024).

3.4 Critical gaps and future research directions

Despite the growing body of literature on online gaming risks, the evidence base remains uneven and insufficient to guide effective, systemic interventions. This section identifies three categories of gaps—geographical, methodological and thematic—and frames them as strategic priorities for future research and policy development.

3.4.1 Geographical gaps

Research remains heavily concentrated in Europe and North America. While these regions provide important insights into platform governance and industry practice, they represent only part of the global picture (ADL, 2023). Rapid market growth is occurring in the Asia-Pacific (APAC) region, Latin America and Africa, where gaming adoption is accelerating but where empirical child-focused research is sparse (Technavio, 2025). The APAC region alone is projected to account for ~44% of global market growth by 2029 (Technavio, 2025). This lack of geographical representation means that very little is known about children playing via shared devices, cybercafés or unofficial and modified game clients that sit outside standard platform safety controls. These children are also poorly captured in the platform safety data. This imbalance also risks obscuring how socio-economic conditions, regulatory contexts and cultural norms in understudied regions shape children's exposure to risks and pathways into harm. For instance, monetisation models and moderation standards may diverge sharply between platforms in Europe and North America, and those dominant in APAC markets, creating blind spots in global policy. Addressing this requires deliberate investment in locally grounded research partnerships to ensure that international debates are informed by diverse and representative evidence.

3.4.2 Methodological gaps

A second weakness lies in how risks are measured and evaluated. Much of the current evidence is drawn from cross-sectional surveys or self-reported industry data, providing only snapshots rather than capturing how risks escalate into harms over time. There is a pressing need for longitudinal studies that can track trajectories of grooming, radicalisation or problematic gaming use across months or years (NSPCC, 2024; Kilmer and Kowert, 2024). Without this, policymakers lack the ability to see how early exposures compound into long-term harm. Independent evaluation of interventions is also rare: transparency reports and pilot programmes often highlight successes, but with little external validation of impact (Roblox, 2024; RUSI, 2025). Finally, children's own voices are still insufficiently integrated into research design and assessment. Studies that do centre youth perspectives highlight critical insights into how risks are experienced and perceived (Thorn, 2024), but these remain small-scale. Systematically embedding child participation is essential for designing safeguards that reflect lived realities rather than external assumptions.

3.4.3 Thematic gaps

Beyond geography and methodology, several thematic blind spots limit the development of comprehensive safeguarding strategies:

• **Intersectional risks:** Research rarely disaggregates findings by identity markers. Yet evidence shows that girls and marginalised groups often face disproportionate risks of harassment, exclusion or targeted exploitation in gaming spaces (Miller-Idriss, 2025; RAN, 2023). Gendered harassment, for example, has been shown to overlap with extremist recruitment pathways, amplifying risk. Without intersectional analysis, safeguards will remain too generalised to protect the most vulnerable. Offline vulnerabilities such as mental health difficulties, special educational needs or family instability further compound these risks. Children experiencing such challenges are not only more likely to encounter harmful behaviours online but also more likely to be specifically targeted by offenders, making their protection a priority for future safeguarding strategies.

• **Children as creators, workers and consumers:** Safeguarding frameworks often treat children as passive users, overlooking their role as creators, streamers monetising audiences or consumers in complex in-game economies. Unlike adults, they lack protections as workers, consumer rights as spenders or IP rights as creators. This regulatory blind spot enables new forms of exploitation that require urgent research and policy attention (Rees, 2025; Hyde and Cartwright, 2023).

• **Platform accountability and transparency:** Industry transparency reports are becoming more common, but they remain inconsistent, selective and unaudited. This data asymmetry, where platforms hold detailed safety data but release it selectively, creates a structural barrier to accountability (ISD, 2024; Roblox, 2024). Developing independent auditing mechanisms and mandating data access for researchers and regulatory authorities are critical steps towards ensuring safety claims can be verified and interventions evaluated with rigour.

• **Emerging technologies:** Immersive environments such as virtual reality (VR) and augmented reality (AR) can intensify both positive and harmful experiences because interactions may feel more invasive, harder to exit and more personally violating for children (Technavio, 2025). Early VR research describes "phantom touch", where some users report feeling tactile sensations in response to virtual interactions, suggesting that certain experiences can be processed as more physically immediate than in non-immersive contexts (Copson and Johnson, 2025). For safeguarding, the practical implication is that harassment or sexualised behaviour in immersive spaces may be experienced as more embodied and distressing, increasing the urgency of safety-by-design protections before VR and AR adoption scales further (Copson and Johnson, 2025).

4. Interview and survey findings

This chapter synthesises the thematic analysis from the stakeholder interviews and the open-ended survey questions, contextualised within the wider evidence base on how online gaming environments shape children's safety, rights and everyday lives. In addition to 13 interviews, 41 WeProtect Global Alliance members took part in an open-ended survey spanning the public sector, intergovernmental bodies, civil society, platforms, safety tech companies, research, the private sector and helplines. Across responses, online gaming was primarily framed as a child protection, well-being and digital safety issue (Survey, Annex 3). Consistent with wider child online safety evidence, the analysis distinguished between higher-severity, lower-prevalence harms (for example, online sexual abuse and severe bullying) and more widespread, lower-severity concerns (for example, time use and everyday peer conflict), rather than treating all named risks as equivalent.

Interview and survey responses converged on two core points:

1. **Gaming is a core form of social infrastructure.** Many interviewees described gaming as a primary social space where friendships are formed and identities are explored. However, how this social life plays out depends on resources. Practitioners based in LMICs emphasised that participation often relies on shared devices and limited connectivity (Stakeholders 1, 7, 11). Crucially, for some marginalised children, gaming provides *"a community and a sense of belonging"*, particularly where physical environments are unsafe (Stakeholder 4). Interviewees warned that blanket bans and poorly targeted restrictions can cut children off from these connections and shift play into less governed spaces, where support and reporting options are weaker (Stakeholders 4, 6).

2. **The visibility gap remains poorly understood.** The most serious harms are not well captured by public debates that focus on screen time or individual bad actors. Stakeholders repeatedly contrasted the visibility of *"time spent"* which parents/caregivers and politicians can easily measure, with the comparative invisibility of grooming, financial sexual extortion of children and image-based abuse as well as persistent sexual and gender-based harassment that filters girls and marginalised groups out of many gaming spaces (Stakeholders 1, 11, 12).

The findings revealed five interconnected themes that show how evolving risks and the benefits of play, privacy rules, unequal infrastructure, parenting expectations and policy responses combine to influence children's risk and protection in gaming spaces. The themes overlapped, but together they showed how risks and responsibilities are currently unevenly distributed across platforms, families, regulators and support services.

Theme 4.1. Evolving risks, the benefits of play, and moving beyond the screen time focus

Interviewees emphasised three linked points. First, they distinguished risk from harm and stressed that not all exposure leads to harm, and some challenge is developmentally useful. Second, they argued that the most serious threats are changing in form, from one-to-one contact risks to more organised and financially motivated abuse, alongside extremist exploitation through game-adjacent spaces and routine gendered and sexual harms. Third, they highlighted that design choices shape who is exposed to what, particularly through monetisation systems and how these interact with age-related vulnerabilities and neurodiversity. Across these strands, interviewees argued that

safety strategies focused mainly on minimising play can miss both developmentally normal challenges[3] and the pathways through which higher-severity harms escalate.

4.1.1 The risk versus harm distinction

Interviewees cautioned that policy debates frequently conflate risk with harm. Risk refers to exposure to something that could lead to negative outcomes. Harm refers to actual damage to a child's well-being, dignity or development. This distinction shaped prioritisation. Interviewees argued that evidence and investment should prioritise higher-harm areas, including CSA, bullying, harassment and financial sexual extortion of children. By contrast, generic concerns about screen time or low-level profanity often lack clear links to harm and can crowd out attention from more serious threats. As explained in an interview, ***"Risks do not always translate into harms. Bullying and CSA are priority areas"*** (Stakeholder 6). Interviewees' scepticism regarding screen time metrics is supported by recent empirical data. While simple duration is a poor proxy for well-being, it remains a strong predictor of risk exposure. The *Levelling up to stay safe* report (eSafety Commissioner, 2024) identified that highly engaged players encounter significantly more toxicity and harmful content than casual players, simply because they spend more time in the "surface area" of risk through voice chats, lobbies and trading systems.

This distinction suggests that prevention and enforcement should focus on severe harms, particularly where risks are structurally enabled by design or commercial incentives. Moreover, not all discomfort or conflict in play should be treated as harm, provided clear boundaries exist and unacceptable harms are actively reduced.

4.1.2 Gaming as social infrastructure and resilience

One interviewee emphasised that stable, positive gaming communities can function as a protective factor (Stakeholder 10). Belonging to guilds, clans or teams where young people feel recognised and valued can reduce the appeal of extremist recruitment and grievance narratives that target isolation and status-seeking: ***"If a young person already feels they belong in a guild, it is harder to peel them off into extremist communities… gaming identity can make you more resilient"*** (Stakeholder 10). In this framing, identity and belonging operate alongside other protective factors, rather than as a standalone safeguard.

For children in high-stress environments, gaming can offer a controlled space to decompress, practise skills, and sustain connection. Interviewees urged a more nuanced view of adversity in play, noting that losing a match, managing disagreement with a teammate or navigating setbacks can support coping skills and emotional regulation (Stakeholder 6). The policy challenge is to distinguish ordinary play-related challenges from unacceptable harms, including CSA and organised financial sexual extortion of children, and to reduce harm without stripping play of ordinary difficulty and exploration.

[3] Refers to everyday, manageable difficulties that commonly arise in play (for example, competition, frustration, disagreement or boundary-testing). These experiences can support learning and resilience when clear boundaries exist and serious harms (for example, harassment, coercion and sexual exploitation) are actively prevented and addressed.

One interviewee argued that society has ***"shut its eyes towards gaming"*** by treating it as a minor hobby, even as it has become central to many children's social lives (Stakeholder 8). They called for games to receive the same level of research, policy attention and infrastructural thinking that has been applied to social media, precisely because of their protective as well as risky potential. As games function as real social infrastructure for many children, the key policy question is not whether children will be in these spaces, but what protections follow them across the pathways where harm escalates.

4.1.3 From individualised grooming narratives to organised, financially motivated abuse

Interviewees cautioned that many safety narratives still assume an individualised model of risk (a lone adult approaching a child within a single game). While such cases still occur, interviewees described a shift towards more organised and financially motivated abuse, including schemes designed to scale across multiple platforms and identities. They stressed that financial motivation is increasingly central, particularly in schemes involving the financial sexual extortion of children that often target boys (Stakeholder 2). Interviewees described patterns where offenders initiate contact in-game or via game-adjacent spaces, then move quickly to encrypted channels to escalate demands and monetise. Survey participants likewise described online grooming and sexual exploitation as among the most severe and urgent risks in gaming contexts, alongside rising concern about financial harms linked to monetisation practices, gambling-like mechanics and cross-border enforcement gaps (Survey, Annex 3).

Interviewees argued that this shift changes what effective prevention looks like: advice focused solely on not speaking to strangers is insufficient on its own. They pointed instead to disrupting networks, sharing signals across services, coordinated action to remove offenders and partnerships that support financial tracing and account recovery (Stakeholders 2, 10). Platforms and policymakers cannot treat grooming or financial sexual extortion of children as isolated incidents. These harms require systems built for repeat offenders who create new accounts quickly and move children between services to escalate harm.

4.1.4 Violence presented through game-like aesthetics in extremist ecosystems

One interviewee raised concerns about the ***"gamification of offline violence"*** (Stakeholder 9). For this report, the concern is not games themselves, but the use of game-like presentation and connected community spaces to circulate violent content and route users towards more explicitly ideological material. This refers to real-world attacks or threats that are filmed or edited to resemble games, for example, through first-person perspectives, heads-up display graphics, scoreboards or livestream overlays that mimic popular interfaces (Stakeholder 9). One participant mentioned how attack footage may be **"designed to look like a game"**, using familiar visual language to make real-world violence feel legible to audiences shaped by gaming cultures.

Another interviewee described similar dynamics operating across connected, ***"game-adjacent social spaces such as Discord"***, including how these spaces can link outward into other forums and communities where harmful content circulates (Stakeholder 4). In this account, the risk is not confined to one service or platform. It sits in the connected pathways through which content, social belonging and escalation can travel across spaces and platforms. This can include fan and modding communities[4] where extremist networks may already be active. One interviewee pointed to ***"this huge white supremacist community of mod makers"*** around them and argued that ***"the use of narratives and myth and motif"*** can matter strategically (Stakeholder 9). In this framing, gaming cultures and adjacent production communities can provide both an audience and a set of shared references that extremist actors can exploit to distribute content and signal belonging.

Interviewees were careful not to claim simple cause-and-effect links between playing games and committing violence. Instead, they described gaming aesthetics and game-adjacent infrastructures being used instrumentally within wider extremist ecosystems (Stakeholder 9). This has implications for detection and moderation across video-sharing platforms, social media and game-adjacent environments.

4.1.5 Sexual and gender-based harms

Sexual and gender-based harms are described as common in many gaming spaces rather than rare shocks. Interviewees stressed that for girls and some other marginalised groups, routine misogyny, racism and sexualised abuse can operate as a barrier to participation and sustained play. One interviewee described an ***"eight-week cliff"***, referring to a pattern in their workshops and research in which some girls who encounter early toxicity and misogyny disengage within the first eight weeks (Stakeholder 8). They stressed that this is not a deficit in resilience. It reflects an environment in which staying requires tolerating harassment, sexualised comments and gendered slurs as a routine cost of participation.

One interviewee drew on long-standing evidence that girls avoid voice chat because ***"as soon as they [were] identified by their gender, they would get bullied"*** (Stakeholder 13). They cited this as a key business and rights case for investing in large-scale voice chat moderation systems aimed at reducing daily abuse. In their view, platforms that fail to address routine misogyny are not neutral. They are actively selecting for players who can tolerate abuse and filtering out those who cannot or will not. From a business perspective, stakeholders framed this as both a rights issue and a retention crisis. Platforms are effectively shrinking their potential player base by tolerating environments that drive away a substantial proportion of girls and marginalised youth. This highlighted the commercial benefits of protecting players in games.

4.1.6 Financial and design harms

Across interviews, stakeholders described monetisation systems and design patterns as ways commercial incentives can increase spending pressure. One interviewee explained how monetisation can be designed to track children's developmental stages. For younger children, they argued that time gating[5] is used because many at this age ***"do not***

[4] Fan or modding communities are groups of enthusiasts who create their own custom content or alter a game's original software to change how it looks or plays.

