ANNEX 3

SCALING TRUST ON THE WEB

RESPECTING CHILDREN AS RIGHTS HOLDERS

COMPREHENSIVE REPORT OF THE TASK FORCE FOR A TRUSTWORTHY FUTURE WEB

The mission of the Digital Forensic Research Lab (DFRLab) is to identify, expose, and explain disinformation where and when it occurs using open-source research; to promote objective truth as a foundation of government for and by people; to protect democratic institutions and norms from those who would seek to undermine them in the digital engagement space; to create a new model of expertise adapted for impact and real-world results; and to forge digital resilience at a time when humans are more interconnected than at any point in history, by building the world's leading hub of digital forensic analysts tracking events in governance, technology, and security.

ISBN: 978-1-61977-279-3

Please direct inquiries to:
Atlantic Council
1030 15th Street, NW, 12th Floor
Washington, DC 20005

For more information, please visit www.AtlanticCouncil.org

June 2023

ANNEX 3
RESPECTING CHILDREN AS RIGHTS HOLDERS

TABLE OF CONTENTS

A TRUSTWORTHY FUTURE WEB MUST REFLECT THE RIGHTS—AND PERSPECTIVES—OF CHILDREN.

Digital technologies, the entities that provide them, and the laws and policies that apply to them are all generally the work of adults. Yet, adults and children alike use (and are otherwise impacted by) those technologies, must operate within those policy frameworks, and must have their views and respective online practices reflected in them. A growing body of research shows that involving children directly in decisions that impact them (in an age-appropriate way) is key to identifying their best interests, and to effectively understanding and addressing children's diverse needs.[1] However, children have traditionally had little direct say in the development of the technology policies and digital products that affect them.[2] Instead, their interests—if they are considered at all—are largely represented by adults (indeed, even in the drafting of this piece).

Policies and products aimed at protecting children's safety have increasingly played a pivotal role in influencing trust and safety practices within companies, as well as driving forward new laws and regulations (some proposed, some adopted) based in protecting children's safety.[3] Combined with the absence of actual children from policy conversations, this singular focus on children's safety has demonstrated a challenge inherent to establishing safety as the foundational premise of tech policy development: it can lead to the violation of other crucial rights that children enjoy, do so without creating any space for their input or involvement in making that tradeoff, lead to counterproductive outcomes, and result in violating the rights of whole other communities of rightsholders.

[1] Emily Weinstein and Carrie James, *Behind Their Screens: What Teens Are Facing (and Adults Are Missing)* (Cambridge, MA: MIT Press, 2022).

[2] Compare tech policy to, say, education for women and girls, or the climate crisis. By contrast, those are policy issues on which young people, who are directly (and negatively) affected by adults' policy choices, have insisted on making their voices heard, turning activists such as Malala Yousafzai and Greta Thunberg into household names while they were still teenagers.

[3] For a brief history and layout of the online-safety ecosystem, see: Anne Collier, "The Child Online Safety Ecosystem: A Look at the History, Education, Content Moderation and Developments around the World" in Kalinda Raina, ed., *Children's Privacy and Safety* (Portsmouth, NH: International Association of Privacy Professionals, 2022).

In our collective time working on these issues, we have noticed four common themes in tech policy discussions involving children: the assumption that digital technologies are (primarily or solely) detrimental to children; that children can be considered as a homogeneous group; that children's safety from physical, mental, and (especially) sexual harms is achieved through limiting children's access to digital media and devices, rather than through empowering harm-reduction approaches or prioritizing the full range of digital rights children hold; and a preference to address online child-safety concerns through criminal law enforcement and increased digital surveillance (by parents, companies, and state agencies such as schools and police).[4]

In this annex, we aim to expand the aperture, giving space to considerations that can otherwise be excluded when the narrative is restricted to child safety, and hopefully illuminating underexplored areas for future attention and investment. We seek to reframe the discussion by centering children's rights and agency, and treating safety as part of a constellation of rights that children enjoy, and that must coexist with the other rights of children as well as the rights of adults. We begin by defining what we mean by these two terms ("children's safety" and "children's rights") and noting the tension between them in existing policy discussions. We then explore lessons that have been learned in both tech and non-tech spaces regarding methods for supporting the participatory inclusion of children safely and effectively. We close with a wide and illustrative range of policy areas in which children's rights can, and should, be considered, and where their inclusion could be operationalized.

