FRONTIERS IN DIGITAL CHILD SAFETY

DESIGNING A CHILD-CENTERED DIGITAL ENVIRONMENT THAT SUPPORTS RIGHTS, AGENCY, AND WELL-BEING

FINAL REPORT OF THE FRONTIERS IN DIGITAL CHILD SAFETY WORKING GROUP

TUM THINK TANK

Technical University of Munich

BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

University of Zurich UZH

June 2025

Frontiers in Digital Child Safety: Designing a Child-Centered Digital Environment That Supports Rights, Agency, and Well-Being

Final Report of the Frontiers in Digital Child Safety Working Group

OVERVIEW

Host and Collaborating Organizations:
TUM Think Tank at the Munich School of Politics and Public Policy at the Technical University of Munich (TUM) in collaboration with the Berkman Klein Center for Internet & Society at Harvard University, and the Department of Communications and Media Research at the University of Zurich.

Core Research and Editorial Team:
Co-Principal Investigators and editors:
Sandra Cortesi and Urs Gasser

Researchers and chapter lead editors (in alphabetic order):
Noha Lea Halim (approaches 3 and 2)
Camila Hidalgo (chapter 2, Foundations and approaches 1, 3, and 4)
Annabel Jones (approaches 1 and 4)
Kirsten Müller-Daubermann (approaches 2, 1, and 3)

Editorial support and additional international contributions (in alphabetic order):
Alexa Hasse (editorial support and review)
Anyu Jiang (additional international contributions)
Madeline McGee (editorial support)

Authorship and Contributors:
This report is the collective work of the Working Group (members listed under "Contributors"), composed of distinguished subject matter experts representing a broad range of disciplinary backgrounds and professional expertise. While individual members contributed in varying capacities and provided critical input throughout the process, the report should be understood as a collaborative product rather than a reflection of unanimous agreement on every point. The group's deliberations, extensive feedback, and collective judgment were instrumental in shaping the overall direction and substance of the report. Final responsibility for the content and conclusions remains with the Co-Principal Investigators.

Project Managers:
Noha Lea Halim, Madeline McGee, and Markus Siewert

Funding and Oversight:
The Working Group gratefully acknowledges institutional resources contributed by the TUM Think Tank as well as support provided by Apple Inc. to the Technical University of Munich (TUM). Any views, opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and should not be interpreted as reflecting the views, policies or position, either expressed or implied, of Apple Inc.

The overall initiative has been governed by German and Bavarian laws and regulations and is subject to the codes of conduct and oversight mechanisms of the Technical University of Munich. For more context, please visit the FAQ at https://tumthinktank.de/project/frontiers-in-digital-child-safety.

Editors Expressions of Gratitude:
We extend our heartfelt gratitude to all members of the Working Group for their invaluable contributions, acknowledging their leadership, scholarly excellence, and, in many cases, decades-long dedication to this issue area. The group's generosity, collegial commitment, scholarly rigor, and varied expertise were vital to the depth and overall quality of this final report. Our sincere thanks also go to all staff members and to all others who contributed their expertise and support to this report. Additionally, we would like to express our gratitude to Ruben van der Dussen, Sandy Parakilas, and Erik Neuenschwander for enabling support and insightful collaboration. Thanks also to Camila Romero Otárola (cnromero.com) for her artistic vision and development of the visualizations and the creative layout and design of this report.

Contact:
TUM Think Tank
Richard-Wagner-Strasse 1
80333 Munich, Germany
Email: tumthinktank@hfp.tum.de

SUGGESTED CITATIONS
APA
Frontiers in Digital Child Safety Working Group. (2025). Frontiers in digital child safety: Designing a child-centered digital environment that supports rights, agency, and well-being (S. Cortesi & U. Gasser, Eds.). TUM Think Tank at the Munich School of Politics and Public Policy at the Technical University of Munich, Berkman Klein Center for Internet & Society at Harvard University, and Department of Communications and Media Research at the University of Zurich. https://tumthinktank.de/en/project/frontiers-in-digital-child-safety

KEYWORDS
Youth, young people, children, digital child safety, risks, child rights, child-centered, child agency, well-being, design approaches, child-protective features, interventions, warnings, help-seeking, reporting, education, skills.

PUBLICATION DATE
June 2025

CONTRIBUTORS

HOSTS

Technical University of Munich (TUM)
Urs Gasser
Noha Lea Halim
Camila Hidalgo
Anna Maria Schneider
Markus Siewert

Department of Communications and Media Research (IKMZ), University of Zurich
Sandra Cortesi
Anyu Jiang
Annabel Jones
Kirsten Müller-Daubermann

Berkman Klein Center for Internet & Society, Harvard University
Christopher Bavitz
Madeline McGee
Leah Plunkett

WORKING GROUP / MEMBERS

Isobel Acquah,
Certa Foundation

Stephen Balkam,
FOSI

Michael Best,
Georgia Tech

Lionel Brossi,
University of Chile; Berkman Klein Center

Ernesto Caffo,
S.O.S.-
Il Telefono Azzurro Onlus; Fodazione Child; University of Modena

Anne Collier,
NetFamilyNews

Sebastian Diaz,
Berkman Klein Center

Diana Freed,
Berkman Klein Center

Nathan Freitas,
Guardian Project; Berkman Klein Center

Alexa Hasse,
Tufts University

Sameer Hinduja, Cyberbullying Research Center, Florida Atlantic University;
Berkman Klein Center

Chelsea Johnson,
ASML;
Berkman Klein Center

Lisa Jones,
Crimes against Children Research Centre (CCRC), University of New Hampshire

Laura Jeanne D'arc Kagina,
Certa Foundation

Daniel Kardefelt Winther,
UNICEF Innocenti

Enkelejda
Kasneci,
Technical University Munich

Claudia Lampert,
Leibniz Institute for Media Research | Hans-Bredow-Institut

Amanda Lenhart,
Joan Ganz Cooney Center

Larry Magid,
ConnectSafely

Meg Marco,
ASML;
Berkman Klein Center

Latifah Mariza,
Certa Foundation

Leigh McCook,
Georgia Tech

Andras Molnar,
OECD; Berkman Klein Center

Riana Pfefferkorn,
Stanford Institute for Human-Centered AI

Maria Jose Ravalli,
UNICEF

Michael Rich,
Harvard Medical School

Fanny Rotino,
ITU

Lara Schull,
ASML;
Berkman Klein Center

Fabio Senne,
Cetic.br

Elisabeth Sylvan,
Brown University;
Berkman Klein Center

Rebecca Tabasky,
Berkman Klein Center

Amanda Third,
Young and Resilient Research Centre, Western Sydney University; Berkman Klein Center

Andrew Zack,
FOSI

TUM THINK TANK

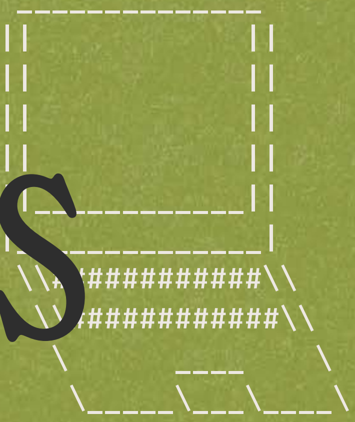
Technical University of Munich

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

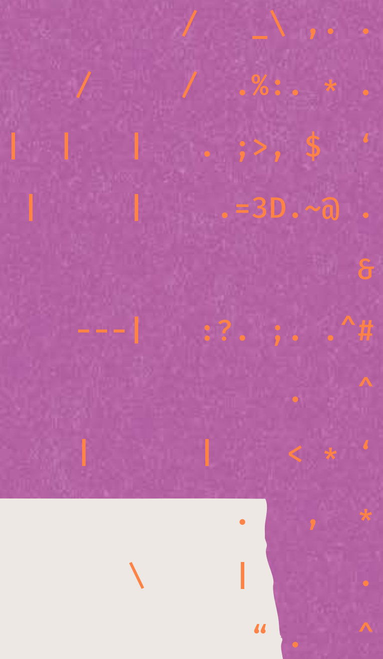
University of Zurich UZH

TABLE OF CONTENTS

TABLE OF CONTENTS



INTRODUCTION



(I)

1. Motivation

Children today navigate a digital environment that offers both unprecedented opportunities and evolving risks. As digital technologies become more embedded in everyday life, concerns about children's digital safety continue to grow. These concerns are shaped by children's age and evolving capacities, which significantly influence their experiences and vulnerabilities.

While there is broad consensus on the importance of safeguarding children in the digital environment, effective responses remain fragmented. Traditional approaches — such as age-based restrictions, bans on smartphone use, screen time limits, or in-app restrictions — often assume specific attitudes and values, and rely on digital knowledge and skills of parents/caregivers and/or fail to match the complexity of children's lived experiences. In many cases, they may even restrict access to valuable opportunities for learning, connection, and expression.

At the same time, the pace of technological innovation — including in areas such as artificial intelligence (AI) and immersive technologies — challenges the ability of research and regulation to respond in a timely fashion. Taken together, these dynamics introduce new forms of content, conduct, contact, and consumer risks, recognizing that cross-cutting risks also emerge from these changes, often in ways that are difficult to anticipate (OECD, 2021a).

Most recently, for instance, the emergence of generative AI is introducing new considerations for digital child safety. This technology, capable of creating personalized content such as text, images, and videos, can enhance children's learning and creativity by offering engaging, tailored experiences. However, it also introduces new risks, including the potential for generating inappropriate, misleading, or harmful content. The capacity of generative AI to simulate human-like interactions further complicates the digital landscape. To address these challenges, future-oriented digital safety strategies must incorporate a nuanced understanding of generative AI's capabilities and its impact on children's rights, agency, and well-being. Children's digital and non-digital lives are deeply intertwined, and harms in one domain can have cascading effects in the other.

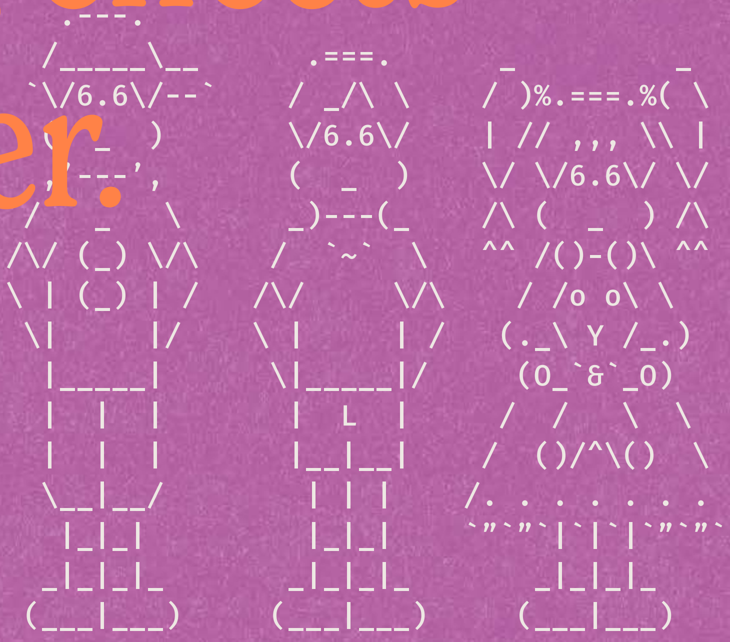
Considering these complex developments, digital child safety must be reimagined not just as a matter of protection, but as a design opportunity. Rather than relying solely on reactive or restrictive measures, safety can be intentionally embedded into the digital environment in ways that promote children's rights, agency, and well-being. This shift draws on earlier design-oriented approaches while pushing toward more proactive, research-informed, and child-centered forms of interventions.

Ensuring digital child safety requires recognizing the complexity of the broader ecosystem. Technology companies, parents/caregivers, educators, policymakers, and children themselves each bring distinct understandings of safety — and differing capacities to support it. Social, cultural, economic, technological, political, and legal conditions vary significantly across contexts. As such, interventions must be adaptable, inclusive, and grounded in the lived experiences of children.

Against this backdrop of diverse contexts and children' s evolving capacities, this report focuses (unless indicated otherwise) on children aged 13–17. This stage is marked by rising autonomy, intensive engagement with the digital environment, and heightened exposure to both opportunity and risks online. The findings presented here may apply differently for younger children and do not translate seamlessly across diverse cultural or geographic contexts. Readers should therefore treat the insights offered in this report as a flexible framework, adapting them to local realities and contexts.

It is amidst this nuanced digital landscape, where transformative potential and emergent vulnerabilities continually intertwine, that the Frontiers in Digital Child Safety project was launched. Convening an international, multi-disciplinary working group, the project sought to move beyond compliance-driven and restrictive approaches toward more innovative, child-centered design strategies. This report presents the outcome of that effort — charting new directions for intervention, surfacing actionable insights, and framing digital child safety as an ongoing process of collaborative learning and adaptive design.

Children's digital and non-digital lives are deeply intertwined, and harms in one domain can have cascading effects in the other.



2. Working Group

This report is a product of the Frontiers in Digital Child Safety project, a one-year initiative led by an academic consortium hosted by the TUM Think Tank at the Munich School of Politics and Public Policy at the Technical University of Munich, the Berkman Klein Center for Internet & Society at Harvard University, and the Department of Communications and Media Research (IKMZ) at the University of Zurich. At the heart of this project was a Working Group – a diverse assembly of researchers, technologists, and practitioners – whose collective work forms the foundation of this report. Through a collaborative and iterative process, the Working Group authored the report, drawing on their wide-ranging expertise to shape its analysis, structure, and conclusions.

Members represented academic institutions such as Harvard Medical School, Stanford University, Georgia Tech, Tufts, the University of Chile, Western Sydney University, and the Hans-Bredow-Institut; international organizations including UNICEF and the OECD; and civil society groups such as the Certa Foundation, Cetic.br, ConnectSafely, the Cyberbullying Research Center, FOSI, Guardian Project, and NetFamilyNews. The group also included technologists from the Applied Social Media Lab and Apple.

With professional backgrounds spanning child protection, digital safety, technology development, and policy, and regional perspectives from North America, Europe, Latin America, and beyond, the Working Group ensured a global, multidisciplinary lens in addressing digital safety challenges.

For further details on the composition of the Working Group see the “Contributors” section.

```
                   ___
                   \ /]
       _          _(_)
    ___))        [  | \___
    ) //o         | |     \
 _ (_    >        | |      ]
(O)  \__<         | | ____/
[/] /   \)       [__|/_
[\]|  ( \        __/___\_____
[/]|   \ \__  ___|           |
[\]|    \___E/%%/|___________|_____
[/]|=====__   (____________________)
[\] \_____ \    |                 |
[/========\ |   |                 |
[\]     []| |   |                 |
[/]     []| |_  |                 |
[\]     []|___) |                 |
```

3. Process

The project was designed to advance knowledge in the field of digital child safety by combining expertise and integrating insights across disciplines and sectors. Over the course of the year, the Working Group participated in a series of three in-person, hybrid, and virtual working meetings, framed by a project launch and concluding meeting session. These convenings served three core purposes:

- To map the current knowledge base, identify gaps, and surface underexplored areas;
- To foster capacity building by promoting information sharing and collaboration among Working Group members and their respective organizations;
- To strengthen the evidence base in support of more effective advancements in digital child safety.

With an emphasis on novel ideas and approaches, the project aimed to complement and enrich existing and emerging efforts led by governments, international organizations, NGOs, and industry actors.

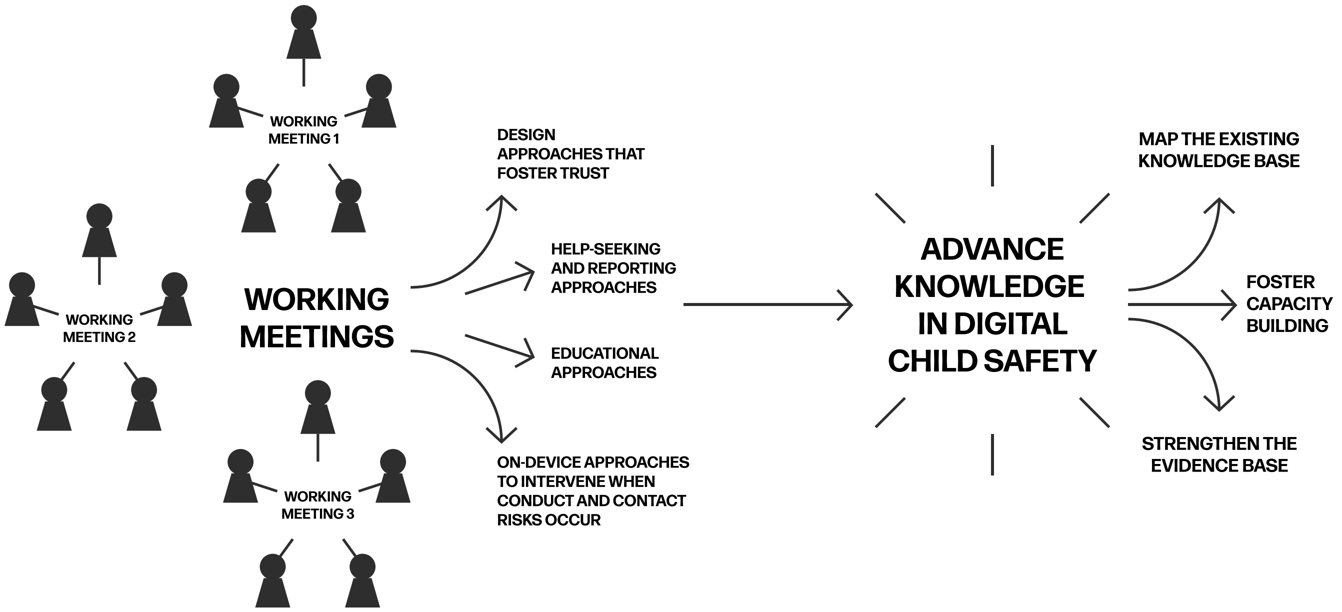
The collaborative work was organized around four key approaches:

1. Design approaches that foster trust
2. Help-seeking and reporting approaches
3. On-device approaches to intervene when conduct and contact risks occur
4. Educational and user interface design approaches

Each approach included a set of guiding questions that informed the project's methodology. Working Group members reviewed relevant literature and expert input in advance of each working meeting, using these materials as the foundation for structured discussion, reflection, and synthesis.

The primary objective of the meetings was to distill and collectively review insights from research and practice, track emerging risks and trends, and explore, map, and benchmark promising interventions. Through deliberative dialogue, the group worked to identify points of convergence, articulate unresolved tensions, and surface actionable strategies.

By identifying points of consensus and drawing from diverse perspectives, the insights collected throughout this process aim to support evidence-based decision-making in a field that remains highly charged and complex.


WORKING MEETING 1
WORKING MEETING 2
WORKING MEETING 3
WORKING MEETINGS
DESIGN APPROACHES THAT FOSTER TRUST
HELP-SEEKING AND REPORTING APPROACHES
EDUCATIONAL APPROACHES
ON-DEVICE APPROACHES TO INTERVENE WHEN CONDUCT AND CONTACT RISKS OCCUR
ADVANCE KNOWLEDGE IN DIGITAL CHILD SAFETY
MAP THE EXISTING KNOWLEDGE BASE
FOSTER CAPACITY BUILDING
STRENGTHEN THE EVIDENCE BASE

4. Structure

This technical report captures key insights from our collaborative process and is structured into four main chapters that highlight the current evidence around digital child safety across the four approaches. We hope this snapshot is helpful to a diverse group of people including policymakers, technology companies, children, educators, parents/caregivers, and others working to enhance digital child safety more broadly.

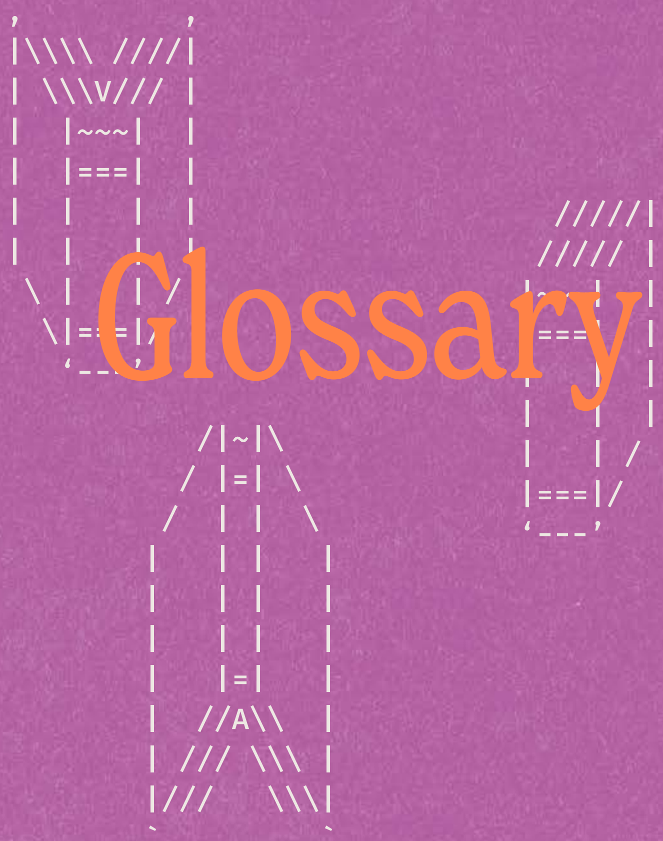
Chapter I begins by establishing the evolving digital landscape where increased digital access presents both opportunities and risks to children's safety. It then details the Frontiers in Digital Child Safety project. This chapter also outlines the report's structure and provides foundational principles. Finally, it offers selective insights highlighting important themes such as the changing nature of child safety, the concept of child safety as a design opportunity, and the shared responsibility of stakeholders.

In chapter II of the report, the background section notes the historical evolution of child safety concerns from the physical environment to the digital age, where children encounter both unprecedented opportunities and complex, rapidly changing risks. To address these challenges, the report adopts "child safety as a design opportunity" as a guiding perspective, emphasizing as foundations the proactive creation of a digital environment that centers children and their rights, agency, and well-being.

Chapter III acknowledges that while digital child safety is a shared priority, reaching consensus on effective strategies is challenging due to various factors. To address this, the report outlines four approaches that aim to intervene and adapt to emerging risks in the digital environment. These approaches explore how design can foster trust, support help-seeking and reporting, embed interventions into devices, and develop educational and user-interface strategies to reduce harm.

In chapter IV, the report brings together cross-sectional insights that highlight the interconnected nature of digital child safety approaches. It reflects on key lessons learned throughout the project, underscoring the importance of designing adaptable solutions that balance protection with empowerment. It concludes by identifying areas for further progress and collaboration, outlining priorities for research, policy development, and practical interventions to address emerging risks and opportunities in the digital environment.

These approaches explore how design can foster trust, support help-seeking and reporting, embed interventions into devices, and develop educational and user-interface strategies to reduce harm.



Glossary

The glossary offers an accessible reference for key terms used in this report, reflecting the evolving landscape of digital child safety. The terms included were specifically chosen for their centrality to the themes and concepts discussed within the “Frontiers in Digital Child Safety” report. We’ve created this detailed section to help all our readers, especially those new to the field, understand the nuanced concepts. Please note that this glossary is intended as a helpful tool for understanding, not as an exhaustive or authoritative legal or academic instrument.

The definitions themselves draw from a blend of sources: established frameworks like the UN Convention on the Rights of the Child and OECD guidelines, concepts developed specifically within this report, and a synthesis of current expert understanding where widely accepted definitions are still emerging. While some terms in this rapidly evolving field might invite further discussion, our aim is to provide clarity and utility for everyone engaging with this report.

(A)

Accountability:
The obligation of stakeholders, including institutions, digital platforms, or individuals to take responsibility for upholding children’s rights and best interests in the digital environment. Accountability implies transparent processes, remedy mechanisms, and enforceable standards for safety and ethical design.

Adaptive design:
An approach to safety interventions that evolves in response to emerging risks, user feedback, and diverse social and cultural contexts. Adaptive design prioritizes flexibility, iteration, and responsiveness to children’s lived experiences, recognizing that effective safety cannot rely on static, one-size-fits-all solutions.

Age-appropriate design:
A principle that calls for tailoring digital products and services as well as policies to the developmental stages and evolving capacities of children. Age-appropriate design may require interoperable and user-friendly technologies, differentiated features, age-appropriate language, content moderation and contact restriction, privacy protection, and support for autonomy appropriate to children’s age, maturity, and rights.

Agency:
The individual and/or collective ability — contextually contingent (i.e., socially embedded and culturally and economically mediated) — of children to make decisions and take action toward their own life and well-being (Cortesi et al., 2021, p. 4; DeJaeghere et al., 2016).

Anonymous reporting tools:
Privacy-preserving and confidential mechanisms that enable children to report harmful digital experiences without fear of retaliation or stigma, thus increasing the likelihood of disclosure and timely intervention.

(B)

Behavioral design:
A field of design that draws from psychology and behavioral economics to guide user choices. In the context of digital child safety, behavioral design techniques — like nudging — are used to support safer habits while respecting autonomy.

(C)

Child-centered design:
A design approach that prioritizes children’s rights, needs, and evolvingcapacities. Child-centered design aims to create digital experiences that enable creativity, emotional regulation, confidence, a sense of purpose, empowerment, social connection, and joy (OECD, 2024).

Child-centered governance:
A framework for policymaking, regulation, and digital platform accountability that puts children’s rights, agency, and well-being at the core. It ensures that governance structures acknowledge children’s evolving capacities, context-specific needs and circumstances, and that those responsible for digital harms to them are held accountable.

Child-protective features:
Features in digital technologies intended to identify and mitigate risks and harms to children. These include tools such as content filters, usage controls, and warning systems. When well designed, they help foster trust while supporting children’s rights, agency, and well-being.

Children:
Refers to every individual below the age of eighteen years acknowledging that different age thresholds may be appropriate in providing certain legal protections (OECD, 2021b)

Child rights:
Rights applying to every child, including political, economic, social and cultural rights, recognised in the UNCRC (UNCRC).

Co-design:
A participatory process in which children, parents/caregivers, and other stakeholders are directly involved in the design and evaluation of digital technologies and safety interventions. Co-design respects children's rights, agency, and well-being and recognizes them as experts in their own experiences.

Collaborative empowerment:
A design and governance approach that shifts from top-down parental/caregiver control toward joint decision-making involving both children and parents/caregivers. Such an approach fosters children's agency and trust in digital safety tools and environments.

Conduct risks:
Risk where children are actors in a peer-to-peer exchange, including when their own conduct can make them vulnerable (OECD, 2021a, p. 9)

Consumer risks:
Children facing risks as consumers in the digital economy (OECD, 2021a, p. 10).

Contact risks:
Risks when children interact in the digital environment (OECD, 2021a, p. 10).

Content risks:
Child exposed to hateful content, harmful content, illegal content, disinformation (OECD, 2021a, p. 7).

Context-aware interventions:
Real-time interventions providing age-appropriate, timely guidance and resources tailored to the context of the user's actions or situation.

Cyberbullying:
Intentional and repeated harm inflicted through the use of computers, cell phones, and other electronic devices (Hinduja & Patchin, 2025).

(D)

Design opportunity:
A framing that treats child safety not as a constraint or reactive measure but as a proactive, creative process embedded in the architecture of digital technologies. This approach draws on principles of design thinking, focusing on adaptability, co-creation, and inclusion to align the digital environment with children's rights, needs, and evolving capacities.

Digital child safety:
A proactive, by-design approach aimed at building a safe digital environment that inherently centers children and prioritizes their rights, agency, and well-being. Achieving this requires embedding safety into the architecture and functionality of each layer of the digital environment and fostering shared responsibility across all stakeholders. The ultimate aim is to create an inclusive ecosystem where children can explore, connect, and thrive safely without unnecessary restrictions.

Digital environment:
A broader context or setting within which digital interactions, processes, and activities occur. It encompasses the technological infrastructure, the socio-technical context, and the conditions enabling or constraining digital operations. This definition also includes children's lived digital experiences — acknowledging, for instance, their social interactions and their engagement with digital content, gaming, and learning platforms. It considers both the opportunities and risks children encounter, as well as the systems designed to safeguard their well-being.

Digital maturity:
The ability to navigate the digital environment with critical understanding, self-awareness, and self-determined use. Digital maturity involves not just practical and physical skills, but also social and emotional, cognitive, and meta-cognitive skills useful for children in the digital environment.

Digital safety by design for children:
An approach to keeping children safe in the digital environment by embedding protective features into products and services from the outset — while still enabling children to explore, connect, and enjoy the benefits of digital life (OECD, 2024, p. 5-7).

Digital skills:
1) Practical (e.g., utilizing new digital technology devices) and physical skills (e.g., using a digital device, such as a tablet or mobile phone, to achieve a specific outcome, like finding information online for a school assignment); 2) social and emotional skills (e.g., collaboration, self-efficacy, empathy); and 3) cognitive and meta-cognitive skills

(e.g., self-regulation, motivation to learn, creativity, and critical thinking) (Cortesi et al., 2021, p. 5; for a comprehensive discussion of the theme, please see Cortesi et al. [2020]).

Digital technologies:
An overarching term encompassing the Internet, mobile technologies, digital networks and databases, digital content, platforms, and services — along with emerging technologies such as AI, robotics, augmented and virtual reality, algorithms, big data, and the Internet of Things (Cortesi et al., 2021, p.4).

(E)

Educational approaches:
Structured pedagogical strategies that equip children with the knowledge and skills to identify, respond to, and report risks and harms in the digital environment, promoting long-term well-being.

Evolving capacities:
A foundational concept in child rights frameworks that acknowledges children's growing ability to make autonomous decisions in accordance with their age and developmental capacities. Digital child safety approaches should reflect and support this developmental progression.

(F)

Forbidden fruit effect:
A psychological phenomenon where warnings or restrictions increase children's curiosity and interest in the prohibited content, sometimes making the intervention counterproductive.

(G)

Guardrails:
Flexible, contextual design features that provide boundaries to reduce harm while preserving children's opportunities including for exploration and learning. Guardrails contrast with rigid restrictions and are intended to evolve with children's needs and the evolving digital environment.

(H)

Help-seeking and reporting approaches: Technology-based interventions designed to support children to seek help from peers or adults after encountering risks, and to report harmful experiences confidentially.

Holistic educational approaches:
Educational strategies that integrate digital safety into broader learning contexts such as mental health, relationships, and bullying prevention, recognizing the interconnected nature of digital and offline well-being.

(I)

Interfaces:
Points of interaction and collaboration between different stakeholder groups, such as parents/caregivers, educators, researchers, designers, children, and companies. Interfaces are not just communication channels but crucial spaces where knowledge, responsibilities, and values intersect and co-shape digital child safety strategies.

Intersectionality:
The understanding that children's digital experiences — and their exposure to risks or exclusion — are shaped by overlapping factors such as age and evolving capacities, circumstances, education (skill level), ethnicity, gender, location, national origin, race, and/or socioeconomic background. Intersectional approaches to safety and design are essential to ensure equity and inclusion in the digital environment.
Intervention: A safety mechanism, feature, or strategy — technical, educational, or policy-based — intended to reduce harm, promote well-being, or reinforce rights in the digital environment. Interventions may range from pop-up nudges and content warnings to algorithmic filters or legal guardrails.

(N)

Nudging:
A behavioral design technique that uses subtle cues — such as interface layout, color, or timing — to influence

user behavior without restricting choice. Nudging can be applied to promote safer digital habits, prevent harmful actions, or guide users toward more informed decisions.

(O)

On-device approaches:
Strategies embedded directly into a user's device (e.g., smartphone, tablet) that aim to detect, prevent, or intervene in real-time when children engage in, or are exposed to, digital risks. These approaches may include AI-based nudges, content moderation filters, privacy-preserving alerts, or in-situ interventions — and are often designed to act locally without requiring constant cloud-based surveillance.

Online:
Refers to digital activities, content, interactions, and experiences that occur via internet-connected platforms, services, or devices. In the context of child safety, "online" encompasses a wide range of spaces, from social media and gaming platforms to educational tools and anonymous sharing apps, amongst others.

(P)

Parental controls:
These comprise software, tools, and features that enable an parents/caregivers to control some or all the functions of a digital device or service used by a child to filter, limit, or otherwise determine their access and use in ways intended to protect their safety (ITU, 2020; UNICEF & ITU, 2014).

Participation:
A child's ability and opportunity to express their views and actively shape decisions affecting them, including those related to digital technologies. Participation is both a fundamental right (UNCRC Article 12) and a design and governance principle that calls for meaningful, age-appropriate mechanisms for engagement.

Peer support networks:
Digital platforms or moderated communities where children can receive emotional support, advice, and reassurance from peers or trained mentors, often serving as a first step before reaching out to adults.

Privacy:
The ability of children and their parents/caregivers to safeguard access to, processing, collection, sharing, and use of children's personal data. In this report, privacy is treated both as a right (grounded in child rights frameworks) and as a design principle, particularly within on-device interventions and data governance strategies. Privacy-preserving tools aim to ensure safety without unnecessary surveillance or erosion of trust

(R)

Resilience:
A child's capacity to cope with, recover from, and learn from challenging experiences. Resilience is fostered through skills, emotional support, agency, and age-appropriate exposure, and should be supported — not undermined — by safety interventions.

Risk:
A factor that has the capacity to significantly affect children's lives in multiple ways and can be mitigated through protective policies, design practices, and regulatory frameworks. Risks are typically categorized as content, contact, conduct, or consumer risks, and often intersect with broader cross-cutting concerns around privacy, advanced technology, health, and well-being (OECD, 2022, p. 4-7)

(S)

Safeguarding:
A broader term used in policy and education contexts to describe systems and practices that protect children from harm — whether online or offline. In the digital environment, safeguarding may involve technical features, reporting systems, educational initiatives, and governance frameworks.