[5] Time gating refers to design that delays progress unless a player waits for a set period or pays to speed it up, for example, waiting several hours for an in-game item to unlock or paying to unlock it immediately.

know how to wait or pause" and want instant gratification, so progress slows or stops unless they wait or pay (Stakeholder 8). As children get older, monetisation can shift towards offers that ***"save you time",*** such as paying to skip repetitive gameplay (Stakeholder 8). They also raised concerns that, in some first-person shooters,[6] systems may ***"manipulate the matchmaking"***[7] after a purchase so that a player is matched against weaker opponents for a period, making it appear that the purchase improved performance and encouraging further spending. They also noted that these techniques are reported most clearly in parts of the mobile gaming market, where monetisation is less consistently regulated and enforced (Stakeholder 8). When a child spends money and loses, they may feel pressure to ***"make up for the loss"***, leading to repeated purchases and, in some cases, attempts to find money through risky or illegal means (Stakeholder 11). Even where countries try to regulate loot boxes, poorly drafted rules can be easy to evade (Stakeholder 1). Companies can repackage the same mechanics under different names or slightly altered formats, preserving the underlying gambling dynamic.

Interviewees also emphasised how these design choices interact with social pressure (Stakeholders 4, 8, 13). Cosmetics and skins[8] were described as a form of social currency. In many games, not spending money is a visible form of exclusion. Those who pay can display rare items or passes, while those who do not may be sidelined from peer activities or viewed as less committed. Interviewees noted that groomers exploit these dynamics, using valued in-game currency or items as leverage to draw children into private chats or build trust, because ***"skins, privileges"*** and in-game currency were described as so valued that they become an easy incentive for adults seeking to gain a child's trust, particularly where families cannot afford regular in-game spending (Stakeholders 6, 8). A participant argued that policy needed to distinguish between benign identity expression through skins and predatory designs (Stakeholder 6).

Many games now rely on a ***"whale hunting"*** model, where a small minority of players produce most of the revenue (Stakeholder 8). Those who spend heavily are then targeted with more offers and marketing. This pressure is intensified for digital breadwinners whose gameplay is tied to income generation for their families. Interviewees stressed that this dynamic ***"affects children as well"*** (Stakeholder 10). They noted that children often have less capacity to regulate spending and may be playing on accounts registered as adults. They also pointed out that ***"the majority of children do not play on child accounts, they play on adult accounts"***, which makes it harder for companies to adopt an explicit child lens in monetisation even when they are motivated to do so (Stakeholder 8).

Interviewees underscored important genre differences (Stakeholders 1, 10). Mobile games emphasise heavy monetisation, short play loops designed to prompt repeat spending and aggressive microtransactions[9]. In large multiplayer titles, social risk and harassment often dominate, with monetisation layered on top. Indie games may raise more questions about content and themes than about monetisation structures (Stakeholders 1, 10). ***"Games are very different. For mobile, we focus more on monetisation; for multiplayer, more on interactions; for indie games, more on content"*** (Stakeholder 1).

One interviewee offered a nuanced view of these status economies. They acknowledged that loot boxes and some cosmetic systems can be predatory but also noted that buying skins and items can be a way for children to express identity, much like wanting ***"new trainers"*** in the offline world (Stakeholder 6). Another interviewee emphasised that the free-to-play model can be an ***"equaliser opportunity"***, expanding access for children who cannot afford upfront costs (Stakeholder 4). For children in low-income households, including many in LMICs, free-to-play titles are often the only gateway to high-quality digital play. A blanket attack on the model risks shutting out those children entirely (Stakeholder 4).

The policy challenge is therefore twofold. First, to curb exploitative design and monetisation practices, especially where they target or disproportionately harm children by exploiting impulse control, loss chasing or social exclusion. Second, to preserve the access benefits that free-to-play models can provide, including opportunities for play and identity expression while ensuring that spending remains transparent, proportionate and compatible with children's rights.

---

[6] First-person shooters are video games centred on weapon-based combat that are played from the perspective of the protagonist, as if seeing through their eyes.

[7] Matchmaking refers to the system a game uses to pair players or teams for a match, typically aiming to create a reasonably balanced game based on factors such as skill level or recent performance.

[8] Skins and cosmetics refer to paid-for visual customisations that change how a character, avatar, weapon or item looks without necessarily changing gameplay.

[9] Microtransactions refer to small in-game purchases, for example, buying virtual currency, items, boosts, subscriptions or cosmetic upgrades.

4.1.7 Neurodiversity and differentiated impact

Interviewees noted that safety and design features do not affect all children in the same way and highlighted neurodivergent children as one group who may be disproportionately exposed to both monetisation pressure and rule-enforcement risks (Stakeholders 3, 6, 9, 12). One interviewee suggested that children with attention differences or higher impulsivity may be more responsive to in-game prompts and offers, and more likely to be flagged by moderation and rule-enforcement systems that rely on behavioural signals (Stakeholder 10). At the same time, interviewees emphasised that clear, consistent cues and simple, predictable rules can be particularly effective supports for some neurodivergent players (Stakeholders 3, 6, 12).

Interviewees cautioned against treating neurodiversity as a catch-all explanation for heavy play or problematic behaviour, noting the risks of stigma and over-medicalisation (Stakeholders 6, 9, 12). Instead, they framed this primarily as a design and usability issue. One interviewee relayed feedback from an autistic child who had played for years without recognising that a small red flag icon was the reporting function: ***"I've been playing [game] for six years, and I didn't actually realise that the red flag meant report. Because my brain doesn't work like that"*** (Stakeholder 3). They argued that this points to the value of more explicit and accessible safety cues and consistent iconography across platforms, and that developers should seek input from neurodivergent children when tools are designed (Stakeholder 3).

These accounts reinforce a wider finding of this report that both risks and protective tools are unevenly distributed. If reporting routes, prompts and defaults are not usable and legible for neurodivergent children, then children who may benefit most from structured support can be least able to access it. This strengthens the case for safety-by-design that is tested with diverse children rather than built around a single "standard user" assumption. Survey respondents also noted the lack of robust, comparable and applied evidence on which prevention approaches work in gaming ecosystems, particularly those grounded in children's lived experience across diverse contexts (Survey, Annex 4).

Theme 4.2. The privacy paradox: legal uncertainty and data sharing

4.2.1 Systems optimised for commercial privacy

Across the interviews, privacy and data protection law shaped what gaming platforms can do to protect children and what they do in practice. Many interviewees described current interpretations of privacy and data protection legislation as being centred on adult privacy and business risk rather than on children's rights. This creates a recurring tension between privacy laws, such as the EU GDPR and the United States Children's Online Privacy Protection Act (COPPA), and the data that safety teams say they need to identify, prevent and respond to risk. As one interviewee summarised, the result is ***"a system optimised for commercial privacy rather than child safety"*** (Stakeholder 5). Many responses to the survey echoed this as an unresolved governance challenge, describing a need to improve detection of harmful behaviour while avoiding surveillance-led approaches that erode trust or enable misuse of children's data (Survey, Annex 3). Under the GDPR, for example, children's data are recognised as requiring specific protection, yet enforcement and industry compliance practices still largely assume an adult default and focus on commercial tracking and advertising rather than child protection outcomes (UN Committee on the Rights of the Child, 2021; Stoilova et al., 2020). Conversely, in South Asia, practitioners noted the gap is not over-regulation but a lack of specific gaming frameworks in legislation (Stakeholders 7, 11). In Nepal, for example, the main legal reference point is general cyber and electronic transactions legislation, including the Electronic Transactions Act, 2063 (2008), alongside broader criminal law, rather than gaming-specific online safety duties or regulatory guidance (Council of Europe, 2020). As one interviewee in Nepal observed, ***"we don't have many interventions [sic] or policies or even the legal frameworks covering gaming… so I think the approach is very general"*** (Stakeholder 7), which left grooming, bullying and gambling-style mechanics largely unaddressed. This lack of tailored legislation leaves little scope to enforce stronger child protection standards across the industry.

Safety practitioners said many of their day-to-day challenges stemmed from privacy restrictions. One interviewee explained that many obstacles relate to what data the gaming industry is permitted to collect or keep (Stakeholder 2).

"If you're not able to store certain data… it makes detection a bigger lift… striking that privacy safety balance has been the eternal struggle" (Stakeholder 2). In their view, limits on storing chat logs, behavioural signals and links across services made detecting grooming and other harms much harder. When platforms cannot keep relevant information, investigations become challenging and much of the evidence may have disappeared by the time police become involved.

Interviewees emphasised that the issue is not privacy itself, but the narrow, adult-centric way that privacy is applied by regulators and internal legal teams. Across the interviews, privacy was described as framed mainly around adults' interests and commercial data flows, while children's privacy was rarely treated as a positive right that must be balanced with protection, participation and access to information (Stakeholders 1, 5, 6; Livingstone et al., 2019). This critique was consistent with the UN Committee on the Rights of the Child's General Comment No. 25 on children's rights in the digital environment, which calls for children's best interests to be a primary consideration in the design and governance of digital services and emphasises the need to reconcile privacy and protection in ways that serve children's rights as a whole (UNCRC, 2021).

Several interviewees pointed to misaligned incentives embedded in platform governance. One participant noted that platforms can fully monetise an account treated as belonging to an adult, whereas verifying that a user is a child may trigger stricter rules, creating ***"no incentive to identify that they are a child"*** because ***"you can fully monetise an adult account"*** without triggering restrictive safety codes (Stakeholder 8). A second participant connected this to wider ecosystem gaps, including limited local age and content rating infrastructure, observing, ***"We don't have an age rating portal for India or a content rating portal for India like…ESRB or PEGI"*** (Stakeholder 11). In this configuration, privacy law functions less as a shield for the child and more as a barrier to safety-by-design, allowing high-risk commercial profiling to continue while hampering the detection of predatory behaviour.

Several interviewees advocated for privacy-preserving approaches that still allow proportionate detection and action, for example, through limited data retention, clear governance and strict purpose limits, rather than restricting protective use of data by default (Stakeholders 2, 3, 4, 6, 10, 13). This analysis was consistent with broader evidence that children's data are routinely exploited for profiling and monetisation, while data governance for child protection and rights impact remains underdeveloped (Livingstone et al., 2019; UNICEF, 2017; Data Protection Commission, 2021). They further argued that this can create the wrong incentives: if a company does not clearly identify child users or track high-risk patterns, it may reduce legal and compliance exposure, even if that weakens protection.

This tension is captured in contrasting observations from different stakeholders. Safety teams frequently described current privacy interpretations as a genuine barrier to retaining logs and sharing high-risk signals, noting that in some regions, even storing risk scores or behavioural flags may be interpreted as incompatible with principles of data minimisation[10] and storage limitation[11] under the GDPR. One interviewee argued that privacy is sometimes invoked to avoid legal risk or investment, with companies often citing reluctance to store detection data for fear of breaching data protection law.

Some interviewees felt that commercial uses of behavioural data often receive more organisational support than child protection uses. This perception fuels suspicion that the main privacy constraint is organisational and political rather than legally mandated.

4.2.2 Privacy as constraint, liability fear and convenient alibi

Interviewees described three overlapping ways in which privacy manifests in practice. First, there are substantive legal constraints. Safety leads described genuine uncertainty about whether storing richer behavioural data or risk scores will breach data protection rules. Those working across multiple regions also highlighted conflicting laws across countries on data retention and sharing, which makes it unclear what can lawfully be stored or transferred when a single child's case data sits within several countries. For example, a grooming case that begins in a European server, moves to a North American messaging app and involves a child physically located in a third region can trigger different, and potentially conflicting, interpretations of permissible data processing. In this environment, risk-averse

[10] Collecting only the minimum data needed
[11] Not keeping the data longer than needed

interpretations of privacy law can lead organisations to delete or avoid collecting data, even when this undermines child protection.

Second, several safety leads described a pervasive climate of liability fear, sometimes referred to as a ***"legal chill"***, where the risk of litigation slows or prevents safety work (Stakeholders 2, 3, 12, 13). Safety proposals may be delayed because in-house legal teams worry that detailed logs or risk assessments could later be used to argue the company ***"knew and did not act"***, creating what one interviewee called a ***"discovery trap"*** (Stakeholder 12). As another put it, ***"the more you know and record, the more it can be used against you"*** (Stakeholder 2). This creates a perverse incentive structure in which ignorance can appear legally safer than knowledge, even when better knowledge would enable earlier interventions to protect children. This maps to survey concerns about fragmented legal and regulatory frameworks, with respondents calling for clearer guidance on responsibilities and proportionate safety action across jurisdictions (Survey, Annex 3). Some legal teams are said to resist ***"deterrence messaging"***, for example, public statements that imply offenders are being actively monitored, on the grounds that this could reveal detection practices to offenders (Stakeholder 2).

Third, several interviewees described situations where privacy is used as a convenient alibi (Stakeholders 5, 10, 12). In these cases, privacy is invoked to resist transparency or cross-industry collaboration or to avoid investments that would be costly or politically sensitive. Several civil society interviewees cited examples where proposals for cross-platform research, transparent data access for independent auditors or participation in shared signal-sharing alliances were rejected primarily on privacy grounds, even though similar data were already processed for commercial purposes.

These three dynamics often coexist within the same organisation. Several interviewees noted that genuine uncertainty about legal boundaries sits alongside a tendency to invoke privacy defensively, particularly when projects involve cross-team collaboration, external scrutiny or costs that are hard to justify within short budget cycles (Stakeholders 10, 11).

From a policy perspective, some interviewees suggested that policymakers and regulators may need to create clearer legal protections for good-faith child safety work (Stakeholders 2, 12). This could include explicit recognition in data protection guidance that proportionate retention and sharing of high-risk safety signals, within tightly governed frameworks, is compatible with data minimisation and purpose limitation,[12] where it is demonstrably in the child's best interests. Without such protection, liability fears were described as risking the erosion of safety innovation and allowing data minimisation to be weaponised against child protection efforts.

[12] Use data only for a clear, stated safety purpose.

4.2.3 Cross-platform abuse and the need for signal-sharing

Interviewees described abuse as often spanning multiple platforms, while signals were hard to share because of privacy and competition constraints. Survey respondents similarly stressed that gaming-related harms cannot be addressed in isolation. Gaming was described as sitting within a wider digital ecosystem, including social media, gambling-adjacent services, advertising and emerging technologies (Survey, Annex 3). Offenders were described as moving fluidly between games, social media and encrypted apps, while most safety tools remain siloed within single platforms. This mirrors wider evidence that abuse often begins in games and escalates elsewhere, with games functioning as an initial contact or recruitment space before escalation (Smahel et al., 2020).

> ***"Most of the game companies, they forbid some sexual words on the games. But the problem is that the predator asks the kids to go to other platforms. And he uses each platform on the limit of the moderation, so no one can kick the guy out and this is a problem … games are an entry point to grooming, but it's difficult for the game company to kick the predator off because he's doing nothing wrong, asking someone to go to other places"*** (Stakeholder 1).

This aligns with wider evidence that social media platforms remain the primary environment where online sexual harms are documented[13]. Several interviewees noted that ***"most documented cases are recorded on social media rather than gaming platforms"***, while cautioning that this likely obscures the role of games as a ***"recruitment or first-contact space"*** (Stakeholders 2, 6).