We would like to acknowledge from the outset that the regulations, legislation, and citations grounding this article reflect expertise in the evolution of these questions within the European Union (EU), the United Kingdom (UK), and the United States. Child safety and children's rights are key issues to be examined and illuminated around the globe. We would like to note the importance of giving greater attention to the evolution of these concepts across other regions, and encourage investments in supporting a more equitable body of research in order to better inform the wide range of stakeholders making policy and product determinations with regards to children's safety and rights.

CHILDREN'S SAFETY

"Child safety" is an umbrella term that can mean different things to different people (children included). The phrase might primarily evoke defense against child sex-abuse material and exploitation (CSEA, a term that includes imagery depicting child sex abuse, child sex trafficking, and sextortion, among other concepts). However, its meaning goes well beyond sex-related harms and can extend to safety from physical violence, threats, hate speech, and harassment (e.g., parental abuse, state violence, cybersecurity-related harms) or psychosocial well-being (e.g., "screen addiction," cyberbullying, eating disorders, self-harm), among other concerns.

"Child safety" is not only a broad term, but also a politically charged, culturally nuanced concept. Because "children's lives have become digital by default," "online safety" for children spurs fierce debate and passionate focus among adults seeking to establish broad and enforceable rules for a constantly shifting, highly personal paradigm. Tradeoffs regarding agency, autonomy, governance, privacy, security, and a range of related values will never be simple or universal. However, the process of identifying and debating tradeoffs (as well as looking for creative ways to avoid tradeoffs and optimize for multiple values) can be grounded in the practices that have been developed over decades to grapple with exactly these complexities and provide a coherent foundation for negotiation and consensus building. That foundation includes long-standing frameworks of human-rights law.

[4] This preference embeds a further assumption: that parents and the state do not harm children's safety or infringe upon their rights.

CHILDREN'S RIGHTS

Children are humans, so children share in the human rights enjoyed by adults, in addition to specific rights based on their status as children. Human rights apply online as well as offline.

To define children's rights, this annex refers to the United Nations (UN) Convention on the Rights of the Child (CRC). The CRC has been ratified by every UN member state except the United States.[5] That means those 196 countries have voluntarily agreed to abide by the CRC, an obligation that extends to these countries' laws and policies regarding digital technologies. In 2021, the UN published guidance, called General Comment 25, to states regarding their CRC obligations in the digital environment. That guidance has had differing impacts in different stakeholder settings, with the greatest traction occurring in European, Australian, and UK governmental debates. Its impact in the United States is limited, in part, because the (typically private-sector) entities that actually create the digital environment are more focused on the binding laws applicable to them than on nonbinding guidance directed to states, though European legislation such as the General Data Protection Regulation (GDPR) and Digital Services Act (DSA) will have growing impact even on US-based companies' child-protection practices.

The CRC recognizes that children have the right to protection from violence, abuse, and exploitation (Articles 19, 34, 35, and the Optional Protocol on the Sale of Children, Child Prostitution, and Child Pornography (OPSC)). Children's rights go well beyond safety, however. Scholars have categorized their rights under what they've characterized as the "three Ps": protection, provision, and participation rights, which, under General Comment 25, have their digital counterparts, including the right to express their views on matters that affect them (in accordance with the child's age and maturity) (Article 12); rights to privacy (including in their correspondence) (Article 16) and freedom of expression (Article 13); the right to seek and receive information (including access to mass media) and the right to education (Articles 13, 17, and 28); freedom from discrimination (Article 2); rights in the justice system (Article 40); and freedom from all forms of violence (physical, mental, sexual), exploitation, and trafficking (Articles 6, 19, 24, 32, 34, 35, 36, and 39).

All of these rights are universal, non-hierarchical, indivisible and interdependent, equal and nondiscriminatory. One set of rights cannot be enjoyed without another, which makes them all equally important—and makes the three categories of protection, provision, and participation a helpful framework for upholding them in balance with one another. Implementing decisions based in human-rights principles is, and remains, a challenging task, but new guidance is emerging and is worth noting. For example, the UN Human Rights B-Tech project provides guidance for technology companies about how to carry out human-rights due diligence, including human-rights impact assessments. UNICEF has worked with UN Human Rights to produce a special briefing on how children's rights can be considered by technology companies as part of the human-rights due-diligence process, which will be published later this year.