Safety signals:
Subtle, interface-based cues designed to reassure users or guide behavior — such as privacy indicators, reporting prompts, or content labels. Safety signals play a role in building trust and encouraging responsible choices without disrupting the user experience.

Skill-based interventions:
Targeted learning activities aimed at building specific skills across areas of life (Cortesi et al., 2020) to help children acquire the skills they need to thrive in today's society and manage online risks meaningfully.

(T)

Transparency:
The degree to which children, parents/caregivers, and other users understand how digital systems operate — including how safety features work, what data is collected, and how decisions are made. Such information can be provided to children in language that is clear, plain, and appropriate to their age and maturity. Transparency can foster trust, accountability, and informed decision-making.

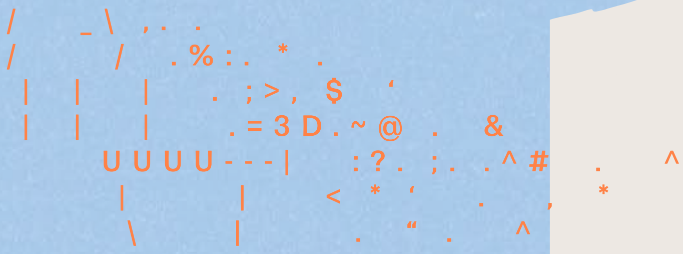
Trust:
In the context of digital child safety, trust refers to a child's (and their parents'/caregivers') confidence that safety features and the digital environment are protective, predictable, transparent, and respectful of their agency. Building trust can involve co-design, positive reinforcement, and transparent communication. Distrust can result from overly restrictive or opaque interventions.

(W)

Well-being:
A multidimensional concept that includes cognitive, psychological, physical, and social aspects of children's development and everyday life. Digital experiences can both enhance and threaten well-being, which is why safety interventions must aim not only to reduce harm but also to support children's ability to thrive across these dimensions.

CONTEXT AND CHALLENGES IN DIGITAL CHILD SAFETY

(II)



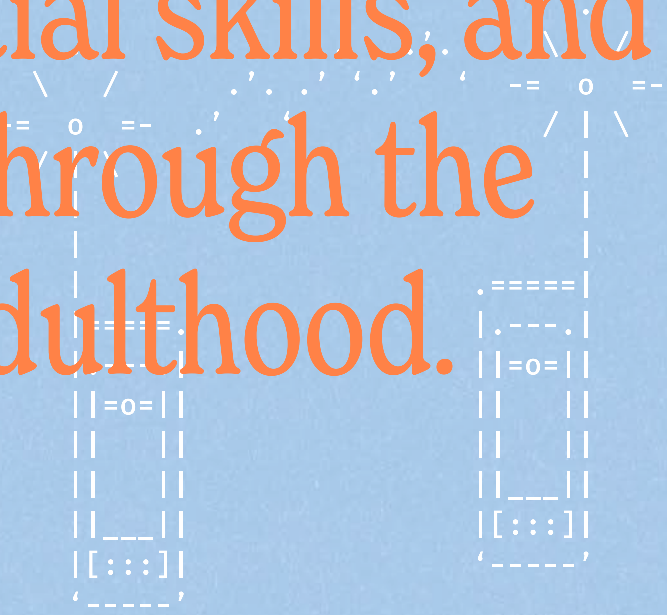
1. Balancing Opportunities and Risks

As of 2025, more than two billion people live in a world where Internet access, digital devices (including smartphones), and a vast digital environment are widely available (Henry & Shannon, 2023). Over the past decade, Internet connectivity among young people – particularly in high- and upper-middle-income countries – has also increased (Faverio & Sidot, 2024; UNICEF, 2025a). However, stark inequalities persist: youth in lower-middle-income countries and those in the poorest households within wealthier nations remain far less connected (UNICEF, 2025a). Nonetheless, research shows that the digital environment can significantly enrich children's lives by enabling children to socialize, communicate, play, and learn — thereby supporting their rights, helping them develop essential skills, and guiding them through the transition to adulthood (Fan et al., 2024; Gasser et al., 2012; Gasser & Cortesi, 2017; Livingstone & Pothong, 2022; Lombana-Bermudez, Cortesi, et al., 2020a; Palfrey & Gasser, 2008, 2016).

At the same time, the digital environment also introduces risks. Concerns about child safety predate the digital era, when protection efforts focused on physical spaces and interpersonal interactions — addressing issues like education, healthcare, exploitation, and community safety (Korbin, 2003).

From the early days of the Internet, risks and harms have remained a central concern (e.g., Barbosa, 2014; Byrne et al., 2016; Gasser, Cortesi, & Gerlach, 2012; Lenhart et al., 2011; Livingstone et al., 2011; Nash, 2014; OECD, 2021a; O'Neill, Staksrud, & McLaughlin, 2013; Palfrey, boyd, & Sacco, 2010; Palfrey & Gasser, 2008, 2016; Smahel et al., 2020). Initial debates focused on contact risks — particularly the "stranger danger" concern (Jones, Mitchell, & Finkelhor, 2013; Palfrey, boyd, & Sacco, 2010). Over time, attention expanded to include peer-to-peer risks like cyberbullying, hate speech, online sexual harassment, digital dating abuse, sextortion, and cyberstalking (Hasse et al., 2019; Hinduja & Patchin, 2021; Hinduja & Patchin, 2010; Levy et al., 2012; Mishna et al., 2023; Obermaier & Schmuck, 2022; Patchin & Hinduja, 2024; Ray & Henry, 2025; Taylor et al., 2021; Ybarra et al., 2012; Wachs et al., 2023; Walsh et al., 2025).

While peer-based forms of aggression in the digital environment remain highly relevant (Hinduja & Patchin, 2024), recent debates have increasingly focused on the broader impacts of smartphones, screen time, and social media use on children's health and well-being (Agha et al., 2023; Bhaimiya, 2024; Cooney & Standley, 2024; Langreo, 2024; Madden et al., 2024; McBain, 2024; Odgers, 2024; Ortutay, 2024; Paul, 2024; Remnick, 2024; Schlott, 2024; Villano, Research shows that the digital environment can significantly enrich children's lives by enabling children to socialize, communicate, play, and learn — thereby supporting their rights, helping them develop essential skills, and guiding them through the transition to adulthood.



2024). These concerns have sparked proposals such as banning smartphones in schools and introducing age restrictions on social media use (Bhaimiya, 2024; Cooney & Standley, 2024; Kaleem, 2024; Langreo, 2024; McGuirk, 2024; OECD, 2021a, 2022; Ortutay, 2024; Paul, 2024).

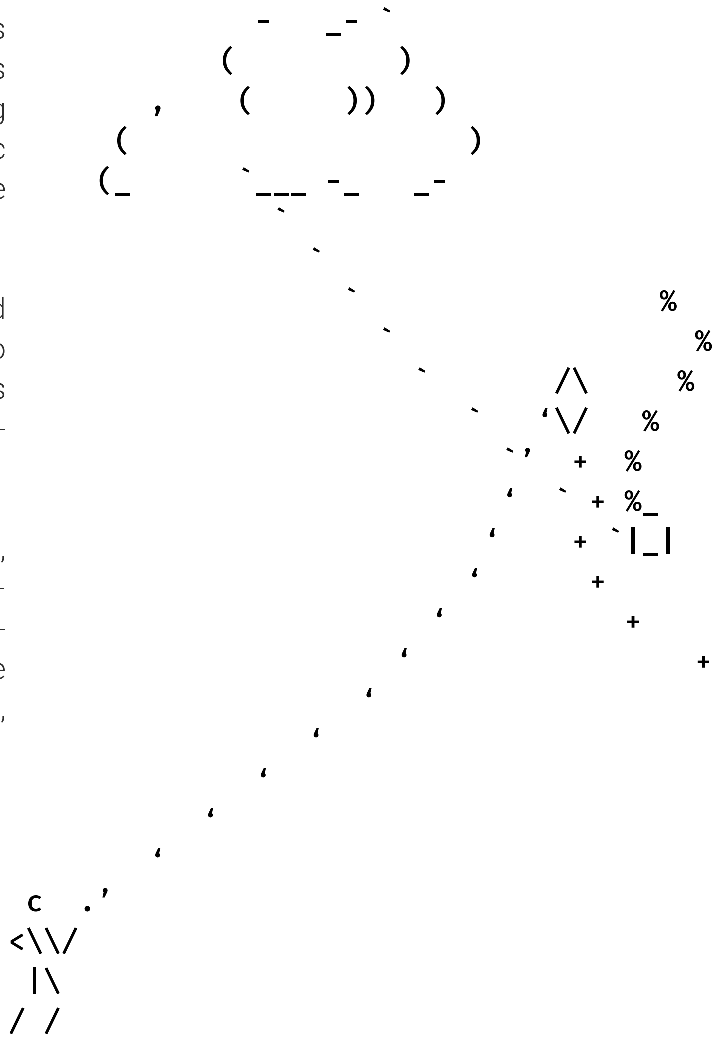
The Typology of Risks developed by the OECD and the Berkman Klein Center for Internet & Society at Harvard University provides structured approaches for understanding these challenges, categorizing risks into content, conduct, contact, and consumer risks, as well as cross-cutting concerns such as privacy, health and well-being, and advanced technology (OECD, 2021a).

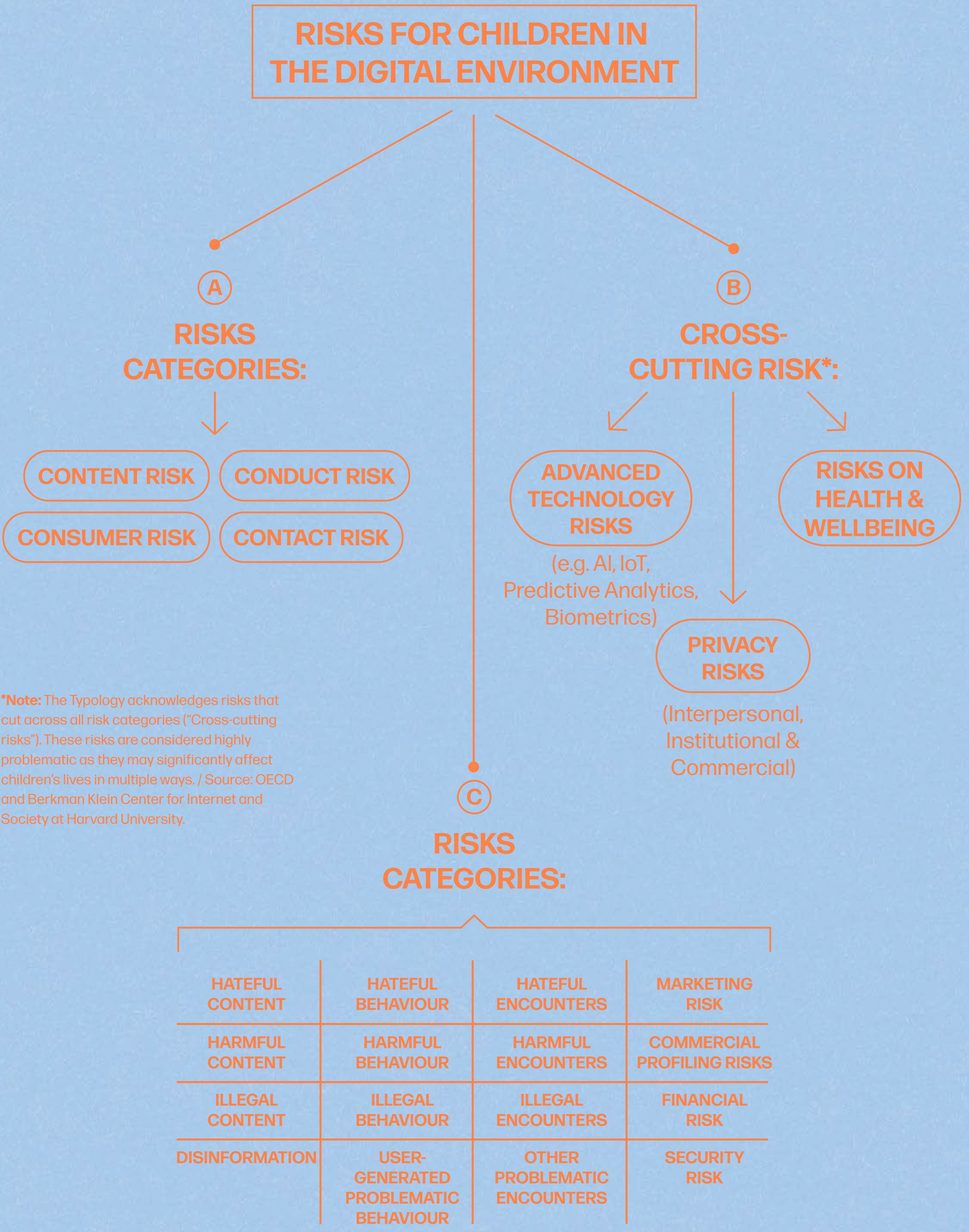
Risks and benefits in the digital environment are not static. The rapid pace of technological innovation demands continuous adaptation in how we conceptualize and address them. Emerging and advanced technologies further contribute to the complexity and dynamism of this space. Generative AI, as mentioned before, serves as a catalyst for both opportunities and risks: it can create personalized content that enhances learning (Kasneci et al., 2023), yet also poses challenges such as generating harmful or illegal content, creating disinformation, and blurring the lines between authentic and artificial experiences (The Alan Turing Institute and LEGO, 2025).

Children's experiences with digital opportunities and risks are shaped not only by technology itself but also by cultural, contextual, and socioeconomic factors. This underscores the importance of inclusive, context-sensitive, and adaptable strategies.

Advancing digital child safety will require international, multi-sectoral, and interdisciplinary collaboration — accounting for diverse values, definitions, and approaches. Crucially, this also means balancing protective measures with those that safeguard children's access, agency, and participation.

Risks and benefits in the digital environment are not static.





RISKS FOR CHILDREN IN THE DIGITAL ENVIRONMENT

Ⓐ RISKS CATEGORIES:

- CONTENT RISK
- CONDUCT RISK
- CONSUMER RISK
- CONTACT RISK

Ⓑ CROSS-CUTTING RISK*:

- ADVANCED TECHNOLOGY RISKS (e.g. AI, IoT, Predictive Analytics, Biometrics)
- PRIVACY RISKS (Interpersonal, Institutional & Commercial)
- RISKS ON HEALTH & WELLBEING

Ⓒ RISKS CATEGORIES:

HATEFUL CONTENT	HATEFUL BEHAVIOUR	HATEFUL ENCOUNTERS	MARKETING RISK
HARMFUL CONTENT	HARMFUL BEHAVIOUR	HARMFUL ENCOUNTERS	COMMERCIAL PROFILING RISKS
ILLEGAL CONTENT	ILLEGAL BEHAVIOUR	ILLEGAL ENCOUNTERS	FINANCIAL RISK
DISINFORMATION	USER-GENERATED PROBLEMATIC BEHAVIOUR	OTHER PROBLEMATIC ENCOUNTERS	SECURITY RISK

***Note:** The Typology acknowledges risks that cut across all risk categories ("Cross-cutting risks"). These risks are considered highly problematic as they may significantly affect children's lives in multiple ways. / Source: OECD and Berkman Klein Center for Internet and Society at Harvard University.

2. Child Safety as a Design Opportunity

This report proposes viewing child safety as a design opportunity, rather than relying on rigid or reactive measures. This shift aligns with design thinking — emphasizing a human-centered, proactive, and iterative approach that integrates ongoing research and adapts to evolving needs. Here, “design” goes beyond product interfaces to encompass the intentional shaping of the entire digital environment to prioritize children’s well-being.

This broader understanding of design is often missing from contemporary policy and industry responses. In a politically charged environment, decision-makers — parents/caregivers, educators, community leaders, technology companies, and policymakers — frequently propose restrictive measures out of genuine concern, but without sufficient evidence or viable alternatives, or sufficiently centering children’s realities and lived experiences. These interventions can unintentionally neglect children’s rights to provision, protection, and participation, while limiting opportunities for digital connection, creativity, and support.

Building on established frameworks such as safety by design, privacy by design, and children’s rights by design (Cavoukian, 2011; Digital Futures Commission, 2023; eSafety Commissioner, n.d.; OECD, 2024; Thaler & Sunstein, 2008), the opportunity-oriented approach proposed in this report expands the scope of ‘by-design’ thinking. It spans the full spectrum of design choices — from risk mitigation to creatively enhancing empowering features — emphasizing learning, adaptability, and contextual relevance.

For example, guardrails, as conceptualized by Gasser and Mayer-Schönberger (2024), provide flexible boundaries that mitigate harm while preserving opportunities. Balancing exposure to risks like hate speech with space for resilience and learning (Ito et al., 2020; Livingstone, 2013) requires context-sensitive guardrails that evolve with children’s realities. Studying edge-case communities, as Kuang and Fabricant (2019) suggest, can inspire more inclusive and universally beneficial design.

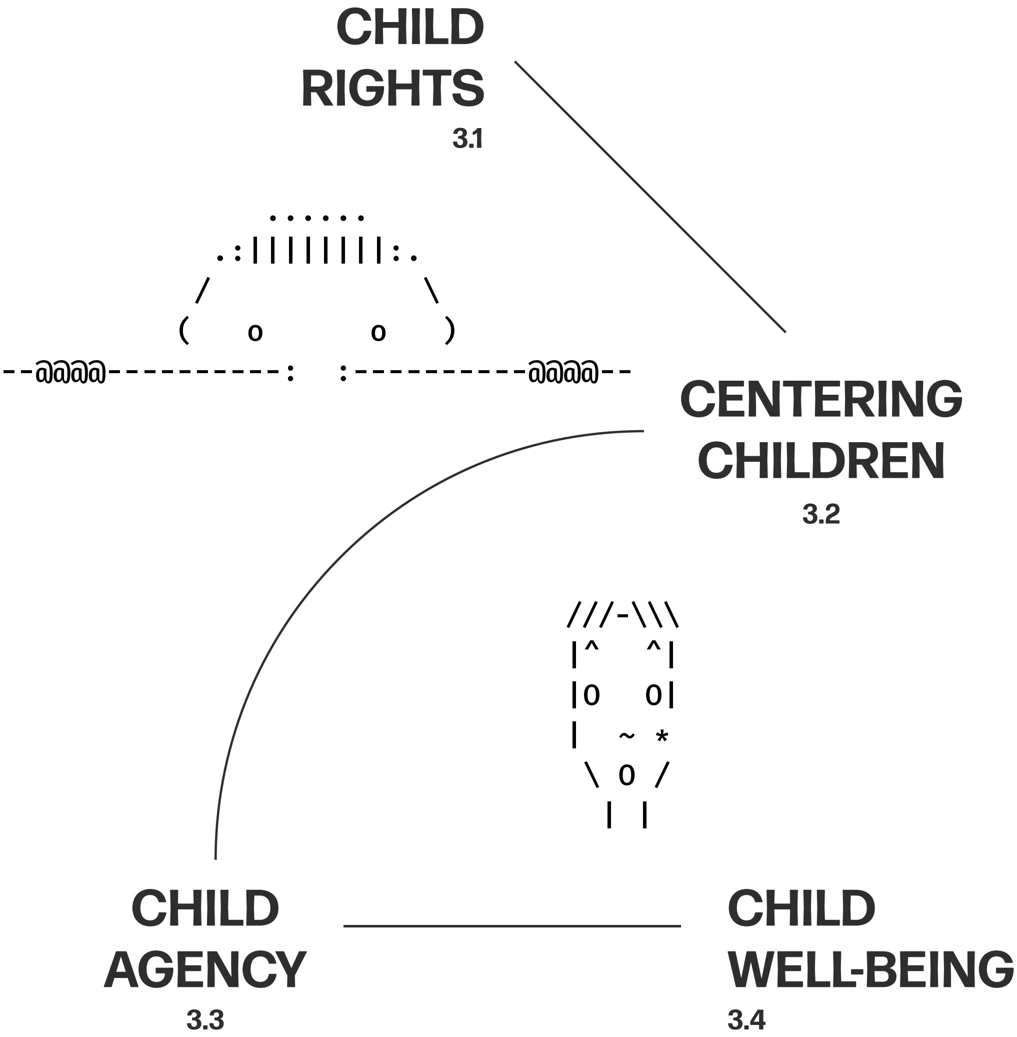
The urgency of this approach is reinforced by the rapid technological shifts outlined earlier, which bring both emerging risks and novel opportunities. These intertwined dynamics highlight the need for adaptive, forward-looking strategies. By reimagining child safety as a design opportunity — reflected in the four approaches introduced in this report — this work contributes to the broader goal of building digital environment that are not only safer, but also more empowering for children.

```
                         ____
                  /^\   / -- )
                 / | \ (____/
                / | | \ / /
               /_|_|_|_/ /
                |     / /
 __    __    __ |    / /__
[  ]__[  ]__[  ].   / /[  ]__
|__            ____/ /___
   |          / .------  )
   |         / /        /
   |        / /        /
```

3. Foundations

Framing digital child safety as a design opportunity also requires clarity about the principles that should guide such an approach. The following section outlines the foundational concepts identified by the working group, which serve as the normative, procedural, and substantive anchors for advancing a child-centered digital environment.

Child rights are the normative foundation for this work. Moreover, it centers children as the procedural foundation, grounding frontiers in Digital Child Safety in a holistic approach that keeps the child at the core of our effort. Finally, child agency and well-being were identified as the substantive foundations for moving forward.



CHILD RIGHTS
3.1

```
                  ......
             .:|||||||||:.
           /               \
         (    o         o    )
--@@@@----------:   :----------@@@@--
```

CENTERING CHILDREN
3.2

```
///-\\\
|^   ^|
|O   O|
|  ~ *
 \ O /
  | |
```

CHILD AGENCY
3.3

CHILD WELL-BEING
3.4

3.1 Child Rights

As highlighted in the previous section, children "have a specific set of needs and rights that are not met by governance regimes designed for everyone" (Livingstone et al., 2016, p. 6). The Convention on the Rights of the Child (CRC) remains the most forceful expression of children's rights, even though it was written in the pre-digital era. With the growing presence of Information and Communication Technologies (ICTs) in children's lives, the Committee on the Rights of the Child explains how children's rights must be upheld in the digital environment in General Comment No. 25. This Comment states that although the digital environment was not designed for children, all actions regarding the provision, regulation, design, management, and use of the digital environment should ensure the child's best interest (Committee on the Rights of the Child, 2021). Moreover, it highlights that risks and opportunities in the digital environment depend, to a great extent, on children's ages and stages of development; therefore, measures that ensure their rights should be designed accordingly (Committee on the Rights of the Child, 2021).

The CRC provides a useful framework for mapping issues related to children in the digital environment and for safeguarding their best interests (Cortesi & Gasser, 2017). Although the CRC itself does not explicitly categorize rights into groups, children's rights scholars (Livingstone & O'Neill, 2014) and practitioners commonly organize the Convention's rights into three clusters: Protection, provision, and participation rights. Despite significant progress in protecting children, these clusters propose a more holistic interpretation of children's safety – ensuring access to the opportunities of the digital environment, safe interaction within it, and strengthening avenues toward provision and participation rights (Committee on the Rights of the Child, 2021). For example, children should have access to information, be equipped to evaluate information quality, and their views should be taken into account when making decisions that affect their lives.

3.2 Centering Children

Children's experiences, perspectives, needs, and voices are crucial when addressing digital safety, which is why centering children is one of this working group's foundations. While this expert group did not directly engage children, their interests and perspectives were represented through the research and daily work of the working group members.

Involving children in the design of the digital environment is key to safeguarding their rights, agency, and well-being. Alongside such involvement, parents/caregivers, and educators play vital roles — parents/caregivers as primary guides who help children navigate the digital environment and educators who foster (practical and physical, social and emotional, and cognitive and meta-cognitive) digital skills. A multi-stakeholder approach that genuinely centers children ensures shared responsibility and collaboration, fostering a digital environment that reflects children's actual strengths and experiences.

Children tend to feel more supported through a holistic approach to digital safety where various stakeholders – including parents/caregivers, educators, community leaders, technology companies, and policymakers, among others – work together to ensure their rights, agency, and well-being (Lala et al., 2022). A trustworthy, supportive ecosystem, inclusive of all stakeholders, should collaboratively address the risks and responsibilities among all users.

Family engagement in digital safety interventions and parental/caregiver mediation are essential for reducing risks in the digital environment (Finkelhor et al., 2021; Hinduja & Patchin, 2022a; Livingstone & Blum-Ross, 2020; Palfrey & Gasser, 2020). However, parents/caregivers often face the challenging task of safeguarding their children's digital safety without sufficient resources (including ability or willingness) or guidance, making it critical for a multistakeholder approach to equip them adequately. Supporting parents/caregivers with specific knowledge and skills can strengthen their ability to safeguard children's digital safety, as well as their rights, agency, and well-being (UNICEF, 2020).

To meet children's needs in the face of digital risks and emotional or academic challenges, it is essential to strengthen the interfaces between stakeholders — including families, educators, researchers, designers, and technology companies. These interfaces are not merely channels for communication, but sites where responsibilities, values, and knowledge must align. The interface between research and design, in particular, plays a critical role in ensuring that evidence meaningfully informs how the digital environment is built and refined.

While design choices by technology companies can support children thrive in the digital environment, this potential must be accompanied by accountability. As primary architects of digital spaces, these actors must be held to robust standards (Costello et al., 2024; Hinduja & Lalani, 2025) — especially when it comes to building safer systems, improving reporting tools, and responding to harm. Real progress in digital child safety requires not only collaboration, but governance frameworks that recognize disparities in power and ensure all actors — particularly those with outsized influence — fulfill their responsibilities in creating a digital environment that is not only safer, but also more just and empowering.

3.3 Child Agency

Although the definition of agency and its operationalization vary across and within disciplines, it is widely recognized as a distinctively human quality that emerges and develops during childhood (Archard, 2015; Brod et al., 2023). For this report, agency is understood as the individual and collective capacity of children to make decisions and take action toward their own lives and well-being. Research within childhood studies demonstrates that children actively exercise their agency through awareness, knowledge and skill acquisition, and strategy (Mühlbacher & Sutterlüty, 2019; Valentine, 2011). **For more definitions and concepts, please visit the Glossary.**

Recent research emphasizes that children are not only aware of their agency but also strongly desire active involvement in managing their digital safety (Third, 2024). This highlights the critical importance of actively integrating children's voices into discussions and decision-making regarding digital safety (Lala et al., 2022). Furthermore, involving children in conversations about their online activities and collaboratively establishing guidelines can foster a more open environment, making it easier for them to share experiences and report issues they encounter (Hasse et al., 2019).

Educational interventions significantly strengthen children's agency by equipping them with essential knowledge and skills to navigate risks in the digital environment (Hasse et al., 2019; Reardon et al., 2017; Reich et al., 2012). According to the Theory of Planned Behavior, improving children's awareness of what constitutes risk, equipping them with appropriate response strategies, and enhancing their confidence in handling difficult situations effectively mitigates harm and reinforces their agency (Bright et al., 2023).

Effective educational interventions should focus on developing skills for risk management and harm reduction, such as recognizing and responding to inappropriate online requests, safely navigating the digital environment, and seeking help when needed (Finkelhor et al., 2021). Evidence shows that children with better digital skills are less likely to experience harm in the digital environment and are more capable of dealing with risky situations (Kardefelt Winther et al., 2023).

Additionally, involving children directly in designing safety features for digital technologies provides a deeper and more nuanced understanding of their specific needs, concerns, and preferences. Crucially, environments deliberately designed to give children a genuine sense of control over their social media and digital interactions correlate positively with increased life satisfaction, enhanced perceived social support, and significantly reduced symptoms of anxiety, depression, and stress (Lee et al., 2024).

Thus, actively supporting and enhancing child agency in digital contexts through education, co-creation of safety practices, and co-design is fundamental to fostering safer and more empowering digital experiences for children.

Children are not only aware of their agency but also strongly desire active involvement in managing their digital safety.

3.4 Child Well-Being

Child well-being has become a “conceptual focal point for assessing the state of children and the discourses on their status” (Ben-Arieh et al., 2014, p. 2). Well-being and safety are interconnected since they are related to the capacity to confront the risks that the digital environment may come with to safeguard one’s well-being (Cortesi et al., 2020). For instance, to reduce risk-taking and victimization, social and emotional skills, such as emotion management, decision-making, and empathy play a vital role (Finkelhor et al., 2021).

The significance of fostering children’s well-being was heightened during the COVID-19 pandemic, which profoundly affected almost every aspect of daily life, including living conditions, education, health, and social relationships (OECD, 2020b; UNESCO, 2023; UNICEF, 2025b). The pandemic disproportionately impacted children, interrupting their education, intensifying mental and physical health issues, and reducing access to essential mentorship and peer support at crucial developmental stages (Hults & Adelsheim, 2020). Furthermore, these challenges were exacerbated by socioeconomic inequalities, systemic discrimination, racism, and race-based violence, highlighting the need for holistic approaches to support children’s well-being (UNICEF, 2024; Yip, 2020).

The literature offers a number of ways to conceptualize well-being, from a focus on physical and/or mental health (Ito et al., 2020; Salam et al., 2016) to a more encompassing concept that integrates multiple dimensions of day-to-day life, from employment to social interactions, and living conditions (European Commission, n.d.; Ross et al., 2020).

The Typology of Risks developed by the OECD and the Berkman Klein Center **(see page 22)** demonstrates how digital experiences directly intersect with various dimensions of well-being, including three critical domains: Social interactions, economic and physical safety, and physical and mental health. For instance, contact risks can affect children’s social interactions, physical safety, and mental health. Consumer risks may influence their economic safety, while conduct risks can have implications for both their physical and mental health. These examples highlight how the risk framework used in this report connects to the broader understanding and prioritization of children’s well-being. Ultimately, this approach underscores the need for an integrated perspective on protecting and promoting children’s well-being in the digital environment.

FOUR APPROACHES TO FRONTIERS IN DIGITAL CHILD SAFETY

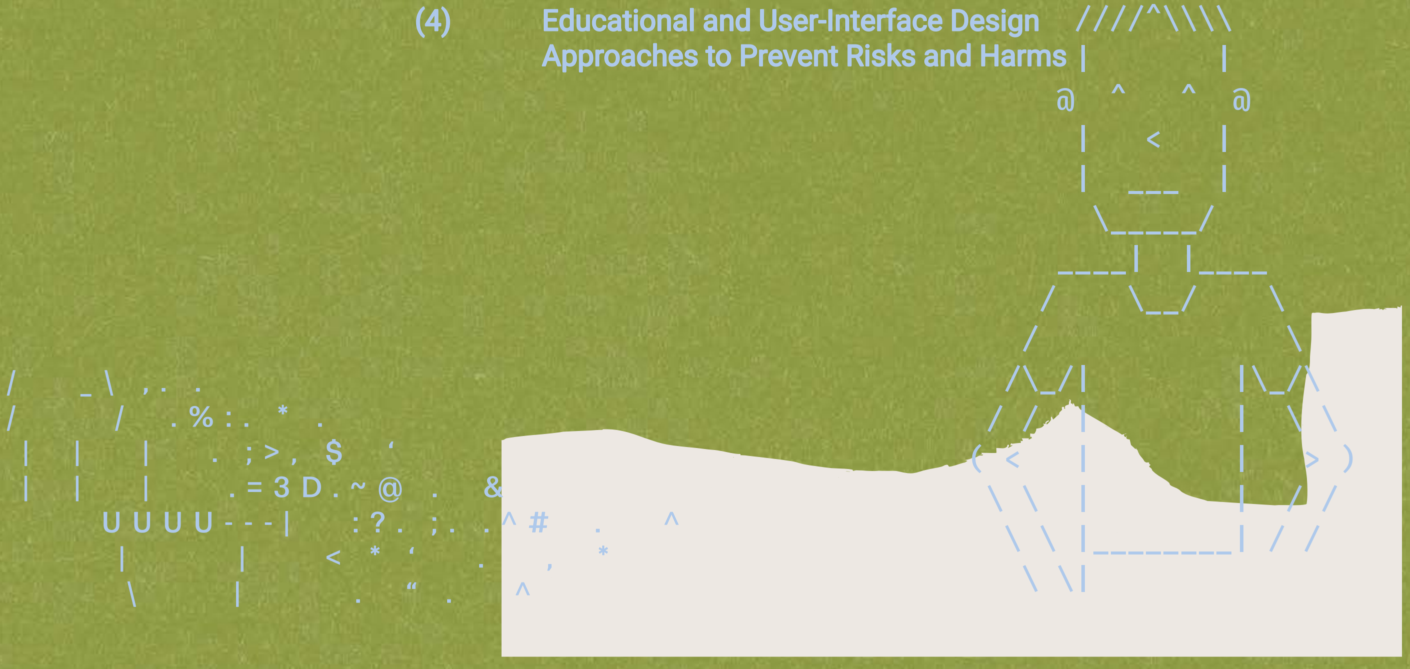
(III)

(1) Design Approaches that Foster Trust

(2) Help-Seeking and Reporting Approaches

(3) On-Device Approaches to Intervene When Conduct and Contact Risks Occur

(4) Educational and User-Interface Design Approaches to Prevent Risks and Harms



Digital child safety is a shared priority, yet reaching consensus on the best ways for how to address it remains a challenge, given competing business models, shifting policies, and evolving user behaviors. Understanding the ever-changing landscape of risks and harms is an important step toward developing effective social, technical, legal, and educational interventions. A nuanced, research-driven approach – Child Safety as a Design Opportunity – can help map opportunities for intervention that adapt to risks and harms in the digital ecosystem.

The Frontiers of Digital Child Safety project complements existing child protection efforts by mapping key knowledge areas and broadening the collective understanding of children's experiences at the intersection of safety and the digital environment.