Several interviewees described this off-platform migration as a widespread pattern, relevant not only to contact abuse but also to the spread of harmful sexual and gender norms (Stakeholders 5, 8). They further highlighted that the inherently social nature of gaming, and the feeling of ***"doing things together"*** as a team, can create strong social foundations that are later exploited to manipulate and recruit children, including into extremist groups and organised crime (Stakeholder 5).

Several stakeholders pointed to cross-service signal-sharing as one of the few approaches that matches the cross-platform reality they described, where an offender can leave a game and reappear elsewhere before a single platform has enough visibility to act. Initiatives such as Lantern, facilitated by the Tech Coalition, were highlighted as valuable for coordinating action against online CSE and abuse (Stakeholder 2). In practical terms, signal-sharing allows one service's detection to become another service's warning, supporting faster identification of repeat offending patterns and more consistent enforcement when offenders attempt to evade bans by moving between platforms.

While signal-sharing initiatives are much needed and valuable, they present structural barriers for smaller services. High-quality safety data are expensive to generate, and the data a company holds are treated as a commercial advantage, which weakens incentives to share it beyond narrow, legally required purposes, leading to reduced transparency across the sector (Stakeholders 2, 12). Smaller platforms, including many games, may lack the technical and legal resources and capacity to integrate with programmes such as Lantern, even when they are most in need of shared intelligence. Tech Coalition's own parameters for signal-sharing, for example, require that signals be legally shareable, necessary, proportionate and aligned with platform policies and privacy frameworks, and that Lantern does not facilitate automated enforcement decisions (Tech Coalition, 2024). For smaller firms with limited counsel, this legal and compliance overhead can be difficult to navigate. The result is that smaller studios, including many in LMICs, are structurally least able to join exactly the kind of scheme that could strengthen protection for their players, reinforcing infrastructure inequalities[14].

[13] A large-scale study found that in 2023, the platforms where minors most frequently reported online sexual experiences were Snapchat (16%), Instagram (14%) and Messenger (13%), with gaming platforms appearing much lower in prevalence (Thorn, 2024).

[14] Lantern Transparency Report (2024). The report describes Lantern's model for sharing high-risk signals (including CSAM URLs and image and video hashes, plus incident-linked identifiers such as usernames and email addresses). It reports that during the 2024 compliance cycle, Lantern-supported workflows contributed to enforcement actions against **102,082** accounts, removal of **7,048** pieces of CSAM and the blocking or removal of **135,077** CSEA URLs.

Finally, cross-platform signal-sharing sits within wider tensions between privacy and transparency. Lantern's governance has been shaped by a human rights impact assessment that warns that even legitimate child protection efforts can put rights such as privacy and freedom of expression under pressure if not carefully managed (Tech Coalition, 2024). Privacy measures designed primarily around adult users, such as hiding friends lists or minimising visible histories, can remove cues children rely on to judge whether a stranger feels safe or risky (Stakeholder 8). This can result in children being more exposed to high-risk contact while companies remain constrained in their ability to share signals about offenders across services.

4.2.4 A "cheating-centred" safety culture and misaligned incentives

How companies interpret privacy and allocate data budgets is heavily shaped by internal incentives. Several interviewees described a consistent resourcing asymmetry, where fraud and anti-cheat risks are treated as core business risks and receive stronger engineering and data support than child safety (Stakeholders 2, 10, 12). Interviewees described this as a ***"cheating-centred"*** safety culture in which stopping cheaters and fraud was treated as core business risks, while child safety is treated as a compliance issue (Stakeholders 2, 12). As one interviewee put it, ***"the gaming industry hates cheaters, that's where the safety culture is centred"***, and another added that anti-cheat work is often ***"quite comprehensive for these major players"*** (Stakeholders 2, 12).

This priority is reflected in the depth and intrusiveness of technical measures. Interviewees described anti-cheat tools that can run at system level on players' devices and inspect broadly for signs of cheating software (Stakeholders 2, 8, 10). By contrast, grooming, harassment and other contextual harms do not require extra software and often look, on the surface, like everyday social interaction until the point of abuse. Detecting them typically relies on language and behaviour analysis that is more subtle, harder to automate (current automated systems struggle to understand context) and historically underfunded. One interviewee summarised the imbalance bluntly: companies ***"invest heavily"*** **in anti-cheat, but it is** ***"much harder to get budget to stop groomers"*** (Stakeholder 10).

Interviewees emphasised that this is not only a technical challenge, but also a choice about priorities. The industry has shown that it is willing to deploy intrusive tools when the perceived threat is to game integrity and revenue, which suggests the reluctance to use data in comparably robust ways for child protection reflects governance and resourcing decisions as well as legal uncertainty. This does not imply that system-level anti-cheat tools can simply be repurposed to detect grooming because the behaviours and signals are different. However, it does illustrate the scale of investment and operational tolerance for intrusiveness when risks are framed as integrity and revenue threats.

Several interviewees linked this to wider resourcing patterns. ***"Significant cuts in trust and safety teams"*** across parts of the tech sector in recent years[15] were noted (Stakeholder 8). In some cases, promising initiatives were lost overnight when key staff were made redundant, and their posts were not replaced. These reductions have often coincided with increased investment in live-service operations, monetisation features and anti-cheat capabilities. Similar concerns are echoed in critiques of digital business models that prioritise engagement and revenue growth over children's rights and well-being (UNICEF, 2017; Stoilova et al., 2020). However, this pattern is not universal. Some studios, particularly those under close regulatory scrutiny or with strong public commitments to child safety, have invested in safety-by-design and grown trust and safety capacity (Stakeholders 4, 13).

Some safety leads respond by deliberately reframing CSA and grooming as integrity or cheating problems internally to access existing budgets and tools. Closer collaboration between anti-cheat and child safety teams was emphasised*:* ***"The anti-cheating team and the child safety teams can work more together"*** (Stakeholder 3). They pointed out that anti-cheat functions already share information about bad actors and techniques across companies when it is in their mutual interest. They argued that a similar model of collaboration and technical rigour is needed for child safety. Others cautioned that while this is a pragmatic way to win resources, it risks reducing sexual abuse to a problem of rule-breaking within the game rather than recognising it as a fundamental violation of children's rights (Stakeholder 6).

[15] For example, reports indicate that X Corp slashed its global trust and safety staff by 30% and reduced its safety engineering team by 80% following its acquisition (McGuirk, 2024).

4.2.5 Age assurance and age verification: contested but unavoidable

Debates about privacy and safety converged sharply around age assurance and the methods used to estimate or verify whether a user is a child or an adult. Interviewees described it as central to the privacy and protection paradox, where privacy, liability and business models collide. Views differed on acceptable approaches, but distinguishing children from adults is central to targeted protections. Age assurance spans methods with different levels of confidence, from self-declaration (widely regarded as ineffective) to age estimation and higher assurance checks such as identity-linked or bank-linked verification[16]. See Annex 4 for a summary of common approaches.

One interviewee emphasised that there is no ***"silver bullet"*** (Stakeholder 5). Higher-assurance methods (for example, passports) offer certainty but raise privacy and exclusion concerns, while lower-assurance methods (for example, estimation from a photo) may be more inclusive but less precise. The policy challenge is matching assurance to risk, for example, requiring higher assurance for age-restricted purchases than for access to a social lobby, without excluding children who lack documentation. Several interviewees argued that without credible age assurance, platforms cannot systematically tailor protections, including restricting gambling-like features (such as loot boxes) or applying stricter default privacy settings for children (Stakeholders 8, 10). Another interviewee suggested the ***"main game changer"*** would be an age verification system that works reliably for younger users and ***"can be trusted by all parties",*** while warning that some early proposals offered limited safety gains alongside ***"a huge invasion of privacy for adults"*** (Stakeholder 12). This aligns with wider concerns that early generations of such schemes can increase privacy and security risks without proportionate child safety gains.

As checks become more sophisticated, so do evasion tactics, including virtual private networks (VPNs), spoofed images[17] for selfie-based checks and tools that manipulate age estimation systems. Rigid, poorly designed mandates risk ***"sabotaging trust"*** and driving children towards workarounds that create new privacy and security risks (Stakeholder 5). Multiple interviewees emphasised that strict age verification can displace children into less governed and less visible spaces (Stakeholders 4, 7, 11, 13). This can deepen inequity because children who cannot pass verification due to documentation gaps or unstable connectivity are often already facing higher offline risks.

Survey data and industry usage figures suggest that many children start playing mainstream online games and connecting socially between the ages of 10 and 12, yet privacy laws in some jurisdictions assume that users under thirteen should not be present at all (Smahel et al., 2020; Ofcom, 2022). This was described as a ***"10 to 12 limbo…They built the system that pushes children to lie about their age"*** (Stakeholder 8). Platforms may pretend these tweens do not exist or allow them to present as older teenagers to keep accessing services, which places them outside the scope of child-specific protections. Without age-appropriate models that recognise this group, regulators and companies risk leaving a large cohort of children structurally invisible to safety tools.

While the recent adoption of international standards[18] introduced the first globally agreed metrics for measuring the accuracy of age estimation tools and protocols for interoperability, regulatory alignment remains fragmented (Age Verification Providers Association, 2024). Different countries still set different rules on what age checks are acceptable, which can force families to repeat verification across services. Stakeholders argued that greater alignment and interoperability would reduce this burden. However, equity remains a challenge. One interviewee noted that even standardised age estimation tools are often more accurate for older teenagers than for younger children and asserted that ***"no system can yet reliably validate whether a user is under 12 or 13 without collecting sensitive data"*** (Stakeholder 12). Documentation gaps, uneven civil registration systems and shared device use can make "one child, one device, one ID" models exclusionary, particularly in LMICs (Stakeholder 4).

Against this backdrop, some companies are moving from voluntary to mandatory age assurance under regulatory pressure. Interviewees referenced established legal regimes, such as the United Kingdom's Online Safety Act and the European Union's Digital Services Act, alongside newer, stricter mandates. Most notably, Australia recently implemented a mandatory age delay on social media for children under 16, requiring platforms to use age assurance

[16] As defined by the Age Verification Providers Association (AVPA) and international standards such as ISO 27566, age assurance is not a single tool but a spectrum of methods, each offering a different "level of assurance".

[17] The use of non-live or manipulated imagery, such as printed photos, digital screens, or deepfakes, to deceive an age verification system into accepting a false identity or age.

[18] Specifically, IEEE 2089.1-2024 (Standard for Online Age Verification) and ISO/IEC 27566-1:2024 (Age assurance systems — Part 1: Framework).

technologies such as government ID checks or facial age estimation (UNICEF Australia, 2025; BBC News, 2025). This reflects a broader international momentum, with a number of countries, including Spain, Singapore, Denmark, Turkey and Mongolia, also considering or drafting legislation to enforce age limits or social media curfews for minors (BBC News, 2025). In response, several large platforms have begun to introduce harder verification for high-risk content such as adult-rated games or explicit user-generated material. One interviewee described layering ID verification for adults with face or parental verification for younger users, then limiting richer communication features to connections that can be shown to know each other offline (Stakeholder 3). Regulators increasingly frame these shifts as part of safety-by-design, where services default younger users into safer settings and limit high-risk features, whatever their declared age.

Multiple interviewees cautioned that age assurance is only one part of protection and cannot replace safer design, moderation and accessible reporting routes (Stakeholders 1, 5, 13). Some argued that services should be built to be safe for children and teenagers by default rather than relying on brittle entry gates (Stakeholders 5, 8). Others favoured stricter enforcement of age limits, including analogies to film classification, and proposals to limit younger users to offline or single-player modes until a certain age (Stakeholders 7, 11, 12). Taken together, interviewees described no consensus on the right balance between verification, design and education.

Theme 4.3. Inequalities in safety tools and infrastructures

Across interviews, participants described how online safety was still often framed as an individual responsibility, where risk is assumed to be manageable if parents and caregivers use the "right" settings and children follow the "right" advice (Stakeholders 1, 4, 6, 8, 10). They argued that many safety tools were built around a ***"wealthy default"***, assuming every child has their own device, stable Wi-Fi and parents and caregivers with the time and skills to supervise their children online (Stakeholder 1). Survey respondents reinforced this equity gap, noting that children and communities in the global majority are often underrepresented in governance, research and safety design, with language accessibility and culturally relevant safeguards repeatedly described as missing (Survey, Annex 3). This theme highlights three linked patterns: (1) high-risk digital environments in low-resource settings, (2) children as "digital heads of household" and (3) social norms and stigma that raise the social cost of reporting. In this section, "caregivers" refers to parents and other adults with day-to-day responsibility for a child. Most interviewees spoke about parents, but the constraints described apply more broadly.

4.3.1 The high-risk, low-resource realities

In many low-resource settings, families may face significant offline constraints, while children still participate in high-risk digital environments. Interviewees highlighted a consistent pattern: children may encounter the same spending pressures (for example, prompts to spend in-game) and grooming behaviours as children in higher-resource settings, but with fewer safety tools, less moderation, less regulation and enforcement and even more limited support services (Stakeholders 1, 7, 11). This reinforces the need for policy and design approaches that account for offline constraints and online risk at the same time.

Shared devices were common in settings with limited resources, shaping what families can realistically do with safety settings. As one interviewee explained, ***"If you have just one device per family, you are not setting a strong parental control because you are the adult using it"*** (Stakeholder 1). One interviewee illustrated how this can look in marginalised communities: ***"Mobile is what it is because now it is affordable. You have cheaper mobiles, so children in marginalised groups share one device a week, so everybody has one per day and then it rotates in the group. Consoles are meant for people who can really afford those"*** (Stakeholder 1).

These dynamics also shape the types of games, monetisation schemes and social interactions children encounter. Children in higher-income households may play in console-based settings with more developed reporting and blocking tools, whereas children in lower-income households are more likely to use free-to-play mobile games, internet cafés or informal servers where protections are weaker and oversight is more limited (Stakeholder 6).

Interviewees highlighted further economic complexity in some contexts: children acting as ***"digital breadwinners"***,[19] where restrictive safety measures can become an economic threat to the household, and caregivers may prioritise access over protection (Stakeholder 4). In some regions, young players generate essential household income through in-game markets, live streaming or esports, altering the power dynamics of supervision and turning some online risks into a price families feel compelled to pay.

A further consequence is data gaps. Where play happens through shared devices, informal accounts, pirate servers or private channels, platforms have weaker signals about who is playing and what is happening, and less ability to detect and act. Automated safety systems are therefore more likely to be built and tested on the behaviour of wealthier, individually registered users, and then applied to contexts they do not adequately represent. One interviewee warned that ***"the next billion gamers will be mobile first, often in Africa or the Middle East, but those user pathways are not very well understood because a lot of gaming research is still centred on Global North, wealthy contexts"*** (Stakeholder 9). The result is that children facing higher risks can be the least visible in the datasets that guide safety investment and enforcement.