TENSION BETWEEN CHILDREN'S SAFETY AND OTHER RIGHTS

Children's rights are co-equal, meaning that, in theory (as General Comment 25 notes), protecting their right to be safe online must not come at the expense of their other rights, for example, safety at the expense of privacy. This is a difficult balancing act in any context, and especially within the context of ever-evolving digital spaces in which the threats and opportunities for children can shift and scale with astonishing speed. While most agree that preventing harms against children in their digital environment is imperative, there is a lack

[5] While the United States has not ratified the CRC, it has ratified the CRC's Optional Protocol on the Sale of Children, Child Prostitution, and Child Pornography. According to Human Rights Watch, children's rights are poorly protected in the United States, prompting scholars to posit an alternative model of parental rights within US law that promotes a broader range of children's present and future interests.

of consensus among stakeholders as to how to prevent harms while protecting children's agency and rights of participation, and often a lack of technological solutions that could immediately address harms even if a policy determination were made. The interpretation of long-standing children's rights also constantly evolves as a component of broader digital transformations—for example, traditional prohibitions against child labor are now being considered within the context of a raging "creator economy." This creates an ongoing tension in tech policymaking, trust and safety tooling, and broader digital-product development.

Children are frequently positioned within policy debates as dependents or subordinates of those who hold rights over them (such as parents or the state). Deep tensions exist around the question of where to set the boundaries of parents' and children's own respective agency and accountability. Moreover, children have historically emerged as the framing (one that parents themselves do not necessarily condone) for much larger normative debates within communities about highly sensitive political topics and evolving cultural norms around sexuality, gender, and religious or ethnic identity. Finally, proponents of approaches that center children's safety as one of many competing rights that must be considered can face a high political cost, as well as personal attack.

Alongside recent debates about cyberbullying, Internet addiction, and data privacy, one of the most significant and recurring threads in tech policy debates involving children centers on how to keep children safe from CSAM (especially CSAI) and impede the vast scaling and potential normalization of such material.[6] This policy priority has inspired legislation and regulatory proposals in the United States, EU, China, and elsewhere, although legal frameworks vary in robustness in terms of definitions and effective criminalization of the perpetrators of CSAM crimes. While most stakeholders agree that preventing harms against children in their digital environment is imperative, they lack consensus about how to do so while protecting children's agency. This disagreement not only creates an ongoing tension in tech policymaking, but also reflects a more fundamental normative challenge. Stakeholders do not necessarily operate from a baseline agreement that children are rights holders in terms of their own rights and agency, nor do they share a baseline agreement on the extent to which adults' rights can, and should, be limited or implicated.

Great opportunity arises when this tension forces a recognition that competing rights can, and must, be balanced. For example, in August 2021, Apple announced a plan for new iPhone child-safety features that drew controversy due to their negative impacts on privacy and other rights of both adult and child iPhone users. An executive director at the National Center for Missing and Exploited Children (NCMEC) privately congratulated Apple for "prioritizing child protection" and described those who expressed concerns as "the screeching voices of the minority." This is an example of the political challenges that rights advocates can face in debates around child safety.

After public pushback, Apple dropped the most controversial (and privacy-intrusive) component of its plan, and revised another feature to give child users more privacy and agency, responding to concerns that the original design would jeopardize the safety of child users with abusive parents. While Apple's original plan subordinated some rights (of children and adults) to children's right to be free from CSAE (especially CSAI), its revised plan reflected a more non- hierarchical approach to its users' rights and an understanding of child protection that includes agency, digital privacy, and safety from family abuse, not just safety from CSAE. (That said, the revised plan was not well received by NCMEC.)

Notably, Apple's review benefited from several key elements: existing normative agreement and legal standards regarding the unacceptability of the production and dissemination of CSAI; knowledge exchanges that reflected deep expertise on critical policy components, as well as complex technological questions; and in-

[6] For a deeper analysis of this topic, see *Annex 4: Deconstructing The Gaming Ecosystem*.

creased opportunity for input by a more extensive group of advocates, companies, and regulators than Apple had originally consulted. However, the perspective of one key set of stakeholders remained notably absent from this collaboration: children.