Through four core approaches and guiding questions, the Working Group charts a proactive path forward exploring how design can foster trust, support help-seeking and reporting, embed interventions into devices when risks arise, and develop educational and user-interface strategies to reduce risks and mitigate harms.

APPROACH	DESCRIPTION	GUIDING QUESTIONS:
1. Design Approaches that Foster Trust	Designers and developers have incorporated various features into digital technologies that are intended to promote user's trust in the technology itself and the ecosystem in which it is embedded. Nonetheless, numerous implementation questions remain open at the operational level of such trust and privacy-enhancing features and tools.	• How should child-protective features approach user enablement and disablement of those features, whether by the parent/caregiver, child or both? • What are the potentially negative consequences of surfacing interventions or warnings to children, and how can those be mitigated?
2. Help-Seeking and Reporting Approaches	Children who have encountered content- or contact-based risks are often reluctant to confide in adults. Technology-based interventions can help lower these barriers, empowering children to seek help while enabling peers and adults to support them in navigating challenging situations.	• When and why are children most likely to ask for help? • How can technology interventions be the most helpful? • How can technology support the role of friends in a child's life when dealing with challenging situations? • How can technology support the role of parents/caregivers, teachers and other supportive adults in a child's life when dealing with challenging situations? • What improvements in reporting methods to child safety organizations and law enforcement would most increase the efficacy of potential child exploitation reports?
3. On-Device Approaches to Intervene When Conduct and Contact Risks Occur	Recent efforts to enhance child safety include on-device approaches that acknowledge smartphones as primary access points among children. These developments raise crucial questions about how to design proactive measures that safeguard child users while respecting their privacy.	• What are optimal on-device approaches to act upon and prevent children from acting in a way that contributes to risky digital content or contact? • What are optimal on-device approaches to interrupt instances where a child is a victim of (or recipient) of an action? • How can it be assured that such approaches preserve the privacy of children and users?
4. Educational and User-Interface Design Approaches to Prevent Risks and Harms	Educational approaches may play a crucial role in the interventions designed to mitigate possible risks and harms building on children's strengths, and user-interface design choices may contribute to safer engagement.	• What educational methods are most effective to help children identify and seek help with specific risks they face? • How should user interfaces, including language, be customized to best speak to different groups (for instance, in terms of age, maturity, or circumstances) of children?

Approach 1: Design Approaches That Foster Trust

```
                      \ | /
                     -- * --
                      / | \
       /’)    ./’)             (‘\.    (‘\
     /’ /.--’’./’’)           (‘’\.’’--.\ ‘\
:--’’  ;    ‘’./’’)           (‘’\.’’    ;  ‘’--:
:     ‘     ‘’./’)             (‘\.’’     ‘     :
:           ‘’./’               ‘\.’’           :
:--’’-..--’’’’                     ‘’’’--..-’’--:
```

Overview

Main Concepts and Definitions

- Child-Protective Features: Features in digital technologies intended to identify and mitigate risks and harms to children. These include tools such as content filters, usage controls, and warning systems. When well designed, they help foster trust while supporting children's rights, agency, and well-being.
- A New Collaborative Approach: Shifting from parent/caregiver-centric control toward joint decision-making involving both parents/caregivers and children, thereby enhancing children's rights, agency, and trust in protective features.
- The "Forbidden Fruit" Effect: A phenomenon where warnings or restrictions inadvertently increase a child's curiosity and interest toward prohibited or restricted content, counteracting the intended protective aim.

Key Findings

- Trust-Building Design: Evidence strongly supports the shift from restrictive parental controls to collaborative approaches that include children's input, showing significant potential to enhance trust.
- Negative Consequences of Overly Restrictive Interventions: Strong evidence suggests that restrictive and unilateral parental controls can erode trust, diminish children's rights and agency, and motivate circumvention behaviors.
- Effectiveness of Positive Reinforcement: Evidence consistently supports designing child-protective features around positive reinforcement, subtle behavioral nudges, and context-sensitive guidance to effectively encourage safe online behavior without causing guilt, shame, or anxiety.
- Lessons from Other Domains: Warnings tend to be more effective when tailored to children's diverse identities and context, use clear and emotionally resonant language, and are visually prominent. Emotional cues can enhance impact, and designs should be regularly updated to prevent habituation. While not specific to digital child safety, these design principles might be adapted and tested for digital contexts.

Questions for Further Research

- Children's Experiences and Reactions: There's a significant lack of research on how children perceive and respond to digital protective features, pointing to an important research gap regarding child-centered design effectiveness.
- Long-Term Effectiveness and Habituation: Research on how the effectiveness of warnings and interventions evolves over time, particularly considering habituation and evolving online risks, remains limited and requires further exploration.
- Cross-Domain Knowledge Transfer: Open questions remain regarding how effectively insights from non-digital domains (like product warnings or public health) translate to digital child safety contexts, requiring targeted validation and research.

Design Approaches That Foster Trust

Designers and developers incorporate various trust-enhancing features into digital technologies in an effort to create a safer digital environment. Yet, questions remain about how to implement these features in ways that are both effective and meaningful (Zieglmeier & Lehene, 2022). While protection is a crucial component of building trust — and this section focuses primarily on child-protective features as one way design can operationalize trust — it is important to situate it within a broader framework. Trust-building efforts should also account for provision, which ensures access to digital experiences and participation and supports children's ability to actively engage, express themselves, and be heard in the digital environment.

Child-protective features aim to identify and mitigate risks and harms — ideally before a child or their parent/caregiver encounters them (National Telecommunications and Information Administration, 2024). Designing these tools requires a careful balance: ensuring children's access and participation, while clearly communicating what the features do and the outcomes they are intended to achieve (Wisniewski et al., 2017; Ghosh et al., 2018; Badillo-Urquiola et al., 2019; Stoilova et al., 2023). When done well, these features can empower children and their parents/caregivers to be actively involved in decision-making processes about children's digital experiences, thereby reinforcing their trust.

Approach 1 explores how design strategies can be used to foster trust in the digital environment, particularly through child-protective features. It examines the challenges of implementing these features in ways that are both effective and respectful of users' autonomy. Key questions include how control over such features should be shared between children and parents/caregivers, and how to minimize potential negative consequences from interventions like warnings or alerts. The approach emphasizes the importance of transparency, user agency, and thoughtful design in building lasting trust.

To the best of our knowledge, the initial literature review did not yield specific findings to directly address the guiding questions, as little research has explored children's experiences with and reactions to child-protective features. Therefore, we took the following approach: (1) review child protective features across selected applications to better understand their design, focusing on the affordances and constraints they offer children in managing their digital experiences, and (2) examine selected literature on historical examples of warnings from various domains, including offline contexts such as child-care product warnings, to translate relevant lessons into the digital environment. By synthesizing insights from these two areas — alongside working meetings and shared foundational knowledge — we applied our findings to the digital child safety domain. Given the limited literature in this field, our proposed responses are necessarily more speculative compared to other approaches in this report.

1.Child-Protective Features

How Should Child-Protective Features Approach User Enablement and Disablement of Those Features, Whether by the Parent/Caregiver, Child, or Both?

1.1 From Parental Control to Collaborative Empowerment

A crucial aspect of designing child-protective features lies in determining how, and by who, they should be enabled or disabled (Wang et al., 2021; Quayyum et al., 2021). The literature advocates for a paradigm shift in how child-protective features are enabled and disabled. A recurring theme is the need to move away from solely parent-centric models toward a more collaborative approach that includes children in the decision-making process. By giving children a sense of agency and predictability, we foster their trust in these features and make it more likely that they will use the tools as intended rather than finding ways to bypass them.

Current Landscape: Current child safety apps predominantly position parents/caregivers as the primary stakeholders, largely because they are both the main purchasers of these services and often legally responsible for their children. In the U.S., parents in nearly three-in-four households have child-protective features set on their children's devices (FOSI, 2025), underscoring just how central parental control is to safety management. As a result, parents/caregivers are frequently granted unilateral control over enabling and disabling child-protective features (Wang et al., 2021; Agha et al., 2025). While this approach may be well-intentioned, it can create an imbalance of power and risk undermining trust between children and their parents/caregivers (Akter et al., 2022; Stoilova et al., 2023; Theopilus et al., 2024). Moreover, it may lead to decisions that do not always align with the child's best interests, particularly when children's perspectives and evolving capacities are overlooked (Livingstone et al., 2024; Özkul et al., 2025).

Proposed Shift: Fostering trust calls for a more nuanced, child-centered approach. Rather than placing primary control over the enablement and disablement of child-protective features in the hands of parents/caregivers, a more collaborative model could engage children as active participants in shaping safer digital experiences (Badillo-Urquiola et al., 2019; Ghosh et al., 2018; Akter et al., 2022; Park et al., 2024; Stoilova et al., 2023; Wisniewski et al., 2017). Such an approach recognizes children's rights and agency, values their input and evolving understanding of risks, and offers a more promising strategy for building trust.

1.2 Strategies for Collaboration

To achieve this shift towards a more collaborative approach, several strategies can be implemented in the design of child-protective features.

Joint Decision-Making: Parents/caregivers and children could collaboratively configure safety settings and determine appropriate levels of protection. This participatory approach, which centers children, can promote open communication and build trust within the family by, for example, incorporating parent-child collaboration features (Badillo-Urquiola et al., 2019; Ghosh et al., 2018; Iftikhar et al., 2021; Lake et al., 2025; Park et al., 2024; Quayyum, 2025; Stoilova et al., 2023; Wisniewski et al., 2017).

Gradual Autonomy: Features could be designed to adapt as children mature, gradually granting them more control and freedom. This approach recognizes children's evolving capacities for responsible online behavior (Badillo-Urquiola et al., 2019; Gnanasekaran & De Moor, 2025; Park et al., 2024; Wisniewski et al., 2017).

Transparency and Education: Children are more likely to accept and adhere to safety measures when they understand the reasoning behind them (Badillo-Urquiola et al., 2019; Ghosh et al., 2018; Park et al., 2024; Wisniewski et al., 2017). Providing clear explanations about why specific features are enabled and how they function might prove essential for fostering digital skills and supporting children's rights, agency, and well-being. Research on users' perspectives on parental control has suggested that flexible parental control solutions designed to facilitate open communication and transparency are promising avenues for cultivating digital skills among children and parents/caregivers (Gnanasekaran & De Moor, 2025). Both parents/caregivers and children appreciate features that offer transparency regarding apps and device permissions, recognizing that such visibility enhances communication (Wang et al., 2021). For instance, an app could provide prompts for discussion topics based on user activities, alongside guidance for parents/caregivers and teens on how to raise sensitive issues or questions with one another (Akter et al., 2022).

By giving children a sense of agency and predictability, we foster their trust in these features and make it more likely that they will use the tools as intended rather than finding ways to bypass them.

2. Interventions or Warnings to Children and Design Mitigation Strategies

What Are the Potentially Negative Consequences of Surfacing Interventions or Warnings to Children, and How Can Those Be Mitigated?

Interventions and warnings may come with unintended pitfalls for children and parents/caregivers who use them, undermining their effectiveness in keeping children and adult users safe. We explore general potential pitfalls that may erode trust in these features and propose design mitigation strategies to overcome them. We also draw on best practices from other domains that can enrich how we design features for a safer digital ecosystem.

2.1 General Potential Pitfalls

While interventions and warnings aim to protect children, they can sometimes have unintended negative consequences. Ignoring these can undermine the effectiveness of child-protective features and erode the trust they are meant to foster. The literature identifies several potential pitfalls and proposes strategies for mitigating them.

The "Forbidden Fruit" Effect: Warnings that frame content as off-limits can pique a child's curiosity, potentially leading to increased interest in the restricted material (Bridgland et al., 2022; Stoilova et al., 2024). This mirrors findings in other domains, such as movie content warnings or screen time limits increasing intentions to consume the restricted media (Gunter, 2018; Prasad & Quinones, 2020). One way to address this is by moving beyond purely restrictive messaging. Engaging children in open conversations about online risks and age-appropriateness (Hasse et al., 2019; Walsh et al., 2024), and fostering critical thinking and AI skills (Bright et al., 2023; Kardefelt Winther et al., 2024) can help them develop their own internal filters and reduce the allure of the forbidden.

Guilt, Shame, and Anxiety: Warnings may provoke negative emotions such as guilt, shame, or anxiety (Prasad & Quinones, 2020). While the Broaden-and-Build Theory of Positive Emotions suggests that positive emotions can expand an individual's views and actions (Fredrickson, 2001) and may also encourage children to think of others and behave in prosocial ways (Stifter et al., 2020), negative emotional responses have been linked to problematic smartphone use, particularly among children (Yadav & Chakraborty, 2022). To avoid this, designers and developers can focus on positive reinforcement — rewarding healthy behaviors — and draw on contextual information, such as user mood or usage patterns, to determine whether and when to issue warnings. Visual cues and digital nudges can help guide positive behavior online (Veretilnykova & Dogruel, 2021). These tools make users aware of potential risks such as misinformation and sharing sensitive content, yet still allow them to make their own choices, giving them the option to ignore warnings (Alghythee et al., 2024; Gnanasekaran & De Moor, 2025). For instance, evidence suggests that using color-coded visual cues to mark information sources can encourage users to think more critically about what they consume on social media and reconsider content sharing, without turning into nagging, triggering discomfort or limiting autonomy (Gnanasekaran & De Moor, 2025; Mirbabaie et al., 2020; Prasad & Quinones, 2020; Stoilova et al., 2024).

Eroding Trust and Agency: Overly restrictive or opaque interventions may make children feel distrusted and controlled. This can strain relationships with parents/caregivers and lead to efforts to circumvent safety measures (Akter et al., 2022; Ghosh et al., 2018; Hashish et al., 2014; Stoilova et al., 2023). To mitigate these risks, designers should emphasize transparency and inclusion. Clearly explaining the purpose of safety settings, involving children in decision-making, and offering options that support their autonomy — such as requesting access to blocked content or negotiating screen time — can reinforce trust and encourage responsible engagement (Hasse et al., 2019; O'Reilly et al., 2022).

Hindering Digital Skills: Child-protective features can be reframed as opportunities for awareness-raising, open dialogue, and skill-building. These interactions can help children strengthen practical skills (e.g., safe device use), social-emotional skills (e.g., self-efficacy and communication), and cognitive skills (e.g., critical thinking andself-regulation) needed to to assess and respond to challenges on their own (Hasse et al., 2019; Hinduja & Patchin, 2017; O'Reilly et al., 2022). Attempting to shield children from all risks — an approach that is arguably not feasible — can impede their ability to develop resilience and hinder the development of key digital skills (Cortesi et al., 2021; Badillo-Urquiola et al., 2019; Wisniewski et al., 2017).

Design Flaws and Inconvenience: Technical issues, false positives, and poor usability can frustrate users and undermine trust in safety features, leading to their abandonment (Ghosh et al., 2018). Addressing this requires a commitment to accessible, user-friendly design. Features should be intuitive, reliable, and thoroughly tested for real-world use (Kuang & Fabricant, 2019). Regular feedback from children and parents/caregivers should inform ongoing improvements to ensure tools are not only functional but also trusted and adopted (Stoilova et al., 2023).

2.2 Insights from Other Safety Domains

As discussed, there is thin evidence on children's experiences with and reactions towards child-protective features. Therefore, we reviewed examples from other, primarily non-digital domains — including product and substance, physical, and media risks — to explore the use and effectiveness of warnings. Studies have shown that responses to physical warnings translate to the digital domain (Jeong & Chiasson, 2020). The review suggests that while warnings are crucial for products with unavoidable hazards, their efficacy hinges on their ability to communicate the appropriate level of danger and elicit desired safety behaviors (Flor et al., 2021; Jeong & Chiasson, 2020; Trommelen, 1997; Zaikina-Montgomery & Silver, 2018). The research emphasizes that warnings cannot replace good design, but can significantly enhance the safety of products with inherent risks (Trommelen, 1997).

In broad terms, our review of research from other domains suggests that effective and trustworthy warnings for children should be designed with specific principles in mind.

Tailored to the Audience: A range of identity-shaping factors — including age and evolving capacities, individual circumstances, education and digital skill level, gender, culture, socioeconomic background, prior experience, and information processing abilities — influence how individuals perceive and respond to warnings (Arrúa et al., 2017; Flor et al., 2021; Gunter, 2018; Hall et al., 2021; Hammond, 2011; Jeong & Chiasson, 2020; Lesch et al., 2016; Morgenstern et al., 2021; Morrongiello et al., 2016; Prasad & Quinones, 2020; Saavedra-Garcia et al., 2022; Sampson et al., 2001; Silic, 2016; Waterson & Monk, 2014; Zaikina-Montgomery & Silver, 2018). Therefore, these factors are crucial when planning interactions between children and parents/caregivers, particularly when children seek help regarding warnings (Flor et al., 2021; Gunter, 2018; Hammond, 2011; Jeong & Chiasson, 2020; Morgenstern et al., 2021; Prasad & Quinones, 2020; Saavedra-Garcia et al., 2022; Waterson & Monk, 2014; Zaikina-Montgomery & Silver, 2018).

Clear and Explicit: Clear instructions on safe product usage and potential hazards are crucial. Studies have shown that explicit warnings improve understanding and recall, especially for individuals with limited prior knowledge (Cabrera et al., 2017; Gunter, 2018; Hammond, 2011; Jeong & Chiasson, 2020; Mirbabaie et al., 2020; Morgenstern et al., 2021; Morrongiello et al., 2016; Ross et al., 2018; Trommelen, 1997; Waterson &Monk, 2014; Zaikina-Montgomery & Silver, 2018). Moreover, specific, short and simple, and vivid language enhances the perceived hazardousness of products and reduces ambiguity, prompting individuals to take warnings seriously (Jeong & Chiasson, 2020; Mirbabaie et al., 2020; Morrongiello et al., 2016; Ross et al., 2018; Waterson & Monk, 2014; Zaikina-Montgomery & Silver, 2018).

While interventions and warnings aim to protect children, they can sometimes have unintended negative consequences. Ignoring these can undermine the effectiveness of child-protective features and erode the trust they are meant to foster.

Design Shapes Perception of Likelihood and Severity: People’s responses to warnings are strongly influenced by how likely and how severe they believe the potential harm (e.g., injury) to be. Determining whether warnings are noticed, understood, and acted upon is crucial. Research shows that design-related factors — such as the appearance of a product or interface, the user’s sense of control, and their familiarity with the technology — can all shape perceived risk (Cabrera et al., 2017; Gunter, 2018; Morgenstern et al., 2021; Morrongiello et al., 2016; Ross et al., 2018; Trommelen, 1997; Waterson & Monk, 2014).

Emotional Appeal: Research supports that warnings are most effective when they generate emotional reactions (Hammond, 2011; Morgenstern et al., 2021; Mirbabaie et al., 2020; Ross et al., 2018). For example, messaging that produces fear, guilt, or shock, can enhance attention to the immediate situation, increase risk perception, and motivate individuals to alter harmful behaviors offline as well asonline (Baumgartner et al., 2021; Cho et al., 2018; Sidhu et al., 2022). Nonetheless, recent research shows gratitude’s potential to trigger cascades of positive effects in motivating behavior change, such as perseverance and prosocial behavior and fostering a focus on long-term relationships and health over immediate gratification (Schneider et al., 2021; Wang et al., 2024).

Prominence and Design: The size, placement, and design of warnings, including the use of contrasting colors and imagery, pop culture references, anthropomorphism, and multimedia elements, significantly impacts their noticeability and effectiveness (APAC Ad Junkie, 2019; Cabrera et al., 2017; Fernandes et al., 2015; Flor et al., 2021; Hall et al., 2021; Hammond, 2011; Jeong & Chiasson, 2020; Li et al., 2022; Liang & Park, 2023; Morgenstern et al., 2021; Mirbabaie et al., 2020; Schneider et al., 2019; Tait et al., 2015; Waterson &Monk, 2014; Zaikina-Montgomery & Silver, 2018). Therefore, visually appealing warnings that utilize bright colors, engaging characters, and simple multimedia imagery can help to capture attention and enhance understanding (Cabrera et al., 2017; Flor et al., 2021; Hammond, 2011; Jeong & Chiasson, 2020; Morgenstern et al., 2021; Mirbabaie et al., 2020; Waterson &Monk, 2014; Zaikina-Montgomery & Silver, 2018).

Iteration: Previous work from other domains emphasize the need for continuous evaluation and improvement of warning designs given habituation the evolving nature of risks and hazards, and the ever-changing state of user behaviors (Flor et al., 2021;Hammond, 2011; Trommelen, 1997; Waterson and Monk, 2014). As such, warnings need to be updated to continue to promote the desired behavior (Hammond, 2011; Ross et al., 2018; Zaikina-Montgomery & Silver, 2018). This can occur through different and dynamic formats, relatable storytelling interwoven throughout, and fresh design elements.

Absent specific research on the guiding questions, these insights from other domains can provide applicable suggestions for the digital child safety context, as outlined above. Again, these statements will need to be critically examined and challenged, given that these are extrapolations based on non-specific research from other domains.

```
  _          _          _          _
 //|\       //|\       //|\       //|\
 c_”/       c_”/       c_”/       c_”/
__/\__       /\__       /\         /\
 /   \      /\ \       /\/\       /\/\
/____\     /____\     /____\     /____\
  /l         ll         l\         ll
```

Evaluating Design Approaches That Foster Trust

Building on the research and practices reviewed in this chapter, the following discussion presents the Working Group's assessment of key concepts and findings related to design approaches that foster trust, with particular emphasis on child-protective features. At the core of this evaluation is the recognition that a safe digital environment must not only protect children effectively but also uphold and strengthen their rights, agency, and overall well-being.

The literature robustly supports a shift away from traditional, unilateral parental control models towards collaborative approaches, which involve joint decision-making by both parents/caregivers and children. This shift is based on substantial evidence indicating that excessively restrictive interventions can unintentionally erode trust, provoke resentment, and motivate children to bypass or circumvent safety measures. Conversely, a collaborative approach that includes children as active participants in configuring and managing safety features consistently results in better engagement and greater trust, ultimately enhancing the effectiveness of these interventions.

The role of positive reinforcement and transparency also emerges as critical, with substantial evidence indicating that clearly communicating the rationale behind child-protective features and rewarding positive digital behaviors significantly improves compliance and reduces negative emotional responses such as guilt, shame, or anxiety. However, the chapter highlights that the explicit implementation of behavioral nudges and adaptive, context-sensitive warnings remains relatively experimental. These nuanced strategies, though promising, have not yet been thoroughly validated in the digital context, signaling a critical need for rigorous empirical investigation.

Furthermore, the chapter identifies significant knowledge gaps, particularly in understanding children's direct experiences and reactions to digital safety warnings. Current literature provides modest insight into children's perceptions and emotional responses to these features, limiting our understanding of their overall impact. Additionally, the efficacy of transferring safety insights from offline contexts — such as warnings used in product or public health safety domains — to digital environments remains mostly speculative given the current state of research. Careful, context-specific evaluation is necessary before confidently integrating these lessons into digital tools. Finally, research on the long-term effectiveness of digital warnings, including potential habituation and diminishing effectiveness over time, remains notably underdeveloped, highlighting an important area for future longitudinal research.

In conclusion, although a collaborative approach and transparency emerge as fundamental baseline principles, the field urgently requires more comprehensive, context-specific research to refine and validate the effectiveness of interventions. Addressing these gaps, particularly through empirical studies of children's experiences, the nuanced impacts of different design approaches, and long-term efficacy, represents a critical frontier for advancing digital child safety.

Approach 2: Help-Seeking and Reporting Approaches

Overview

Main Concepts and Definitions

- Help-Seeking and Reporting Approaches: Technology interventions designed to lower barriers for children seeking help and support from peers and adults after encountering online risks or harms.
- Peer Support Networks: Digital platforms and moderated environments where children receive emotional support, reassurance, and guidance from peers or trained peer mentors, often acting as an initial trusted source before reaching out to adults.
- Anonymous Reporting Tools: Digital mechanisms allowing confidential disclosure of online risks and harms, encouraging reporting without fear of stigma or retaliation.

Key Findings

- Importance of Emotional Distress and Perceived Severity: Strong evidence indicates children are most likely to seek help immediately following incidents causing significant emotional distress or perceived as highly severe.
- Effectiveness of Anonymous Reporting: Evidence supports the effectiveness of anonymity in reporting tools, which significantly increases the likelihood that children and others will report incidents without fear of retaliation or judgment.
- Influence of Peer Networks: Clear evidence shows peer influence significantly impacts children's willingness to seek help, with peer encouragement often facilitating further help-seeking behaviors from adults.
- Critical Role of User-Friendly Interfaces: High-confidence evidence supports the design of simple, intuitive, and accessible reporting interfaces to ensure effective use and higher reporting rates.
- Technology Supporting Adult Intervention: Emerging evidence suggests that communication tools and educational resources can significantly enhance adults' capacity to support children navigating online risks; however, empirical research evaluating their long-term impact and ethical implications remains limited.

Questions for Further Research

- Children's Perceptions and Long-Term Effects: There is a notable gap regarding children's own perceptions and emotional experiences with help-seeking tools, emphasizing a critical need for longitudinal studies on long-term efficacy and emotional impacts.
- Cross-Domain and Cultural Applicability: Research remains sparse on the transferability of insights from offline or other non-digital domains into the digital reporting contexts, particularly considering diverse cultural and socio-demographic contexts.
- Ethical Implications of AI-driven Monitoring: The accuracy, potential biases, and ethical implications of AI-based monitoring tools remain underexplored, especially concerning false positives and their psychological and relational impacts.

1. Factors of Influence: When and Why Children Seek Help

When and Why Are Children Most Likely to Ask for Help?

Current research suggests several factors that influence both temporally and situationally when a young person may reach out for help when facing content or contact risks online (OECD, 2020a). The research does not clearly explicate a hierarchy of likelihood in given scenarios, but does offer evidence that the following indicators are most likely to motivate a young person to ask for help.

Emotional Distress: Children are more likely to seek help immediately after an incident if they are experiencing intense emotions such as fear, anxiety, or sadness. A systematic review highlights that children often seek help when they perceive a high level of emotional distress, which can be linked to immediate incidents that trigger strong emotional responses (Pekárková et al., 2022; Reardon et al., 2017).

Perceived Severity: Children who perceive a situation as highly dangerous or severe are more likely to seek help. Research on post-traumatic stress disorder (PTSD) in children indicates that those who experience or witness severe traumatic events are more likely to seek help due to the intense impact of such experiences (Boston Children's Hospital, 2023).

Physical Evidence: When there is physical evidence of an incident, such as cyberbullying screenshots or screen recordings, children may feel an urgent need to share it with someone. This urgency is often driven by the tangible nature of the evidence, which can heighten the child's awareness of the incident's severity and the need for intervention (World Health Organization, 2025; Dempsey et al., 2011; Fenaughty & Harre, 2013).

Repeated Incidents: Children who recognize that they are experiencing a pattern of harmful incidents are more likely to seek help. This is supported by research showing that repeated exposure to trauma or distress can lead to increased awareness and a stronger motivation to seek assistance to stop the pattern (Boston Children's Hospital, 2023).

After Peer Influences: Friends who have had positive experiences with seeking help can significantly influence their peers to do the same. The impact of peer encouragement is well-documented, as children often turn to their social circles for advice and support (Reardon et al., 2017). Children might confide in friends first before approaching adults or using technology-based tools. This initial step of sharing with peers can build the confidence needed to seek further help from adults or professionals.

After Educational Interventions: Programs in schools or communities that educate children about risks and harms and encourage help-seeking can prompt children to reach out. Such interventions are crucial in raising awareness and providing the necessary tools for children to seek help when needed (Hasse et al., 2019; Reardon et al., 2017).

2. How Technology Interventions Can Be the Most Helpful

How Can Technology Interventions Be the Most Helpful?

Our review of the research suggests that multiple factors should be considered when designing technology interventions to support help-seeking.

Accessibility and Ease of Use: Ensuring accessibility and ease of use is essential for fostering an environment where individuals can seek help or report issues promptly. Digital platforms should offer user-friendly interfaces and be operational around the clock, recognizing that distressing incidents may occur at any time (Fenaughty & Harre, 2013; (Livingstone et al., 2012).

Automated Alerts and Monitoring: Leveraging AI for automated alerts and monitoring can enable timely interventions. For example, features like Apple's Communication Safety detect when an incoming image may contain nudity, blur it, and display a prompt to the user — offering a moment of reflection and an opportunity to seek support. Similarly, parental controls, implemented with the child's consent, help oversee activities without compromising trust. These strategies aim to ensure a balanced approach, maintaining privacy while prioritizing safety (Garaigordobil & Machimbarrena, 2017).

Building Support Networks and Communities: Establishing support networks and communities can foster a sense of belonging and mutual support. Safe digital communities and mentorship programs can connect children with trained mentors, offering guidance and reassurance (UNICEF, n.d.; Dempsey et al., 2011). Additionally, incorporating crisis intervention features, such as emergency contacts and real-time counselor support, ensures immediate assistance when needed (Boston Children's Hospital, 2023).

Anonymity and Privacy: Safeguarding anonymity and privacy is crucial, with resources like anonymous reporting apps and confidential online counseling services enabling children to seek assistance without fear of judgment or retaliation (Cross et al., 2015; Pekárková et al., 2022).

Educational Resources and Interactive Content: By providing access to engaging educational resources and interactive content, children can build both the knowledge and skills needed to recognize, interpret, and respond to risks and harms. Tools such as information hubs and gamified learning modules not only raise awareness but can also foster coping strategies and digital skills (Reardon et al., 2017). For further reading, see chapter IV, section 2.

3. How Technology Can Support Friendships in Helping Children Navigate Challenges

How Can Technology Support the Role of Friends in a Child’s Life When Dealing with Challenging Situations?

Recent research underscores the significant role of technology in facilitating the support that friends can provide to children facing challenging situations. Below, we outline some existing technology, its strengths and weaknesses, and general findings from the current literature. Many of the popular technologies currently used by children, however, take the form of apps or online tools. Their helpful components can serve as a starting point for discussion when assessing their relevance and potential for inclusion when considering solutions for user-interfaces or on-device approaches.

Anonymous Sharing Platforms: Anonymous sharing platforms can enable children to disclose their experiences without the stigma often associated with seeking help. Anonymity has been shown to enable young people to share more openly, allowing peers and professionals to recognize and address issues more effectively (Livingstone & Smith, 2014). Anonymous school safety tip lines are another example of how anonymity can serve as a critical mechanism for surfacing concerns, helping schools identify and respond to issues such as bullying or potential harm (Pfefferkorn et. al, 2025). However, general-purpose anonymous platforms can also foster environments where cyberbullying and toxic content emerge. This dual nature highlights a key challenge: the same features that encourage openness can also lead to higher risk and harm. Examples (selection): Project YES offers anonymous, self-guided interventions tailored for youth mental health, while Togetherall provides a moderated, anonymous online community for peer support.

Peer Support Networks: Moderated environments where young people can connect with trained peers and mental health professionals can foster emotional support and community. These platforms demonstrate scalable, youth-centered approaches to mental health support, emphasizing accessibility, peer connection, and community engagement. Examples (selection): Crisis Text Line offers free, 24/7 text-based support from trained volunteer counselors, providing immediate assistance to individuals in crisis. Headspace, Australia’s National Youth Mental Health Foundation, delivers integrated online and in-person services for individuals aged 12–25, including peer-led group chats and one-on-one support. Sanvello combines cognitive behavioral therapy tools with community discussion boards and live coaching sessions, helping users manage stress, anxiety, and depression. Calm Collective Asia facilitates culturally sensitive peer support through its Calm Circles program, creating safe spaces for open conversations among youth..In Germany, JUUUPORT is a free of charge nationwide online counseling platform where trained teen and young adult volunteers provide confidential, data-protection-compliant support for youth facing online challenges such as cyberbullying, media addiction, and sexual harassment.

Private Messaging Apps and Moderated Online Spaces: Private messaging apps featuring end-to-end encryption (E2EE), such as Signal, WhatsApp, and iMessage, can help children maintain confidentiality when discussing sensitive or distressing topics and seeking peer support (Jones et al., 2013). However, these platforms also pose significant risks. E2EE messaging apps can inadvertently facilitate harmful interactions, such as the grooming or exploitation of children, or serve as channels for adults sharing child sexual abuse material (CSAM) (OECD, 2023). Moderated online spaces — such as supervised private group chats on platforms like Discord or monitored Facebook groups — may offer safer environments, but risks persist, including groomers directing young users from discoverable platforms toward encrypted apps where harmful interactions become harder to detect. Therefore, the beneficial privacy provided by encrypted platforms should be accompanied by robust education, parental/caregiver guidance, and informed intervention strategies to mitigate associated risks (Reich et al., 2012).

Educational Resources and Interactive Content: Research is beginning to show that some interactive digital tools can help children navigate personal and social challenges together (Calvin et al., 2024). Apps simulating situations like bullying and online harassment may allow friends to collaboratively practice empathy and supportive strategies (Cortesi et al., 2021). Examples (selection): Mindfulness applications tailored for children, such as Smiling Mind and Breathe, Think, Do with Sesame, enable peers to build shared emotional resilience and stress-management techniques and are already showing promising results (Dix, 2024). Additionally, tools like ReThink, which prompt reconsideration of harmful messages, and digital social-emotional programs like FRIENDS, can help friends develop stronger communication and emotional skills. However, continued research remains essential to determine the long-term impact of these digital supports (Livingstone & Smith, 2014; Slonje et al., 2013).