One interviewee reported that gaming is widespread in Nepal, but that many caregivers focus mainly on time spent online and have limited visibility of online risks*: **"Almost all children I talk to, they play games … but the language children use, the way they interact and all these things ... most parents only worry about the time children spend online, so they do not know the traces, the content, contact and conduct"*** (Stakeholder 7). Several interviewees resisted narratives that frame these gaps as failures of parental responsibility. They argued that built-in safety tools are often mismatched to high-risk, low-resource environments, forcing families into blunt trade-offs between child safety and household needs, especially when the same device is required for daily life (Stakeholders 1, 6). Safety must be built directly into games, rather than assuming households can install and manage protections themselves (Stakeholder 13).

4.3.2 Children as "digital heads of household"

These gaps reshape family roles. Interviewees described children as ***"digital heads of household"****,* particularly in contexts where children are the family member most confident managing devices, apps and accounts (Stakeholders 1, 7, 11). Adults may rely on children for troubleshooting, while children manage passwords, settings and game mechanics (Stakeholder 11). Participants argued that these inequalities are not only about devices and bandwidth, but also about who is expected to manage safety settings in practice (Stakeholders 3, 4, 5, 6, 7, 10, 13). As one interviewee put it, there is ***"a huge divide in the technology quotient … way higher in children and non-existent in parents"*** (Stakeholder 7), raising questions about how safety systems function ***"when parents are not [digitally] literate, let alone game literate"*** (Stakeholder 3).

Several interviewees described the COVID-19 pandemic as a turning point. As schooling moved online, families that had not previously owned devices were suddenly required to provide connectivity, often in contexts where parents were digitally illiterate. One interviewee recalled: ***"In the pandemic, it was almost mandatory for parents to provide tools for education, but many parents were digitally illiterate, so children had to navigate everything"*** (Stakeholder 9). Interviewees noted that this reduced adult visibility into which games are installed and with whom children interact, because adults do not control the accounts where play takes place (Stakeholder 11).

4.3.3 The "gaming literacy" gap

The interview findings suggest a critical distinction between digital literacy and gaming literacy. Digital literacy refers to the functional use of devices, for example, using messaging platforms, using social media or accessing a bank app. Gaming literacy refers to understanding the cultural and social context of games: mechanics, language, norms and risk signals. Some caregivers are comfortable using phones and apps but are less familiar with how online games or gaming platforms work and what is normal within them (Stakeholder 7). Cooperative raids may be mistaken for interpersonal violence, fast-paced banter confused with bullying and guild structures treated as undifferentiated online strangers, regardless of whether they are peers from school or genuinely unknown adults.

[19] Bringing income into the household through gaming-related activity.

For many children, particularly those in precarious social or physical environments, gaming can provide opportunities for social connection, skill-building and identity formation (Stakeholder 6). Risk increases when caregivers lack gaming literacy and cannot distinguish protective benefits from actual harm. One interviewee suggested this is one reason parental control dashboards have low uptake: ***"Parents who do not understand the underlying activity struggle to decide which features to limit and which to leave alone"*** (Stakeholder 13).

In lower socio-economic contexts, time poverty further constrains parental engagement. Interviewees encouraged co-play and open conversations as best practice but acknowledged that for many families, this advice does not align with economic realities, as caregivers may have multiple jobs and will not have the time for that level of engagement (Stakeholders 1, 10). Messaging that tells caregivers to supervise more, co-play or configure elaborate controls is structurally flawed when adults cannot confidently use the device or spare the time. Interviewees argued that safety cues and interventions must reach children directly in-product, in language and formats they understand, rather than relying on an imagined parental filter that often does not exist (Stakeholders 1, 3, 5, 13).

4.3.4 Cultural norms, shame and the social cost of reporting

Even when reporting tools are available, interviewees emphasised that they are often underused because of social norms, shame and fundamental lack of trust that the platform will investigate or respond. One interviewee noted that in some peer groups, ***"reporting toxic behaviour is perceived as weakness or betrayal"*** (Stakeholder 11). As relayed by a participant on how children described report buttons: ***"Reporting is for children who have no self-respect, not able to cope with the situation or distress"*** (Stakeholder 3). In these contexts, using platform tools can mark a child as oversensitive or unable to "handle" it, particularly in competitive games where taunting is treated as normal. In such environments, openly challenging harassment may invite further targeting, while formal complaints can feel futile or dangerous.

These dynamics intersect with parental responses. Interviewees consistently reported that children fear disclosing grooming, financial sexual extortion of children or bullying because they expect punishment, such as losing access to games or devices, rather than support (Stakeholders 1, 5, 7, 11). Survey respondents echoed this, identifying reporting and moderation failures, slow or inconsistent enforcement and barriers for third-party advocates as persistent reasons children and supporters perceive that ***"nothing is done"*** when harm is reported (Survey, Annex 3). Interviewees also raised an additional tension around child privacy. When platforms share details of a child's report with caregivers (for example, through email notifications or dashboards), this can expose sensitive information about the child. One interviewee emphasised that designers must balance involving caregivers with protecting young people from potential backlash at home, because automatically forwarding every report to a parent can raise the stakes of speaking up rather than lowering them (Stakeholder 13). It would be better to create systems that ***"promote dialogue between parents and children"*** (Stakeholder 13).

Police and justice system capacity also shapes whether reporting feels worthwhile. Interviewees described overloaded cybercrime units where child protection is deprioritised relative to financial crime, terrorism and other crimes, which consequently shaped young people's perceptions of report buttons as cosmetic compliance symbols rather than real pathways to help (Stakeholders 4, 10, 11). From a policy perspective, interviewees argued that an effective reporting ecosystem must lower the social and relational cost of seeking help.

Promising initiatives were described as emerging to address this deficit of trust. Some platforms were described as moving beyond simple reporting tools to implement comprehensive ***"reputation systems"*** that made the consequences of behaviour visible (Stakeholder 5). One example cited was a tiered ***"Honour System"*** that restricts communication privileges for toxic accounts and rewards prosocial conduct in a popular game (Stakeholder 5). Alongside these reputation mechanics, interviewees stressed the need for broader technical measures, such as anonymous reporting, reporting by witnesses and options to seek support without automatically escalating to caregivers or law enforcement. It also requires child-centred messaging that frames reporting as an act of strength and care for others rather than as ***"snitching"*** as well as responses that do not punish victims by default, such as through bans or confiscation that entrench silence (Stakeholders 3, 11). One interviewee also highlighted the value of nudging and proactive outreach when users show signs of distress rather than waiting for a complaint (Stakeholder

3). Without these shifts, reporting tools risk functioning as compliance features rather than credible pathways to safety.

Theme 4.4. The parenting gap: from “control” to “participation”

Across interviews, caregivers were described as one part of a wider protection network that includes siblings, peers, gaming communities, schools and platforms. However, interviewees described a profound gap between the policy expectation that caregivers will manage safety and the reality of how families navigate trust, identity and risk. Several practitioners argued that parenting guidance often treats adults as gatekeepers who restrict, monitor or switch off, rather than as trusted supporters of children whose social lives are often tied to games (Stakeholders 1, 5, 7, 11).

4.4.1 The "time-use" blind spot: misunderstanding developmental needs

Interviewees repeatedly highlighted that many caregivers (and policymakers) treat gaming mainly as a time-use problem, expressed through worries about ***“too much gaming or addiction”*** (Stakeholders 4, 6, 8, 10). This lens misses what several interviewees described as the developmental core of gaming in adolescence. As one interviewee put it: ***“Most parents only worry about the time children spend online, so they do not know the content, contact and conduct”*** (Stakeholder 7). Participants stressed that in South Asia, gaming is often framed as ***“wasting time”*** rather than legitimate play, which leaves caregivers and children talking past one another and turns games into a flashpoint for arguments rather than a shared understanding (Stakeholders 7, 11). When adults intervene mainly to limit or ban without recognising what gaming gives young people, children experience this as rejection rather than a conversation about safety.

4.4.2 The trust erosion loop: why bans backfire

Many caregivers struggle to interpret the game environments they see on their children's screens, leading them to misread complex social play as addiction or danger. Interviewees described this interpretative gap as the catalyst for a ***“trust erosion loop”*** (Stakeholder 7).

The result is a structural inversion where the person responsible for safety (the parent) lacks the access needed to support the child at risk. One practitioner warned that in households where the device is framed as a privilege that can be withdrawn for misbehaviour, a child is significantly less likely to report grooming, financial sexual extortion of children or harassment (Stakeholder 1). Children, in turn, experience caregivers as ***“clueless”*** or hostile to their hobby (Stakeholder 7), severing the line of communication needed for safety.

As one interviewee put it, ***“Part of the reason is parents freak out... that is why children will not disclose”*** (Stakeholder 12). Another interviewee added that this can be reinforced by ***“blanket blind trust on gaming platforms, not questioning and not reporting, and a lack of awareness of signs and symptoms of red flags”*** (Stakeholder 7). In practice, the main casualty of low gaming literacy is relational trust, not only technical safety outcomes. Stakeholders stressed that this is not simply parental failure. It is a predictable outcome when messaging tells adults to take control without helping them understand gaming, and when platforms make blocking and limiting easier than tools that support conversation and help-seeking.

Unable to navigate the nuance of the digital space, caregivers often default to blunt, reactive measures such as total bans, device confiscation or severing internet access. Interviewees described this loop in three stages:

1. **Reaction:** Caregivers hear alarming stories about grooming or addiction and react with strict limits. ***"Parents create confrontation instead of education ... children don't talk about their interactions [because parents] only worry about the screen time"*** (Stakeholder 7).

2. **Silence:** Children hide problems because they fear losing access to their gaming lives. As one interviewee noted, ***"the adult in real life's first reaction is to take away the device",*** leading children to believe that ***"parents will take the games ... away"*** rather than help (Stakeholder 10).

3. **Reinforcement:** This silence leads caregivers to assume the worst, reinforcing punitive measures and further eroding trust.

4.4.3 From gatekeeper to guide: reframing parental control

Several interviewees argued that the dominant parental control framing can be part of the problem (Stakeholders 6, 7, 13). It casts caregivers mainly as people who limit access, rather than as guides who help children make sense of risks. A stakeholder recommended a shift in language from control to participation: ***"Change parental control to parent participation, rephrase this as guidance or participation to help children make safe choices"*** (Stakeholder 11).

Interviewees suggested that strict control can work for younger children, but it often fails as the primary approach for adolescents seeking independence, for whom gaming is a key social space. For older age groups, a guidance model is often more realistic and more protective. Other interviewees agreed that strict parental control is appropriate for very young children, but that guidance and participation work better for adolescents (Stakeholders 3, 6). They suggested that adults show curiosity about what children play, ask open questions about their interactions and discuss how to block or report other players without using these tools as punishment.

Participants noted that this kind of participatory mediation often relates to better disclosure, earlier detection of problems and higher child confidence in seeking help (Stakeholder 13). For instance, when children were asked what they want from digital parenting, many said they ***"want their parents to play more games with them"*** (Stakeholder 8). This desire for shared experience supports the idea that children want to narrate their experiences to trusted adults, provided the adult enters the space with interest rather than judgement. Interviewees acknowledged that parental participation and mediation are resource-intensive. It requires time, emotional bandwidth and at least minimal gaming literacy. It cannot simply be mandated in generic parenting tips. As several participants noted, expecting all caregivers to become co-players ignores the economic and time constraints described earlier (Stakeholders 1, 7, 8, 11). For this reason, they argued that participation should be treated as a supported goal enabled by community education and platform support, rather than a baseline that all caregivers can meet.

4.4.4 Structural impossibility: why “supervise more” is not an answer

A large number of caregivers, particularly many in low- and middle-income country contexts, are already stretched thin by economic insecurity, long working hours and limited access to digital skills. Treating parental supervision as the main line of defence can deepen inequities because it assumes time, skills and stability that many families do not have.

Interviewees identified three intersecting constraints:

1. **Literacy and language.** Some caregivers struggle to read interface text or interpret warnings, spending prompts or in-game chat, even in their first language.

2. **Cultural taboos.** In many settings, discussing sexuality, financial sexual extortion of children or certain kinds of harassment is difficult or unacceptable across generations. Children may refrain from raising concerns to avoid embarrassment or perceived shame for the family.

3. **Economic dependency.** In some contexts, children earn money through gaming-related activity. Where this income supports the household, caregivers have less freedom to restrict play.

Survey respondents highlighted parent and caregiver knowledge gaps, including limited awareness of what children play, low familiarity with in-platform tools and intergenerational digital literacy divides, reinforcing that responsibility cannot sit primarily with families in the absence of systemic support (Survey, Annex 3). One interviewee reinforced this by highlighting the disconnect between design assumptions and the real conditions in marginalised communities (Stakeholder 11).

4.4.5 The hazard light analogy: cognitive overload in practice

Stakeholders also highlighted that even highly motivated caregivers who want to use safety tools are often defeated by fragmentation and poor interoperability. A comparison used by interviewees was the hazard light symbol in cars: drivers can switch between vehicles and still immediately recognise the triangle icon that signals emergency flashing lights (Stakeholder 13). Just as a hazard light is universal across cars, online safety needs shared iconography. Caregivers cannot be expected to relearn safety protocols for every new app. In online services, there is no equivalent shared language for safety. An interviewee suggested this solution, ***"If children are playing multiple games ... the idea of having a universal reporting system with icons that are similar so that parents don't have to relearn"*** (Stakeholder 13).

For caregiving practice, this fragmentation has three key effects:

1. **Increases mental load.** Caregivers supporting a child who plays on a console, uses mobile games and chats through separate apps may need to manage several disconnected safety dashboards, each with its own account system and terminology (Stakeholders 1, 13).

2. **Reduces caregiver confidence.** For caregivers already facing a steep gaming literacy gap, every new interface reinforces the sense that they are out of their depth. This can make conversations about games feel risky or embarrassing for adults, which in turn reduces the likelihood that they will initiate them (Stakeholder 11).

3. **Increases the likelihood that caregivers disengage.** Practitioners described caregivers who genuinely want to set boundaries but become overwhelmed by inconsistent interfaces and give up. Safety becomes something they hope platforms will take care of on their behalf, rather than something they feel equipped to manage (Stakeholder 8).

In practice, caregivers are asked to manage multiple services using tools that are inconsistent and time-consuming. Stakeholders argued that a basic step would be more consistent design standards, including common icons for reporting and blocking, clearer language and fewer steps to reach help across games and devices.

4.4.6 Caregivers within a systemic model: partners, not scapegoats

"There is no one-size-fits-all for every family; they do not necessarily want regulators or companies telling them exactly how things should be done" (Stakeholder 13). Interviewees therefore argued that support for caregivers should be flexible and optional, not prescriptive scripts (Stakeholders 1, 7, 11). This might include short explainers about specific games, simple prompts for conversations or culturally adapted resources co-designed with local communities. Closing the parenting gap is less about insisting adults take control and more about redesigning systems so participation is possible, trust is not punished and caregivers are neither idealised nor blamed for failures that occur far upstream of the household (Stakeholder 11).