APPROACHES TO YOUTH ENGAGEMENT

In some countries, youths are actively engaged, to some extent, on relevant issues such as privacy, consent, and data sharing in the digital realm, although ample opportunities remain to expand such engagement throughout the world. Many children are adaptable, nimble, tech-fluent digital natives, often with well-formed opinions about the digital environment's impact on their rights. Their perspectives and solutions-based thinking can vary widely from those of policymakers who may overwhelmingly represent a significant generational divide from them. Youths' voices not only deserve to be elevated within debates around their rights and safety in the digital realm, but their lived experience with social media is needed for adult intelligence gathering and policymaking; they may also add significantly to the richness and creativity brought to bear in identifying strong paths forward.[7] This is true for decision-making spaces across the ecosystem, be it in informing governmental debate, within corporate decision-making processes, or setting research and advocacy priorities. The UNICEF and Lego RITEC project, which sets out a framework for integrating the well-being of children into digital design, centers even young children's own participation in defining what it takes to maximize children's well-being.

In considering approaches to youth engagement on the subject of children's rights and safety online, it is imperative to recognize the diverse experiences and perspectives of young people, which can be shaped by innumerable factors and may include geography, race/ethnicity, religion/faith, socioeconomic status, sexuality, gender identity, cognitive development, and (significantly) levels of trauma, among many others. While there is no one-size-fits-all approach, there are existing methodologies and practices for operationalizing youth engagement effectively and ethically in tech design and policymaking (both in and outside of the technology sector), as well as innovative new public-, private-, and cross-sector initiatives.

One established consideration in youth engagement is to examine the intended goals and outcomes of youth-engagement efforts. Should such efforts be activated to raise youth awareness of online harms, which might include different learning and teaching strategies, such as social and emotional learning? Should such efforts prioritize youth engagement, which might include youth development, collective empowerment, and/or systems changes? Different goals necessitate different approaches.

When engaging with children on topics involving online harm, it is also important to accommodate children's own history of, or ongoing experiences with, violence. This should be based in an understanding of the typology of different harms (e.g., armed conflict, familial, racial, or sexual violence, etc.), children's possible experiences of multiple overlapping harms, and these harms' potential impacts on the child. For example, how might a youth-engagement strategy approach a child or teen who has been sexually abused, and who, as a result of this trauma, acts out sexually inappropriately toward other youth online? Trauma-informed frameworks can provide an additional lens through which such learning or engagement strategies can be developed.

Depending upon which strategy is most effective for determined outcomes, it is valuable to include educators, mental-health professionals, and child-development specialists alongside youth (or to review resources in these areas and request consultations, if it is not possible to include youth-serving professionals during actual discussions). Such inclusion will help ensure any discussions are appropriately scaffolded for youth (using the abovementioned strategies and frameworks) depending upon youths' cognitive development and existing experiences with trauma, among other factors.

[7] See, supra note 2

KEY CONCEPTS IN ONLINE CHILDREN'S RIGHTS AND SAFETY POLICY

The following concepts are key areas of current inquiry in tech policy and product-development debates that hold particular salience for children's rights. Each raises underexplored avenues for future research, presenting an opportunity for ongoing, innovative investment framed through a rights-centric lens. For each concept, a full discussion of all of the rights and tradeoffs involved is beyond the scope of this summary-level document. However, the CRC-enumerated rights listed above can serve as a jumping-off point, particularly the four principles grounding the CRC: non-discrimination; a child's best interests as a primary consideration; survival and development; and the right to be heard.[8]

ARTIFICIAL INTELLIGENCE (AI)

AI holds great promise for helping children stay safe and exercise their rights. AI can be used to locate information efficiently, and for generating (or translating) text, sound, video, and images. Thus, children can use AI to learn, express themselves, create art, play, and socialize with others. AI-based tools are already crucial to the automated detection and review of potentially abusive content online at massive scale (although algorithms can replicate bias, and the use of CSAI to train AI tools is controversial and legally dubious). Offline, AI is starting to be used in cancer detection and drug discovery, as well as to aid the work of professions that help protect children, such as legal services and refugee services.