Confidential Peer-to-Peer Mentoring: Research by Yeo et al. (2023) found that digital peer support platforms significantly improved users' mental health and sense of connection, especially when confidentiality and user trust were carefully maintained. Similarly, Douglas et al. (2019) demonstrated that peer mentoring among at-risk youth enhanced emotional processing and self-worth, particularly when framed within trauma-informed digital communities. Furthermore, online models of support have shown promise beyond adolescence; Shah et al. (2022) reported that empathetic peer interactions on platforms like 7Cups fostered user engagement and reduced isolation. Examples (selection): MentorHub and newer interventions such as the "Mood" app offer structured environments for children to share experiences, gain guidance, and build coping strategies (Estwick, 2025).

4. How Technology Can Support Adults in Helping Children Navigate Online Challenges

How Can Technology Facilitate the Role of Parents/Caregivers, Teachers and Other Supportive Adults in a Child's Life When Dealing With Challenging Situations?

Technology can be designed and deployed to empower adults — such as parents/caregivers, educators, and mentors — in supporting children when dealing with challenging situations, as some of the reviewed sources suggest. At the same time, Working Group members pointed out that the same technology, if misused by abusive or overly controlling adults, can enable surveillance and restriction rather than genuine support.

There are a number of technologies designed for adults that enable them to coordinate, monitor, or learn how to navigate digital child safety — often operating from a more top-down perspective, including the following.

Communication Channels: Technology that supports communication between parents/caregivers – such as apps and virtual meeting platforms – can enhance the frequency and quality of interaction, allowing parents/caregivers and educators to collaborate on addressing children's safety (Kraft & Bolves, 2021; Olmstead, 2013). Open communication can support early identification of issues and timely intervention, preventing problems from escalating (Chicote-Beato et al., 2024). However, practicing effective communication requires a strong foundation of trust among parents/caregivers, teachers and other adults — something that is not universally present (Bordalba & Bochaca, 2019). Meanwhile, the growing trend of technology bans at school can also unintentionally discourage parent/caregivers reaching out to educators — not only due to fear of disciplinary action (Tavernise, 2024), but because these bans may signal to parents/caregivers that digital engagement is inherently negative or unwelcome in an educational setting. Examples (selection): Apps such as ClassDojo, Remind, and Bloomz provide streamlined channels for sharing updates, raising concerns, and discussing a child's emotional or educational needs. These tools are particularly valued for facilitating regular, low-barrier communication, which can be especially beneficial for early identification of challenges and coordinated intervention (Livingstone & Smith, 2014). Additionally, virtual meeting platforms like Zoom and Microsoft Teams have expanded access to parent/caregiver-teacher conferences, allowing for more timely and flexible conversations regardless of scheduling or geographic constraints. These technologies can reduce logistical barriers for working parents/caregivers and offer new opportunities for consistent engagement in a child's school experience (Dowling & Rickwood, 2013).

Monitoring Tools: Real-time monitoring tools and alert systems are used by parents/caregivers and schools to detect and respond to online risks children may face, including cyberbullying or exposure to harmful content (Chan et al., 2025). In school settings, these systems may be used to flag concerning activity and alert staff, while at home, monitoring apps allow parents/caregivers to track screen time, location, and digital communication. However, their effectiveness and ethical acceptability depend heavily on how they are implemented and by whom. Excessive or poorly explained monitoring may compromise children's sense of privacy and trust, particularly if it signals that they are not trusted — potentially harming the parent/caregiver–child relationship (Slonje et al., 2013). Additionally, the use of these tools remains controversial in many regions. In parts of Europe, for example, strict data protection laws — including in Germany — may prohibit school-based monitoring due to concerns about student surveillance and the potential for rights violations. These tensions underscore the need for context-sensitive approaches that balance safety, privacy, and trust. Examples (selection): Bark uses AI to analyze children's online activity — such as social media interactions, messages, and searches — for signs of cyberbullying, self-harm, or emotional distress. Tools like GoGuardian and Lightspeed Systems are used to monitor students' digital activity, potentially identifying early signs of self-harm, bullying, or emotional distress through behavioral analytics and emotion recognition tools (Collins et al., 2021). Other tools like Qustodio and Norton Family rely primarily on rule-based monitoring systems to flag predefined behaviors and content (Livingstone & Smith, 2014). These tools can issue alerts to parents/caregivers, allowing for timely intervention.

Educational Resources: Digital platforms offering webinars and courses can equip parents/caregivers, teachers and other supportive adults with the knowledge and skills necessary to promote children's digital safety (Dowling & Rickwood, 2013). While these resources can offer greater flexibility compared to in-person events, their effectiveness also depends on adults' ability to engage — something that can be constrained by time, stress, and competing responsibilities. Even when resources are accessible, practical barriers such as caregiving duties, work schedules, or digital literacy may limit participation. This suggests the need for ongoing consideration of how educational content is delivered and supported. Examples (selection): A range of websites and organizations — such as Common Sense Media, Youth and Media, NetSmartz, and Children and Screens — offer digital safety resources for parents/caregivers and educators. Resource hubs like Cyberbullying.org, StopBullying.gov, and the Child Mind Institute, provide practical tools, research, and guidance for responding to both online and offline challenges children may face. Additional initiatives such as the Digital Wellness Lab offer family-focused guides to help foster healthy digital habits. While many of these efforts are U.S.-based, valuable international resources also exist, including the Safer Internet Centers in Europe (e.g., klicksafe.de, BEE SECURE in Luxembourg), the UK's NSPCC, Germany's Gutes Aufwachsen mit Medien, and multilingual platforms like elternguide.online. These digital platforms enable adults across different contexts to engage with up-to-date, expert-backed guidance, often at their own pace and from the convenience of home.

There are also technologies designed to facilitate engagement between adults and children, often incorporating more collaborative approaches in addressing digital child safety.

Anonymous Reporting Tools: Anonymous reporting tools allow children to disclose issues that arise on or offline, across apps, social media, or physical school spaces. By routing alerts to designated adults, they may help flag risks early and potentially prevent harm. However, their effectiveness appears to depend heavily on factors such as the responsiveness of adults, school climate, and students' perceptions of trust and confidentiality (Payne et al., 2024). Although marketed as anonymous, many of these systems collect metadata such as IP addresses, which law enforcement or schools could use to identify the reporter — raising concerns about the limits of confidentiality (CISA & USSS, 2021). Furthermore, research and real-world accounts suggest that some tips may be dismissed or mishandled, leading to missed interventions and eroding student trust in the system (Pfefferkorn et. al, 2025). Examples (selection): Stop!t, SpeakUp, Safe2Tell, and TipTxt offer students a discreet way to report incidents of bullying (Messman et al., 2024).

Support and Counseling Services: Online support and counseling platforms may offer an important source of help for children, especially in communities underserved by traditional services (Asare et al., 2023; Dwyer et al., 2021). Integrating features like emergency contacts or crisis hotlines could enhance their safety and responsiveness, though these functions are still being refined (Jones et al., 2012). While these tools expand access, concerns about uneven quality persist. For example, a 2024 report found that most teens rated their experiences as only moderately helpful, with youth of color more likely to use these services — raising questions about accessibility, relevance, and equity (Calvin et al., 2024). Ensuring that expanded access does not come at the cost of quality remains a key challenge. Examples (selection): Services such as BetterHelp, Teen Counseling, and Talkspace may offer valuable support options for children seeking mental health resources (Asare et al., 2023; Dwyer et al., 2021). These services could be more accessible when parents/caregivers or educators help facilitate initial engagement, especially for younger users (Karcher, 2009). In parallel, apps like Sanvello and Calm are being explored in school settings as tools for managing stress and anxiety, often in coordination with teachers or school counselors.

Interactive and Engaging Tools: Interactive digital tools — particularly those that incorporate storytelling and gamification — can offer creative ways for children to explore and communicate their online experiences, while also involving adults in meaningful conversations. Examples (selection): Platforms like MyLife (formerly Stop, Breathe & Think) and Storybird encourage emotional expression through guided journaling, digital storytelling, and reflection activities. These can be shared with parents/caregivers or educators, helping trusted adults better understand what children are feeling or facing online. Similarly, gamified programs such as ReThink (David & Fodor, 2023) — which prompts users to pause before posting potentially harmful messages —, aim to build digital empathy, kindness, and critical thinking. While these tools are engaging and accessible, their effectiveness is likely to be enhanced when paired with strong pedagogical support that encourages deeper discussion and contextual understanding. Adults who engage with children during or after these experiences — whether at home or in the classroom — can help translate digital lessons into real-world values and coping strategies.

5. Reporting Methods for Effective Response to Child Exploitation

What Improvements in Reporting Methods to Child Safety Organizations and Law Enforcement Would Most Increase the Efficacy of Potential Child Exploitation Reports?

Recent research highlights several key areas where improvements in reporting methods to child safety organizations and law enforcement could significantly enhance the efficacy of child exploitation reports. The advancements in technology-based reporting systems carry the potential to address many existing challenges, although gaps remain that necessitate further exploration and expert input.

Simplified Reporting Processes: Designing reporting tools with intuitive, user-friendly interfaces is essential for ensuring that children and concerned individuals can report incidents easily. The usability of these tools is paramount, as complex process, and a lack of training of the educators or other adults on the other end of such tools could potentially deter reporting (Pfefferkorn, 2025). Research underscores the importance of simplicity and accessibility in designing these interfaces, as the difficulty of triaging the volume of reports received, which vary in quality, can limit law enforcement's ability to respond (Grossman et al., 2024) Additionally, providing multilingual support ensures that non-English speakers can report incidents effectively, thus broadening the reach and inclusivity of reporting systems.

Enhanced Data Collection and Analysis: Detailed reporting forms that capture comprehensive information about the incident, including the nature of the harm and any available evidence, are essential for effective data collection (Coiera, 2018). However, as mentioned throughout this report, given the sensitivity of the data collected, preserving privacy should remain an important consideration. The implementation of AI and machine learning tools is developing quickly, and could help analyze reports for patterns, prioritize cases based on severity, and quickly identify high-risk situations (Livingstone et al., 2018). Despite these opportunities, concerns about potential biases and the accuracy of AI-driven tools highlight the need for further research as these technologies expand in both capability and application.

Anonymous Reporting Options: Allowing anonymous reporting can protect the identity of the reporter, which encourages more individuals to come forward without fear of retaliation or exposure. . Recent studies have found that secure communication channels and encryption are crucial for maintaining the confidentiality and safety of reporters (Petronio, 2002; Thorn, 2024b). However, while anonymity can promote disclosure, it also presents challenges — particularly the risk of abusive reporting, where individuals misuse the system to harass or falsely accuse others. Further research is needed to explore how to balance the benefits of anonymity with safeguards that ensure accountability and enable appropriate follow-up in investigations.

Integrated Reporting Systems: Centralized platforms that collate reports from social media, email, direct submissions, and other sources streamline the reporting process — reducing the risk of oversight and improving data coordination across agencies. Cross-agency collaboration through these integrated systems is vital for responding effectively to online child exploitation. Recent initiatives illustrate this: Lantern, launched by the Technology Coalition in 2023, enables secure cross-platform signal sharing (e.g., hashes, usernames, URLs) among participating companies — enhancing detection, linking fragmented indicators, and accelerating reporting to law enforcement, with over 1 million signals shared and 7,000 CSAM items removed in 2024. Similarly, Stanford Internet Observatory's 2024 analysis highlighted improvements needed in the U.S. CyberTipline system, noting that better integration of triage tools across platforms would significantly reduce the burden on law enforcement (Grossman et al., 2024).

Real-Time Reporting and Response: Real-time capabilities — such as live chat support — can improve crisis responsiveness by enabling immediate interaction and ensuring 24/7 availability (Livingstone & Smith, 2014). However, real-time environments like livestreams can also facilitate harassment and trolling, which complicates moderation efforts (Gorwa & Thakur, 2025). A recent study analyzing over 57,000 chat messages across more than 8,000 safety incidents found significant inconsistency in the quality of emotional support; it suggested that LLM-powered tools might help standardize responses (Liu et al., 2024). Yet, questions remain around system reliability, escalation protocols, and the comparative effectiveness of automated versus human-led interventions. Further research is needed to understand how such systems can balance immediacy with user safety, particularly in child-specific contexts.

Feedback and Follow-Up: Providing feedback and follow-up to reporters is important for maintaining transparency and trust. Regular updates about the status of their report can reassure reporters that their concerns are being taken seriously. Informing reporters about the outcomes of investigations, when appropriate, can further enhance trust, while maintaining confidentiality and privacy requirements (Berson & Berson, 2005). Additionally, research has found that victims often feel re-violated due to the insensitive or inadequate response of those who are supposed to come to their aid when they do report (Campbell & Raja, 1999; Patterson, 2011). This occurs when platforms fail to recognize the gravity of the offense or display empathy toward the victim's experience. Infrequent contact or incomplete follow-up with the victim also can produce high levels of uncertainty and a deep lack of trust (Wemmers, 2002; Wemmers, 2013), which can result in the victim backing out of the case and choosing not to report any future incidents

Public Awareness Campaigns: Public awareness campaigns can play an important role in educating communities about the importance of reporting child exploitation and how to do so effectively. Leveraging digital platforms — such as social media — can help promote reporting tools and encourage users to speak up about harmful content (Livingstone & Haddon, 2018). Campaigns that emphasize users' digital rights can be especially impactful. For example, the HateAid initiative Unser Internet highlighted individuals' rights not to be exposed to abusive content, including unsolicited explicit images, while raising awareness of how to assert and protect those rights online (HateAid, 2024).

Enhanced Support for Victims: Developing reporting systems that prioritize the needs and safety of victims, and providing direct links to support services, including counseling, medical assistance, and legal help, are crucial for ensuring that victims receive appropriate support throughout the reporting and investigation process (Finkelhor, 2008).

Collaboration with Technology Companies: Collaboration with technology companies can play an important role in improving how child exploitation is detected and reported. These companies have access to data and technical expertise that may support the development of automated tools to flag and escalate suspected content to relevant authorities However, enabling direct reporting to law enforcement — whether by technology companies or users — raises a number of complex questions. As the OECD (2024) points out, such mechanisms must be carefully designed to avoid risks such as over-reporting, misclassification, or inadvertently penalizing vulnerable users. Concerns about privacy, due process, and transparency are also central. The OECD further highlights the need for robust governance and oversight, particularly where automated systems are involved, and calls for safety-by-design approaches that account for children's rights and agency, and involve them in the process of design from the outset. Additional OECD research also notes that transparency reporting on child sexual exploitation and abuse (CSEA) remains inconsistent across the tech sector, making it difficult to evaluate how technology companies handle reports, what actions are taken, or how effective those actions are (OECD, 2023). In light of these gaps, it may be worth exploring how digital reporting systems can be complemented by stronger connections to offline support — such as schools, counselors, and community justice services — to ensure responses are contextual, proportionate, and oriented toward child well-being.

Evaluating Help-Seeking and Reporting Approaches

While assessing the research and practices presented in this section related to help-seeking and reporting approaches in children's digital safety, the Working Group emphasizes that effective digital interventions require careful consideration of the various and often interacting psychological, social, and practical obstacles children encounter when deciding to report online harms, such as emotional reluctance, privacy fears, and issues of trust.

Current research consistently underscores key factors influencing children's help-seeking behaviors, notably emotional distress, perceived severity, and peer influence. Strong evidence indicates children are more inclined to seek immediate assistance when faced with highly distressing incidents or scenarios perceived as significantly harmful. Moreover, anonymity in digital reporting systems emerges as an especially effective feature, strongly supported by evidence showing it reduces children's fear of stigma, judgment, or retaliation, thus significantly improving reporting rates and enabling timely and effective responses. Additional features that significantly enhance the effectiveness of these tools include ensuring accessibility and ease of use, availability of educational resources that provide practical guidance and coping strategies, and integrating crisis intervention features for immediate assistance.

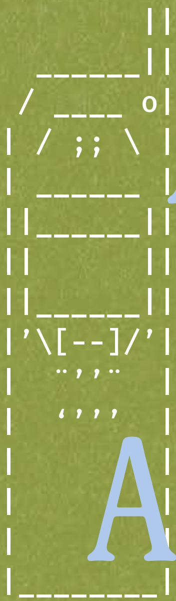
Peer support networks, facilitated by digital platforms and moderated online environments, also represent an important and promising approach. Peers frequently serve as initial confidants, offering children trusted initial points of contact that help bridge the gap toward seeking further help from adults or professionals. Furthermore, robust evidence highlights the critical role of intuitive, accessible user interfaces, underscoring the necessity of user-friendly designs to promote consistent and effective usage of help-seeking and reporting tools.

Technology also offers substantial potential for supporting adults, such as parents/caregivers and educators, in effectively responding to children's online challenges. Digital interventions like real-time monitoring systems, parental controls, communication platforms facilitating direct parent/caregiver-teacher interactions, and targeted educational resources have shown promise. However, while these tools demonstrate considerable practical value, current empirical evidence evaluating their long-term impact and addressing ethical concerns related to surveillance and privacy remains relatively limited and calls for further investigation.

Additionally, findings suggest several possible improvements in reporting methods for responding effectively to child exploitation. Simplifying and streamlining reporting processes through intuitive interfaces, providing multilingual and anonymous reporting options, and integrating real-time support features are among the recommendations strongly supported by existing evidence. Effective collaboration with technology companies and law enforcement is also highlighted as essential for improving the efficiency and responsiveness of reporting systems. Nevertheless, continued empirical research is necessary to validate the efficacy and practical implementation of these improvements.

Despite these positive findings, significant research gaps and complexities remain. Direct insights into children's personal experiences, perceptions, and emotional responses toward digital safety interventions are notably limited, creating uncertainty around their psychological and relational impacts. Additionally, transferring insights from offline domains — such as traditional public health warnings — to digital contexts, particularly across diverse cultural and socio-demographic groups, requires careful validation and further empirical investigation. Ethical considerations related to AI-driven monitoring, including the accuracy, risks of false positives, potential biases, and psychological impact on children and families, remain an essential but relatively under-explored frontier.

Addressing these critical knowledge gaps, specifically children's direct experiences, ethical implications of automated systems, and cross-contextual applicability, represents crucial areas for future research. Continued empirical work in these domains will significantly enhance the effectiveness, fairness, and appropriateness of digital child safety interventions.



Approach 3: On-Device Approaches to Intervene When Conduct and Contact Risk Occur

Overview

Main Concepts and Definitions

- On-Device Safety Tools: Real-time, proactive interventions implemented directly on users' devices to monitor digital behavior and intervene without compromising privacy.
- Privacy-Preserving Approaches: Methods like on-device detection, that adhere to data minimization and transparency aimed at protecting children's safety while upholding their right to privacy.
- Digital Resilience: The capacity to confidently and independently navigate online risks, supported by thoughtful interventions that guide rather than restrict online experiences.
- Context-Aware Interventions: Real-time interventions providing age-appropriate, timely guidance and resources tailored to the context of the user's actions or situation.

Key Findings

- Effectiveness of Real-Time Detection: Real-time, AI-driven monitoring and early-warning systems may effectively prevent harm by intervening promptly when risks or harmful behaviors are detected.
- Positive Behavioral Reinforcement: Subtle, positive feedback (e.g., badges, nudges) designed to encourage safer behaviors is effective, improving children's online safety and reinforcing positive engagement.
- Privacy Protection and Trust: Privacy-preserving tools, especially on-device detection, ensure child safety interventions can effectively operate without transmitting sensitive data off-device, thus preserving autonomy and user trust.
- Safety by Default: Automatically enabling protective features without requiring user intervention significantly improves children's safety online, immediately reducing risks.

Questions for Further Research

- Bias Mitigation in AI Systems: How can AI-driven detection tools effectively and systematically address biases, ensuring fairness, inclusivity, and accurate detection across diverse groups?
- Long-Term Impact Assessment: What are the long-term effects of on-device safety interventions on children's digital resilience and independence, and how can their effectiveness be reliably measured over time?
- Contextual Optimization of Interventions: What strategies and methodologies can optimize context-aware interventions, ensuring they are consistently relevant, timely, and appropriately adapted to individual maturity and cultural backgrounds?

On-Device Approaches to Intervene When Conduct and Contact Risk Occur

Among the recent developments to enhance child safety are a series of on-device approaches, which take into account that smartphones are major access points among young people. Designing on-device approaches requires carefully balancing monitoring children's online activities with protecting their privacy, in order to prevent, detect and intervene effectively.

On-device approaches can help monitor peer-to-peer exchanges to detect and intervene upon conduct and contact risk. However, they may also pose privacy risks. These approaches process children's personal information and data, whether knowingly shared or inadvertently generated through their online activities or even disclosed by parents/caregivers and peers, that must be safeguarded. The OECD's and Berkman Klein Center's Typology of Risks acknowledges privacy risks cut across all risk categories, potentially affecting children's lives in multiple ways (OECD, 2021). Research suggests that these approaches can prevent and interrupt harmful digital conduct and contact by striking the right balance while keeping children safe without compromising their privacy (Gernand, 2022).

Creating trustworthy on-device approaches that center children involves seamlessly integrating devices into their social fabric, reinforcing positive behavior while signaling harmful risks, and doing so within a safe environment. Children who engage in risky peer-to-peer exchanges, including when their own actions make them vulnerable, also have internal strengths to be encouraged and further cultivated in meaningful ways while learning about harmful risks. Technology can play an integral role in recognizing and encouraging these strengths, helping children develop resilience while learning about potential harms in a supportive way.

Software and devices could embed these lessons subtly, through design patterns, interactions, and feedback loops that feel intuitive to both children and trusted adults in their lives. By embedding positive values, clear guidance, and engaging cues into the design of devices, we enable children to see — and learn to trust — how the technology works. Parents/caregivers can then step in at opportune moments, protecting both the child's growth and privacy. By fostering safe exploration, encouraging healthy behavior, and nurturing a sense of well-being and agency, technology use can center children — giving them the confidence to engage with devices designed to keep them safe. In this section, we explore on-device approaches to prevent, detect, and intervene when a child contributes to or faces risky digital conduct or contact while preserving their privacy.

It is important to note that while on-device approaches offer promising avenues for real-time intervention and user-tailored support, their feasibility varies across global contexts. In particular, access to the Internet and to individual digital devices among children is uneven, with significant disparities across and within countries.

1. On-Device Approaches to Prevent Children from Acting in a Way That Contributes to Risky Digital Conduct or Contact

What Are Optimal On-Device Approaches to Act Upon and Prevent Children from Acting in a Way That Contributes to Risky Digital Conduct or Contact?

Current scenarios tend to focus on protecting children as recipients of harmful content (OECD, 2020a; Stoilova et al., 2023). However, there is a need for two-way protection to address situations where children might engage in harmful behavior, such as producing or sharing CSAM (OECD, 2020a). Although framed here within conduct and contact risks, such cases may also intersect with other risk dimensions. Given these examples of conduct and contact risks, the following approaches could offer useful starting points for addressing such challenges.

1.1 Prevention

Early Warning Systems: Develop on-device systems that analyze a child's online behavior over time to identify potential warning signs of risky behavior or victimization. This could involve monitoring factors like changes in typing patterns, emotional sentiment in messages, and engagement with specific types of content.

Positive Behavior Reinforcement: On-device approaches that track text, audio, photos, and video for risky behavior can not only be used to intervene in risky situations, but also to reward positive behavior contributing to a safe online experience. Some evidence suggests that interventions to prevent risky actions may generate negative feelings that paradoxically further problematic behavior (Prasad & Quinones, 2020), but that positive reinforcement like badges or awards awarded to the child may encourage the child towards safer environments and actions, for example, rewarding child users who respect screen limits set in agreement with their parents/caregivers (Prasad & Quinones, 2020). Badges or awards could be interactive and tailored to the child using machine learning (Hinduja, 2023). These incentive structures could be extended beyond mere app avoidance to promote positive engagement within apps themselves, for example through built-in reward systems in mobile health or social well-being apps.

This aligns with findings from screen time research, which suggest that instead of relying on negatively framed warnings or interruptions, positive incentives for safe behavior, such as rewards for not opening certain apps, can be more effective (Prasad & Quinones, 2020). These incentive structures could be extended beyond mere app avoidance to promote positive engagement within apps themselves, for example through built-in reward systems in mobile health or social well-being apps. This opens the door to designing on-device mechanisms that not only deter risky behavior but actively guide children toward healthier and safer digital habits.

1.2 Detection

Content-Based Detection: While current systems primarily screen images and videos, expanding detection capabilities to encompass audio, text, and emerging formats like deepfakes and VR/AR content could prove beneficial (Ayub Khan et al., 2025; Gómez-Quintero, 2024; Lee et al., 2020).

Multi-Signal Analysis: Moving beyond single-signal analysis (e.g., an isolated image) to incorporate multiple signals could enable more proactive intervention

(Edwards et al., 2021; Steel, 2024). For instance, a system could analyze a combination of factors such as the presence of multiple images of another user, escalating conversation tone in messages, and the recent download of an app known to create nude deep fakes.

Real-Time Detection for Various Communication Channels: Expanding real-time content analysis and intervention beyond Messages, AirDrop, and FaceTime, providing programmatic access or toolkits for developers of other popular communication apps like WhatsApp, social media platforms, and Generative AI models (or tools) could provide more comprehensive protection (Agha et al., 2025; Stoilova et al., 2023).

Victim-Specific Algorithms: Victim-specific algorithms could detect grooming language patterns or coercive language, which are common in harmful interactions. These algorithms help detect predatory behavior early by analyzing the back-and-forth communication between the predator and the victim, focusing on linguistic exchanges that subtly push the victim towards compliance (Cook et al., 2023).

1.3 Intervention

Machine Learning Intervention: On-device machine learning models can be used to detect potentially harmful encounters or behaviors as they occur. They can also be used to monitor the emotional state of the child (Guo et al., 2024). By training algorithms to recognize risky actions, like sending explicit images or sharing personal information, or emotional states, these systems can provide immediate warnings tailored in a way the child will respond best to, helping to prevent risky behavior in real-time (Prasad & Quinones, 2020).

Privacy-Preserving Machine Learning Techniques for On-Device Approaches: Tools that provide immediate, privacy-preserving alerts when potentially harmful behavior is detected can help intervene while safeguarding trust on devices (Brumen et al., 2023). Such nudges can offer contextual guidance to dissuade children from harmful behavior (Information Commissioner's Office, n.d.; Veretilnykova & Dogruel, 2021).

Improve Image Classification: Research suggests that one important step is improving how technology recognizes (i.e., classifies) images. By analyzing images in new ways — such as looking closely at unusual patterns (e.g., frequency) or subtle errors (i.e., error-level analysis) — embedded tools could better identify risky content, particularly deep fakes (Frank et al., 2020; Rafique et al., 2023). However, it is important to note that advances in deepfake-generating technology (e.g., Generative Adversarial Networks or GANs) have reduced visible artifacts significantly since 2020 (Gupta et al., 2023; Mirsky & Lee, 2020). Therefore, methods effective in earlier research may need continuous updating or supplementation with newer techniques.

Expanding Health Notifications: Physical health is currently monitored and encouraged by digital providers, notifying users to track their steps and reach their fitness goals throughout the day. Some providers have also introduced options to expand that tracking and encouragement towards mental health, such as daily mindfulness goals or well-being audits through platforms like Woof (Khameneh, 2023). However, experts argue there is little evidence that quantifying these metrics, such as emotion and experiences, may help to decrease children's mental health problems (Khameneh, 2023). Given the crucial role that social and emotional skills can play in managing the difficult emotions that may spur problematic digital behavior, expanding the tools available and incorporating them into notifications like physical health measurements, can support emotional resilience and ultimately reduce the propagation of risky digital content (Finkelhor et al., 2021; Hasse et al., 2019; Livingstone, 2013).

Context-Aware Interventions: Alternatively to blocking or blurring content, context-aware interventions that provide age-appropriate explanations, guidance, and resources to help children understand and navigate risky situations could be explored. For example, if a child attempts to share a sexually suggestive image, the system could provide a warning about the potential consequences and offer resources on healthy relationships and online safety.

2. On-Device Approaches to Interrupt Instances Where a Child Is a Victim (Or Recipient) of an Action

What Are Optimal On-Device Approaches to Interrupt Instances Where a Child Is a Victim (Or Recipient) of an Action?

Effective on-device approaches to interrupt when a child is victim (or recipient) of a harmful action requires a careful balance — stepping in to protect them when necessary while gradually giving them the skills to protect themselves. When a child faces a situation online, whether as a victim or an unknowing participant, immediate intervention can prevent further harm.

Child's safety depends on more than just interruption; it encompasses the capacity to confront the risks that the digital world may come with to protect one's well-being (Cortesi et al., 2020). Due to their evolving capacities, children cannot be expected to handle the complexities of the digital world alone. They need guidance, not just protection, to build the resilience and skills required to stay safe.

On-device approaches can play a crucial role in this balance. They can step in at critical moments, offering protection when children are being targeted or unknowingly engaging in harmful interactions, in ways that are both user-friendly and effective in preventing harm. More importantly, such approaches should be designed to not only prevent harm but also teach children how to respond — ensuring that over time, they gain the confidence and ability to protect themselves.

While the focus is set on conduct and contact risks, some of the on-device strategies discussed also engage with content-related and privacy risks, reflecting the reality that risk categories often overlap in practice.

2.1 Prevention

Considering Digital Maturity and Agency: Research suggests that children's digital maturity — defined as the ability to navigate the digital environment with understanding and self-determined use — is positively associated with their sense of well-being (Laaber et al., 2023). Recent research on developmental studies suggests that digital maturity and care for oneself and others while using technology for beneficial purposes plays a crucial role in digital habits to foster children's agency and well-being while attending to their developmental stages (Konrath et al., 2025; Laaber et al. 2024). This raises the question of how on-device approaches can be tailored to support this developmental progression for children's flourishing.

Safety by Default: Safety by Default ensures that protective features are automatically enabled, minimizing risk without requiring user action. By integrating safeguards directly into device settings and design, children receive immediate protection, supporting safe use from the start. This approach proactively reduces the likelihood of exposure to harmful behaviors or encounters and enhances overall user safety by incorporating features such as real-time monitoring and content filtering directly within the device. Aligned with the principles of Safety by Design, this method shifts responsibility towards the technology, ensuring that safety measures are embedded within the digital experience from the outset (eSafety Commissioner, n.d.).

2.2 Detection

Expand Communication Safety Features: Broaden the scope of current on-device features to address other forms of harmful encounters beyond nudity, including violent or hateful material. Extend its application to encompass a wider range of communication channels.

Address Text-Based Risks: Develop on-device mechanisms to detect and address harmful or manipulative text that can be identified as grooming behaviors, sexually explicit language, and links to unsafe websites.

Real-Time Detection of Grooming Behaviors: Train on-device algorithms to detect patterns of language and interaction common in online grooming attempts (Borj et al., 2023a; Prosser & Edwards, 2024).

Contextual Analysis of Communication: Develop systems that analyze communication within the context of the relationship between individuals (Borj et al., 2023b). For instance, a system could differentiate between age-appropriate communication among peers and potentially inappropriate interactions between an adult and a child.

Improve Age Verification: Implement more robust and privacy-protecting on-device age verification methods (Hinduja & Lalani, 2025), potentially by analyzing data reports and usage patterns (Hogg & Schwarztrauber; Liu & Scheffler, 2025; Tam & Denham, 2025)], to prevent children from bypassing safety features by misrepresenting their age.

Real-time Agents: Using deep learning techniques for image, audio, and text processing, these agents analyze real-time screen captures and audio signals to identify and minimize children's exposure to risky peer-to-peer exchange such as cyberbullying, pornography, and induction to self-harm (Jevremovic et al., 2021).

2.3 Intervention

Proactive Intervention and Reporting: Instead of reacting to harmful peer-to-peer exchange, explore proactive interventions such as warning a child about potentially suspicious behavior or automatically reporting serious threats to parents/caregivers (Dempsey et al., 2022).

3. Privacy-Preserving On-Device Approaches

How Can It Be Assured That Such Approaches Preserve the Privacy of Children and Users?

Maintaining a child-centric perspective is crucial, ensuring that any technological intervention prioritizes their rights, agency, and well-being. Such a perspective requires careful consideration of the potential impact of these approaches on children's lives and a commitment to ongoing evaluation and refinement.

3.1 General Privacy

Independent Auditing and Oversight: Establish mechanisms for independent auditing and oversight of on-device safety systems to ensure transparency, accountability, and compliance with privacy regulations. This could involve regular reviews by security researchers, privacy advocates, and ethical experts to identify and mitigate potential risks, as well as existing regulatory redress mechanisms such as a children's rights impact assessment (Fosch-Villaronga et al., 2023; Mantelero & Esposito, 2021).

Addressing Potential Biases: Rigorously test and evaluate on-device algorithms and datasets to identify and mitigate potential biases that could disproportionately impact certain groups of children, such as those from marginalized communities or with diverse cultural backgrounds (Araujo et al., 2017).

On-Device Detection: In situations where children are victims or recipients of hateful, harmful, illegal, or problematic user-generated content (e.g., sextortion attempts or grooming messages), on-device detection can offer protection. This can include automatically blurring sensitive images or using recordings and transcriptions to enable violent language detection in near-real time, all while safeguarding their agency and privacy by keeping data stored on the device rather than transmitting it off-device (Anwar & Kanjo, 2023).

3.2 On-Device Privacy

Default Child Safety: Design systems with child safety features enabled by default, enabling immediate protection for children, which might prevent them from acting in a way that contributes to risky digital conduct and contact. While parental controls are valuable, requiring them for activation (vs. by default) can create vulnerabilities, including security and privacy issues such as exposing monitored children's data to third parties (Ali et al., 2020; Anderson et al., 2015). Moreover, research highlights that children perceive these safety features positively when they feel they afford them more agency, improve their relationship with parents/caregivers, and see direct benefits, such as managing unhealthy behavior (Ghosh et al., 2018; Gnanasekaran & De Moor, 2025).