A child-rights approach emerging from the interviews sees caregivers as partners within a broader safety ecosystem, rather than assigning them the primary responsibility for managing risks on commercial platforms. In this model:

- **Platforms** are responsible for safety by design, meaningful defaults, usable tools and responding quickly when harm occurs.
- **Regulators** are responsible for setting enforceable standards, requiring transparency and addressing barriers in access and education.
- **Schools and community organisations** can provide shared learning and spaces where gaming is discussed without stigma.
- **Caregivers** are responsible for building trust, understanding what gaming means to the child and supporting disclosure.

Theme 4.5. Policy mismatches: bans, displacement and green zones

One interviewee described a persistent mismatch between how governments, regulators, and, at times, platforms respond to public pressure—often driven by moral panic—and how children actually move through gaming ecosystems (Stakeholder 1). They argued that reactive bans and blunt restrictions can displace play into less visible spaces, including offshore or pirate servers, private networks or encrypted channels, where children may remain active but are harder for moderation systems and support services to reach. An alternative response could focus on keeping children within moderated parts of mainstream services and using graduated, in-product steps to reduce escalation, rather than ejecting them altogether. Interviewees emphasised that these approaches do not claim any space is safe, but aim to keep children within reach of moderation, reporting tools and escalation routes. Across both, interviewees cautioned that moral panic, poorly drafted rules and uncritical enthusiasm for technologies such as AI can reproduce harms if they are not grounded in children's lived realities.

4.5.1 Blunt bans and displacement into less governed spaces

Several interviewees expressed concern that blanket bans on games or devices can be introduced in response to public outcry, without a clear evidence base or assessment of likely effects. One interviewee described a case where a highly popular game was banned after intense media and parental pressure: ***"Parents demanded they banned [a specific game] ... decision-makers did not investigate. It was a blanket ban, a reactive approach"*** (Stakeholder 7). Another interviewee argued that such measures may not remove demand but instead change where and how children play (Stakeholder 5). The interviewee suggested the game may reappear under a different name, through a different platform or in a stripped-down copy of the game shared illegally. Interviewees raised similar concerns beyond formal game bans. One interviewee observed that when schools implement strict mobile phone bans, children often ***"get them in there somehow"*** and show each other harmful content anyway (Stakeholder 13). As another interviewee put it, ***"Blanket regulations, kids will find another way to access these games"*** (Stakeholder 10).

Reinforcing the concerns raised in Section 4.2.5 regarding age assurance, interviewees suggested that displacement has two practical consequences. First, children may lose access to local-language moderation, usable reporting tools and clearer pathways to help available on mainstream platforms (Stakeholders 3, 9). Second, play may shift into environments with weaker governance, fewer duty-of-care expectations and less reliable escalation routes (Stakeholder 1). Several interviewees linked this to reduced data visibility. When children move to unlicensed servers

or encrypted, offshore ecosystems, their activity may be less likely to appear in the datasets used to train safety tools and guide investment, potentially reducing the ability of systems to detect risk and trigger support.

4.5.2 Alternatives to reactive bans

As an alternative to blunt bans, one interviewee described approaches that keep children within moderated parts of mainstream services and use in-product interventions to address risky behaviour, such as restricting specific features when there are signs of harm or escalation (Stakeholder 3). This was framed as keeping users in the ***"green zone"***, where moderation, reporting and escalation routes still apply, so children remain visible to systems and communities that can intervene (Stakeholder 3).

> ***"There's loads of those kinds of intervention points where you can prevent them overstepping it and becoming that bad actor. So how do we intervene? We keep them in the green zone is how we talk about it … recognising those flags, those changes in behaviour and making those interventions in that moment … Sometimes [this] can be way more [effective] than us banning somebody for a day because that in itself can become a bit of a badge of honour"*** (Stakeholder 3).

Interviewees stressed that community norms can sometimes be more effective than sanctions alone. They argued these interventions ***"help to build that positive community"***. They described situations where the community calls out bullying behaviour directly: ***"Somebody will start being rude or a bit of a bullying experience, and the community will call them out. That's the best way of tackling things is when they're like, hey, we do not do that round here, you know, that is not cool"*** (Stakeholder 3). In this view, the goal is not only to adjust individual behaviour, but also to support communities that set and enforce their own positive norms. Another interviewee underscored the importance of simple design tweaks that make it easier for children to act when something feels wrong. For example, prompting users who block someone to consider reporting them in the same flow: ***"When people block others, do you want to report them too? People thought this was helpful for getting them to report"*** (Stakeholder 2). Participants saw this kind of prompt as both a confidence builder and a way to surface more information about harmful actors.
Other actors extend this approach into community-level practice. As one practitioner noted, the ***"digital safeguarders"*** initiative, for example, trains trusted moderators and community managers on popular online gaming servers to act as a first line of support for children who disclose mental health concerns, grooming or extremist content (Stakeholder 8). This lowers the threshold for seeking help by locating support within the spaces where harm occurs, rather than requiring children to step outside their communities.

Interviewees also noted limits to what in-platform interventions can and should do. One interviewee stressed that serious sexual harms, repeat predatory behaviour and criminal offences require firm consequences and, where appropriate, referral to law enforcement (Stakeholder 11). The policy challenge is to distinguish between these situations and avoid systems where minor misbehaviour triggers major sanctions.

Interviewees provided examples:

- Targeted restrictions, such as muting voice chat with non-friends or limiting access to higher-risk modes while still allowing lower-risk play, so children stay within spaces where safety tools still apply (Stakeholder 2).
- Friction and warnings when behaviour escalates, such as prompts that remind players of community standards or flag language that has crossed a line (Stakeholders 2, 3).
- Short time-outs that temporarily restrict features such as chat or matchmaking, rather than suspending accounts entirely (Stakeholders 2, 3, 13).

4.5.3 Regulatory moral panic and informed, co-designed rules

Practitioners distinguished sharply between evidence-based regulation and regulation that emerges from moral panic dynamics (Stakeholders 1, 4). The former is seen as an essential lever for child protection; the latter is seen as a recurrent source of policy misalignment.

Practitioners described moral panic regulation as having several recurring characteristics (Stakeholders 1, 4):

- Regulators move quickly under political and media pressure, with limited time to understand monetisation flows, cross-platform dynamics or how particular mechanics work in games, often mirroring wider gaps in adult gaming literacy.
- Definitions are often vague or conceptually weak. For example, early attempts to regulate loot boxes are said to have relied on narrow or easily circumvented definitions of chance-based rewards.
- Rules are drafted without adequate consultation with child-rights advocates, gaming industry representatives or young people themselves, which leads to misalignment between legal categories and technical realities.

Stakeholders argued that reactive rules often lean on what is easiest to count or symbolically ban, rather than addressing the industrialised harms that are hardest to measure. As summarised: ***"Regulators are not well informed and are moved by moral panic"*** (Stakeholder 1). In these interviewees' view, such rules invite evasion. Companies can comply with the letter while preserving the spirit of monetisation through minor tweaks, such as re-packaging

loot boxes[20] as battle passes[21] or using pseudo-random (gambling-like) mechanics that sit just outside a narrow legal definition.

By contrast, informed, co-designed regulation is described as:

- Grounded in a realistic understanding of game design, business models and the ways that risk travels across platforms (Stakeholder 5).
- Developed through structured engagement between regulators, industry and child-rights actors, with space for technical feedback and iteration (Stakeholders 2, 10, 11).
- Centred on children's rights, including the right to play, freedom of expression and protection from exploitation, rather than on abstract notions of decency or national reputation (Stakeholders 4, 6).

4.5.4 AI honeypots and proactive detection: policy, safeguards and deterrence

The use of AI was described as a new frontier where policy is playing catch-up. While stakeholders raised concerns about deepfakes and automated grooming, some argued that defensive AI is now necessary for detection at scale. A stakeholder noted, ***"We can't talk about safety without AI – it works both ways, the good and the bad"*** (Stakeholder 3). One stakeholder described projects where AI-generated or semi-automated avatars of children are deployed in tightly governed environments to detect offenders who initiate sexual contact (Stakeholder 10). No real child is involved. The approach is intended to detect, document and refer credible criminal approaches to appropriate authorities. Interviewees noted that such approaches are debated and depend heavily on legal basis and safeguards. ***"We can use AI to ensure a real child is never in that room, it is a honeypot, not a victim"*** (Stakeholder 10).

In addition to honeypots,[22] AI models can be used to educate and deter. A practitioner described prototypes where conversational agents teach children about grooming tactics, boundary-setting and help-seeking, embedded within platforms or educational campaigns (Stakeholder 10). They further suggested that visible AI presence can itself have a deterrent effect, signalling that the environment is monitored and that deceptive approaches may be detected: ***"We can use an AI model to really educate children, a deterrent to predators"*** (Stakeholder 10). However, interviewees described this technical capability as sitting within a legal vacuum.

[20] Virtual containers that provide a randomised selection of in-game items (e.g., skins or weapons), often compared by stakeholders to gambling due to the element of chance.
[21] A monetisation system that rewards players for consistent play over a fixed period. Companies use this to drive retention or repackage revenue models.
[22] A controlled decoy environment or avatar designed to attract offenders and gather evidence without putting real children at risk.

NGOs and experts warned that without clear frameworks, these tools risk entrapment or mission creep:

- **Entrapment and due process**. There are concerns that poorly governed decoy operations could blur lines between investigation and entrapment (Stakeholder 10).
- **Cross-border legality.** AI-driven detection efforts often operate across jurisdictions with different legislation, data protection regimes and evidentiary thresholds, which complicates prosecution and heightens the risk of rights violations (Stakeholder 10).
- **Mission creep and surveillance.** Stakeholders worry that tools developed for child protection could be repurposed for broader surveillance of children's behaviour, chilling legitimate play and expression or disproportionately targeting certain groups (Stakeholder 13).

Enthusiasm for AI solutions can distract from more mundane but essential investments in human moderation, cross-border law enforcement cooperation and support services for victims (Stakeholder 7). The interviewee cautions against treating AI as a standalone solution for harms that are rooted in inequalities, weak institutions and under-resourced justice systems. Supporters of proactive detection responded that the industrialised nature of abuse networks leaves few alternatives (Stakeholders 2, 3, 10). Manual detection alone was described as struggling to keep pace with offenders who use automation, multiple identities and cross-platform tactics. From this perspective, refusing to explore AI-driven defences risks leaving offenders with a capability advantage over those already using similar tools offensively.

In line with the broader argument of this report, interviewees stressed that AI should be seen as one tool within a multi-layered safety ecosystem, not as a substitute for structural reform. If used carefully, it may help reduce the need for the most vulnerable children to put themselves in risky situations in order for harms to be detected. Used carelessly, it risks expanding surveillance, reinforcing existing biases and deepening mistrust in institutions that are already struggling to earn children's confidence.

5. Recommendations

This section presents recommendations arising from the analysis. The proposals are grounded in a child-rights framework consistent with the United Nations Convention on the Rights of the Child (UNCRC), emphasising children as rights-holders and legitimate actors in shaping the conditions of their online play. Recommendations address different stakeholder groups — industry, governments, private sector, CSOs and researchers — and conclude with cross-cutting priorities that address the structural conditions identified throughout the analysis.

5.1 Industry (game developers, platforms, publishers) and private sector companies

Industrially scaled threats require a systemic and rights-based response that places children's interests at the centre of design. Industry actors should treat children not merely as users or customers but as persons with the right to protection (Arts. 19, 34), to participate in decisions affecting them (Art. 12), to non-discrimination (Art. 2), to play safely (Art. 31) and to seek remedy (Art. 39).

Industry should participate in, and contribute data to, cross-platform infrastructures for detecting and disrupting high-risk harms. Where serious threats operate across services, companies should no longer treat safety-relevant data as proprietary competitive advantage. Instead, high-risk signals should be handled as a public-health-like category of information subject to well-governed data exchange practices. This signal-sharing is essential to understanding how risk becomes harm in the gaming ecosystem and what can be done to prevent further harm in the specific instances where offenders shift between platforms to evade controls. Ideally, coalitions can support the smaller platforms with the legal counsel required to engage in such signal-sharing activities.

Industry should recognise that business models directly constrain achievable safety and must therefore be reformed where they conflict with children's rights. Monetisation systems that rely on compulsion, the exploitation of developmental processes or financial vulnerability place hard limits on safety outcomes. Rather than attempting to overlay interface-level protections, companies should evaluate and redesign revenue systems that drive aggressive upselling, peer pressure and exploitative reward structures. Safety measures in gameplay should take a privacy-preserving approach that allows early detection and prompt action, acknowledging that these are good business decisions that also drive revenue. The gap that exists between anti-cheat strategies and child safety strategies should also be addressed, possibly by reframing child safety to make it the most viable option.

Industry should establish and fund permanent child and youth advisory mechanisms with the space, voice, audience and influence to inform decisions about the development of games and safety features. These bodies should function as reference points or youth councils that can review safety policies, provide early warnings about risky features and explicitly influence design decisions. The WeProtect Global Alliance Youth Participation Guidelines[23] can be applied to ensure that participation is ethical, safe, empowering and impactful rather than tokenistic. This model emphasises learning and seeking feedback from children as a continuous process, with participation types that range from consultations to leadership. Such participation requires careful planning, safeguarding, flexibility and respect for participants' autonomy. Moving beyond symbolic consultation towards shared governance would address the recurrent blind spots that emerge when gaming platforms rely on imagined users rather than the children who navigate their systems. As the 2025 Global Threat Assessment states, "Partnerships with child-led and child-focused organisations can promote safe participation, detect early risks and harms and inform effective, child-centred intervention" (WeProtect Global Alliance, 2025).

Industry should prioritise visibility and remedy by designing systems that keep children within governed environments. Companies should emphasise graduated responses, accessible reporting pathways and processes that allow children to track outcomes when they seek help. Gaming platforms should have explicit age and content rating information to help children and caregivers understand the type of game and platform in an accessible way. They should also create clearer safe harbours for children with in-built safety in games so that children can be reached within the game. Reporting should be reframed as positive gameplay or an act of care for others, actively encouraged

[23] WeProtect-Participation-Guidance-Upload.pdf

and, as much as possible, have no social or relational costs. This includes anonymous reporting, reporting by witnesses and support-seeking options that are not automatically escalated to caregivers or law enforcement.

Industry should make platform data available for research and online safety practices. Data can be useful to understand both the most common risks and the most dangerous ones. Anonymised and disaggregated data can help the platform understand where the gameplay and associated functionalities can be exploited and where gaming experiences can be improved to support positive and safe experiences. Understanding the diversity of gamers and their experiences would also be useful to ensure that all players have access to sufficient support.

Industry and private sector companies should come together to establish universal safety design features to make parental mediation easier. Such features could facilitate parents' and caregivers' experiences of monitoring and mediating their children's gaming experiences. Having common symbols for features, such as managing in-app purchases, setting time limits, granting access to specific content and monitoring accounts, would require caregivers to learn what the symbols mean once and apply that knowledge across the different platforms their children use for gaming. Several icons (such as the burger menu icon) are now recognised widely for their functions and are evidence of how such iconography can be used for digital safety.