At the same time, AI's perils are already well documented. Children's photos or other personal information may be ingested into (or output from) AI systems. Children may be exploited or bullied by using their images in deepfakes. AI-generated misinformation and disinformation may mislead and misinform kids; synthetic media might fool them into falling prey to scams or hacks. Depending on how they are used, generative AI tools can support and/or undermine a child's education. In the juvenile-justice system, decisions may be made using algorithmic tools that replicate systemic biases. Given these risks, claims about AI require exceptional vigilance, especially claims of "AI that will help children." Stakeholders should approach AI project proposals with skepticism, and must demand satisfactory answers about ethics, privacy, safety, and risk mitigation.

CHILD SEXUAL EXPLOITATION AND ABUSE

Perhaps uniquely as types of content go, depictions of child sexual abuse are illegal basically everywhere, making CSAI a common compliance concern worldwide. Online CSAI is an enduring problem, despite decades of concerted technological and political interventions. This is tragic and yet unsurprising, since the instrumentalities for CSAE offenses are the core functions of the open Internet: transmitting, storing, and accessing files, and communicating with other users.

Online CSAI sharing and solicitation is a deeply complicated and multifaceted issue. For starters, the problem eludes precise quantification, with debates over the reliability of official numbers of CSAI reports by online services and the reasons why they have ballooned lately. Plus, it is not only pedophiles who share CSAI. The sharer's motivation may be humor or outrage, or the child depicted may have produced and shared the image initially. Imagery created by the child depicted, called self-generated CSAI (SG-CSAI), raises a range of complex issues. Some SG-CSAI involves a teenager who is of age and consenting. Some is harmful, as with grooming, sextortion (someone extorting the child into creating and sending CSAI), or the nonconsensual

[8] For a deeper analysis of this topic, see *Annex 1: Current State of Trust and Safety*; *Annex 2: Building Open Trust and Safety Tools*; and *Annex 4: Deconstructing The Gaming Ecosystem*.

sharing of an image that was originally created and sent consensually (nonconsensual intimate imagery, or NCII, also called "revenge porn"). SG-CSAI is still illegal, so consenting teenagers may risk prosecution for documenting legal sexual activity.

Service providers at multiple levels of the tech stack have devoted significant resources to detecting CSAI. Many major tech companies scan their users' files with so-called "hash" technologies (such as Microsoft's PhotoDNA). New, unknown images and livestreamed abuse pose greater technical challenges for automated detection than do known images. Similarly, exploitation offenses (e.g., grooming, enticement, and solicitation) pose a different technical challenge than the detection of known CSAI. While text-based classifiers for CSEA have been developed, finding matches for known imagery is more straightforward than determining whether a text or voice interaction is abusive. Different CSAI and CSEA scanning technologies vary as to their accuracy and the implications for user privacy.

Through automated scanning for known CSAI, major online services detect large volumes of CSAI every year. While this technology aids children, it has a significant privacy impact. Many tech companies scan voluntarily, but policies to compel scanning for CSAI have been proposed in jurisdictions such as the EU and India. These proposals can conflict with constitutional (e.g., Fourth Amendment) or statutory (e.g., General Data Protectino Regulation, European Convention on Human Rights) privacy protections; even voluntary scanning may run afoul of some privacy laws absent a carveout. Likewise, privacy laws may impede platforms from sharing information with each other about users who use multiple services to share CSAI or exploit children.

CSAI scanning helps to detect and stymie horrific sex abuse of children. However, scanning all online content and communications is a significant privacy intrusion that also has effects on other rights such as free expression and association. Scanning mandates for CSAI also pose a threat to human rights (including children's) because authoritarian governments could also require online services to scan for other types of content: political dissent, religious expression, LGBTQ+ (lesbian, gay, bisexual, transgender and queer) content, etc. Because automated scanning for CSAI does not work in end-to-end encrypted environments, which are relied upon for safety and privacy around the world, legislative proposals that overtly or effectively require scanning for CSAI pose a risk to online service providers' ability to legally offer end-to-end encryption.

Policy proposals to combat CSAI and child sexual exploitation often reflect an attitude of "tech solutionism": expecting technical interventions to fix the larger, sticky societal problem of CSAE. Anti-CSAE policy proposals tend to focus on law enforcement and criminal punishment. Comparatively few have focused on prevention and support—for example, by investing in child-protection systems, sex-education and CSAE-awareness programs, and housing stability for at-risk youth. A comprehensive response requires a victim-centered criminal-justice system, coupled with prevention measures and support for victims. Victim-focused policies are more rights respecting than those that center the interests of criminal law-enforcement agencies. However, policymakers may prefer approaches that rely on self-funded tech-company initiatives, rather than government support. Prevention- and support-centric policy models for mitigating CSAE are, thus, an area in which philanthropic investment could expect to meet with a favorable response from both governments and tech companies.