Transparency and User Control: Provide users, including children, with transparent and understandable information about how on-device safety features function, what data is accessed, and how it is used, and offer user-friendly controls to adjust settings and manage privacy preferences, which will support children's ability to maintain their data agency based on their age and evolving capacities (UNICEF, 2021b; Wang et al., 2021). Data processing such as web tracking and data shared for secondary process should be transparent to the user and held to the minimum (UNICEF, 2021b).

Online Safety Nudges: Research on online safety nudges in the context of adolescent online safety emphasizes that on-device tools can provide contextual guidance without fully removing content or censoring sensitive content (like images) with an option for teens to decide whether to view it or not (Obajemu et al., 2024). Moreover, Obajemu and colleagues (2024) emphasize that teens believe that nudges can be better optimized by, for example, providing actionable guidance or personalization controls that create less disruption in their digital experience.

Minimizing Data Collection and Sharing: Adhere to the principle of data minimization, collecting and storing only the essential data required for the specific safety feature and for the shortest duration necessary (Committee on the Rights of the Child, 2021; UNICEF, 2021a; UNICEF, 2021b). Avoid unnecessary sharing of data with third parties, including law enforcement, without explicit user consent or a legally mandated requirement (Assis & Valença, 2024).

Secure Data Handling and Encryption: Implement strong encryption protocols to protect user data both during processing on the device and during any transmission or storage, even if minimal, to external servers. Prioritize privacy-preserving techniques like differential privacy or federated learning to minimize the risk of data exposure (Assis & Valença, 2024).

Involving Children in the Design Process: Prioritizing children's perspectives and lived experiences is essential when designing and developing on-device safety features. This includes creating spaces, programs, and methodologies that actively support their meaningful participation (Cortesi et al., 2021). Involving children in key design decisions — such as when and how they receive alerts — can make on-device solutions more responsive to their specific needs and increase the likelihood that they will engage with these features proactively (Lehnert et al., 2022; Quayyum, 2025). For further discussion, see Approach 4, Section 2: User Interface Design for Diverse Children.

Communication of Safeguards to Build Trust. Trust plays an important role in the interactions between the device and the user. If a user has no trust in the platform that is giving advice, the prompts may be more readily dismissed (Von Der Linde et al., 2025). Being transparent about the measures the device is taking to help could help establish or maintain trust and make children to follow rather than bypass them (Wang et al., 2021) For further reading, see approach 1, Design Approaches that Foster Trust.

3.3 Intervention Verification Considerations

Data Collection and Privacy Risk: Age verification processes often require personal data, which can expose children to privacy risks if collected, stored, or shared insecurely (UNICEF, 2021a). This raises concerns under privacy frameworks like the European Union's General Data Protection Regulation and the Digital Service Act (Beltrán & de Salvador, 2024; Hinduja & Lalani, 2025; Livingstone & Third, 2017).

Children's Right to Privacy: International agreements like the UN Convention on the Rights of the Child emphasize the importance of protecting children's privacy in all settings, including online. Age verification must balance protecting children with respecting their privacy (Committee on the Rights of the Child, 2021; UNICEF, 2021a).

Technical Limitations and Overreach: Some age verification methods, such as biometric analysis or third-party data collection, can lead to excessive data collection that infringes on privacy rights (Sas & Mühlberg, 2024). This disproportionate data collection raises ethical and legal concerns (Leaton Gray, 2018) that can potentially fragment user trust in the device ecosystem.

Informed Consent and Transparency: Children often lack the capacity to fully understand the implications of privacy-related choices. Ensuring that age verification tools are transparent and that children's assent is meaningful proves to be essential to protecting their rights (Van der Hof, 2017).

Evaluating On-Device Approaches

Considering the review of research and practices summarized in this chapter, the concluding discussion critically evaluates current and emerging on-device approaches aimed at enhancing children's online safety, focusing particularly on proactive measures that respect and preserve user privacy. Central to this analysis is the increasing recognition of on-device safety tools, which emphasize real-time threat detection and intervention as promising avenues for preventing harm.

There is a broad variety of technical and socio-technical interventions spanning prevention, detection, and intervention strategies. Promising examples include real-time AI-driven monitoring, subtle behavioral nudges, positive reinforcement techniques, and adaptive context-aware warnings. Some interventions, such as proactive machine learning-driven detection of grooming behaviors or sophisticated nudging methods, remain largely in the experimental phase. In contrast, others, like on-device detection for explicit images or default-enabled safety settings, are already more widely deployed. Despite this diversity, the current evidence supporting the efficacy of many specific interventions is very limited, if available at all, highlighting a critical need for robust empirical validation.

Nonetheless, certain baseline practices — particularly privacy-preserving approaches like on-device detection — stand out as a promising current technique. This method has demonstrated initial results in protecting children's privacy while allowing for effective real-time interventions. However, transparency and user control mechanisms, though frequently recommended, currently have less substantial evidence supporting their long-term effectiveness and user engagement, marking an important area for future research.

The concept of digital resilience highlights the importance of supporting children's independent and confident navigation of online environments. While initial interventions show potential, rigorous long-term studies examining their sustained impact remain notably limited.

Context-aware interventions, whether platform-specific, device-specific or embedded at the app level, represent an important frontier. Grounded in behavioral design principles, these interventions leverage positive reinforcement and subtle nudges to encourage safer online behavior. Early evidence suggests effectiveness; however, more detailed empirical validation across diverse cultural contexts is required to fully understand and optimize their impact.

An important tension emerges between the proactive nature of safety interventions and maintaining children's privacy and autonomy, presenting an ongoing challenge in designing effective safety measures. Addressing this tension thoughtfully remains a critical focus for both practice and research.

Key knowledge gaps and promising research frontiers identified in this chapter include addressing biases in AI-driven detection tools, evaluating the long-term impacts of safety interventions on digital resilience, and optimizing context-aware interventions across diverse user demographics and cultural contexts.

Approach 4: Educational and User-Interface Design Approaches

```
   __________ __
   |    ___  _|__|_
   |  ,’   ‘. {‘\)
   | :  O O  : )(
   | :  ._\. :/ )\
   |  ‘.___\\/ / |
   |        ‘-’  |
   |_________|   |
      /-----\|___|
jrei /       \|||
             .:;;
```

Overview

Main Concepts and Definitions

- Educational Approaches: Structured interventions that teach children essential knowledge, skills, behaviors, and competencies to identify, respond to, avoid, and report online risks, ultimately enhancing their agency and overall digital well-being.
- Child-Centered Design: A participatory approach that actively involves children in developing digital safety tools and educational content, addressing their specific developmental stages, diverse backgrounds, and real-life contexts.
- Holistic Approaches: Comprehensive educational strategies integrating online safety within broader topics such as mental health, bullying prevention, and relationship education, emphasizing the role of parents/caregivers, educators, and community members.
- Skill-Based Interventions: Structured training methods explicitly aimed at enhancing practical and critical competencies and digital literacy skills, enabling children to better recognize, evaluate, and manage online risks independently.

Key Findings

- Integration and Holistic Education: Strong evidence indicates that integrating online safety into broader educational contexts (e.g., mental health, bullying prevention) significantly enhances program effectiveness compared to treating digital safety as an isolated topic.
- Participatory Co-Design: Robust evidence supports actively involving children in designing educational tools and content, significantly increasing relevance, engagement, and effectiveness by leveraging children's insights and creativity.
- Repeated and Early Education: Clear evidence supports the effectiveness of initiating digital safety education early and delivering it through frequent, brief, repeated sessions across diverse venues (schools, community centers, homes, digital platforms) to ensure sustained knowledge and behavior change.
- Critical Role of Clear, Specific, and Positive Communication: Strong evidence supports educational methods that use clear, relatable, and empowering language, significantly enhancing children's comprehension, agency, and adherence.

Questions for Further Research

- Optimal Integration of Digital Safety in Broader Education: There is limited empirical evidence regarding the most effective methods to integrate online safety into wider curricula on mental health, bullying, relationships, and sexuality education.
- Effectiveness and Ethical Implications of Personalization and Algorithmic Interventions: Significant open questions remain regarding the potential, ethical risks, and long-term efficacy of personalized safety tools leveraging algorithms and behavioral design techniques.
- Cross-Cultural Applicability and Accessibility: Research gaps persist on how effectively culturally tailored, multilingual, and accessible educational and user-interface interventions work across diverse user groups and contexts, highlighting the need for targeted, inclusive studies.

Educational and User-Interface Design Approaches

Educational approaches – including prevention, detection, and effective interventions – are essential for teaching children how to identify, respond to, avoid, and disclose severe risks and harms to children (Finkelhor et al., 2021). These knowledge, skills, and competencies also foster children's sense of agency and well-being, equipping them to navigate the digital environment and protect themselves from potential harms.

Adults play a pivotal role in implementing educational strategies that center children. Parents/caregivers and educators can help children cultivate digital literacy and social-emotional skills such as problem-solving, empathy, and resilience (Hinduja & Patchin, 2017; Richardson & Milovidov, 2019; Hinduja & Patchin, 2022b). Meanwhile, business developers and technology companies can design interfaces to connect with children from diverse backgrounds, enhancing communication and helping them recognize risks and seek support. Grounding this approach in a foundation of centering children empowers child users to navigate the digital environment safely, while being supported by their wider network of parents/caregivers, educators, community leaders, healthcare professionals, technology companies, and policymakers..

1. Educational Methods to Help Children Identify Risks and Seek Help

What Educational Methods Are Most Effective to Help Children Identify and Seek Help with Specific Risks They Face?

Educational initiatives, both within and outside of schools, play a critical role in teaching children digital safety (Palfrey et al., 2010; Jones et al., 2023). Schools, as formal institutions of education, address digital child safety because it is connected to their traditional responsibility for offline safety. In principle, school environments can be promising venues for intervention: students arrive primed to learn, already engaged in a structured environment, and familiar with their instructors (Finkelhor, 2007; Bright et al., 2023). In practice, however, schools vary widely within and across countries, so effective educational methods may differ across locations, age groups, demographics, and other contextual factors (Chicote-Beato et al., 2024). Additionally, teachers contend with packed schedules, leaving little room for in-depth digital safety lessons — an often-overlooked constraint that limits both the scope and impact of school-based efforts (Hedderich et al., 2024).

Recognizing the shifting landscape of digital education and rising public concerns about online predators, Internet addiction, and cyberbullying, schools in the mid-to-late 2000s initially prioritized Internet Safety Education (ISE) in response to (Palfrey & Gasser, 2008; Jones et al., 2012; Cortesi et al., 2020; Jones et al., 2023). By the early-to-mid 2010s, however, a policy shift emerged. Rather than focusing solely on risks, educators and policymakers began advocating for a more holistic approach — one that emphasized critical, effective, and responsible online engagement (UNESCO, 2016; Cortesi et al., 2020).

Despite the absence of a standardized online safety curriculum, given the ever-evolving risks and harms in this space, surveys indicate that approximately 46% of U.S. teachers incorporate digital citizenship materials into their classrooms, with cyberbullying ranking among the most frequently addressed topics (Fredrick et al., 2023; Lauricella et al., 2020; Vega & Robb, 2019). Notably, much of the existing research on cyberbullying has focused on the online experiences of children from majority groups (Modecki et al., 2014; Brochado et al., 2017; Jain et al., 2020; Polanin et al., 2022). Expanding this body of work to encompass more diverse populations would yield valuable insights to advance cyberbullying programs.

With a focus on U.S.-based curricula, metrics, and academic literature, this section explores key insights into the effectiveness of a school-based cyberbullying curriculum. As a definitional matter, the effectiveness of school-based cyberbullying prevention and intervention programs is determined by their ability to demonstrate a statistically significant decrease in the rates of online bullying and/or victimization when comparing the outcomes of the experimental to the control group in a pre/post evaluation (Hasse et al., 2019). In essence, a program is deemed successful if it can show that it leads to a measurable reduction in cyberbullying incidents among the students who participate in it as compared to those who do not (Hasse et al., 2019; Lukács et al., 2023). Consistent with this benchmark, Walsh et al. (2015) found programs to increase knowledge, protective behaviors, and reporting when reviewing 24 evaluations of school-based sexual abuse prevention programs. Drawing from these findings and complementary literature, several key factors emerge as predictors of success for school-based interventions, offering valuable guidance for the design of future programs.

1.1 Overarching Framing

Holistic Approaches to Online Safety and Well-being: Young people favor a holistic approach to online safety in which families, educators, and policymakers work together to support their well-being and equip them with the tools to protect themselves (Lala et al., 2022; Marsden, 2022). Research highlights the importance of family involvement, particularly in addressing Internet overuse and suicide prevention, with parental/caregiver mediation playing a crucial role in mitigating online risks (Finkelhor et al., 2020; Hilliard, Batanova, & Bowers, 2015). Rather than treating digital safety in isolation, integrating it into broader programs focused on bullying, dating abuse, sexual abuse prevention, and mental health leads to more effective outcomes (Collier, 2013; Finkelhor et al., 2020). Studies suggest that combining digital safety education with well-established, evidence-based programs about offline harms and other curricula is more effective than creating new, untested initiatives (Bickham et al., 2018; Collier, 2013; Finkelhor et al., 2020; Finkelhor et al., 2021). Digital harms often stem from deeper issues such as mental health struggles, family conflicts, peer rejection, and offline trauma exposure, underscoring the need for interventions that address these root causes alongside digital safety education (Finkelhor et al., 2020). Successful interventions, particularly in areas like sexting and influencing teen sexual behavior, often involve multi-session and multi-element programs. These programs should allow children to contribute their views, explore values, discuss relationships, and practice interpersonal skills (Finkelhor et al., 2020).

Children's Needs, Contexts, and Agency: Research on child preferences and attitudes reveals a strong sense of agency and a desire to be actively involved in their own digital safety. They want to be included in discussions and decision-making processes that shape their digital experiences (Lala et al., 2022). Although definitions of agency and its operationalization vary across and within disciplines, it is widely recognized as a fundamental human quality that develops during childhood and should be nurtured through education (Brod et al., 2023). Effective interventions empower young people by equipping them with the knowledge and skills to make informed decisions and navigate online risks with confidence (Finkelhor et al., 2021). Their concerns about digital safety often stem from their everyday experiences, underscoring the need for tailored, context-sensitive approaches for educational interventions (Lala et al., 2022; Marsden et al., 2022). Real-life examples, relatable scenarios, and age-appropriate language can enhance engagement and make content more relevant and impactful.

Factors such as age and evolving capacities, circumstances, education (skill level), ethnicity, gender, location, national origin, race, and/or socioeconomic background can shape young people's experiences of digital risks and harms, highlighting the need for inclusive, culturally sensitive safety strategies addressing the diverse needs and experiences of children (Blumenfeld, 2020; Heller & Magid, 2019; Ito et al., 2020; ITU, 2023; Lala et al., 2022). Interventions targeting online risks such as bullying should provide situation-specific guidance rather than generic advice, which often fails to address the nuances of individual experiences (Dinakar et al., 2012). By tailoring interventions to the specific context of the bullying incident, victims and bystanders are more likely to find the advice relevant and applicable to their situation (Chen et al., 2024).

Research also challenges the assumption that all risk exposure is harmful. Instead, some degree of risk-taking can foster resilience and adaptability, particularly when children engage in online communities and peer support networks, even while encountering potential threats such as hate speech (Ito et al., 2020; Livingstone, 2013; Rideout & Robb, 2018).

Effective interventions must go beyond knowledge-sharing and focus on developing practical skills for risk management and harm mitigation. Key competencies include recognizing and responding to inappropriate online requests, navigating the digital environment safely, and seeking help when needed (Finkelhor et al., 2021). Childhood studies emphasize that children exercise agency through awareness, competence, and strategic decision-making, enabling them to assert their rights and actively shape their digital interactions (Valentine, 2011; Mühlbacher & Sutterlüty, 2019). Research further shows the importance of focusing (albeit not exclusively) on skill-building, demonstrating that children with stronger digital skills are less likely to experience harm online and are better equipped to manage risky situations (Kardefelt Winther et al., 2023).

Leveraging Children's Existing Knowledge and Skills: Key to fostering a safe and supportive school environment for children is not only protecting them from harm, but also cultivating their positive attributes and developing their character (Dailey & Roche, 2025; Hilliard et al. 2015; Hilliard et al., 2014). Research supports this strength-based approach. Studies on bullying emphasize that rather than focusing solely on reducing problematic behavior, a more effective approach combines prevention with efforts to promote children's thriving through resource alignment (includingpeers, extracurricular activities, etc.) (Dailey & Roche, 2025; Hilliard et al., 2015). Emphasizing young people's well-being as a whole is vital given that children are multifaceted and can not be reduced to victim/bully binaries (Dailey & Roche, 2025; Hilliard et al., 2015).

1.2 School Context

Demographics: Given that demographic factors such as age and evolving capacities, circumstances, education (skill level), ethnicity, gender, location, national origin, race, and/or socioeconomic background may pose unique challenges for children digitally, educational initiatives that account for children's individual differences and different susceptibilities to risk are needed (Blumenfeld, 2020; Heller & Magid, 2019; Hilliard et al., 2015; Ito et al., 2020; Lala et al., 2022; Madden et al., 2024; Valkenburg & Peter, 2013; Abades Barclay & Banaji, 2024). Research shows that, for example, already vulnerable children — for instance, those who struggle with family issues, peer rejection, trauma, social isolation, and/or mental health challenges — are most at risk of being influenced by online self-harm content or being the target of bullying or mean behavior online (Hinduja & Patchin, 2007; Blumenfeld, 2020; Finkelhor et al., 2020). Furthermore, social and psychological factors, including mental health struggles and sensation seeking, are primary risk factors for online harm (Livingstone, 2013). These findings underscore the need for targeted educational interventions that equip these children with the tools to navigate the digital environment safely. To address these vulnerabilities, schools must implement strong screening and referral systems to identify at-risk children and connect them with appropriate support services (Finkelhor et al., 2020).

School Climate: Research suggests that school climate may play a role in the occurrence of bullying, cyberbullying, and the sense of victimization students experience as a result (Hinduja & Patchin, 2020; Zacharia & Yablon, 2022). For example, a study of middle and high schools found that school climate moderates the relationship between students' victimization and their sense of safety at school (Zacharia & Yablon, 2022). Aspects of positive school climate that may contribute to the prevention of bullying include small size, a sense of community and belonging, student engagement and participation, and a flexible and diverse educational ideology (Izadi & Hart, 2023). Additionally, research on "authoritative school climate" — a framework that combines clear rules with strong student support — links this approach to lower rates of bullying, cyberbullying, and improved academic achievement (Hinduja & Patchin, 2020). Students in authoritative schools report fewer instances of bullying, suggesting that a structured yet supportive environment fosters safer peer interactions. These findings underscore the value of school-based interventions that reinforce firm behavioral expectations while providing emotional support. By cultivating a positive, well-regulated school culture, educators can promote both student well-being and academic success.

Peers, Family Members, Trusted Adults, andProfessionals: Children want their parents/caregivers and educators to be well-informed about online safety so they can offer guidance and support (Lala et al., 2022). Studies show that children facing issues such as suicide and self-harm are more likely to seek help from peers, family members, and trusted adults than from healthcare professionals, emphasizing the need to target these networks for prevention (Finkelhor et al., 2020). Yet, parents/caregivers often underestimate their children's online activities, including their involvement in cyberbullying. Educating parents/caregivers about technological tools and platforms and digital risks can help them recognize warning signs and equip their children with essential online safety skills (Hasse et al., 2019; Quayyum et al., 2021). Social Control Theory further supports this approach, showing that strong social bonds help prevent deviant behaviors, including bullying (Hasse et al., 2019; Hinduja, 2017; Hirschi, 1969).

1.3 Material Content Topics

Norm Setting: Educational interventions should equip children with the tools to recognize and challenge harmful behaviors (Abades Barclay & Banaji, 2024). Studies suggest that these interventions should address and model what sort of speech or behavior is considered mean, hurtful, or beneficial; what a healthy relationship or good digital citizenship looks like; and which norms about sexuality are perpetuated by peers, siblings, and family members, and how they might be shifted (Dinakar et al., 2012; Finkelhor et al., 2020; Gallagher & Magid, 2019; Gallagher et al., 2017; Heller & Magid, 2019; Hilliard et al., 2015; Kardefelt Winther et al., 2023; Thorn, 2024b). Recent survey data reveals that many minors perceive harmful online experiences, including the sharing of explicit content, as "normal" (Thorn, 2024b). This normalization underscores the need for educational initiatives that challenge these perceptions, emphasize the potential consequences and risks associated with such behaviors, and promote critical reflection. Another study indicates that children are most likely to be exposed to hate messages and violent images online, with European children finding this violent content the most distressing and harmful to see (Kardefelt Winther et al., 2023). These findings suggest the need for effective programs that equip child users with strategies to critically assess digital content, build resilience, and seek support when needed.

Skill-Building: Students will benefit from developing skills to recognize, navigate, and respond to situations that give rise to cyberbullying (Mishna et al., 2009; Collier, 2012; Polanin et al., 2022). According to the Theory of Planned Behavior, strengthening three key psychological factors may enhance children's ability to make safer choices, focusing on recognition (understanding what is risky and how to act), perceptions (believing in their ability to take action) and intentions (developing the motivation to avoid harm and disclose concerns) (Bright et al., 2023). These skills are vital for students to act independently and safely in real-time, making skill-building a key component of cyberbullying prevention since adults and peers are likely to be absent in these online interactions (Mishna et al., 2009; Collier, 2012; Polanin et al., 2022).

Broader social and emotional skills — such as collaboration, self-efficacy, empathy — as well as broader digital citizenship education may help reduce cyberbullying, online risk-taking, and victimization and increase self-efficacy in solving online problems (Collier, 2012; Finkelhor et al., 2021; Gallagher & Magid, 2019; Hilliard et al., 2015; Hilliard et al., 2014; Jones et al., 2023). Self-esteem, self-efficacy, and resilience play a crucial role in protecting young people from online harm. Research suggests that self-esteem acts as a protective factor, helping to mediate digital risks (Livingstone, 2013). Similarly, self-efficacy and self-esteem enhance resilience and protective factors against bullying, which strengthen children's ability to cope with bullying and reduce their likelihood of victimization (Hasse et al., 2019). Empathy-building also emerges as a key strategy for addressing cyberbullying and self-harm. Programs such as KiVa and NoTrap! incorporate empathy into their curricula, encouraging students to recognize and respond to the emotions of others, thereby fostering safer and more supportive school environments (Hasse et al., 2019; Polanin, 2022).

As technology evolves, so too do the behaviors of children, demanding that education curricula keep pace with topics that are relevant for students (Abades Barclay & Banaji, 2024). Emerging risks, such as the misuse of "nudify" or "undress" apps — which enable children to digitally manipulate images to victimize peers — highlight the need for proactive and ongoing updates to material topics in digital child safety. Integrating topics like digital consent, privacy, and bodily autonomy into school curricula ensures that students are supported and develop the critical awareness needed to navigate these issues responsibly. Given the sensitive nature of such discussions, educators, rather than law enforcement, are generally better suited to foster meaningful and supportive conversations. By integrating these competencies into digital safety education, programs can more effectively equip children with the tools to navigate the digital environment with awareness, agency, and resilience (Collier, 2013; Finkelhor et al., 2021; Gallagher & Magid, 2019).

Digital Safety Tools: Training both students and educators in the effective use of digital safety tools is essential for navigating today's digital environment. A 2023 survey revealed that minors were nearly twice as likely to seek help from online platform tools during an online sexual interaction than to confide in parents/caregivers or peers (Thorn, 2024b). This finding underscores the importance of integrating well-designed digital safety tools into educational interventions , a point further emphasized by the fact that half of young people express a desire for more information about online safety and how these tools function (Thorn, 2024b). Educators, too, require training to facilitate meaningful discussions and address emerging digital issues. Schools that prioritize teacher scaffolding – facilitating in-depth discussions and connecting learning to real-life scenarios – tend to see more engaged students and stronger learning retention, suggesting that well-trained educators enhance teaching effectiveness on digital safety curricula (Abades Barclay & Banaji, 2024).

1.4 Material Methods

Specificity and Clarity: Research emphasizes the importance of specificity in educational interventions. Focusing on concrete online harms (e.g., sexual exploitation andcyberbullying) and their associated risk factors, rather than abstract concepts like privacy, may lead to more effective learning and behavior change (Abades Barclay & Banaji, 2024). Clearly linking specific online behaviors to potential risks helps children develop a more accurate understanding of digital dangers (Finkelhor et al., 2021). Overly broad messages, by contrast, can create confusion, foster skepticism, and instill a false sense of security, ultimately weakening the impact of online safety education (Collier, 2013; Finkelhor et al., 2021). Research highlights the importance of focusing on concepts uniquely conveyed through the educational program rather than those already understood by children from other sources (Bright et al., 2023). By emphasizing clarity and specificity, online safety programs can better equip children to navigate the digital world, prioritizing teaching new and essential safety skills and knowledge.

Fostering Child-Adult Dialogue: Previous literature reviews suggest that open and effective communication between educators, parents/caregivers, and children is essential in addressing the issue of cyberbullying (Abades Barclay & Banaji, 2024; Blumenfeld, 2020; Heller & Magid, 2019). Many parents/caregivers struggle to fully understand their children's online activities, creating a disconnect that can hinder their ability to recognize and respond to digital risks (Hasse et al., 2019). By contrast, parents/caregivers who engage in regular, open conversations about their children's favorite platforms, online interactions, and digital experiences are better equipped to identify warning signs and offer support. Engaging children in conversations about their online activities and encouraging collaborative rule-setting around Internet use can make them feel more comfortable discussing and reporting any issues they encounter (Gallagher & Magid, 2019; Hasse et al., 2019). Simple, non-intrusive questions such as "What is your favorite app?" or "What do you do on it?" help build trust and make children more likely to report online concerns (Hasse et al., 2019).

This need for open dialogue extends to online sexual interactions to identify potential conduct and contact risks. A recent survey found that one in six minors who experienced an online sexual encounter did not disclose it because they believed it was not a "big deal" (Thorn, 2024b). To counter this, educational interventions should normalize discussions about online sexual interactions, reinforcing that no experience is too minor to be shared. Research suggests that children are more receptive to sex and relationship education — an essential tool in preventing online sexual exploitation — when they have input into the curriculum over multiple sessions (Finkelhor et al., 2020).

Early Start: Research underscores the need for early intervention in online safety education, particularly in addressing risks such as sexting and bullying. Strategies to prevent sexting — including discouraging the creation and sharing of sexual images — should begin before sexual exploration begins (Finkelhor et al., 2020) as well as teach safe texting skills focused on harm reduction (Patchin & Hinduja, 2020). Similarly, bullying prevention efforts are most effective when introduced before high school, as early education can help shape healthier peer interactions and reduce long-term harm (Hasse et al., 2019). A study on child safety curricula further supports this approach, demonstrating that programs designed for kindergarten through second grade can successfully teach protective behaviors when tailored to young learners' developmental levels (Bright et al., 2023). These findings highlight the importance of age-appropriate content and delivery methods in fostering engagement, retention, and meaningful learning.

Frequency and Duration of Lessons: Research on a leading child safety program suggests that shorter, more frequent lessons are more effective than fewer, longer sessions in improving children's understanding of risky and unsafe situations (Bright et al., 2023; Gentile & Gentile, 2008). This approach enhances retention, reinforces key safety concepts over time, and sustains engagement without overwhelming learners.

Engaging Formats: Children prefer educational content that is creative, engaging, and visually dynamic, such as videos and animations (Lala et al., 2022). These findings suggest that incorporating interactive elements, multimedia content, and serious games can enhance both the appeal and effectiveness of digital safety education. Previous meta-analyses found the most significant effects of serious games in the context of educational health interventions in knowledge acquisition, while other researchers found evidence for the positive effects on cognitive, perceptual, behavioral, affective, and motivational aspects (Tolks et al., 2020). For example, Re-Mission, a game designed for young cancer patients, has been shown to improve disease management, self-efficacy, disease-related knowledge, and communication behaviors (Tolks et al., 2020). Storytelling and narrative-based activities also prove to be a powerful teaching tool, particularly in combating cyberbullying (Hedderich et al., 2024). By engaging with narratives that depict real-life scenarios, young people can reflect on peer interactions, discuss what constitutes acceptable behavior, and develop a deeper understanding of online risks (Dinakar et al., 2012). Blending interactive digital tools with narrative-based learning creates a more impactful educational experience, increasing knowledge, protective behaviors, and reporting.

2. User Interface Design for Diverse Children

How Should User Interfaces, Including Language, Be Customized to Best Speak to Different Groups (For Instance, in Terms of Age, Maturity, or Circumstances) of Children?

Building a child-centered digital environment requires designing user interfaces that resonate with children and reflect their age and evolving capacities, circumstances, education (skill level), ethnicity, gender, location, national origin, race, and/or socioeconomic background. However, the digital environment is shaped not only by its particular design but also by how users interact with and adapt technologies to serve their own purposes (Costanza-Chock, 2020; Miller et al., 2021). Research consistently shows that children are not a monolithic group — they differ across intersecting identities, including race, gender, context, and age, each bringing distinct needs and experiences (Moreno et al., 2022). For example, studies have found that gender, sexuality, race, and ability can significantly shape how children experience online risks and harms (Ito et al., 2020; Lala et al., 2022).

Children's concerns about online safety are often connected to their everyday experiences and the specific risks they face in their online interactions (Marsden et al., 2022). Therefore, tailoring user interfaces to adapt detection and interventions to their particular needs and concerns, including real-life examples and age-appropriate language in different cultural and contextual settings, can have a greater impact on them. Evidence provides some guidance on key considerations for customizing user interfaces to effectively engage diverse groups of children.

2.1 Child-Centered Design

Design Approaches for Children's Specific Needs: Research suggests that understanding children's digital technology use requires design approaches that capture distinct profiles within study populations — encompassing differences in age, interests, capabilities, and needs (Blomberg et al., 2022) — while also balancing the risks and benefits of digital engagement (Moreno et al., 2022). For instance, designers can craft culturally relevant scenarios and integrate multilingual support when introducing new features or teachable moments, fostering a positive environment for learning and open communication (Kim et al., 2021). Moreover, portable and interactive devices such as tablets serve as particularly effective tools for specific population groups, such as those in early childhood and children with autism (Papadakis et al., 2018; Peña et al., 2024; Tamaral et al., 2025). These devices enable these groups of children to conduct daily activities independently and acquire new skills, particularly in communication and emotional learning (Groba et al., 2021; Tamaral et al., 2025).

Co-Designing with and for Children: Several studies have suggested that involving children in the co-design of digital technologies provides opportunities for children and adults to act as more equal partners, positioning children as essential participants in the development process (Blomberg et al., 2022). Narrative design activities – like storytelling, gaming, and drawing – help children act out, draw, and talk about their life experiences during the design process of devices and apps (Blomberg et al., 2022; Uğraş et al., 2022; Quayyum, 2025). Through these stories, children express how they want to design interfaces that support their learning about cybersecurity and privacy through game narratives (Uğraş et al., 2022). They also explore online security scenarios, consider possible outcomes, and reflect on the role of adults in creating a safer digital environment (Quayyum, 2025).

Several studies emphasize the importance of incorporating children's perspectives in the design of digital technologies, advocating for a more participatory approach where children are seen as active contributors rather than passive users (Boston Children's Digital Wellness Lab, n.d.; Thabrew et al., 2018; Kumar et al., 2023). Child users perform as testers and informants, providing direct feedback on usability, identifying challenges in navigation, and suggesting improvements to enhance engagement and accessibility (Blomberg et al., 2022; Druin, 2002). For instance, when co-designing parental controls, children expressed preferences for designs that not only enforce security but also educate them on risk management and encourage open communication with parents/caregivers (Quayyum, 2025). The iterative co-design process ensures that their input shapes both usability and narrative elements, leading to more effective and age-appropriate digital products (Blomberg et al., 2022). Moreover, developers can leverage user data to complement usability testing, tailoring personalized safety advice and optimizing engagement strategies to make security and privacy education more meaningful for young users.

Stakeholder-Centric Approach: Designing technology for children demands a stakeholder-centered approach that acknowledges the roles of parents/caregivers, educators, and peers while keeping children's needs and preferences front and center (Kumar et al., 2023; Quayyum, 2025). Designers must strike a balance between these influences to create digital tools that are both functional and empowering. For example, interactive applications help children connect with others, share personal stories, and build a sense of community, sometimes extending to interactions with researchers and social workers (Blomberg et al., 2022). In addition to learning and socialization, children rely on adult support to navigate online safety, particularly regarding cybersecurity risks and protective measures (Quayyum, 2025). By integrating stakeholder perspectives while prioritizing children's agency, designers can craft meaningful, inclusive, and developmentally appropriate digital experiences.

2.2 Interactive Interfaces

Customizable and Co-Designed Digital Interfaces For Children: Designing digital interfaces for children requires flexibility and customization, allowing each child to engage with digital interfaces that best support their learning and well-being. Interfaces should adapt to different user groups, offering tailored designs and language styles that align with children's learning preferences and cognitive abilities. Co-design features further enhance user engagement by enabling children to provide context cues that influence both content and delivery — for instance, selecting local dialects, adjusting interface settings, or modifying accessibility features to better meet their needs. Research underscores the importance of customization, identifying key modifiable elements such as images, sound, reinforcement strategies, text, colors, screen layout, and content (Groba et al., 2021). Tailored interfaces also serve to integrate protective measures to safeguard children's privacy and well-being. Features such as cool-down periods for negative messaging and visually appealing time-limit reminders can help mitigate risks while fostering healthier digital habits (OECD, 2024; Yadav & Chakraborty, 2022). By combining co-design principles, personalized learning pathways, and embedded safety features, designers can create digital experiences that are engaging, educational, and developmentally appropriate while ensuring a secure and supportive online environment.