5.2 Governments

States have a duty to ensure that children's rights can be realised in the digital environment and that commercial incentives do not override these rights. As the UN General Comment No. 25 details, child rights also apply in the digital environment (UN Committee on the Rights of the Child, 2021). Legislation and policy should therefore create enabling conditions for children's protection.

Governments should embed children's participation into regulatory design and oversight. Regulators should consult directly with diverse groups of young players when developing safety codes or approving high-risk features. This may include statutory youth advisory structures with a clear mandate to influence regulatory outcomes. Such participation, planned according to the WeProtect Global Alliance Youth Participation Guidelines, could allow better distinctions between normal play-related challenges and unacceptable harms and ensure that actions or sanctions are of an appropriate measure. It can also provide deeper insight into how gaming aesthetics are used for purposes such as propaganda, as well as ways to strengthen educational efforts, as suggested by the young people themselves.

Governments should establish legal frameworks that enable cross-platform safety infrastructures and data sharing. Treating safety-relevant intelligence as a form of public health data would permit legally governed sharing between platforms, supported by appropriate safe harbours and privacy protections. Without this, networked predation will continue to outpace protection mechanisms trapped within single-company silos. Governments can also consider incentives for those companies that engage in signal-sharing and data-sharing to acknowledge their contribution to child safety.

Regulation should address the structural incentives created by business models that exploit developmental vulnerabilities. Policymakers should recognise that commercial design and monetisation practices set a ceiling on safety and must therefore be subject to direct scrutiny, including the discouragement of designs that rely on compulsion, comparison and financially motivated exploitation of minors. This requires infrastructural thinking and the development of informed regulation that requires age-appropriate game design so that games fit around a healthy childhood rather than create conditions where children are overspending, becoming dependent or being shamed. Legislation should limit or ban manipulative monetisation for children, make spending transparent and controllable and request regular audits to establish independent oversight. Additionally, legislation could make safe design the economically rational choice by establishing penalties when companies profit from harmful practices and rewarding safe design models.

Legislation should mandate proportionate, privacy-preserving age assurance to tailor protections to children's developmental needs, while ensuring alignment across platforms and devices. Regulators should require services to implement age assurance systems that distinguish between adult and child users, and between different stages of childhood, so that safeguards, commercial pressures and interaction settings can be adjusted accordingly. These systems must be designed to minimise data collection, prevent them from being repurposed beyond child safety and

avoid excluding children. Alignment across platforms and devices is essential so that protections travel with the child rather than being fragmented by service boundaries. Consistent standards would also prevent a race to the bottom, reduce the burden on families and limit opportunities for offenders to exploit weaker environments. Age assurance should therefore be treated as enabling infrastructure for children's rights, supporting participation, privacy and protection, rather than as a gating mechanism that pushes vulnerable children into less visible spaces.

5.3 Civil society organisations

CSOs are central to enabling children to exercise their rights but cannot be expected to compensate indefinitely for structural deficits elsewhere. However, their role should still be recognised as essential rather than supplementary.

CSOs should elevate children's lived experiences and ensure these perspectives reach industry and policymakers. Programmes should incorporate children's accounts as evidence of harms, barriers to remedy and gaps in existing safety infrastructures, particularly for marginalised groups who remain largely invisible in official datasets. These accounts should be understood not only in relation to what happens inside gaming environments, but also in the context of children's offline lives, including poverty, family stress, time pressures, limited access to safe play spaces, language barriers, disability, migration status, experiences of discrimination and broader social exclusion. These offline constraints often shape when, how and why children engage with games, how vulnerable they may be to harm and the extent to which they can access support. Recognising these intersecting realities ensures that safety policies, reporting systems and remedies are grounded in the everyday circumstances of the children most at risk, rather than those with the greatest visibility or resources.

CSOs should reframe education and outreach away from discussions primarily on screen time and toward structural risks and help-seeking. Focusing on what children are doing, with whom and under which governance conditions offers a more accurate reflection of risk than simple time-based approaches, especially in contexts where gaming is tied to income or education.

CSOs should develop strategies to reach children outside formal access points. Attention must focus on children in cybercafés, on shared devices and in situations where bans have pushed them into privacy-shielded environments. Programmes should account for low literacy, limited connectivity and the absence of parental mediation.

CSOs should also recognise that some children will circumvent safeguards, and ensure that education, support and accessible routes to help remain central. Even the strongest age assurance and safety controls will not completely prevent children from seeking out restricted content or bypassing protections. For some children, this behaviour reflects normal developmental curiosity, experimentation or peer influence; for others, it is shaped by external pressures and the offline environments in which they live. Treating circumvention only as rule-breaking risks obscuring these underlying drivers. A child-rights approach assumes that systems will sometimes fail and that children will sometimes take risks and therefore prioritises equipping them with realistic knowledge about harms, self-protection strategies and trusted avenues for help when environments are poorly governed. This education should be embedded across schools, youth programmes and community settings to ensure that children who are gaming in cybercafés, on shared devices or in informal spaces can still access support without fear of punishment or exclusion. This approach reframes education not as a substitute for safety-by-design, but as a necessary backstop for the moments when safeguards are absent, bypassed or insufficient.

CSOs should provide education and support to parents and caregivers to help them understand online gaming and equip them to support their children in these spaces. While the burden of protecting their children online should not be primarily placed on parents and caregivers, positive relationships and support are important protective factors that should be promoted.

CSOs should contribute systematically to shared safety knowledge. By sharing anonymised pattern-level insights from helplines and frontline services, CSOs can support cross-platform safety learning and help identify emerging harms while preserving confidentiality.

5.4 Researchers

Researchers have a unique role in making hidden dynamics visible and in finding evidence that allows regulators and industry to evaluate whether progress is real and meaningful from a child-rights perspective.

Researchers should prioritise the visibility of groups who are currently missing from official datasets. Research designs should include children using shared devices, modified or pirated versions of games that bypass official protections or informal access points, as well as those in low-resource settings, recognising that these children face the highest risks, yet remain least visible. Researchers should identify innovative and inclusive research designs to ensure that, as much as possible, research represents the contexts and diversity of children playing games.

Researchers should co-develop studies and indicators with children and young people. Children's perspectives should inform the questions being asked, the methods used and the interpretation of results. Participatory approaches are necessary to ensure that research captures harms, barriers to remedy and the everyday realities of diverse young players.

Researchers should examine how business models and monetisation strategies shape safety outcomes. Independent research should analyse the effects of compulsion-based design, aggressive upselling and variable reward structures on children's rights, highlighting where commercial incentives directly undermine protection.

Researchers should develop child-centred outcome measures that can inform regulatory targets and platform accountability. Without indicators such as the disruption of grooming or access to timely help, progress will remain difficult to assess, and transparency will continue to prioritise volume metrics over meaningful outcomes.

Researchers should advocate for data from gaming platforms to be made available for research. Without compromising privacy and anonymity, such data can provide useful information about the pathways to risk and what game mechanics are leveraged to expose children to harm, what functions as protective factors and where platforms can strengthen safety mechanisms.

5.5 Cross-cutting recommendations

Certain conditions must change at a structural level if games are to become environments that uphold children's rights. These include recognising and responding to industrialised threats, enabling prevention, correcting business models that conflict with children's rights and ensuring that young players can meaningfully influence design and governance.

All actors should design for real children and not imagined ones. Policies, products and protections must reflect diverse lived realities rather than assumptions of individual devices, high connectivity or unlimited parental mediation. To be effective, safety regimes must centre those children currently rendered invisible.

Child participation must become institutional rather than symbolic. Children's right to be heard requires governance structures that allow them to influence safety decisions at every level, including industry, regulation and research, ensuring that their expertise and lived experience guide the design of gaming environments.

Safety must be approached as a networked challenge rather than an individual or household responsibility. The threats children face in online games are already industrialised and networked, moving fluidly across platforms, devices and jurisdictions. By contrast, most protections remain fragmented, dependent on the policies of individual companies, the capacity of specific regulators or the unpaid labour of families and NGOs. A child-rights approach suggests that safety should not depend on where a child happens to log in, how digitally literate their caregivers are or whether they have access to well-resourced services. Instead, protection should be built as layered infrastructure: strong platform-level safeguards, cross-platform signal-sharing and early-warning systems, coordinated regulatory oversight and well-funded independent support services that children can reach when systems fail. Mandatory, privacy-preserving information-sharing and interoperable safety tools are essential so that offenders cannot simply migrate to weaker environments, and children remain visible to those who can help them across the whole

ecosystem. In this model, caregivers are still partners in children's safety, but they are no longer expected to carry the burden alone or to compensate for gaps created by commercial and regulatory design choices.

Finally, transparency must evolve into verifiable accountability. Claims of progress should be subject to independent scrutiny, audits, adversarial testing and child-centred outcome indicators. Only under such conditions can online games function as spaces of play, social connection and learning consistent with children's rights, without imposing unacceptable risks as the cost of participation.

6. Conclusion

The evidence presented throughout this report illustrates that online gaming environments have evolved into complex social ecosystems where children play, learn, build relationships and increasingly encounter networked forms of exploitation and harm. Industrialised predation operates across platforms and jurisdictions, while the systems intended to safeguard children remain fragmented, voluntary and unequal. The result is an ecosystem in which the most vulnerable children are often the least visible, least protected and least able to exercise their rights to participation, protection and remedy.

Addressing these challenges requires more than incremental change or additional guidance for parents. It demands structural reform grounded in children's rights: business models that no longer rely on exploiting developmental processes, regulatory frameworks that enable prevention rather than commercial enclosure and participatory mechanisms that treat young players as governance actors rather than occasional consultees. Crucially, safety must be understood as a shared responsibility that cannot be outsourced to households, NGOs or those children already burdened by inequality.

If games are to fulfil their enormous potential as spaces of creativity, social connection and joy, they must be designed for the real children who inhabit them, not the imagined ones for whom many current systems are implicitly built. Realising this vision will require sustained commitment, new legal and commercial approaches and the meaningful participation of young people in decision-making. Only under these conditions can online gaming environments truly reflect a child-rights approach focused on dignity, protection and participation.

7. References

1. ACAMS Today. (2023). *Gaming, loot boxes, and laundering risks.* Association of Certified Anti-Money Laundering Specialists.

2. Age Verification Providers Association (2024) *Standards for Age Verification*. Age Verification Providers Association.

3. American Psychiatric Association (2023) 'Internet Gaming', *Psychiatry.org (Patients and Families).* Physician review: James Sherer, M.D., January 2023.

4. Anti-Defamation League (ADL) (2021) *Hate is no game: Harassment and positive social experiences in online games*. New York: Anti-Defamation League. Available at: *https://www.adl.org/resources/report/hate-no-game-harassment-and-positive-social-experiences-online-games* (Accessed: 9 December 2021).

5. Anti-Defamation League (ADL) (2023) *Hate is no game: Harassment and extremism in online games*. New York: Anti-Defamation League.

6. BBC News (2025) 'Australia has banned social media for kids under 16. How will it work?', *BBC News*, 10 December.

7. Copson, N. and Johnson, S.D. (2025) 'A scoping review of metaverse-facilitated child sexual abuse', *Social Sciences & Humanities Open*, 12, p.101667.

8. Council of Europe (2020) *Nepal: country profile*. Octopus Cybercrime Community.

9. Data Protection Commission (2021) C*hildren front and centre: fundamentals for a child-oriented approach to data processing*. Dublin: Data Protection Commission.

10. eSafety Commissioner (2024) *Levelling up to stay safe: Young people's experiences navigating the joys and risks of online gaming*. Canberra: Australian Government.

11. Fry, D., Krzeczkowska, A., Ren, J., Lu, M., Fang, X., Anderson, N., Jin, W., Liu, W., Vermeulen, I., McFeeters, A. and Harker-Roa, A. (2025) 'Prevalence estimates and nature of online child sexual exploitation and abuse: a systematic review and meta-analysis', *The Lancet Child & Adolescent Health*, 9(3), pp. 184–193.

12. Gottfried, J. and Sidoti, O. (2024) *Teens and video games today.* Washington, D.C.: Pew Research Center.

13. Hindenburg Research (2024) *Roblox: inflated key metrics for Wall Street and a pedophile hellscape for kids.* Hindenburg Research.

14. Hyde, R. and Cartwright, P. (2023) 'Exploring Consumer Detriment in Immersive Gaming Technologies', *Journal of Consumer Policy*, 46, pp. 335-361.

15. IEEE (2024) *IEEE Standard for Online Age Verification (IEEE Std 2089.1-2024)*. New York: The Institute of Electrical and Electronics Engineers.

16. Institute for Strategic Dialogue (2021) *Gaming and extremism series: Discord, Steam, Fortnite, Roblox.* London: Institute for Strategic Dialogue.

17. Institute for Strategic Dialogue (2024) *Regulation of gaming and online safety: gaps and opportunities*. London: Institute for Strategic Dialogue.

18. Kilmer, E. D. and Kowert, R. (2024) 'Grooming for violence: similarities between radicalisation and grooming processes in gaming spaces', *Global Network on Extremism and Technology*, 8 February.

19. Király, O., Koncz, P., Griffiths, M.D. and Demetrovics, Z. (2023) 'Gaming disorder: a summary of its characteristics and aetiology', *Comprehensive Psychiatry*, 122, art. 152376.

20. Lakhani, S. (2021) *Video gaming and (violent) extremism: an exploration of the current landscape, trends, and threats.* Luxembourg: Publications Office of the European Union.

21. Lamphere-Englund, G. (2025) *Gaming and Violent Extremism in Africa*. United Nations Interregional Crime and Justice Research Institute (UNICRI).

22. Life After Hate (2024) '*Life After Hate releases Discord channel, engaging with today's violent extremists where they digitally live*', Life After Hate, 23 May.

23. Livingstone, S. and Stoilova, M. (2021) *The 4Cs: classifying online risk to children* (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI). Available at: https://doi.org/10.21241/ssoar.71817

24. Livingstone, S., Stoilova, M. and Nandagiri, R. (2019) *Children's data and privacy online: growing up in a digital age. An evidence review*. London: London School of Economics and Political Science. Available at: https://eprints.lse.ac.uk/101283/

25. McGuirk, R. (2024) 'X Corp. has slashed 30% of trust and safety staff, an Australian online safety watchdog says', *AP News*, 10 January.

26. Marinoni, Carlo & Rizzo, Marco & Zanetti, Maria Assunta. (2024) 'Social Media, Online Gaming, and Cyberbullying during the COVID-19 Pandemic: The Mediation Effect of Time Spent Online', *Adolescents,* 4, 297-310. 10.3390/adolescents4020021.

27. Miller-Idriss, C. (2025) 'Misogyny incubators: how gaming helps channel everyday sexism into violent extremism', *Frontiers in Psychology*, 16, art. 1537477.