DUTY OF CARE

The concept of the duty of care is a common-law term that comes from the law of negligence and is used to impose a standard of care on technology companies to avoid careless acts that could foreseeably harm children. Duty-of-care obligations, some specific to minors, are a key part of the EU's new Digital Services Act and the UK's pending Online Safety Bill. The services covered by such duties are typically those accessible to, or likely to be accessed by, children. Definitions of the duty of care vary, but typically require only collecting data on young users that are required for a service to function, setting strong default privacy settings, and avoiding deceptive or addictive design features.

In practice, the duty of care effectively requires age assurance to determine which users are children (or, alternatively, verifying all users' ages to ensure compliance, potentially imperiling most users' privacy), meaning it implicates the same rights as age assurance (see Annex 2: Building Open Trust and Safety Tools). Duty-of-care bills have also been criticized for letting regulators and tech companies decide what is in children's best interests, which jeopardizes children's access to online content that might, in fact, be crucial for actual children's well-being (mental or sexual health resources, LGBTQ+ resources, etc.). It is unclear how duty-of-care obligations will apply to end-to-end encryption (E2EE). Fear of liability could dissuade providers from offering E2EE services to children, or at all; conversely, duty-of-care policies that require the most privacy-protective settings by default for children's accounts could be read to mandate default E2EE for child users.

END-TO-END ENCRYPTION

End-to-end encryption is a technology for protecting data privacy and security by encoding data so that they can only be decoded by the sender and intended recipient(s). A recent report by Child Rights International Network (CRIN) and Defend Digital Me (DDM) analyzes the complicated interactions between E2EE and children's various rights—some positive, some negative. As the report describes, E2EE messaging enables children to exercise their rights (free expression, privacy, etc.), and protects children's lives and safety by keeping outsiders (such as an oppressive government or abusive parents) from monitoring their communications.

However, E2EE can also be used to violate children's rights, because it complicates the detection of abusive interactions with, or involving, children (although research shows that providers have other means of detecting abuse in E2EE settings). E2EE's usage in CSAE offenses, in particular, has prompted regulators and other stakeholders to call for regulating E2EE to enable investigatory access to users' communications. These proposals implicate users' rights (including children's rights), and would undermine E2EE systems' privacy, security properties, and user expectations. The question of how to mitigate the use of E2EE services for CSAE, without detriment to children's and all users' rights or harming digital security, is an ongoing area of intense and emotionally charged debate.

EPHEMERALITY

Ephemerality is a functionality of multiple popular messaging and social media services commonly used by children, such as Snapchat, Instagram, and WhatsApp. Ephemerality can be fundamental to how an app works, or it can be an optional function that users can choose to turn on. Ephemeral messages, photos, videos, or collages (e.g., Instagram Stories, WhatsApp Disappearing Messages, Snapchat Snaps and Chats) disappear after a set amount of time. Whether that period is triggered upon send or upon receipt/viewing depends on the app. The time length may be set by the app, or the app may give the user a range of time periods to choose from. Generally, ephemeral content disappears for everyone including the sender, but the sender may have the option to save the content.

Ephemeral functionalities enable children to express themselves spontaneously, and to communicate privately and securely, with less worry that what they share will stay online, spread, or linger in someone else's message history. However, ephemerality can enable abuse: it is harder to determine after the fact whether an account posted harmful content, and there is a limited time period in which a user can report content before it disappears. Plus, ephemerality is not bulletproof: recipients may screenshot or otherwise preserve ephemeral content (although some apps try to defeat screenshotting or notify the sender about it). This may pose a risk to children who share sensitive information without realizing it could still be preserved and shared (e.g., with a parent or school).