Interactive Games: Interactive games and multimedia tools provide engaging and effective ways to teach children about online safety, privacy, and security. Using an inductive, embodied learning approach, these resources help child users recognize digital risks and develop critical thinking skills in real-world online interactions. A range of educational tools — including games, comics, mobile apps, and social media simulators — enhance children's understanding of cybersecurity (Kumar et al., 2023). For example, interactive storytelling programs introduce Internet of Things (IoT) security concepts, while mobile apps and online games offer hands-on experiences that reinforce safe online behaviors. Designed to accommodate diverse learning styles and cognitive levels, these tools make online safety education more accessible and engaging. By integrating interactive elements and real-world scenarios, digital safety resources empower children to navigate online risks, protect their personal information, and develop responsible digital habits (Quayyum, 2025).

2.3 Language

Intuitive and Familiar Language: Digital devices and software should use familiar language — words, phrases, and concepts drawn from children's real-world experiences — to enhance comprehension and encourage engagement with safety features. Poorly designed or unfamiliar terminology can deter children from using security tools, potentially leaving them vulnerable to risks. Research shows that when graphic design elements reflect familiar attributes, children more easily and accurately recognize situations and tasks (Ahmad, 2020; Stålberg et al., 2016). Likewise, using relatable characters and real-life examples in educational resources boosts intrinsic motivation and engagement (Quayyum, 2025). Incorporating diverse languages and dialects further improves accessibility, ensuring inclusivity for children from different linguistic and cultural backgrounds. Given that privacy and security are complex, context-specific concepts (Kumar et al., 2023), designing with linguistic and cultural familiarity in mind makes these critical topics more approachable and actionable.

Uplifting Language: Instead of relying solely on warnings and "red alerts," digital safety education should incorporate empowering strategies that emphasize strengths. Framing security measures as "power-ups" rather than threats fosters agency, helping children feel more in control of their online experiences (Wong-Villacres et al., 2020). Research reinforces this perspective, recommending that successful interventions, particularly for children with autism, focus on individual strengths (Groba et al., 2021). Studies also show that children often view online safety in extremes — either implementing protective measures successfully or suffering severe consequences — without fully grasping the complexities and trade-offs involved (Quayyum, 2025). To bridge this gap, online safety education should not only teach children about risks but also help them recognize opportunities for agency, informed decision-making and digital resilience (Hinduja, 2020; Patchin & Hinduja, 2013). By shifting from fear-based messaging to an approach centered on empowerment and problem-solving, digital safety initiatives can equip children to navigate the digital environment with confidence and competence.

Age-Appropriate Language: To help children navigate digital safety features effectively, interfaces should use age-appropriate language and tailored reporting mechanisms (Livingstone & Sylwander, 2025). For instance, chat boxes and reporting tools should be designed with language and interaction styles suited to different age groups, ensuring clarity and engagement. Beyond accessible reporting, children also need to understand how digital systems work — their functionalities, how information spreads, what data is collected, and how data is stored (Quayyum et al., 2021). To support this, designers should present privacy policies in child-friendly formats, simplify user guidelines (Information Commissioner's Office, n.d.), and integrate instructional notes within the system and/or as separate manuals.

2.4 Accessibility

Minimal and Direct Interface: While offering children and parents/caregivers options is important, digital interfaces should avoid overwhelming users with excessive choices that lead to decision fatigue and distraction. A streamlined, intuitive design allows children to report harmful behavior easily — by tapping a feature directly on the screen — without unnecessary complexity. Research suggests that a child's ability to navigate digital interfaces depends more on prior experience with technology than on age (Stålberg et al., 2016). Among touch-screen devices, iPads have become increasingly ubiquitous in education, widely used by preschoolers and older children alike (Kaulanov & Kazimova, 2024; Mann et al., 2025; Otterborn et al., 2018). Their customizable, intuitive nature synthesizes multiple technologies into a single, accessible activity center (Markopoulos & Bekker, 2003; Papadakis et al., 2018; Peña et al., 2024). Case studies highlight their value as alternative communication platforms for children with autism and apraxia, where traditional methods may fall short (Flewitt et al., 2014; Tamaral et al., 2025). To enhance usability, interface design should prioritize simplicity, ensuring that information is displayed clearly and accessibly (Groba et al., 2021). By creating direct, easy-to-navigate reporting tools and minimizing cognitive load, digital platforms can empower children to engage safely and confidently with technology.

Multimodal Approaches: Safety tools should give children the flexibility to choose their preferred method based on their communication abilities and comfort level. While text-based reporting offers clarity, it may present language barriers, whereas images and icons are more universally recognizable but still carry cultural limitations. Effective reporting mechanisms should enable children to upload screenshots, record audio or video explanations, or select from pre-set response options to ensure clarity and ease of use. Providing both quick-reporting features and deeper explanations accommodates varying levels of understanding and urgency. Safety tools should also be intuitive and designed to minimize reliance on others. A dedicated button for reporting risky situations, for instance, allows children to act immediately without navigating complex menus (Muthu et al., 2025). Accessibility considerations are essential, particularly for children with disabilities, who may benefit from multimodal communication tools that combine audio, visual, and tactile elements (Groba et al., 2021). Evidence suggests that platforms such as Autcraft, a Minecraft community for autistic children, highlight the importance of inclusive design tailored to sensory preferences (Tamaral et al., 2025). Multimodal displays — incorporating both visual and auditory elements — enhance comprehension and engagement, making digital safety features more effective for all users. At the same time, designers must address notification fatigue to ensure that children remain responsive to important alerts.

Visibility, Confidentiality, and Trust in Reporting Mechanisms: Effective interfaces for children must balance visibility and confidentiality in reporting mechanisms. Children need a safe space to seek help and report harmful risks and harms without fearing shame, yet they also deserve clear communication about how their reports are handled. Transparent user interfaces should use friendly language to explain how reports are processed, how long information is retained, and what actions may follow. Providing updates on report analysis and outcomes reassures children that their concerns matter, bolstering their confidence in the system. Reporting mechanisms should also facilitate parent/caregiver-child communication. Children prefer a collaborative approach in which parents/caregivers act as guides rather than as enforcers of rigid monitoring systems (Akter et al., 2022; Badillo-Urquiola et al., 2019; Ghosh et al., 2018; Stoilova et al., 2023; Park et al., 2024; Theopilus et al., 2024; Wisniewski et al., 2017). While children seek parental/caregiver help in navigating online risks, they also value their independence and want to take an active role in their own digital safety (Quayyum, 2025). Reporting mechanisms that encourage parent/caregiver-child collaboration empower children to develop autonomy while maintaining open communication with their parents/caregivers.

Safety Centers: A robust digital safety framework should include clear, age-appropriate Safety Center articles designed for both children and parents/caregivers, explaining product tools and how they address specific risks. Machine learning tools can enhance protection by analyzing interactions, identifying potential dangers, and directing children to relevant resources. In addition to child-focused support, these Safety Center resources should provide parents/caregivers with guidance on digital safety, equipping them with strategies to support their children effectively. Educational materials should cover key topics such as using appropriate language when discussing online risks, reporting concerns in a way that preserves children's privacy and agency, and involving parents/caregivers or law enforcement in a way that is proportionate and non-aggravating (AlShabibi & Al-Suqri, 2021). To further strengthen support, platforms should offer direct access to local Safer Internet Centers (Apple, 2025), allowing children and parents/caregivers to connect with trained professionals who can provide advice, intervention, and additional resources.

Evaluating Educational and UI-Design Approaches

Drawing upon the comprehensive review of findings presented throughout this chapter, the Working Group is well-positioned to offer an assessment of core educational approaches, identifying robustly supported interventions and outlining significant knowledge gaps.

Effective digital safety education extends beyond mere knowledge transmission, emphasizing active engagement, practical skill-building, and the promotion of children's agency through participatory frameworks that accommodate diverse developmental trajectories, socio-cultural backgrounds, and real-world contexts. Educational approaches play a crucial role not only in prevention but also in empowering children to detect risks effectively and respond appropriately when facing online harms.

Current practices strongly supported by evidence include holistic educational programs that integrate digital safety into broader contexts such as mental health, bullying prevention, relationship education, and sexuality education. Integrated approaches consistently demonstrate higher efficacy compared to isolated interventions because they resonate deeply with children's daily experiences, developmental needs, and social realities. Evidence specifically highlights programs that connect digital safety with essential life skills such as emotional intelligence, critical thinking, and interpersonal relationships, significantly enhancing children's understanding, skill acquisition, and agency to adopt safer online behaviors.

Participatory co-design approaches emerge as another robustly supported strategy. Active involvement of children in creating digital safety tools and educational content has proven highly effective in enhancing relevance, usability, and sustained engagement. By engaging children as active contributors rather than passive recipients, these approaches effectively leverage children's insights and creativity and reinforce agency and long-term adoption of safety behaviors. Additionally, open dialogue facilitated by parents/caregivers and trusted adults have demonstrated significant value, providing supportive contexts for children to express concerns, ask questions, and collaboratively address online challenges.

Moreover, substantial evidence supports initiating digital safety education early in childhood, emphasizing repeated and consistent exposure delivered through multiple shorter sessions rather than single or infrequent interventions. Educational venues include diverse environments such as schools, community centers, homes, and digital platforms, underscoring the importance of accessible, consistent messaging across various settings. Frequent reinforcement, interactive engagement, and clear, positively framed communication significantly improve knowledge retention, reduce anxiety, and promote intrinsic motivation, thereby fostering healthier and safer online interactions.

The chapter notably emphasizes evidence regarding vulnerable populations, indicating that educational interventions must explicitly account for heightened risks faced by marginalized or at-risk groups, including children with disabilities or those experiencing social isolation. Tailored and inclusive interventions designed specifically to address these vulnerabilities are essential for equitable safety outcomes.

Despite these robust findings, several important research gaps remain. Firstly, empirical evidence is limited regarding the most effective methods for integrating digital safety education and digital skills programs into existing curricula, such as mental health programs or sexuality and relationship education, posing significant practical challenges for broad implementation. For instance, while many stakeholders have developed frameworks and accompanying materials for digital skill programs intended for formal or informal learning settings, there is little published data measuring the actual efficacy of these efforts (Cortesi et al., 2020). Such a gap highlights a critical need for more rigorous evaluation of such initiatives.

Secondly, personalized interventions utilizing algorithmic decision-making and advanced behavioral analytics offer considerable potential but raise substantial ethical concerns. The efficacy, potential biases, and privacy implications of personalized approaches remain insufficiently understood and warrant rigorous empirical examination.

Lastly, cross-cultural applicability, multilingual accessibility, and inclusivity in educational design remain underexplored, highlighting the urgent need for culturally sensitive, contextually appropriate, and inclusive research. Addressing these critical knowledge gaps — particularly curricular integration methods, ethically sound personalized interventions, and cross-cultural inclusivity — constitutes an essential frontier requiring collaborative efforts among diverse stakeholders, including parents/caregivers, educators, community leaders, technology companies, and policymakers.

EXPANDING THE FRONTIERS OF DIGITAL CHILD SAFETY

(IV)

EXPANDING THE FRONTIERS OF DIGITAL CHILD SAFETY

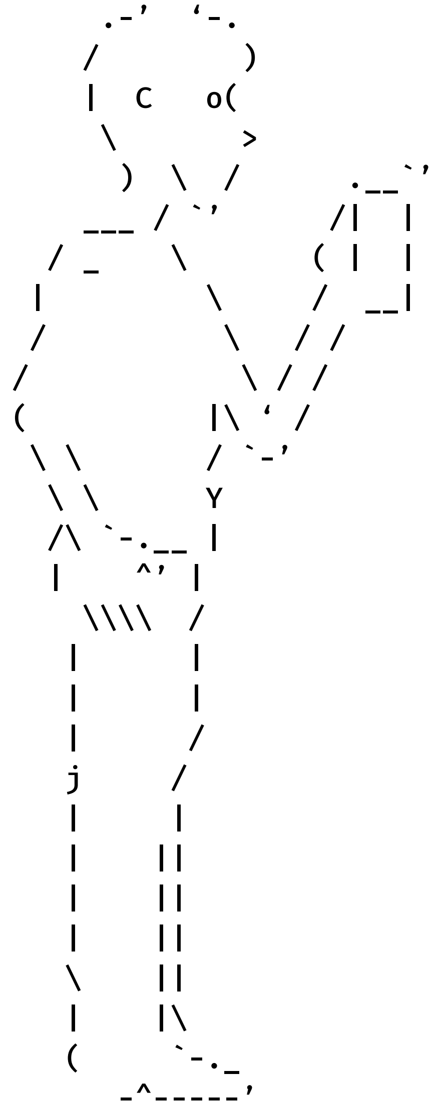
The Frontiers in Digital Safety Working Group set out to critically reevaluate conventional approaches to children's digital safety — shifting the focus from restrictive, control-based approaches to a proactive design framework that embeds safety into the very fabric of the digital environment.

Traditional measures — such as smartphone bans, age limits, and strict screen time limits — often fail to reflect the diversity and complexity of children's digital lives. These approaches are frequently based on assumptions that lack robust empirical evidence (Mansfield et al., 2025; Vuorre & Przybylski, 2024) and can unintentionally erode trust between children and parents/caregivers by placing the full burden of responsibility on families while leaving technology developers largely unaccountable (Citron & Waldman, 2025).

Grounded in the normative, substantive, and procedural principles outlined in this report, digital child safety should not be imposed from the outside. Instead, it should be embedded into the design of the digital environment in ways that promote access, empower children to develop relevant skills, and equip them with the tools and resources needed to manage risks and harms effectively. Moving beyond rigid, one-size-fits-all solutions, an adaptive, participatory, and context-sensitive design process offers a more sustainable and equitable path toward digital child safety.

This perspective guided the Working Group's exploration of four complementary approaches to digital child safety: (1) design approaches that foster trust, (2) help-seeking and reporting approaches, (3) on-device approaches, and (4) educational and user-interface design approaches. All four approaches are grounded in shared foundational principles — child rights, agency, and well-being, and actively centering children's perspectives throughout development and implementation.

In conclusion, the following sections synthesize cross-cutting themes identified across the four approaches, highlight promising directions for future research, and propose actionable priorities for advancing digital child safety. By integrating these insights into design strategies and fostering collaborative partnerships, digital child safety efforts can move beyond control-based approaches toward solutions that genuinely enable children to thrive in the digital environment.



1. Cross-Cutting Insights in Digital Child Safety

While each approach in the report is unique and distinct, several overarching themes emerge that cut across all four approaches to digital child safety. These shared insights could provide a promising path forward for parents/caregivers, educators, community leaders, technology companies, policymakers, and children seeking to identify actionable takeaways from this report.

Participation: Children are not merely passive users within the digital environment but active participants whose engagement varies in depth and form across different contexts. Effective digital child safety measures — such as parental controls, real-time warnings, and AI-driven threat detection — become more impactful when combined with genuine dialogue and meaningful, dynamic participation by children themselves. Recognizing participation as an inclusive and dynamic process rather than a linear or hierarchical one, it is essential to involve children in shaping their own safety, respecting their diverse experiences, perspectives, and degrees of engagement. By fostering open, ongoing adult-child communication that actively seeks children's perspectives, we promote their agency, enabling them to navigate the digital environment confidently and responsibly. Integrating digital safety education within broader discussions around bullying prevention, mental health, and digital citizenship ensures that participation is contextualized and relevant, further empowering children in their digital experiences.

Proactive and Adaptive Design: A proactive approach to digital safety — leveraging behavioral design, real-time safety nudges, and accessible reporting tools — would ensure that children receive timely and tailored support. Asset-based approaches, which focus on reinforcing positive behaviors rather than just mitigating risks, empower children to engage with the digital environment safely. Strengthened reporting systems, peer support networks, and trusted adult guidance help create a robust safety net, making it easier for children to seek help when needed. By continuously refining tools and interventions based on user feedback, digital child safety solutions can remain dynamic, effective, and responsive to evolving online risks.

Inclusivity and Culturally Relevant Strategies: Ensuring that digital child safety measures are accessible to all children — regardless of background, ability, or context — is essential for equitable safety. Inclusive design practices, such as offering diverse interface options (e.g., audio, visual, and tactile elements), can broaden the range of children who are able to navigate online spaces safely. Culturally relevant and multilingual support further enhances the effectiveness of digital child safety education by making guidance more relatable, trustworthy, and impactful across diverse communities. By designing digital tools that respect different cultural norms and lived experiences, we can create inclusive, adaptive solutions that meet the needs of more children.

Early Intervention and Prevention: Preventing harm before it occurs or escalates is a core principle of effective digital child safety strategies across all four approaches. Whether through education, real-time detection tools, or user-friendly reporting mechanisms, early intervention helps to increase the likelihood that risks are identified and addressed proactively. AI-driven early-warning systems and privacy-conscious safety tools offer protection while maintaining children's autonomy. Simplified, intuitive safety features — developed in collaboration with educators, policymakers, and technology companies — make protective measures more accessible and user-friendly. Drawing insights from diverse safety domains, including those offline, helps refine interventions, ensuring they are adaptable, culturally relevant, and responsive to emerging risks and harms.

Shared Responsibility and Collective Action: Digital child safety requires a collective commitment from all stakeholders, including children, parents/caregivers, educators, policymakers, and technology companies. Co-designing safety solutions with children ensures that protective tools are effective, engaging and relevant to their lived experiences. Public awareness initiatives and school-based programs play a critical role in fostering digital skills, helping to equip children with the ability to navigate the digital environment safely. Beyond just protection, these initiatives empower children to explore, connect, and thrive in the digital environment.

2. Sustaining Progress in Child Safety

As the digital environment continues to expand, so do the opportunities to develop new, forward-looking strategies that empower children and strengthen their digital resilience. Research, policymaking, education, and design all play a role in shaping the future of digital child safety, ensuring that interventions are not only effective but also accessible, adaptable, and aligned with children's real-world experiences. Several key areas that cut across all four approaches stand out as particularly important for further exploration and innovation.

Empowering Children Through Digital Tools: A key challenge is understanding how children engage with safety tools — not just whether they use them, but how these tools can support their agency rather than limit their experiences. Many current safety measures focus on restricting access rather than equipping children with the skills and resources to navigate risks. There is growing interest in developing safety features that provide real-time guidance, reinforce positive behaviors, and encourage safe decision-making rather than simply blocking content or enforcing rigid controls. Exploring how digital tools can better support resilience, autonomy, and self-protection while maintaining necessary safeguards is an important focus for future research, policy, and design.

Strengthening Co-Creation and Collaboration: Digital child safety is most effective when it reflects the needs and perspectives of those it aims to protect. Engaging children directly in the design of child safety tools — through participatory design processes — can help create solutions that are intuitive, relevant, and widely used. Beyond children, collaboration across sectors — parents/caregivers, educators, community leaders, technology companies, and policymakers — can lead to more holistic and adaptable approaches. There are opportunities to strengthen these partnerships, ensuring that safety strategies are informed by diverse perspectives and that families are supported in fostering open conversations about online risks and responsibilities.

Fostering Trust: Trust is central to the success of any digital child safety interventions. Children and parents/caregivers need clear information about how safety tools work, what data is collected, and how decisions are made — especially when AI-driven interventions are involved. At the same time, there is increasing interest in making safety tools more responsive to individual needs by adapting features based on age, digital experience, or cultural background. While personalization can enhance safety and engagement, it also raises important questions about privacy, fairness, and inclusivity. Finding the right balance between transparency, adaptability, and data protection remains an ongoing area of exploration.

Preparing for Emerging Risks and Evolving Technologies: As new technologies like AI-generated content (e.g., deepfakes) and extended reality become more common, embedding proactive safety-by-design features will be essential to ensure these tools foster positive digital engagement rather than amplifying harm. These innovations bring novel challenges, from misinformation and manipulation to new forms of online interaction that may not fit traditional safety frameworks. Long-term studies and cross-cultural perspectives can help anticipate emerging risks and ensure that safety tools remain relevant across different contexts. Ensuring that AI-driven detection and intervention systems are both fair and effective will also be critical in reducing harm while respecting children's rights.

Ongoing Evaluation and Adaptation: No single solution will remain effective indefinitely. As the digital environment changes, so must the strategies used to protect and support children. There is a growing need for continuous evaluation of safety interventions, ensuring they remain practical, inclusive, and aligned with real-world challenges. Cross-sector collaboration can help identify what works, what needs to evolve, and how different approaches can be refined over time. By fostering a culture of ongoing learning and adaptation, digital child safety efforts can better support children in an ever-changing online world.

3. Onward: Strengthening Digital Child Safety by Design

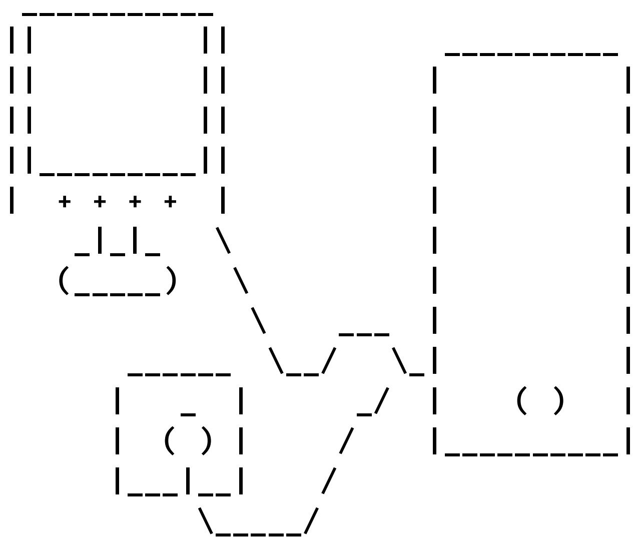
This report has examined the current challenges and opportunities in digital child safety, highlighting the need for approaches that go beyond restrictions and focus on supporting children's rights, agency, and well-being in the digital environment. Rather than relying on measures that attempt to control children's digital experiences, there is value in designing environments that help them navigate risks, build resilience, and make informed decisions.

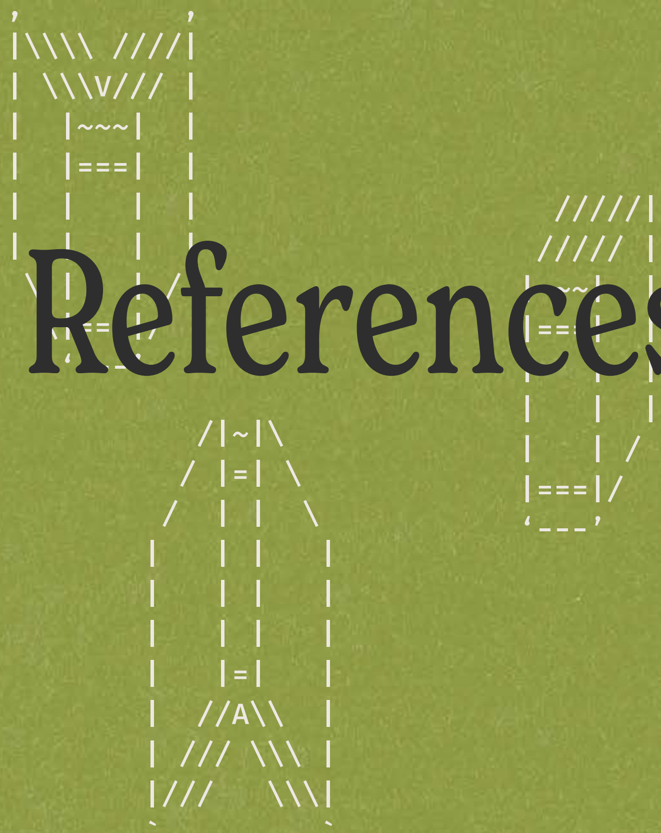
Digital child safety should be seen as an ongoing process that adapts to the realities of children's lives. Centering their perspectives in research, policy, and design ensures that safety tools are both effective and relevant. When integrated thoughtfully into technology, education, and regulation, these measures can provide meaningful support rather than imposing limitations that may not address the root challenges.

Grounding digital child safety efforts in strong evidence is essential. Ongoing research is needed to better understand how children interact with online risks and protective measures, ensuring that interventions are both effective and responsive to their needs. Policies and design choices should be informed by data, real-world testing, and interdisciplinary collaboration, rather than assumptions or reactive measures that may have unintended consequences.

Creating a safer digital environment requires collaboration across sectors, including parents/caregivers, educators, technology companies, and policymakers. As new technologies continue to reshape online interactions, continuous evaluation and adaptation will be essential in keeping safety strategies effective and responsive.

This report offers insights to guide future work in digital child safety, emphasizing the importance of transparency, shared responsibility, and proactive design. By focusing on these principles and grounding action in research, the digital environment can become places where children are both protected and supported in their growth, exploration, and participation.





References

Abades Barclay, F., & Banaji, S. (2024). Teaching digital citizenship in the UK: London School of Economics evaluation of the Common Sense digital citizenship curriculum. Common Sense Media & London School of Economics. https://www.commonsensemedia.org/sites/default/files/uploads/pdfs/files/lse-digital-citizenship-study-common-sense.pdf

Agha, Z., Ali, N. S., Park, J., & Wisniewski, P. J. (2025). A systematic review on design-based nudges for adolescent online safety. International Journal of Child-Computer Interaction, 43, 100702. https://doi.org/10.1016/j.ijcci.2024.100702

Ahmad, J. (2020). Children's recognition of words in familiar logos. International Journal of Early Years Education, 29(1), 75–87. https://doi.org/10.1080/09669760.2020.1779042

Akter, M., Godfrey, A. J., Kropczynski, J., Lipford, H. R., & Wisniewski, P. J. (2022). From parental control to joint family oversight: Can parents and teens manage mobile online safety and privacy as equals? Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1), 1–28. https://doi.org/10.1145/3512904

Alghythee, K. K. A., Hrncic, A., Singh, K., Kunisetty, S., Yao, Y., & Soni, N. (2024). Towards understanding family privacy and security literacy conversations at home: Design implications for privacy-literacy interfaces. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1–12). https://doi.org/10.1145/3613904.3641962

Ali, S., Elgharabawy, M., Duchaussoy, Q., Mannan, M., & Youssef, A. (2020). Betrayed by the guardian: Security and privacy risks of parental control solutions. In Proceedings of the 36th Annual Computer Security Applications Conference (pp. 69–83). https://doi.org/10.1145/3427228.3427287

Anderson, C., Crete-Nishihata, M., Dehghanpoor, C., Ronald, J. D., McKune, S., Ottenheimer, D., & Scott-Railton, J. (2015). Are the kids alright? University of Toronto.

Anwar, A., & Kanjo, E. (2023). User-centred detection of violent conversations on mobile edge devices. In Distributed, Ambient and Pervasive Interactions (pp. 335–346). https://doi.org/10.1007/978-3-031-34609-5_25

APAC Ad Junkie. (2019, November 8). Korean Air inflight safety video goes K-Pop with SuperM and BoA. Branding in Asia. https://www.brandinginasia.com/korean-air-inflight-safety-video-goes-k-pop-with-super-m-and-boa

Apple. (2023, February 7). Apple marks Safer Internet Day by spotlighting features and tools to protect children online. Apple Newsroom. https://www.apple.com/uk/newsroom/2023/02/apple-spotlights-free-resources-to-protect-children-online/

Araujo, C. S., Magno, G., Jr, W. M., Almeida, V., Hartung, P., & Doneda, D. (2017). Characterizing videos, audience and advertising in YouTube channels for kids. ArXiv.cs.SI. https://doi.org/10.48550/arXiv.1707.00971

Archard, D. (2015). Children, adults, autonomy and well-being. In A. Bagattini & C. Macleod (Eds.), The nature of children's well-being: Theory and practice (pp. 3–14). Springer. https://doi.org/10.1007/978-94-017-9252-3_1

Arrúa, A., Machín, L., Curutchet, M. R., Martínez, J., Antúnez, L., Alcaire, F., Giménez, A., & Ares, G. (2017). Impact of front-of-pack nutrition information and label design on children's choice of two snack foods: Comparison of warnings and the traffic-light system. Appetite, 116, 139–146. https://doi.org/10.1016/j.appet.2017.04.012Asare, S., Agyeman, E. P., Ahmoah, J. D., & Asare, B. O. (2023). A systematic review of the role of social media in providing guidance and support to adolescents: A case study of online counseling platforms. American Journal of Multidisciplinary Research and Innovation, 2(6), 42–48. https://doi.org/10.54536/ajmri.v2i6.2213

Assis, J. V., & Valença, G. (2024). Is my child safe online? On requirements for parental control tools in apps used by children. Journal on Interactive Systems, 15(1), 823–838. https://doi.org/10.5753/jis.2024.4240

Badillo-Urquiola, K., Smriti, D., McNally, B., Golub, E., Bonsignore, E., & Wisniewski, P. J. (2019). Stranger danger!: Social media app features co-designed with children to keep them safe online. In Proceedings of the 18th ACM International Conference on Interaction Design and Children (pp. 394–406). https://doi.org/10.1145/3311927.3323133

Barbosa, A. F. (2014). ICT Kids Online Brazil 2014: Survey on the internet use by children in Brazil. EU Kids Online. https://www.lse.ac.uk/media-and-communications/assets/documents/research/eu-kids-online/participant-countries/latin-america/BR-ICT-2014.pdf

Baumgartner, T., Lobmaier, J. S., Ruffieux, N., & Knoch, D. (2021). Feeling of guilt explains why people react differently to resource depletion warnings. Scientific Reports, 11(1), 11988. https://doi.org/10.1038/s41598-021-91472-0

Beltrán, M., & de Salvador, L. (2024). Implications of age assurance on privacy and data protection: A systematic threat model. In Privacy Technologies and Policy (pp. 1–22). https://doi.org/10.1007/978-3-031-68024-3_1

Ben-Arieh, A., Casas, F., Frønes, I., & Korbin, J. (2014). Multifaceted concept of child well-being. In A. Ben-Arieh, F. Casas, I. Frønes, & J. Korbin (Eds.), Handbook of child well-being: Theories, methods and policies in global perspective (pp. 1–27). Springer. https://doi.org/10.1007/978-90-481-9063-8_134

Berson, I. R., & Berson, M. J. (2005). Challenging online behaviors of youth: Findings from a comparative analysis of young people in the United States and New Zealand. Social Science Computer Review, 23(1), 29–38. https://doi.org/10.1177/0894439304271532

Bhaimiya, S. (2024, July 17). A growing number of parents are refusing to give their children smartphones — and the movement is going global. CNBC. Retrieved December 13, 2024, from https://www.cnbc.com/2024/07/17/a-smartphone-free-childhood-a-global-movement-is-growing.html

Bickham, D. S., Hswen, Y., Slaby, R. G., & Rich, M. (2018). A preliminary evaluation of a school-based media education and reduction intervention. The Journal of Primary Prevention, 39(3), 229–245. https://doi.org/10.1007/s10935-018-0510-2

Bickham, D. S., & Rich, M. (2006). Is television viewing associated with social isolation? Roles of exposure and motivations. Developmental Psychology, 42(2), 491–501.

Blomberg, H., Östlund, G., Lindstedt, P. R., & Cürüklü, B. (2022). Children helping to co-construct a digital tool that is designed to increase children's participation in child welfare investigations in Sweden. Qualitative Social Work, 21(2), 367–392. https://doi.org/10.1177/1473325021990864

Blumenfeld, W. J. (2020, November 3). LGBTQ cyberbullying: Guide for youth, parents & educators. Connect Safely. https://connectsafely.org/lgbtq/Bordalba, M. M., & Bochaca, J. G. (2019). Digital media for family-school communication? Parents' and teachers' beliefs. Computers & Education, 132, 44–62. https://doi.org/10.1016/j.compedu.2019.01.006

Borj, P. R., Raja, K., & Bours, P. (2023a). Detecting online grooming by simple contrastive chat embeddings. In Proceedings of the 9th ACM International Workshop on Security and Privacy Analytics (pp. 57–65). https://doi.org/10.1145/3579987.3586564

Borj, P. R., Raja, K., & Bours, P. (2023b). Online grooming detection: A comprehensive survey of child exploitation in chat logs. Knowledge-Based Systems, 259, 110039. https://doi.org/10.1016/j.knosys.2022.110039

Boston Children's Digital Wellness Lab. (n.d.). The inspired internet pledge. https://inspiredinternet.org

Boston Children's Hospital. (2023). Post-traumatic stress disorder (PTSD). Boston Children's Hospital.

Bridgland, V. M. E., & Takarangi, M. K. T. (2021). Danger! Negative memories ahead: The effect of warnings on reactions to and recall of negative memories. Memory, 29(3), 319–329.

Bridgland, V. M. E., Bellet, B. W., & Takarangi, M. K. T. (2022). Curiosity disturbed the cat: Instagram's sensitive-content screens do not deter vulnerable users from viewing distressing content. Clinical Psychological Science, 11(2), 1–18. https://doi.org/10.1177/21677026221097618

Bright, M. A., Huq, M. S., Miller, M. D., Patel, S., Li, Z., & Finkelhor, D. (2023). Randomized control trial of a school-based curriculum that teaches about multiple forms of abuse. Child Maltreatment, 29(2), 364–374. https://doi.org/10.1177/10775595231152623

Brochado, S., Soares, S., & Fraga, S. (2016). A scoping review on studies of cyberbullying prevalence among adolescents. Trauma, Violence, & Abuse, 18(5), 523–531. https://doi.org/10.1177/1524838016641668

Brod, G., Kucirkova, N., Shepherd, J., Jolles, D., & Molenaar, I. (2023). Agency in educational technology: Interdisciplinary perspectives and implications for learning design. Educational Psychology Review, 35(1), 1–23. https://doi.org/10.1007/s10648-023-09749-x

Brumen, B., Göllner, S., & Tropmann-Frick, M. (2023). Aspects and views on responsible artificial intelligence. In G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, P. Pardalos, G. Di Fatta, G. Giuffrida, & R. Umeton (Eds.), Machine Learning, Optimization, and Data Science (pp. 384–398). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25599-1_29

Brummernhenrich, B., Paulus, C. L., & Jucks, R. (2025). Applying social cognition to feedback chatbots: Enhancing trustworthiness through politeness. British Journal of Educational Technology, Advance online publication. https://doi.org/10.1111/bjet.13569

Byrne, J., Kardefelt Winther, D., Livingstone, S., & Stoilova, M. (2016). Global Kids Online research synthesis, 2015–2016. http://globalkidsonline.net/wp-content/uploads/2016/11/Synthesis-report_07-Nov-2016.pdf

Byron, P., Robards, B., Hanckel, B., Vivienne, S., & Churchill, B. (2019). "Hey, I'm having these experiences": Tumblr use and young people's queer (dis)connections. International Journal of Communication, 13, 2239–2259.