28. NSPCC (2024) 'Online grooming crimes against children increase by 89% in six years', *NSPCC*.

29. Ofcom (2022) *Children and parents: media use and attitudes report 2022*. London: Ofcom.

30. Radicalisation Awareness Network (RAN) (2022) *Gamification and extremism*.

31. Radicalisation Awareness Network (RAN) (2023) *Building resilience in digital youth cultures.*

32. Rees, F. (2025) 'Famous at five: risk assessing digital child labour', *Information & Communications Technology Law*.

33. Roblox Corporation (2024) *Transparency and safety interventions report*. Roblox Corporation.

34. Royal United Services Institute (RUSI) (2025) *Implementing positive gaming interventions: a toolkit for practitioners*. London: Royal United Services Institute.

35. Satapathy, P., Khatib, M.N., Balaraman, A.K., Kaur, M., Srivastava, M., Barwal, A., Prasad, G.S., Rajput, P., Syed, R., Sharma, G. and Kumar, S. (2025) 'Burden of gaming disorder among adolescents: a systematic review and meta-analysis', *Public Health in Practice*, 9, art. 100565.

36. Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K. *et al.* (2020) *EU Kids Online 2020: survey results from 19 countries*. Prague: Faculty of Social Sciences, Charles University.

37. Stoilova, M., Livingstone, S. and Nandagiri, R. (2020) 'Digital by default: children's capacity to understand and manage online data and privacy', *Media and Communication*, 8(4), pp. 197–207.

38. Tech Coalition (2024) *Lantern transparency report 2024*. San Francisco: Tech Coalition.

39. Technavio (2025) *Online gaming market forecast, 2025–2029*. Technavio.

40. Thorn (2024) *Youth perspectives on online safety, 2024*.

41. UN Committee on the Rights of the Child (2021) *General comment No. 25 (2021) on children's rights in relation to the digital environment*. Geneva: United Nations Committee on the Rights of the Child.

42. UNICEF (2017) *The state of the world's children 2017: children in a digital world*. New York: UNICEF.

43. UNICEF Australia (2025) *Social media ban explainer*. UNICEF Australia.

44. Unity (2023) *Toxicity in online multiplayer games: 2023 report*. Unity Technologies.

45. WeProtect Global Alliance (2023) *Global threat assessment 2023: fighting child sexual exploitation and abuse online*. London: WeProtect Global Alliance.

46. WeProtect Global Alliance (2025) *Global threat assessment 2025*. London: WeProtect Global Alliance.

Annex 1. Participant list

Interviewees represented the following organisations or participated in an independent capacity:

1.	**Stakeholder 1**	Academia	South America (Brazil)*
2.	**Stakeholder 2**	Trust and Safety (Industry)	Global
3.	**Stakeholder 3**	Trust and Safety (Industry)	Global
4.	**Stakeholder 4**	International Organisation	Global
5.	**Stakeholder 5**	Academia	North America (USA)*
6.	**Stakeholder 6**	International Organisation	Global
7.	**Stakeholder 7**	Civil Society Organisation	Asia (Nepal)*
8.	**Stakeholder 8**	Civil Society Organisation	Europe
9.	**Stakeholder 9**	Civil Society Organisation	Oceania
10.	**Stakeholder 10**	Civil Society Organisation	North America
11.	**Stakeholder 11**	Civil Society Organisation	Asia (India)*
12.	**Stakeholder 12**	Trust and Safety (Industry)	Global
13.	**Stakeholder 13**	Trust and Safety (Industry)	Global

*Denotes stakeholders who provided insights specific to a national context in addition to their broader regional perspective.

Annex 2. Interview guide

Introductory Questions

1. Could you briefly introduce yourself and describe your current role and how it relates to online gaming or child online safety?

Risks

2. From your perspective, what are the main risks that children face in online gaming environments?
3. Which groups of children are more vulnerable to these risks, and why is that so?
4. How are these risks evolving (e.g., new features, platforms, technologies)?

Responses and Current Interventions

5. What responses, policies or initiatives are currently in place to address these risks?
6. In your experience, which approaches have been most effective?
7. Are there examples of promising practices you have observed in your field or region?

Gaps and Challenges

8. Where do you see the biggest gaps in protection for children in online gaming?
9. What challenges hinder more effective prevention or responses? (e.g., technological, legal, resource-related, cultural).
10. What is the role children have in shaping safer gaming environments?

Opportunities

11. What changes or innovations would make the biggest difference in improving children's safety in gaming environments?
12. If you could recommend one action to policymakers/industry/frontline services, what would it be?

--

Stakeholder-specific questions: (*choose according to stakeholder type)*

For researchers in gaming, digital rights, or child safety

13. What does the existing evidence tell us about risks in online gaming, and where are the key knowledge gaps?
14. Are there methodological or ethical challenges in researching risks in gaming that need to be addressed?

For NGOs, service providers and frontline practitioners (Child Protection, Education, Healthcare)

15. In your work with (e.g. children and families), what concerns about gaming are most raised?
16. How well equipped are practitioners to respond to risks in gaming compared with other online risks?
17. Can you share your experiences of what has worked and what hasn't with respect to prevention and awareness about risks in gaming?

For industry or regulatory stakeholders

18. What measures has your organisation (or the industry more broadly) implemented to reduce risks in gaming?
19. What barriers do you face in balancing safety with commercial or user-experience priorities?
20. How do you engage with children, parents, or child safety experts when designing safety features or policies?

21. How does your organisation measure the effectiveness of its child safety initiatives? Are there particular indicators or metrics you rely on?
22. Are there specific risks (e.g., grooming, exposure to harmful content, financial exploitation, harassment) that are particularly difficult to address within gaming environments?
23. Looking ahead, what innovations or industry-wide collaborations could strengthen protections for children in gaming?

Conclusion

24. Is there anything we have not discussed that you feel is important to highlight?

Annex 3. Information letter and consent form

You are invited to take part in a research study about the risks to children in online gaming. The research project is being carried out by WeProtect Global Alliance, which has appointed Trilateral Research to support with delivery of key activities in the project, including this research. This research is funded by the Global Cybersecurity Forum (GCF), which supports WeProtect Global Alliance in advancing global understanding and collective responses to risks faced by children in online gaming environments. Your participation is voluntary, and you are free to withdraw at any time. Before you decide whether to take part, please take time to read the following information carefully. Be sure you understand why the research is being done and what it will involve. Feel free to ask questions.

The project

The purpose of this project is to generate a comprehensive understanding of the risks faced by children in online gaming and how responses to these risks can be improved. The project aims to synthesise existing knowledge, identify critical gaps in evidence and provide clear, evidence-informed recommendations to strengthen child safety in gaming environments. As part of this project, a literature review and interviews with key stakeholders are being conducted to explore challenges and opportunities in mitigating risks to children in online gaming and gain insights into the initiatives and practices being used by professionals in the field.

What will I be asked to do?

You will be asked to participate in an online interview, where you will be invited to discuss your professional role, views on risks in online gaming, current responses and initiatives, gaps in protection and insights on challenges and opportunities in safeguarding children in online environments.
The only data being collected during interviews will be your opinions. No identifying personal data (such as names or organisations) will be included in interview notes or any associated reports that we may produce when discussing the findings. You will remain anonymous and anything you share will remain strictly confidential throughout the research process.

Where will the research take place?

The interviews will be conducted online via Zoom or Microsoft Teams, at a time that is convenient for you. Each session will last approximately 60 to 90 minutes and may be facilitated by up to two interviewers. A team member from WeProtect Global Alliance may also be present in some interviews. You can choose whether or not to have your camera on during the interview, based on your comfort level.

What will we record or document?

The interview will be audio-recorded by the researchers and will be transcribed for analysis purposes. All audio recordings will be securely stored until they are transcribed, after which they will be permanently deleted. Transcripts will be pseudonymised to protect your identity. You can review any documentation upon request by contacting the researchers (contacts below).

What will we use your participation for?

Your participation will be used to provide a better understanding and therefore improve responses to risks children face in online gaming. Your insights will inform tailored recommendations for policy, industry, and frontline practice

in child online safety, and will contribute to a research report that brings together findings from the interview analysis and the literature review.

All information that could either directly or indirectly identify you will be anonymised. With your consent, anonymised quotes may be shared with WeProtect Global Alliance, referring to your professional role (e.g. professional working in child protection, policymaker, law enforcement expert, gaming industry professional, digital rights researcher, etc.).

The results may be published in reports or shared with organisations working in child protection and policy, but your identity will remain confidential, unless you choose to be identified.

Why have you been chosen?

You have been selected because of your relevant expertise in this field.

Do you have to take part?

No. Your participation is entirely voluntary. You are free to leave at any time, without giving a reason and without any consequences for you or your future participation in the project. You are free to refuse to answer any questions or provide any information. Whether you choose to participate or not will not affect your relationship with WeProtect Global Alliance or any associated organisations. You have the right to ask questions and receive understandable answers before making any decision.

What are the possible disadvantages of taking part?

We do not envisage any disadvantages or risks related to your participation; however, some topics discussed in the interview may be sensitive in nature. You may choose not to answer any question or pause/stop the interview at any time.

What are the possible benefits of taking part?

This work contributes to a better understanding of the risks children face in online gaming and the responses currently available, including identifying possible service gaps, challenges and opportunities. By sharing your knowledge and experience, you will help shape evidence-based recommendations for child online safety in gaming contexts.

Right to withdraw

You may withdraw your consent from this project at any time without giving a reason. To do so, simply contact the researchers whose details are included below. You will be asked whether you would like us to delete your data or whether you are fine with these data to continue to be processed. You may be asked why you have decided to withdraw, but you are under no obligation to give a reason.

Storing personal data

Personal data will not be retained or stored. Personal identifiers will be removed from transcripts, and pseudonyms will be used in all reports and outputs. Following transcription, audio recordings will be promptly deleted. Data will be stored for up to 2 years.

The record of your participation will be kept in a file separate from the research data. All data will be stored on password-protected shared drives managed by Trilateral Research, accessible only to authorised personnel working directly on the project and the designated client team. These shared drives are secured with robust encryption and stringent access controls and are monitored through strict data protection protocols in line with UK GDPR compliance. Trilateral's internal policies ensure all data are protected from unauthorised access, loss or misuse, consistent with its enterprise-wide commitment to privacy and information security.

Your rights

This project has been reviewed and approved by an IRB to ensure that your rights and welfare as a participant are protected.

You have the right to information regarding what is collected and processed, to access your data being processed, to delete or make any changes to this information and to restrict processing. You have the right to receive requested information in a time-limited fashion. If you are concerned or have questions about how your personal data are being processed, or if you wish to exercise any of these rights, please contact us using the contact details below.

Contact for questions, concerns or further information

If you have any questions about this research or your prospective involvement in it, please contact:

Dr Lorleen Farrugia lorleen.farrugia@a1-research.com
Srivatsan Raj srivatsan.raj@trilateralresearch.com

If you have any questions about the wider context of this research, please contact:

Dr Bethany Jennings bethany@weprotectga.org

Statement of informed consent

By signing this form, you agree to take part in the research project on risks to children in online gaming. The nature of the research, your involvement in it and your rights regarding your participation in the project are explained in the information sheet accompanying this form.

Tick 'Yes' to affirmatively consent to the following statements. Tick 'No' to dissent.

I have reviewed the participant information sheet and understand my participation in the research project.	☐ Yes	☐ No
I have had an opportunity to ask questions about the research.	☐ Yes	☐ No
I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason and without penalty.	☐ Yes	☐ No
I consent to have my participation used to better understand the risks children face in online gaming and the current responses to those risks, including identifying possible gaps, challenges, and opportunities.	☐ Yes	☐ No

I understand that the interview will be conducted online and I may choose whether or not to use my camera.	☐ Yes	☐ No
I consent to the audio recording of the interview.	☐ Yes	☐ No
I understand that a WeProtect Global Alliance team member may be present during the interview for quality and transparency purposes.	☐ Yes	☐ No
I consent to my anonymised quotes being shared with WeProtect Global Alliance, referring to the name of my professional role only.	☐ Yes	☐ No
I consent to have anonymised quotes be used in presentations and publications.	☐ Yes	☐ No
I understand that the project may retain my anonymised data for 2 years.	☐ Yes	☐ No
I understand that data will be stored securely in accordance with UK GDPR and ethical research standards.	☐ Yes	☐ No
I am willing to take part in the research.	☐ Yes	☐ No
I understand that the researcher may follow up with me if clarification or additional input is needed.	☐ Yes	☐ No
I understand that I can request access to, correction, or deletion of my personal data at any time.	☐ Yes	☐ No