Cabrera, M., Machín, L., Arrúa, A., Antúnez, L., Curutchet, M. R., Giménez, A., & Ares, G. (2017). Nutrition warnings as front-of-pack labels: Influence of design features on healthfulness perception and attentional capture. Public Health Nutrition, 20(18), 3360–3371. https://doi.org/10.1017/S136898001700249X

Calvin, A., Hasse, A., Madden, M., & Lenhart, A. (2024). Getting help online: How young people find, evaluate, and use mental health apps, online therapy, and behavioral health information. Common Sense Media. https://www.commonsensemedia.org/sites/default/files/research/report/2024-getting-help-online-hopelab-report_final-release-for-web.pdf

Campbell, R., & Raja, S. (1999). Secondary victimization of rape victims: Insights from mental health professionals who treat survivors of violence. Violence and Victims, 14(3), 261-275.

Cavoukian, A. (2011). Privacy by design: The 7 foundational principles. Information & Privacy Commissioner, Ontario, Canada. https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdfChan, J., Kishore, S., & Yang, X. (2025). A real-time cyberbully checker. Software Impacts, 23, 100710. https://doi.org/10.1016/j.simpa.2024.100710

Chicote-Beato, M., González-Víllora, S., Bodoque-Osma, A. R., & Navarro, R. (2024). Cyberbullying intervention and prevention programmes in Primary Education (6 to 12 years): A systematic review. Aggression and Violent Behavior, 77, 101938. https://doi.org/10.1016/j.avb.2024.101938

Cho, Y. J., Thrasher, J. F., Yong, H.-H., Szklo, A. S., O'Connor, R. J., Bansal-Travers, M., Hammond, D., Fong, G. T., Hardin, J., & Borland, R. (2018). Path analysis of warning label effects on negative emotions and quit attempts: A longitudinal study of smokers in Australia, Canada, Mexico, and the US. Social Science & Medicine, 197, 226–234. https://doi.org/10.1016/j.socscimed.2017.10.003

CISA & USSS. (2021). Improving school safety through bystander reporting: A toolkit for strengthening K-12 Reporting Programs. Department of Homeland Security. https://www.secretservice.gov/sites/default/files/reports/2023-05/cisa-usss-k-12-bystander-reporting-toolkit-508_final_0.pdf

Coiera, E. (2003). Guide to health informatics. Arnold Publishers.

Coiera, E. (2018). Guide to health informatics (2nd ed.). Arnold Publishers.

Collier, A. (2012). A 'Living Internet': Some context for the cyberbullying discussion. In J. W. Patchin & S. Hinduja (Eds.), Cyberbullying prevention and response: Expert perspectives (pp. 1–12). Routledge.

Collier, A. (2013, September 4). Challenging 'Internet safety' as a subject to be taught. Net Family News. https://www.netfamilynews.org/challenging-internet-safety-as-a-subject-to-be-taught

Collins, S., Park, J., Reddy, A., Sharifi, Y., & Vance, A. (2021). The privacy and equity implications of using self-harm monitoring technologies: Recommendations for schools. Future of Privacy Forum. https://publicinterestprivacy.org/privacy-equity-self-harm-monitoring/

Committee on the Rights of the Child. (2021). General comment No. 25 (2021) on children's rights in relation to the digital environment. United Nations. https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights

Cook, D., Zilka, M., DeSandre, H., Giles, S., & Maskell, S. (2023). Protecting children from online exploitation: Can a trained model detect harmful communication strategies? In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 5–14). https://doi.org/10.1145/3600211.3604696

Cooney, C., & Standley, N. (2024, February 19). Schools in England give new guidance on stopping phone use. BBC News. Retrieved December 13, 2024, from https://www.bbc.com/news/uk-68334602

Correia, N., Aguiar, C., & Amaro, F. (2023). Children's participation in early childhood education: A theoretical overview. Contemporary Issues in Early Childhood, 24(3), 313–332. https://doi.org/10.1177/1463949120981789

Cortesi, S. (2021). Youth and the Participatory Promise [Doctoral dissertation, University of Basel]. https://edoc.unibas.ch/84346/

Cortesi, S., & Gasser, U. (2017). Children's rights and digital technologies: Introduction to the discourse and some meta-observations. In M. D. Ruck, M. Peterson-Badali, & M. Freeman (Eds.), Handbook of children's rights: Global and multidisciplinary perspectives (pp. 417–436). Routledge.

Cortesi, S., Hasse, A., Lombana-Bermudez, A., Kim, S., & Gasser, U. (2020). Youth and digital citizenship+ (Plus): Understanding skills for a digital world. https://doi.org/10.2139/ssrn.3557518

Cortesi, S., Hasse, A., Eigen, M., Maddens Toscano, P., Malik, M., & Gasser, U. (2021). Youth and extended reality (XR): An initial exploration of augmented, virtual, and mixed realities. Youth and Media, Berkman Klein Center for Internet & Society. https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37369907

Cortesi, S., Hasse, A., & Gasser, U. (2021). Youth participation in a digital world: Designing and implementing spaces, programs, and methodologies. Youth and Media, Berkman Klein Center for Internet & Society. https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37367566

Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we need. MIT Press. https://doi.org/10.7551/mitpress/12255.001.0001

Costello, N., Almassian, M., Sutton, R., Jones, M., Diamond, S., Ojumu, O., Raffoul, A., Salvia, M., Kavanaugh, J. R., & Austin, S. B. (2024, March 27). How to hold social media platforms accountable: A roadmap for state policymakers & advocates for legislation to require independent algorithm risk audits. Strategic Training Initiative for the Prevention of Eating Disorders (STRIPED), Harvard T.H. Chan School of Public Health. https://www.hsph.harvard.edu/striped/social-mediaalgorithm-auditing/

Cross, D., Lester, L., Barnes, A., Cardoso, P. & Hadwen, K. (2015). If it's about me, why do it without me? Genuine student engagement in school cyberbullying education. International Journal of Emotional Education, 7(1), 35-51.

Dailey, S. F., & Roche, R. R. (2025). The SHIELD framework: Advancing strength-based resilience strategies to combat bullying and cyberbullying in youth. International Journal of Environmental Research and Public Health, 22(1), 66. https://doi.org/10.3390/ijerph22010066

David, O. A., & Fodor, L. A. (2023). Preventing mental illness in children that experienced maltreatment: The efficacy of REThink online therapeutic game. npj Digital Medicine, 6(1), 106. https://doi.org/10.1038/s41746-023-00849-0

Dempsey, A. G., Sulkowski, M. L., Dempsey, J., & Storch, E. A. (2011). Has cyber technology produced a new group of peer aggressors? Cyberpsychology, Behavior, and Social Networking, 14(5), 297–302. https://doi.org/10.1089/cyber.2010.0108

Dempsey, J., Sim, G., Cassidy, B., & Ta, V.-T. (2022). Children designing privacy warnings: Informing a set of design guidelines. International Journal of Child-Computer Interaction, 31, 100446. https://doi.org/10.1016/j.ijcci.2021.100446

Digital Futures Commission. (2023). Child rights by design toolkit. https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf

Dinakar, K., Jones, B., Havasi, C., Lieberman, H., & Picard, R. (2012). Common sense reasoning for detection, prevention, and mitigation of cyberbullying. ACM Transactions on Interactive Intelligent Systems, 2(3), Article 19. https://doi.org/10.1145/2362394.2362400Dix, K. (2024, September 23). The results are in: How smiling mind is helping primary students. ACER. https://www.acer.org/gb/discover/article/wellbeing-program-helps-students-develop-new-skills-and-focus

Douglas, L., Jackson, D., Woods, C., & Usher, K. (2019). Peer-to-peer mentoring for and by at-risk young people. Mental Health Practice, 22(6), 21–27. https://doi.org/10.7748/mhp.2019.e1401

Dowling, M., & Rickwood, D. J. (2013). Online counseling and therapy for mental health problems: A systematic review of the research literature. Journal of Technology in Human Services, 31(1), 118–135.1–21. https://doi.org/10.1080/15228835.2012.728508

Druin, A. (2002). The role of children in the design of new technology. Behaviour & Information Technology, 21(1), 1–25. https://doi.org/10.1080/01449290210147484

Dwyer, A., de Almeida Neto, A., Estival, D., Li, W., Lam-Cassettari, C., & Antoniou, M. (2021). Suitability of text-based communications for the delivery of psychological therapeutic services to rural and remote communities: Scoping review. JMIR Mental Health, 8(2). https://doi.org/10.2196/19478

European Commission. (n.d.). Beyond GDP: Delivering sustainable and inclusive wellbeing. Retrieved June 25, 2025, from https://joint-research-centre.ec.europa.eu/projects-and-activities/beyond-gdp-delivering-sustainable-and-inclusive-wellbeing_en

eSafety Commissioner. (n.d.). Safety by design. Retrieved March 18, 2020, from https://www.esafety.gov.au/industry/safety-by-design

Estwick, M. (2025). "Mood."- the development and evaluation of a digital peer support intervention for adolescents with social, emotional, and Mental Health (SEMH) needs. [Doctoral dissertation, University College London]. UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10206305/

Fan, S., Yecies, B., Zhou, Z. I., & Shen, J. (2024). Challenges and opportunities for the Web 3.0 metaverse turn in education. IEEE Transactions on Learning Technologies, 17, 1989–2004. https://doi.org/10.1109/tlt.2024.3385505

Fenaughty, J., & Harre, N. (2013). Factors associated with young people's successful recovery from cyberbullying. [Unpublished manuscript].

Fernandes, S., Arriaga, P., & Esteves, F. (2015). Using an educational multimedia application to prepare children for outpatient surgeries. Health Communication, 30(12), 1190–1200. https://doi.org/10.1080/10410236.2014.909454

Finkelhor, D. (2007). Prevention of sexual abuse through educational programs directed toward children. Pediatrics, 120(3), 640–645. https://doi.org/10.1542/peds.2007-0754

Finkelhor, D. (2008). Childhood victimization: Violence, crime, and abuse in the lives of young people. Oxford University Press.

Finkelhor, D., Jones, L., & Mitchell, K. (2021). Teaching privacy: A flawed strategy for children's online safety. Child Abuse & Neglect, 117, Article 105064. https://doi.org/10.1016/j.chiabu.2021.105064

Finkelhor, D., Walsh, K., Jones, L., Mitchell, K., & Collier, A. (2020). Youth Internet safety education: Aligning programs with the evidence base. Trauma, Violence, & Abuse, 22(5), 1233–1247. https://doi.org/10.1177/1524838020916257

Flewitt, R., Kucirkova, N., & Messer, D. (2014). Touching the virtual, touching the real: iPads and enabling literacy for students experiencing disability. The Australian Journal of Language and Literacy, 37(2), 107–116.

Flor, L. S., Reitsma, M. B., Gupta, V., Ng, M., & Gakidou, E. (2021). The effects of tobacco control policies on global smoking prevalence. Nature Medicine, 27, 239–243. https://doi.org/10.1038/s41591-020-01210-8

Faverio, M., & Sidot, O. (2024, December 12). Teens, social media and technology 2024. Pew Research Center. https://www.pewresearch.org/internet/2024/12/12/teens-social-media-and-technology-2024/

Fosch-Villaronga, E., Van Der Hof, S., Lutz, C., & Tamò-Larrieux, A. (2023). Toy story or children story? Putting children and their rights at the forefront of the artificial intelligence revolution. AI & Society, 38(1), 133–152. https://doi.org/10.1007/s00146-021-01295-w

FOSI. (2025). Connected and protected: Insights from FOSI's 2025 online safety survey. https://fosi.org/wp-content/uploads/2025/05/Connected-and-Protected-Insights-from-FOSIs-2025-Online-Safety-Survey.pdf

Frank, J., Eisenhofer, T., Schönherr, L., Fischer, A., Kolossa, D., & Holz, T. (2020). Leveraging frequency analysis for deep fake image recognition. arXiv. https://doi.org/10.48550/arXiv.2003.08685

Fredrick, S. S., Coyle, S., & King, J. (2023). Middle and high school teachers' perceptions of cyberbullying prevention and digital citizenship. Psychology in the Schools, 60(6), 1958–1978. https://doi.org/10.1002/pits.22844

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American Psychologist, 56(3), 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Gallagher, K., & Magid, L. (2019, June 10). The Parent's guide to educational technology. Connect Safely. https://connectsafely.org/parents-guide-to-education-technology/

Gallagher, K., Magid, L., & Pruitt, K. (2017, May 4). The educator's guide to student data privacy. Connect Safely. https://connectsafely.org/wp-content/uploads/2016/05/Educators-Guide-Data-.pdf

Garaigordobil, M., & Machimbarrena, J. (2017). Stress, competence, and parental educational styles in victims and aggressors of bullying and cyberbullying. Psicothema, 29(3), 335–340. https://doi.org/10.7334/psicothema2016.258

García-Moya, I., Bunn, F., Jiménez-Iglesias, A., Paniagua, C., & Brooks, F. M. (2019). The conceptualisation of school and teacher connectedness in adolescent research: A scoping review of literature. Educational Review, 71(4), 423–444.

Gasser, U. (2019, November 26). AI innovators should be listening to kids. Wired. https://www.wired.com/story/ai-innovators-should-be-listening-to-kids/

Gasser, U., & Cortesi, S. (2017). Children's rights and digital technologies: Introduction to the discourse and some meta-observations. In M. Ruck, M. Peterson-Badali, & M. Freeman (Eds.), Handbook of children's rights: Global and multidisciplinary perspectives (pp. 417–436). Routledge. https://perma.cc/4259-YXJ6

Gasser, U., Cortesi, S., Malik, M., & Lee, A. (2012). Youth and digital media: From credibility to information quality. Berkman Klein Center for Internet & Society at Harvard University. https://doi.org/10.2139/ssrn.2005272

Gasser, U., & Mayer-Schönberger, V. (2024). Guardrails: Guiding human decision in the age of AI. Princeton University Press.

Gentile, D. A., & Gentile, J. R. (2008). Violent video games as exemplary teachers: A conceptual analysis. Journal of Youth and Adolescence, 37(2), 127–141. https://doi.org/10.1007/s10964-007-9206-2

Gernand, T. (2022). Scanning iPhones to save children: Apple's on-device hashing algorithm should survive Fourth Amendment challenge. Dickinson Law Review, 127(1), 307–338.

Ghosh, A. K., Badillo-Urquiola, K., Guha, S., LaViola Jr., J. J., & Wisniewski, P. J. (2018). Safety vs. surveillance: What children have to say about mobile apps for parental control. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1–14). https://doi.org/10.1145/3173574.3173698

Gnanasekaran, V., & De Moor, K. (2025). Providing user perspectives to the next generation of parental controls. International Journal of Child-Computer Interaction, 44, 100735. https://doi.org/10.1016/j.ijcci.2025.100735

Gorwa, R., & Thakur, D. (2025, April 2). Real time threats: Analysis of trust and safety practices for child sexual exploitation and abuse (CSEA) prevention on livestreaming platforms. Center for Democracy and Technology. https://cdt.org/insights/real-time-threats-analysis-of-trust-and-safety-practices-for-child-sexual-exploitation-and-abuse-csea-prevention-on-livestreaming-platforms

Groba, B., Nieto-Riveiro, L., Canosa, N., Concheiro-Moscoso, P., Miranda-Duro, M. del C., & Pereira, J. (2021). Stakeholder perspectives to support graphical user interface design for children with Autism Spectrum Disorder: A qualitative study. International Journal of Environmental Research and Public Health, 18(9), Article 9. https://doi.org/10.3390/ijerph18094631

Grossman, S., Pfefferkorn, R., Thiel, D., Shah, S., DiResta, R., Perrino, J., Cryst, E., Stamos, A., & Hancock, J. (2024). The strengths and weaknesses of the online child safety ecosystem. Stanford Digital Repository. https://doi.org/10.25740/pr592kc5483

Gunter, B. (2018). Predicting movie success at the box office. Palgrave Macmillan.

Guo, R., Guo, H., Wang, L., Chen, M., Yang, D., & Li, B. (2024). Development and application of emotion recognition technology—a systematic literature review. BMC Psychology, 12. https://doi.org/10.1186/s40359-024-01581-4.

HateAid. (2024). Unser Internet: Denn Menschenrecht gilt auch digital [Campaign summary]. HateAid. https://hateaid.org/unser-internet/

Hall, M. G., Lazard, A. J., Grummon, A. H., Higgins, I. C., Bercholz, M., Richter, A. P. C., & Taillie, L. S. (2021). Designing warnings for sugary drinks: A randomized experiment with Latino parents and non-Latino parents. Preventive Medicine, 148, 106562.

Hammond, D. (2011). Health warning messages on tobacco products: A review. Tobacco Control, 20, 327–337. https://doi.org/10.1136/tc.2010.037630

Hartung, P. (2020). The children's rights-by-design standard for data use by tech companies. UNICEF. https://www.unicef.org/innocenti/media/1096/file/UNICEF-Global-Insight-DataGov-data-use-brief-2020.pdf

Hashish, Y., Bunt, A., & Young, J. E. (2014). Involving children in content control: a collaborative and education-oriented content filtering approach. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1797–1806. https://doi.org/10.1145/2556288.2557128

Hasse, A., Cortesi, S., Lombana-Bermudez, A., & Gasser, U. (2019). Youth and cyberbullying: Another look (SSRN Scholarly Paper No. 3477297). Social Science Research Network. https://papers.ssrn.com/abstract=3477297

Hedderich, M. A., Bazarova, N. N., Zou, W., Shim, R., Ma, X., & Yang, Q. (2024). A piece of theatre: Investigating how teachers design LLM chatbots to assist adolescent cyberbullying education. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1–17). https://doi.org/10.1145/3613904.3642379

Heller, B., & Magid, L. (2019, October 27). Parent's and educator's guide to combating hate speech. Connect Safely. https://connectsafely.org/hatespeech/

Henry, C. D., & Shannon, L. (2023). Virtual natives: How a new generation is revolutionizing the future of work, play, and culture. Wiley.

Hilliard, L. J., Batanova, M., & Bowers, E. P. (2015). A positive youth development approach to bullying: Promoting thriving and reducing problem behaviors. In E. P. Bowers, G. J. Geldhof, S. K. Johnson, L. J. Hilliard, R. M. Hershberg, J. V. Lerner, & R. M. Lerner (Eds.), Promoting positive youth development: Lessons from the 4-H study (pp. 249–272). Springer. https://doi.org/10.1007/978-3-319-17166-1_13

Hilliard, L. J., Bowers, E. P., Greenman, K. N., Hershberg, R. M., Geldhof, G. J., Glickman, S. A., Lerner, J. V., & Lerner, R. M. (2014). Beyond the deficit model: Bullying and trajectories of character virtues in adolescence. Journal of Youth and Adolescence, 43(6), 991–1003. https://doi.org/10.1007/s10964-014-0094-y

Hinduja, S. (2016, February 26). Parental monitoring apps and cyberbullying – Our review and recommendations. Cyberbullying Research Center. https://cyberbullying.org/parental-monitoring-apps-cyberbullying-review-recommendations

Hinduja, S. (2017). The social bond: A practical way for schools to reduce bullying. Cyberbullying Research Center. https://cyberbullying.org/social-bond-practical-way-schools-reduce-bullying

Hinduja, S. (2020). Digital resilience. Cyberbullying Research Center. https://cyberbullying.org/digital-resilience

Hinduja, S. (2023). Generative AI risks and harms: The role of platforms and users. Cyberbullying Research Center. https://cyberbullying.org/generative-ai-risks-harms-platforms-users

Hinduja, S., & Lalani, F. (2025). Empowering and protecting European youth online: Streamlining legislation and promoting positive digital experiences. ThinkYoung. https://www.thinkyoung.eu/_files/ugd/efc875_e33ace5403974df6be6c9f5bff7d28be.pdf

Hinduja, S., & Patchin, J. W. (2009). Bullying beyond the schoolyard: Preventing and responding to cyberbullying. Corwin.

Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. Archives of Suicide Research, 14(3), 206–221. https://doi.org/10.1080/13811118.2010.494133

Hinduja, S., & Patchin, J. W. (2017). Cultivating youth resilience to prevent bullying and cyberbullying victimization. Child Abuse & Neglect, 73, 51–62. https://doi.org/10.1016/j.chiabu.2017.09.010

Hinduja, S., & Patchin, J. W. (2020). Authoritative school climate. Cyberbullying Research Center. https://cyberbullying.org/authoritative-school-climate

Hinduja, S., & Patchin, J. W. (2021). Digital dating abuse among a national sample of U.S. youth. Journal of Interpersonal Violence, 36(23–24), 11088–11108. https://doi.org/10.1177/0886260519897344

Hinduja, S., & Patchin, J. W. (2022a). Bullying and cyberbullying offending among US youth: The influence of six parenting dimensions. Journal of Child and Family Studies, 31(5), 1454–1473. https://doi.org/10.1007/s10826-021-02208-7

Hinduja, S., & Patchin, J. W. (2022b). Bias-based cyberbullying among early adolescents: Associations with cognitive and affective empathy. The Journal of Early Adolescence, 42(9), 1204–1235. https://doi.org/10.1177/02724316221088757

Hinduja, S., & Patchin, J. W. (2024). Metaverse risks and harms among US youth: Experiences, gender differences, and prevention and response measures. New Media & Society, 26(4), 1–22. https://doi.org/10.1177/14614448241284413

Hinduja, S., & Patchin, J. W. (2025). Social media, cyberbullying, and online safety glossary. Cyberbullying Research Center. https://cyberbullying.org/social-media-cyberbullying-and-online-safety-glossary

Hitlin, P. (2018). The rise of digital parenting: How does technology impact parental involvement in education? Journal of Technology in Human Services, 36(2–3), 211–230. https://doi.org/10.1080/15228835.2018.1454925

Hirschi, T. (1969). Causes of delinquency. University of California Press.

Hults, R. V., & Adelsheim, S. (2020, June 20). Opinion: More than ever, we must prioritize the mental health and well-being of children. Stanford Medicine Children's Health. Retrieved from https://www.stanfordchildrens.org/en/services/child-adolescent-psychiatry/supporting-mental-health-covid-19

Information Commissioner's Office. (n.d.-a). Age appropriate design: A code of practice for online services. Retrieved March 27, 2025, from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/

Information Commissioner's Office. (n.d.-b). Protect children's privacy by default. Retrieved March 27, 2025, from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/designing-products-that-protect-privacy/childrens-code-design-guidance/protect-children-s-privacy-by-default/

Iftikhar, Z., Haq, Q. R. ul, Younus, O., Sardar, T., Arif, H., Javed, M., & Shahid, S. (2021). Designing parental monitoring and control technology: A systematic review. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Eds.), Human–Computer Interaction – INTERACT 2021 (pp. 676–700). Springer International Publishing. https://doi.org/10.1007/978-3-030-85610-6_39

Ito, M., Arum, R., Conley, D., Gutiérrez, K., Kirshner, B., Livingstone, S., Michalchik, V., Penuel, W., Peppler, K., Pinkard, N., Rhodes, J., Salen Tekinbaş, K., Schor, J., Sefton-Green, J., & Watkins, S. C. (2020). The connected learning research network: Reflections on a decade of engaged scholarship. Connected Learning Alliance. https://clalliance.org/publications/the-connected-learning-research-network-reflections-on-a-decade-of-engaged-scholarship/

ITU. (2020). Guidelines for industry on child online protection. Retrieved from https://www.unicef.org/media/90796/file/ITU-COP-guidelines%20for%20industry-2020.pdf

ITU. (2023). Measuring digital development: Facts and Figures 2023. International Telecommunication Union. https://www.itu.int/hub/publication/d-ind-ict_mdd-2023-1/

ITU & UNICEF. (2014). Guidelines for industry on child online protection. Retrieved from https://www.itu.int/en/cop/Documents/bD_Broch_INDUSTRY_0909.pdf

Izadi, M., & Hart, R. (2023). The influence of the physical environment on social behavior, school climate, and bullying in schools. Children's Geographies, 22(1), 66–81. https://doi.org/10.1080/14733285.2023.2232751

Jain, O., Gupta, M., Satam, S., & Panda, S. (2020). Has the COVID-19 pandemic affected the susceptibility to cyberbullying in India? Computers in Human Behavior Reports, 2, 100029. https://doi.org/10.1016/j.chbr.2020.100029

Jeong, R., & Chiasson, S. (2020). "Lime," "open lock," and "blocked": Children's perception of colors, symbols, and words in cybersecurity warnings. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–13). https://doi.org/10.1145/3313831.3376611

Jevremovic, A., Veinovic, M., Cabarkapa, M., Krstic, M., Chorbev, I., Dimitrovski, I., Garcia, N., Pombo, N., & Stojmenovic, M. (2021). Keeping children safe online with limited resources: Analyzing what is seen and heard. IEEE Access, 9, 132723–132732. https://doi.org/10.1109/ACCESS.2021.3114389

Jones, L., Doces, M., Swearer, S., & Collier, A. (2012, April 16). Implementing bullying prevention programs in schools: A how-to guide. SSRN. http://dx.doi.org/10.2139/ssrn.2197498

Jones, L. M., Mitchell, K. J., & Beseler, C. L. (2023). The impact of youth digital citizenship education: Insights from a cluster randomized controlled trial outcome evaluation of the Be Internet Awesome (BIA) curriculum. Contemporary School Psychology, 28(4), 509–523. https://doi.org/10.1007/s40688-023-00465-5

Jones, L. M., Mitchell, K. J., & Finkelhor, D. (2013). Online harassment in context: Trends from three Youth Internet Safety Surveys (2000, 2005, 2010). Psychology of Violence, 3(1), 53–69. https://doi.org/10.1037/a0031358

Kaleem, J. (2024, December 2). Confiscation. Calls home. Sealed pouches. Why schools struggle to ban cellphones. Los Angeles Times. Retrieved December 13, 2024, from https://www.latimes.com/california/story/2024-12-02/why-california-schools-are-struggling-to-enforce-cellphone-ban

Kaplan, D. E., & Agarwal, R. (2015). Understanding the adoption, use, and impact of integrated reporting systems. Journal of Information Systems, 29(1), 1–20.

Karcher, M. J. (2009). Increased academic achievement by pairing students with mentors. Research in Middle Level Education Quarterly, 33(4), 1–11.

Kardefelt Winther, D., Stoilova, M., Moritz, B., Twesigye, R., Šmahel, D., Bedrosová, M., Kvardová, N., & Livingstone, S. (2023). Children's exposure to hate messages and violent images online. UNICEF Office of Research – Innocenti. https://www.unicef.org/innocenti/documents/childrens-exposure-hate-messages-and-violent-images-online

Kaulanov, M., & Kazimova, D. (2024). Impact of iPads on secondary school children's learning from the teachers' perspective. World Journal on Educational Technology: Current Issues, 16(4), 327–340. https://doi.org/10.18844/wjet.v16i4.8895

Khameneh, A. (2023, June 5). Calibrating the classroom. MIT Technology Review, 126(3), 26–32.

Kim, Y., Marx, S., Pham, H. V., & Nguyen, T. (2021). Designing for robot-mediated interaction among culturally and linguistically diverse children. Educational Technology Research and Development, 69, 3233–3254. https://doi.org/10.1007/s11423-021-10051-2

Konrath, S., James, C., Weinstein, E., & Tench, B. (2025). Development and validation of the Tech With Care Index for teens. Psychology of Popular Media. https://doi.org/10.1037/ppm0000593

Korbin, J. E. (2003). Children, childhood, and society: An anthropological perspective. [Publication details not fully provided in original entry.]Kraft, M. A., & Bolves, A. J. (2022). Can technology transform communication between schools, teachers, and parents? Evidence from a randomized field trial. Education Finance and Policy, 17(3), 479–510. https://doi.org/10.1162/edfp_a_00344

Kuang, C., & Fabricant, R. (2019). User friendly: How the hidden rules of design are changing the way we live, work, and play. MCD.

Kumar, P. C., O'Connell, F., Li, L., Byrne, V. L., Chetty, M., Clegg, T. L., & Vitak, J. (2023). Understanding research related to designing for children's privacy and security: A document analysis. In Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (pp. 335–354). https://doi.org/10.1145/3585088.3589375

Laaber, F., Florack, A., Koch, T., & Hubert, M. (2023). Digital maturity: Development and validation of the digital maturity inventory (DIMI). Computers in Human Behavior, 143, 107709. https://doi.org/10.1016/j.chb.2023.107709

Lake, C., Snell, A., Gormley, C., Vinyard, I., Gillett, M., Anderson, K. S., O'Neil, E., Alles, D., Collins Coleman, E., & Robb, M. (2025). The state of kids and families in America, 2025. Common Sense Media. https://www.commonsensemedia.org/sites/default/files/research/report/2025-common-sense-summit-report-full-web.pdf

Lala, G., Chandra, S., Ogun, N., Moody, L., & Third, A. (2022). Online safety perceptions, needs, and expectations of young people in Southeast Asia: Consultations with young people in Indonesia, Malaysia, Thailand, and Vietnam. Western Sydney University. https://doi.org/10.26183/TZ74-EV38

Langreo, L. (2024, March 27). New Florida law aims to get kids off social media. Will it work? Education Week. https://www.edweek.org/leadership/new-florida-law-aims-to-get-kids-off-social-media-will-it-work/2024/03

Lauricella, A. R., Herdzina, J., & Robb, M. (2020). Early childhood educators' teaching of digital citizenship competencies. Computers & Education, 158, 103989. https://doi.org/10.1016/j.compedu.2020.103989

Leaton Gray, S. (2018). Biometrics in schools: The role of authentic and inauthentic social transactions. In J. Deakin, E. Taylor, & A. Kupchik (Eds.), The Palgrave International Handbook of School Discipline, Surveillance, and Social Control (pp. 405–424). Palgrave Macmillan.

Lee, A., Schreus, L., Liu, S., & Hancock, J. (2024). What makes social media use enhancing or harmful?: Understanding social media use and youth well-being. In V. Harrison, A. Collier, & S. Adelsheim (Eds.), Social media and youth mental health (pp. 105–122). American Psychiatric Association Publishing.

Lenhart, A., Madden, M., Smith, A., Purcell, K., & Zickuhr, K. (2011). Teens, kindness, and cruelty on social network sites. Pew Research Center. https://www.pewresearch.org/internet/2011/11/09/teens-kindness-and-cruelty-on-social-network-sites/

Levy, N., Cortesi, S., Gasser, U., Crowley, E., Beaton, M., Casey, J. A., & Nolan, C. (2012). Bullying in a networked era: A literature review. Berkman Klein Center for Internet & Society at Harvard University. http://ssrn.com/abstract=2146877

Li, Y., Zhu, Y., Zhang, G., Zhou, J., Liu, J., Li, Z., & He, B. (2022). The effects of anthropomorphism, message framing, and voice type on unhealthy sleep behavior in young users: The mediating role of risk perception. International Journal of Environmental Research and Public Health, 19(15), 9570. https://doi.org/10.3390/ijerph19159570

Liu, Y., Li, Y., Mayfield, R., & Huang, Y. (2024, September 24). Improving emotional support delivery in text-based community safety reporting using large language models. arXiv. https://arxiv.org/abs/2409.15706

Liang, X., & Park, H. (2023). Effects of anthropomorphized virus warnings and perceived cuteness on compliance intention. Asia Pacific Journal of Marketing and Logistics, 35(5), 1045–1061. https://doi.org/10.1108/APJML-11-2022-0949

Livingstone, S. (2013). Online risk, harm and vulnerability: Reflections on the evidence base for child Internet safety policy. ZER: Journal of Communication Studies, 18(35), 13–28.

Livingstone, S., & Blum-Ross, A. (2020). Parenting for a digital future: How hopes and fears about technology shape children's lives. Oxford University Press.

Livingstone, S., Carr, J., & Byrne, J. (2016). One in three: Internet governance and children's rights. UNICEF Office of Research. https://www.researchgate.net/publication/292989247_One_in_Three_Internet_Governance_and_Children's_Rights

Livingstone, S., & Gözig, A. (2012). 'Sexting': The exchange of sexual messages online among European youth. In S. Livingstone & L. Haddon (Eds.), Children, risk and safety on the internet: Research and policy challenges in comparative perspective (p. 151–164). Policy Press. https://doi.org/10.1332/policypress/9781847428837.003.0012

Livingstone, S., & Haddon, L. (2009). EU Kids Online: Final report. LSE, London: EU Kids Online.

Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2011). Risks and safety on the internet: The perspective of European children. Full findings. EU Kids Online. http://eprints.lse.ac.uk/33731/1/Risks%20and%20safety%20on%20the%20internet%28lsero%29.pdf

Livingstone, S., & O'Neill, B. (2014). Children's rights online: Challenges, dilemmas and emerging directions. In S. van der Hof, B. van den Berg, & B. Schermer (Eds.), Minding minors wandering the web: Regulating online child safety (pp. 19–38). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-005-3_2

Livingstone, S., & Sylwander, K. R. (2025). There is no right age! The search for age-appropriate ways to support children's digital lives and rights. Journal of Children and Media, 19(1), 6–12. https://doi.org/10.1080/17482798.2024.2435015

Livingstone, S., & Smith, P. K. (2014). Annual research review: Harms experienced by child users of online and mobile technologies: the nature, prevalence and management of sexual and aggressive risks in the digital age. Journal of child psychology and psychiatry, and allied disciplines, 55(6), 635–654. https://doi.org/10.1111/jcpp.12197

Livingstone, S., & Stoilova, M. (2021). The 4Cs: Classifying online risks to children. Journal of Digital Safety. (Exact volume/issue not provided.)

Livingstone, S., & Third, A. (2017). Children's rights in the digital age: A download from children around the world. New Media & Society, 19(5), 675–690. https://doi.org/10.1177/1461444816686318

Livingstone, S., & Pothong, K. (2022). Imaginative play in digital environments: Designing social and creative opportunities for identity formation. Information, Communication & Society, 25(4), 485–501. https://doi.org/10.1080/1369118X.2022.2046128

Livingstone, S., Cantwell, N., Özkul, D., Shekhawat, G., & Kidron, B. (2024). The best interests of the child in the digital environment. Digital Futures for Children Centre, LSE & 5Rights Foundation. https://www.digital-futures-for-children.net/digitalfutures-assets/digitalfutures-documents/Best-Interests-of-the-Child-FINAL.pdf

Lombana-Bermudez, A., Cortesi, S., Fieseler, C., Gasser, U., Hasse, A., Newlands, G., & Wu, S. (2020a). Youth and the digital economy: Exploring youth practices, motivations, skills, pathways, and value creation. Berkman Klein Center for Internet & Society at Harvard University. https://perma.cc/2KC4-K5RX

Lukács, J. Á., Takács, J., Soósné Kiss, Z., Kapitány-Fövény, M., Falus, A., & Feith, H. J. (2023). The effects of a cyberbullying intervention programme among primary school students. Child & Youth Care Forum, 52(4), 893–911. https://doi.org/10.1007/s10566-022-09714-9

Madden, M., Calvin, A., Hasse, A., & Lenhart, A. (2024). A double-edged sword: How diverse communities of young people think about the multifaceted relationship between social media and mental health. Common Sense Media. https://www.commonsensemedia.org/research/double-edged-sword-how-diverse-communities-of-young-people-think-about-social-media-and-mental-health

Mann, S., Calvin, A., Lenhart, A., & Robb, M. B. (2025). The Common Sense census: Media use by kids zero to eight, 2025. Common Sense Media.

Mantelero, A., & Esposito, M. S. (2021). An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. Computer Law & Security Review, 41, 105561. https://doi.org/10.1016/j.clsr.2021.105561

Marsden, L., Moody, L., Nguyen, B., Tatam, L., Welland, L., & Third, A. (2022). Reimagining online safety education through the eyes of young people: Co-design workshops with young people to inform digital learning experiences. Young and Resilient Research Centre, Western Sydney University. https://doi.org/10.26183/3bz3-r451

Mascheroni, G., & Siibak, A. (2021). Datafied childhoods: Data practices and imaginaries in children's lives. Peter Lang.

McBain, S. (2024, March 21). The anxious generation by Jonathan Haidt – A pocket full of poison. The Guardian. Retrieved December 13, 2024, from https://www.theguardian.com/books/2024/mar/21/the-anxious-generation-by-jonathan-haidt-a-pocket-full-of-poison

McGuirk, R. (2024, November 7). Australia plans a social media ban for children under 16. AP News. https://apnews.com/article/australia-social-media-age-limit-e8259408c0b1456f41967decd474782a

McMellon, C., & Tisdall, E. K. M. (2020). Children and young people's participation rights: Looking backwards and moving forwards. The International Journal of Children's Rights, 28(1), 157–182. https://doi.org/10.1163/15718182-02801002

Messman, E., Heinze, J., Hsieh, H.-F., Hockley, N., Pomerantz, N., Grodzinski, A., Scott, B., Goldstein, N., & Zimmerman, M. (2024). Anonymous reporting systems for school-based violence prevention: A systematic review. Health Education & Behavior, 51(1), 62–70. https://doi.org/10.1177/10901981211073734

Miller, D., Rabho, L. A., Awondo, P., de Vries, M., Duque, M., Garvey, P., Haapio-Kirk, L., Hawkins, C., Otaegui, A., Walton, S., & Wang, X. (2021). The global smartphone: Beyond a youth technology. UCL Press.

Mirbabaie, M., Ehnis, C., Stieglitz, S., Bunker, D., & Rose, T. (2020). Digital nudging in social media disaster communication. Information Systems Frontiers, 23, 1097–1113. https://doi.org/10.1007/s10796-020-10062-z

Mishna, F., Milne, E., Cook, C., Slane, A., & Ringrose, J. (2021). Unsolicited sexts and unwanted requests for sexts: Reflecting on the online sexual harassment of youth. Youth & Society, 55(4), 630–651. https://doi.org/10.1177/0044118X211058226

Mishna, F., Saini, M., & Solomon, S. (2009). Ongoing and online: Children and youth's perceptions of cyber bullying. Children and Youth Services Review, 31(12), 1222–1228. https://doi.org/10.1016/j.childyouth.2009.05.004

Modecki, K. L., Minchin, J., Harbaugh, A. G., Guerra, N. G., & Runions, K. C. (2014). Bullying prevalence across contexts: A meta-analysis measuring cyber and traditional bullying. Journal of Adolescent Health, 55(5), 602–611. https://doi.org/10.1016/j.jadohealth.2014.06.007

Moreno, M. A., Binger, K., Zhao, Q., Eickhoff, J., Minich, M., & Uhls, Y. T. (2022). Digital technology and media use by adolescents: Latent class analysis. JMIR Pediatrics and Parenting, 5(2), e35540. https://doi.org/10.2196/35540

Morgenstern, M., Dumbili, E. W., Hansen, J., & Hanewinkel, R. (2021). Effects of alcohol warning labels on alcohol-related cognitions among German adolescents: A factorial experiment. Addictive Behaviors, 117, 1–7. https://doi.org/10.1016/j.addbeh.2021.106868

Morrongiello, B. A., Cox, A., Scott, R., & Sutey, S. (2016). Children's understanding of no diving warning signs: Implications for preventing childhood injury. International Journal of Environmental Research and Public Health, 13(7). https://doi.org/10.3390/ijerph13070669

Mühlbacher, S., & Sutterlüty, F. (2019). The principle of child autonomy: A rationale for the normative agenda of childhood studies. Global Studies of Childhood, 9(3), 249–260. https://doi.org/10.1177/2043610619860999

Muthu, J. R., Mounish, A., & Praveen, B. S. (2025). Advanced child safety device with GPS tracking and alert messaging. AIP Conference Proceedings, 3279(1), 020155. https://doi.org/10.1063/5.0263101

National Telecommunications and Information Administration. (2024, November 7). Industry's role in promoting kids' online health, safety, and privacy: Recommended practices for industry. https://www.ntia.gov/report/2024/kids-online-health-and-safety/online-health-and-safety-for-children-and-youth/taskforce-guidance/recommended-practices-for-industry#fr323

Nash, V. J. (2014). The politics of children's internet use. In M. Graham & W. H. Dutton (Eds.), Society and the internet (pp. 67–80). Oxford University Press.

Nielsen, J. (1994). Usability engineering. Morgan Kaufmann.

Obajemu, O., Agha, Z., Chowdhury, F. A., & Wisniewski, P. J. (2024). Towards enforcing good digital citizenship: Identifying opportunities for adolescent online safety nudges. Proceedings of the ACM on Human-Computer Interaction, 8(CSCW1), 1–37. https://doi.org/10.1145/3637413

Obermaier, M., & Schmuck, D. (2022). Youths as targets: Factors of online hate speech victimization among adolescents and young adults. Journal of Computer-Mediated Communication, 27(4). https://doi.org/10.1093/jcmc/zmac012

Organisation for Economic Co-operation and Development. (2020a). Protecting children online: An overview of recent developments in legal frameworks and policies. https://doi.org/10.1787/9e0e49a9-en

Organisation for Economic Co-operation and Development. (2020b). Combatting COVID-19's effect on children. https://doi.org/10.1787/2e1f3b2f-en

Organisation for Economic Co-operation and Development. (2021a). Children in the digital environment: Revised typology of risks. OECD Publishing. https://doi.org/10.1787/9b8f222e-en

Organisation for Economic Co-operation and Development. (2021b). Recommendation of the Council on Children in the Digital Environment. OECD Publishing. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389

Organisation for Economic Co-operation and Development. (2022). Companion document to the OECD recommendation on children in the digital environment. OECD Publishing. https://doi.org/10.1787/a2ebec7c-en

Organisation for Economic Co-operation and Development. (2023), "Transparency reporting on child sexual exploitation and abuse online", OECD Digital Economy Papers, No. 357, OECD Publishing, Paris, https://doi.org/10.1787/554ad91f-en.

Organisation for Economic Co-operation and Development. (2024). Towards digital safety by design for children (OECD Digital Economy Papers No. 363). OECD Publishing. https://doi.org/10.1787/c167b650-en

Odgers, C. L. (2024, March 29). The great rewiring: Is social media really behind an epidemic of teenage mental illness? Nature News. https://www.nature.com/articles/d41586-024-00902-2

Ohajionu, U. C., & Matthews, S. (2015). Advertising on social media and benefits to brands. Journal of Social Sciences and Humanities, 10(2), 335–351.

Olmstead, C. (2013). Using technology to increase parent involvement in schools. TechTrends, 57(6), 28–37. https://doi.org/10.1007/s11528-013-0699-0

O'Neill, B., Staksrud, E., & McLaughlin, S. (Eds.). (2013). Towards a better Internet for children? Policy pillars, players, and paradoxes. Nordicom.

O'Reilly, M., Levine, D., Donoso, V., Voice, L., Hughes, J., & Dogra, N. (2022). Exploring the potentially positive interaction between social media and mental health: The perspectives of adolescents. Clinical Child Psychology and Psychiatry, 28(2), 668–682. https://doi.org/10.1177/13591045221106573

Ortutay, B. (2024, July 31). What to know about the Kids Online Safety Act that just passed the Senate. AP News. https://apnews.com/article/congress-social-media-kosa-kids-online-safety-act-parents-ead646422cf84cef0d0573c3c841eb6d

Otterborn, A., Schönborn, K., & Hultén, M. (2019). Surveying preschool teachers' use of digital tablets: General and technology education related findings. International Journal of Technology and Design Education, 29, 717–737. https://doi.org/10.1007/s10798-018-9469-9

Özkul, D., Vosloo, S., & Baghdasaryan, B. (2025). Best interests of the child in relation to the digital environment. UNICEF Innocenti – Office of Global Insight and Policy. https://www.unicef.org/innocenti/media/10571/file/UNICEF-Innocenti-Best-interests-child-digital-environment-brief-2025.pdf

Palfrey, J., boyd, D., & Sacco, D. (2010). Enhancing child safety and online technologies: Final report of the Internet Safety Technical Task Force to the multi-state working group on social networking of State Attorney Generals of the United States. Carolina Academic Press. https://cyber.harvard.edu/sites/cyber.law.harvard.edu/files/ISTTF_Final_Report.pdf

Palfrey, J., & Gasser, U. (2008). Born digital: Understanding the first generation of digital natives. Basic Books.

Palfrey, J. G., & Gasser, U. (2016). Born digital: How children grow up in a digital age (2nd ed.). Basic Books.

Palfrey, J., & Gasser, U. (2020). The connected parent: An expert guide to parenting in a digital world. Basic Books.

Pangrazio, L., & Selwyn, N. (2018). 'It's not like it's life or death or whatever': Young people's understandings of social media data. Social Media + Society, 4(3), 1–9. https://doi.org/10.1177/2056305118787808

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). The effectiveness of computer and tablet assisted intervention in early childhood students' understanding of numbers: An empirical study conducted in Greece. Education and Information Technologies, 23, 1849–1871. https://doi.org/10.1007/s10639-018-9693-7

Park, J. K., Akter, M., Wisniewski, P., & Badillo-Urquiola, K. (2024). It's still complicated: From privacy-invasive parental control to teen-centric solutions for digital resilience. IEEE Security & Privacy, 22(5), 52–62. https://doi.org/10.1109/MSEC.2024.3417804

Patchin, J. W., & Hinduja, S. (2013). Words Wound: Delete Cyberbullying and Make Kindness Go Viral. Free Spirit Publishing.

Patchin, J. W., & Hinduja, S. (2020). It is time to teach safe sexting. Journal of Adolescent Health, 66(2), 140–143. https://doi.org/10.1016/j.jadohealth.2019.10.010

Patchin, J. W., & Hinduja, S. (2024). The nature and extent of youth sextortion: Legal implications and directions for future research. Behavioral Sciences & the Law, 42(4), 401–416. https://doi.org/10.1002/bsl.2667

Patterson, 2011: Patterson D. (2011). The linkage between secondary victimization by law enforcement and rape case outcomes. Journal of interpersonal violence, 26(2), 328–347. https://doi.org/10.1177/0886260510362889

Paul, K. (2024, January 31). Zuckerberg tells parents of social media victims at a Senate hearing: "I'm sorry for everything you've been through." The Guardian. https://www.theguardian.com/us-news/2024/jan/31/tiktok-meta-x-congress-hearing-child-sexual-exploitationPayne, C., Cornell, D., & Konold, T. (2024). Anonymous reporting systems, school climate, and student threat reporting. Journal of Threat Assessment and Management, 11(3), 133–148. https://doi.org/10.1037/tam0000228

Pekárková, S., Novák, R., & Kubát, J. (2022). Barriers and facilitators to seeking help for mental health problems among Czech children and adolescents: A qualitative study. Journal of European CME, 11(1), Article 40894. https://doi.org/10.1007/s40894-022-00203-7

Peña, M., Vásquez-Venegas, C., Cortés, P., Pittaluga, E., Herrera, M., Pino, E. J., Escobar, R. G., Dehaene-Lambertz, G., & Guevara, P. (2024). A brief tablet-based intervention benefits linguistic and communicative abilities in toddlers and preschoolers. npj Science of Learning, 9(38). https://doi.org/10.1038/s41539-024-00249-3

Petronio, S. (2002). Boundaries of privacy: Dialectics of disclosure. SUNY Press.

Pfefferkorn, R., Grossman, S., & Liu, S. (2025, May 29). AI-generated child sexual abuse material: Insights from educators, platforms, law enforcement, legislators, and victims. Stanford Digital Repository. https://doi.org/10.25740/mn692xc5736

Polanin, J. R., Espelage, D. L., Grotpeter, J. K., Ingram, K., Michaelson, L., Spinney, E., Valido, A., Sheikh, A. E., Torgal, C., & Robinson, L. (2022). A systematic review and meta-analysis of interventions to decrease cyberbullying perpetration and victimization. Prevention Science, 23(3), 439–454. https://doi.org/10.1007/s11121-021-01259-y

Prasad, A., & Quinones, A. (2020). Digital overload warnings – "The right amount of shame"? In Lecture Notes in Computer Science (Vol. 12183, pp. 117–134). https://doi.org/10.1007/978-3-030-49065-2_9

Prensky, M. (2001). Digital natives, digital immigrants part 1. On the Horizon, 9(5), 1–6. https://doi.org/10.1108/10748120110424816

Prosser, E., & Edwards, M. (2024). Helpful or harmful? Exploring the efficacy of large language models for online grooming prevention. In European Interdisciplinary Cybersecurity Conference (pp. 1–10). https://doi.org/10.1145/3655693.3655694

Quayyum, F. (2025). Co-designing cybersecurity-related stories with children: Perceptions on cybersecurity risks and parental involvement. Entertainment Computing, 52, 100753. https://doi.org/10.1016/j.entcom.2024.100753

Quayyum, F., Cruzes, D. S., & Jaccheri, L. (2021). Cybersecurity awareness for children: A systematic literature review. International Journal of Child-Computer Interaction, 30, 100343. https://doi.org/10.1016/j.ijcci.2021.100343

Quayyum, F., Bueie, J., Cruzes, D. S., Jaccheri, L., & Vidal, J. C. T. (2021). Understanding parents' perceptions of children's cybersecurity awareness in Norway. In Proceedings of the Conference on Information Technology for Social Good (pp. 236–241). https://doi.org/10.1145/3462203.3475900

Rafique, R., Gantassi, R., Amin, R., Frnda, J., Mustapha, A., & Alshehri, A. H. (2023). Deep fake detection and classification using error-level analysis and deep learning. Scientific Reports, 13, 7422. https://doi.org/10.1038/s41598-023-34629-3

Ray, A., & Henry, N. (2025). Sextortion: A scoping review. Trauma, Violence, & Abuse, 26(1), 138–155. https://doi.org/10.1177/15248380241277271

Reardon, T., Harvey, K., Baranowska, M., O'Brien, D., Smith, L., & Creswell, C. (2017). What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? A systematic review of qualitative and quantitative studies. European Child & Adolescent Psychiatry, 26(6), 623–647. https://doi.org/10.1007/s00787-016-0930-6

Reich, S. M., Subrahmanyam, K., & Espinoza, G. (2012). Friending, IMing, and hanging out face-to-face: Overlap in adolescents' online and offline social networks. Developmental Psychology, 48(2), 356–368. https://doi.org/10.1037/a0026980

Remnick, D. (2024, April 20). Jonathan Haidt wants you to take away your kid's phone. The New Yorker. https://www.newyorker.com/news/the-new-yorker-interview/jonathan-haidt-wants-you-to-take-away-your-kids-phone

Richardson, J., & Milovidov, E. (2019). Digital citizenship education handbook: Being online, well-being online, rights online. Council of Europe.

Richarz, A. (2023, July 25). The 'darting children' that helped Japan win its war on traffic accidents. Atlas Obscura. https://www.atlasobscura.com/articles/tobita-kun-japanese-road-safety-signs

Rideout, V., & Robb, M. (2018). Social media, social life: Teens reveal their experiences. Common Sense Media. https://www.commonsensemedia.org/research/social-media-social-life-teens-reveal-their-experiences-2018

Ross, D. A., Hinton, R., Melles-Brewer, M., Engel, D., Zeck, W., Fagan, L., Herat, J., Phaladi, G., Imbago-Jácome, D., Anyona, P., Sanchez, A., Damji, N., Terki, F., Baltag, V., Patton, G., Silverman, A., Fogstad, H., Banerjee, A., & Mohan, A. (2020). Adolescent well-Being: A definition and conceptual framework. The Journal of Adolescent Health, 67(4), 472–476. https://doi.org/10.1016/j.jadohealth.2020.06.042

Ross, B., Jung, A.-K., Heisel, J., & Stieglitz, S. (2018). Fake news on social media: The (in)effectiveness of warning messages. In Thirty-Ninth International Conference on Information Systems (ICIS), San Francisco (Presentation 16, pp. 1–17). https://aisel.aisnet.org/icis2018/social/Presentations/16

Saavedra-Garcia, L., Moscoso-Porras, M., & Diez-Canseco, F. (2022). An experimental study evaluating the influence of front-of-package warning labels on adolescents' purchase intention of processed food products. International Journal of Environmental Research and Public Health, 19(3), Article 31094. https://doi.org/10.3390/ijerph19031094

Salam, R. A., Das, J. K., Lassi, Z. S., & Bhutta, Z. A. (2016). Adolescent health and well-Being: Background and methodology for review of potential interventions. The Journal of Adolescent Health, 59(4), 4–10. https://doi.org/10.1016/j.jadohealth.2016.07.023

Sampson, J., Witte, K., Morrison, K., Liu, W.-Y., Hubbell, A. P., & Murray-Johnson, L. (2001). Addressing cultural orientations in fear appeals: Promoting AIDS-protective behaviors among Mexican immigrant and African American adolescents and American and Taiwanese college students. Journal of Health Communication, 6(4), 335–358.

Sas, M., & Mühlberg, J. T. (2024). A risk-based evaluation of available and upcoming age assurance technologies from a fundamental rights perspective. Retrieved from https://www.greens-efa.eu/en/article/study/trustworthy-age-assurance

Schlott, R. (2024, March 21). Kids' phone-based existence is 'inhuman' — and 10 times worse than any middle-school horrors: Author. New York Post. https://nypost.com/2024/03/21/lifestyle/author-jon-haidt-has-a-solution-for-social-media-addiction/

Schneider, S., Häßler, A., Habermeyer, T., Beege, M., & Rey, G. D. (2019). The more human, the higher the performance? Examining the effects of anthropomorphism on learning with media. Journal of Educational Psychology, 111(1), 57–72. https://doi.org/10.1037/edu0000273

Schneider, C. R., Zaval, L., & Markowitz, E. M. (2021). Positive emotions and climate change. Current Opinion in Behavioral Sciences, 42, 114–120. https://doi.org/10.1016/j.cobeha.2021.04.009

Schopen, K., Otgaar, H., Howe, M. L., & Muris, P. (2022). Effects of forewarnings on children's and adults' spontaneous false memories. European Journal of Developmental Psychology, 19(2), 177–197.

Shah, R. S., Holt, F., Hayati, S. A., Agarwal, A., Wang, Y.-C., Kraut, R. E., & Yang, D. (2022). Modeling motivational interviewing strategies on an online peer-to-peer counseling platform. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 527:1-527:24. https://doi.org/10.1145/3555640

Sidhu, A. K., Johnson, A. C., Souprountchouk, V., Wackowski, O., Strasser, A. A., & Mercincavage, M. (2022). Cognitive and emotional responses to pictorial warning labels and their association with quitting measures after continued exposure. Addictive Behaviors, 124, Article 107121. https://doi.org/10.1016/j.addbeh.2021.107121

Silic, M. (2016, May). Understanding colour impact on warning messages: Evidence from US and India. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 2954–2960).

Slonje, R., Smith, P. K., & Frisén, Å. (2013). The nature of cyberbullying, and strategies for prevention. Computers in Human Behavior, 29(1), 26–32.

Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). EU Kids Online 2020: Survey results from 19 countries. EU Kids Online. https://doi.org/10.21953/lse.47fdeqj01ofo

Staksrud, E., & Livingstone, S. (2009). Children and online risks: The balance between empowerment and protection. Children & Society.

Stålberg, A., Sandberg, A., Söderbäck, M., & Larsson, T. (2016). The child's perspective as a guiding principle: Young children as co-designers in the design of an interactive application meant to facilitate participation in healthcare situations. Journal of Biomedical Informatics, 61, 149–158. https://doi.org/10.1016/j.jbi.2016.03.024

Stifter, C. A., Cipriano, E. A., Conway, A., & Goetz, M. (2020). Positive emotions and their regulation in early childhood: A developmental perspective. Developmental Psychology, 56(3), 564–577. https://doi.org/10.1037/dev0000875

Stoilova, M., Bulger, M., & Livingstone, S. (2023). Do parental control tools fulfil family expectations for child protection? A rapid evidence review of the contexts and outcomes of use. Journal of Children and Media, 18(1), 29–49. https://doi.org/10.1080/17482798.2023.2265512

Tait, A. R., Voepel-Lewis, T., & Levine, R. (2015). Using digital multimedia to improve parents' and children's understanding of clinical trials. Archives of Disease in Childhood, 100(6), 589–593.

Tamaral, C., Hernandez, L., Baltasar, C., & Martin, J. S. (2025). Design techniques for the optimal creation of a robot for interaction with children with autism spectrum disorder. Machines, 13(1), 67. https://doi.org/10.3390/machines13010067

Tavernise, S. (Host). (2024, September 3). The push to ban phones in school [Audio podcast episode]. In The Daily. The New York Times. https://www.nytimes.com/2024/09/03/podcasts/the-daily/phone-ban-school.html

Taylor, B. G., Liu, W., & Mumford, E. A. (2019). Profiles of youth in-person and online sexual harassment victimization. Journal of Interpersonal Violence, 36(13–14), 6769–6796. https://doi.org/10.1177/0886260518820673

Thabrew, H., Fleming, T., Hetrick, S., & Merry, S. (2018). Co-design of eHealth interventions with children and young people. Frontiers in Psychiatry, 9, 481. https://doi.org/10.3389/fpsyt.2018.00481

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

The Alan Turing Institute and LEGO (2025). Understanding the Impacts of Generative AI Use on Children. Retrieved online from https://www.turing.ac.uk/sites/default/files/2025-05/combined_briefing_-_understanding_the_impacts_of_generative_ai_use_on_children.pdf

Theopilus, Y., Al Mahmud, A., Davis, H., & Octavia, J. R. (2024). Digital interventions for combating internet addiction in young children: Qualitative study of parent and therapist perspectives. JMIR Pediatrics and Parenting, 7, e55364. https://doi.org/10.2196/55364

Third, A. (2024). Youth agency, rights, and the promise of a well-designed digital world. In V. Harrison, A. Collier, & S. Adelsheim (Eds.), Social media and youth mental health (pp. 235–258). American Psychiatric Association Publishing.

Thorn. (2024a, August 14). Deepfake nudes and other trends in youth behavior online in 2023: New research from Thorn. Retrieved from https://www.thorn.org/blog/deepfake-nudes-and-other-trends-in-youth-behavior-online-in-2023/

Thorn. (2024b). The state of the issue. Retrieved from https://www.thorn.org/research/state-of-the-issue/

Tolks, D., Lampert, C., Dadaczynski, K., et al. (2020). Spielerische Ansätze in Prävention und Gesundheitsförderung: Serious Games und Gamification. Bundesgesundheitsblatt, 63, 698–707. https://doi.org/10.1007/s00103-020-03156-1

Trommelen, M. (1997). Effectiveness of explicit warnings. Safety Science, 25(1–3), 79–88. https://doi.org/10.1016/S0925-7535(97)00019-2

Uğraş, T., Rızvanoğlu, K., & Gülseçen, S. (2022). New co-design techniques for digital game narrative design with children. International Journal of Child-Computer Interaction, 31, 100441. https://doi.org/10.1016/j.ijcci.2021.100441

UNESCO. (2023, June 21). UNESCO's education response to COVID-19. Retrieved June 12, 2025, from https://www.unesco.org/en/covid-19/education-response/initiatives?hub=80

UNICEF. (n.d.). Protecting children online: Every child must be protected from violence, exploitation and abuse on the internet. Retrieved from https://www.unicef.org/protection/violence-against-children-online

UNICEF. (2020). COVID-19 and its implications for protecting children online. Retrieved from https://www.unicef.org/sites/default/files/2020-04/COVID-19-and-Its-Implications-for-Protecting-Children-Online.pdf

UNICEF. (2021a). The case for better governance of children's data: A manifesto. Retrieved from https://www.unicef.org/innocenti/media/1031/file/UNICEF%20Global%20Insight%20Data%20Governance%20Manifesto.pdf

UNICEF. (2021b). Policy guidance on AI for children 2.0. Retrieved from https://www.unicef.org/innocenti/media/1341/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf

UNICEF. (2024). The state of the world's children 2024: The future of childhood in a changing world. Retrieved from https://www.unicef.org/media/165156/file/SOWC-2024-full-report-EN.pdf

UNICEF. (2025a). Childhood in a digital world: Screen time, digital skills and mental health. UNICEF Innocenti. Retrieved from https://www.unicef.org/innocenti/reports/childhood-digital-world

UNICEF. (2025b). Innocenti report card 19: Child well-being in an unpredictable world. Retrieved from https://www.unicef.org/innocenti/media/11111/file/UNICEF-Innocenti-Report-Card-19-Child-Wellbeing-Unpredictable-World-2025.pdf

Valentine, K. (2011). Accounting for agency. Children & Society, 25(5), 347–358. https://doi.org/10.1111/j.1099-0860.2009.00279.x

Valkenburg, P. M., & Peter, J. (2013). The differential susceptibility to media effects model. Journal of Communication, 63(2), 221–243. https://doi.org/10.1111/jcom.12024

Van der Hof, S. (2017). I agree, or do I? A rights-based analysis of the law on children's consent in the digital world. Wisconsin International Law Journal, 34(2), 409–445. Retrieved from https://wilj.law.wisc.edu/wp-content/uploads/sites/1270/2017/12/van-der-Hof_Final.pdf

Vega, V., & Robb, M. B. (2019). The Common Sense census: Inside the 21st-century classroom. Common Sense Media. Retrieved from https://www.commonsensemedia.org/research/the-common-sense-censusinside-the-21st-century-classroom-2019

Veretilnykova, M., & Dogruel, L. (2021). Nudging children and adolescents toward online privacy: An ethical perspective. Journal of Media Ethics, 36(3), 128–140. https://doi.org/10.1080/23736992.2021.1939031

Villano, M. (2024, April 16). How cell phones are killing our kids, and what we can do about it. CNN. Retrieved December 13, 2024, from https://edition.cnn.com/2024/04/16/health/cell-phones-jonathan-haidt-wellness/index.html

Von Der Linde, M., Göcke, M., Hirschfeld, G., & Thielsch, M. T. (2025). Check or reject? Trust and motivation development in app-based warning systems. Safety Science, 185, 106724. https://doi.org/10.1016/j.ssci.2024.106724

Vuorre, M., & Przybylski, A. K. (2024). Global well-being and mental health in the internet age. Clinical Psychological Science, 12(5), 917–935. https://doi.org/10.1177/21677026231207791

Wachs, S., Krause, N., Wright, M. F., & Gámez-Guadix, M. (2023). Effects of the prevention program "HateLess. Together against Hatred" on adolescents' empathy, self-efficacy, and countering hate speech. Journal of Youth and Adolescence, 52(6), 1115–1128. https://doi.org/10.1007/s10964-023-01753-2

Walsh, K., Zwi, K., Woolfenden, S., & Shlonsky, A. (2015). School-based education programmes for the prevention of child sexual abuse. Cochrane Database of Systematic Reviews, 2015(4), CD004380. https://doi.org/10.1002/14651858.CD004380.pub3

Walsh, W. A., Finkelhor, D., & Turner, H. (2025). Characteristics and dynamics of cyberstalking victimization among juveniles and young adults. Violence Against Women, 31(5), 1306–1327. https://doi.org/10.1177/10778012231225238

Wang, K., Rees, V. W., Dorison, C. A., Kawachi, I., & Lerner, J. S. (2024). The role of positive emotion in harmful health behavior: Implications for theory and public health campaigns. Proceedings of the National Academy of Sciences, 121(28), e2320750121. https://doi.org/10.1073/pnas.2320750121

Wang, G., Zhao, J., Van Kleek, M., & Shadbolt, N. (2021). Protection or punishment? Relating the design space of parental control apps and perceptions about them to support parenting for online safety. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–26. https://doi.org/10.1145/3476084

Waterson, P., & Monk, A. (2014). The development of guidelines for the design and evaluation of warning signs for young children. Applied Ergonomics, 45(5), 1353–1361. https://doi.org/10.1016/j.apergo.2013.03.015

Wemmers, J. A. (2002). Restorative justice for victims of crime: A victim-oriented approach to restorative justice. International Review of Victimology, 9(1), 43-59. https://doi.org/10.1177/026975800200900104

Wemmers, J. A. (2013). Victims' experiences in the criminal justice system and their recovery from crime. International Review of Victimology, 19(3), 221-233. https://doi.org/10.1177/0269758013492755

Willard, N. (2007). Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress. Research Press.

Winner, L. (1980). Do artifacts have politics? Daedalus, 109(1), 121–136.

Wisniewski, P., Ghosh, A. K., Xu, H., Rosson, M. B., & Carroll, J. M. (2017). Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety? In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (pp. 51–69). https://doi.org/10.1145/2998181.2998352

World Health Organization. (2025). European health report 2024: Keeping health high on the agenda. WHO Regional Office for Europe. Retrieved from https://iris.who.int/bitstream/handle/10665/380382/9789289061728-eng.pdf

Xiang, L., & Park, H. J. (2023). Effects of anthropomorphized virus warnings and perceived cuteness on compliance intention. Asia Pacific Journal of Marketing and Logistics, 35(12), 2897–2911. https://doi.org/10.1108/APJML-11-2022-0949

Yadav, S., & Chakraborty, P. (2022). Child–smartphone interaction: Relevance and positive and negative implications. Universal Access in the Information Society, 21(3), 573–586. https://doi.org/10.1007/s10209-021-00807-1

Ybarra, M. L., boyd, d., Korchmaros, J. D., & Oppenheim, J. K. (2012). Defining and measuring cyberbullying within the larger context of bullying victimization. Journal of Adolescent Health, 51(1), 53–58. https://doi.org/10.1016/j.jadohealth.2011.12.031

Yeo, G. H., Loo, G., Oon, M., Pang, R., & Ho, D. (2023). A digital peer support platform to translate online peer support for emerging adult mental well-being: Randomized controlled trial. JMIR Mental Health, 10, e43956. https://doi.org/10.2196/43956

Yip, T. (2020, September 9). Addressing inequities in education during the COVID-19 pandemic. Society for Research in Child Development. Retrieved from https://www.srcd.org/research/addressing-inequities-education-during-covid-19-pandemic-how-education-policy-and-schools

Zaikina-Montgomery, H., & Silver, N. C. (2018). An examination of icons, signal words, color, and messages in warnings for children on the internet. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62(1), 251–255. https://doi.org/10.1177/1541931218621058

Zieglmeier, V., & Lehene, A. M. (2022). Designing trustworthy user interfaces. In OzCHI '21: Proceedings of the 33rd Australian Conference on Human-Computer Interaction (pp. 182–189). https://doi.org/10.1145/3520495.3520525

Zhu, Y., Feng, X., Li, H., Huang, Y., Chen, J., & Xu, G. (2017). A randomized controlled trial to evaluate the impact of a geo-specific poster compared to a general poster for effecting change in perceived threat and intention to avoid drowning 'hotspots' among children of migrant workers: Evidence from Ningbo, China. BMC Public Health, 17(1), 1–9. https://doi.org/10.1186/s12889-017-4