Name: _____

Organisation: _____

Date: _____

Signature: _____

Annex 4. Survey charts

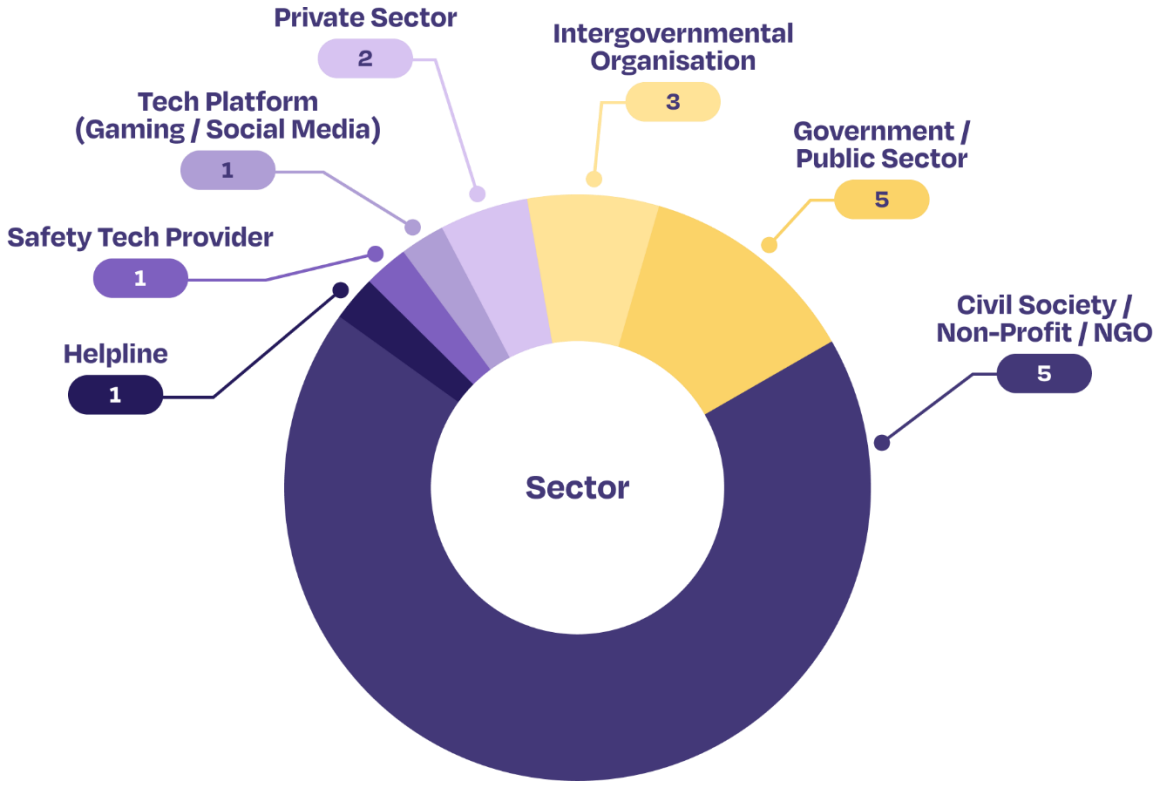


Figure 1 – Sectors to which survey participants belong

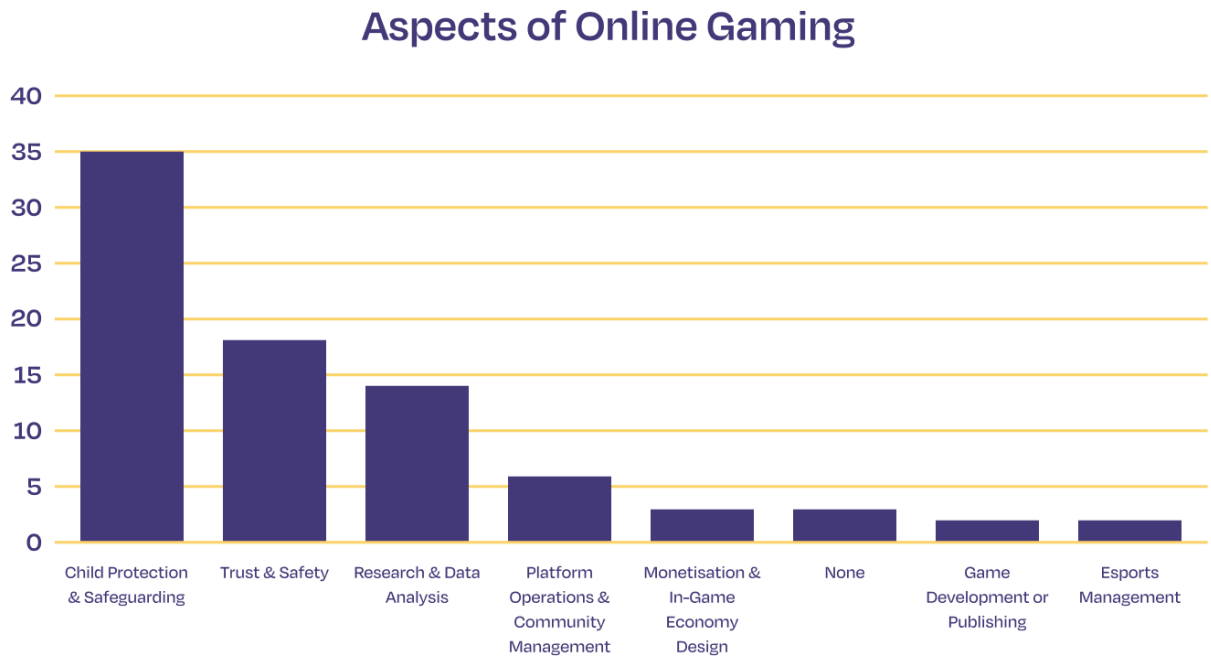


Figure 2 – Aspects of online gaming addressed by the survey participants

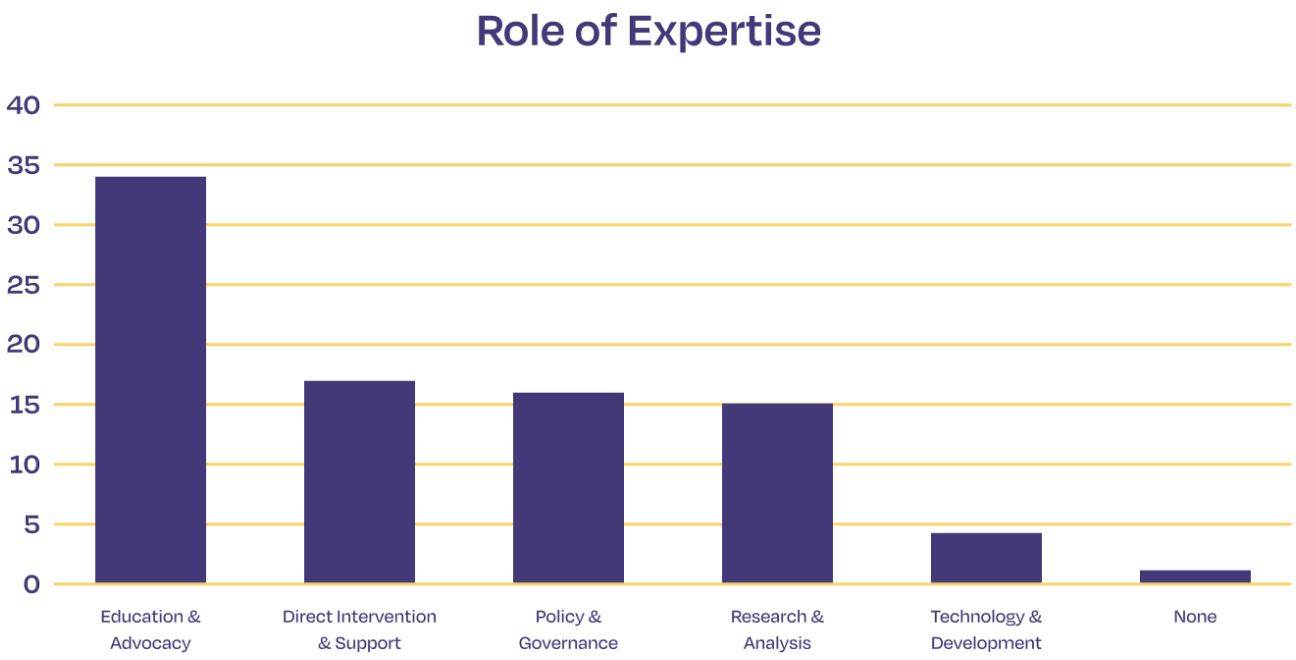


Figure 3 – Role or expertise of survey participants

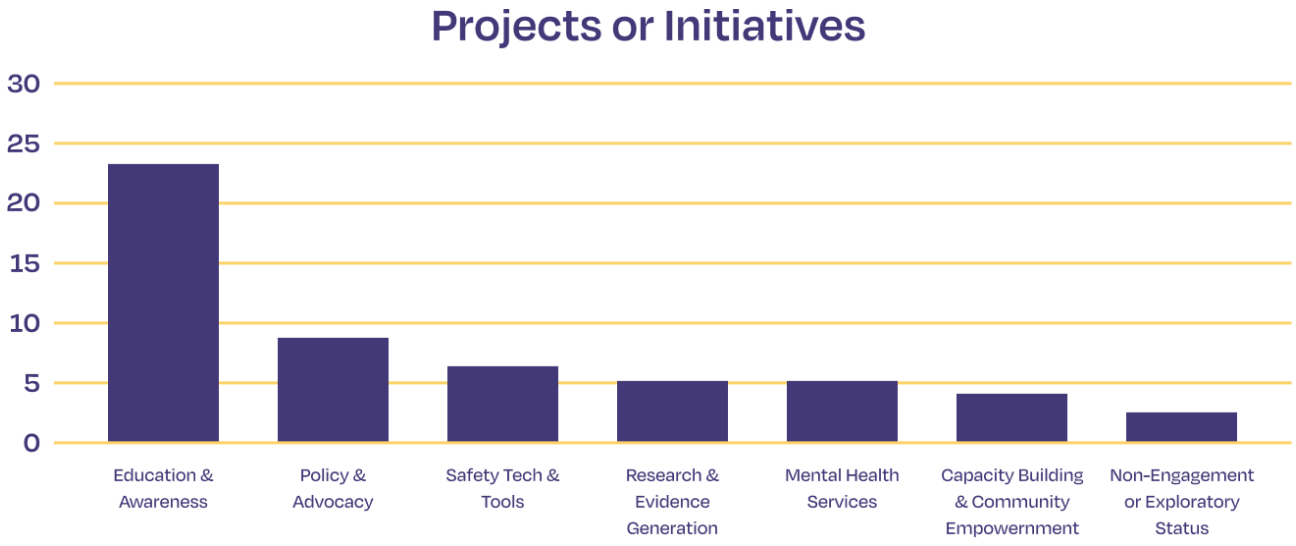


Figure 4 – Types of projects or initiatives undertaken by survey participants

Figure 5 – Survey participants' rating of expertise in relation to online gaming

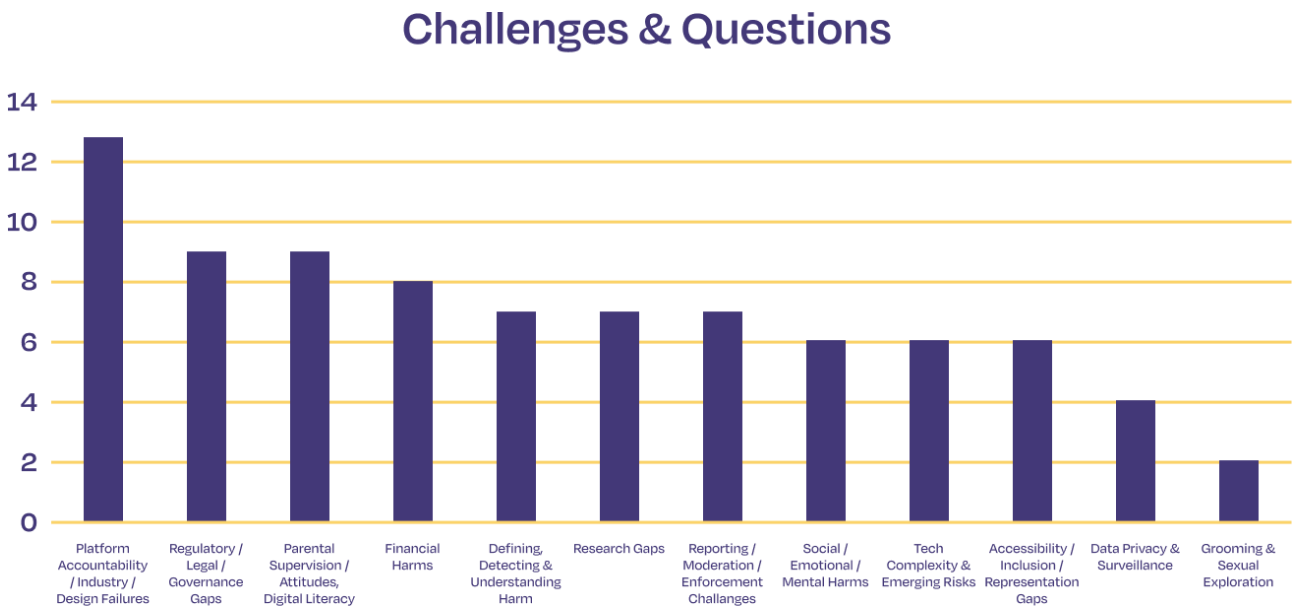


Figure 6 – Challenges and questions that remain unresolved in online gaming

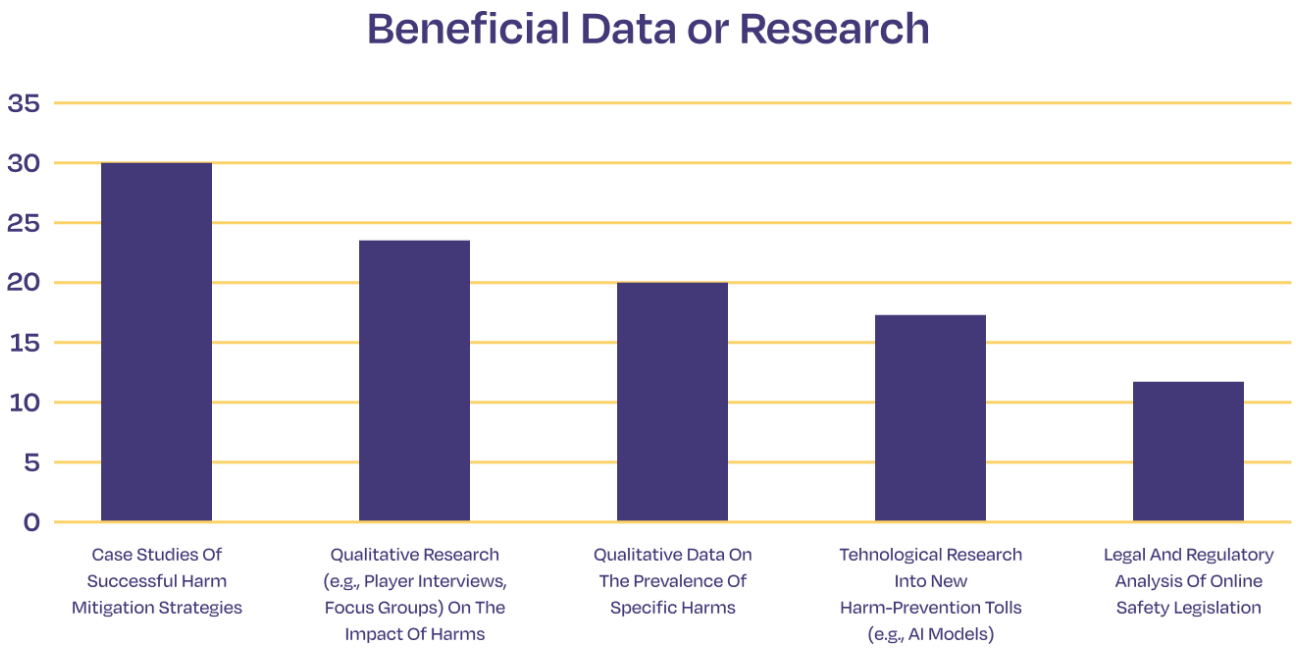


Figure 7 – What data or research would survey participants find useful

Annex 5. Age assurance approaches

Common approaches for age assurance

Category	Examples (AVPA)	How it works	Level of assurance
Self-declaration	• Attestation/self-declaration	User simply ticks a box or enters a date of birth.	**Very Low** (Easily circumvented)
Estimation & profiling	• Age estimation (Biometric) • Social proofing/algorithmic profiling	AI analyses a facial image/voice to estimate age range or analyses user history to infer age likelihood.	**Medium** (Effective for broad age gating; privacy-preserving)
Data checks	• Mobile phone account records • Credit reference agencies • Account holder confirmation	Checks the user's details against existing private sector databases or confirms parental consent.	**High** (Relies on third-party data accuracy)
High confidence checks	• Government ID documents • Open banking/credit card • Biometric ID verification	Verifies official documents (passport, driving licence) or banking credentials, often combined with a liveness check (a process that verifies a biometric identifier, for example, a face scan, comes from a live person rather than a static image) to prevent fraud.	**Very High** (Highest certainty; required for high-risk harms)