MENTAL HEALTH

There is significant debate and public concern about social media's effect on children's mental health and brain development. Research findings to date do not paint a consistent picture. Some research has shown significant differences by geographic region in social media's relationship with youth well-being, and (for reasons that are not yet understood) social media's impacts on individual youths' mental health can be highly heterogeneous. In the United States, studies indicate that teenagers' mental-health crisis is not clearly caused by social media; indeed, some research has shown positive effects. Other research has examined behavioral interventions' impact on teen mental health, such as limiting screentime, also with mixed results.

There is ample opportunity for additional research, with regard to both geographic distribution (many studies to date are from the United States and Europe) and age (outcomes for teens should not be extrapolated to younger children). Additional research is important because, despite the complex picture that existing research has painted, policy proposals commonly assume that social media is an unalloyed harm to children. It also suggests that technological or product fixes can solve problems that originate entirely outside of a platform—for example, within a school. That assumption leads to laws that impact children's rights online, such as by imposing age-related restrictions on children's access to social media (with largely arbitrary age thresholds). Social media's role in children's mental health is an area in which philanthropic investment in research, especially in the Global Majority, could make a significant impact.

PARENTAL CONTROLS

Parental controls are intended to give parents some say in shaping the digital environment their child experiences. Controls may apply to the form and/or the substance of a child's Internet use: both what information they access online (content restrictions) and how they access it (what times of day, for how long, how frequently). Over time, parental controls have evolved from something that policymakers incentivized (in the early years of the Internet) to something that policymakers are increasingly attempting to mandate for online services in jurisdictions such as the United States, UK, and EU. Requirements vary widely, but often include a dashboard that provides parents or guardians easy access to set default privacy and security settings, set screentime limits, and/or filter out certain content for their children. In some cases, laws require platforms to give parents or guardians access to their children's account to view their activity and communications, or to track their location.

The issue of parental controls and the emphasis on parent interests illustrate how policy discussions can be misaligned—if not at odds—with children's interests. Ideally, parental controls help supportive parents keep their children safe online and guide them into healthy habits. However, from a children's-rights perspective, parental controls may also catalyze certain harms, depending on a given child's situation (as noted in the Apple example above). They may exacerbate harm to vulnerable children, such as those in abusive home environments or who would be at risk if a parent learned of their sexuality, gender identity, religious views, etc. More broadly, the effectiveness of parental controls—and their effects on children's development, cognition, and other interests—are not yet well understood. Nor is the impact that parental access to a child's online activity has on children's understanding and exercise of their privacy, free expression, and other rights. For example, how might the proliferation of parental controls shift an expectation of privacy among generations of youth? What impact does a reduced sense of privacy have on the child's cognitive and social development, civic engagement, and collective empowerment? As policy mandates multiply, parental controls will increasingly present a fertile opportunity for philanthropic investment in research.

TRANSPARENCY REPORTING

Lately, regulators often propose making major online services periodically report to the government about their internal policies, practices, and design features for keeping users safe on their services. Some proposals are child specific, whereas others apply to all users. Some, like the EU's new DSA law, create programs to fund studies on technology's effects on children, and/or require covered services to grant researchers access to their internal data for independent study.

Transparency policies could greatly improve public understanding of popular online services' effects on child users. Enabling third-party research could reveal important insights, inform best practices, drive evidence-based policymaking, or improve online services' often-disappointing policy enforcement against abuse. However, there are risks: inadequate safeguards for researcher access could affect children's privacy. Requiring detailed public reports on services' internal practices could give abusive users a roadmap to circumvent them. These mandates also risk creating a vector for state censorship if services feel pressured to change their content-moderation and child-safety programs to better suit the state's preferences (a particular risk for content such as LGBTQ+ content or sexual-health information that some states consider harmful to children).

CONCLUSION

Ensuring children's safety and upholding children's rights require a holistic approach that encompasses the full range of children's rights, including participation, provision, and protection. Policymakers, researchers, advocates, and industry leaders must not only involve young people in decision-making, but also consider children's lived experiences within digital transformations, now and in the future. It is imperative that a wider range of stakeholders be supported to build models that can lower the barrier to incorporating rights-respecting frameworks and inclusionary models in the development of policies and products that impact children. This is particularly critical with regard to Global Majority communities, where existing inequities within policy, product development, and research risk being doubly visited upon children.

Emphasizing the consideration of the entirety of children's rights does not diminish in any way the critical importance of protecting them from harm within a quickly evolving threat landscape. It does, however, improve the likelihood that future iterations of online spaces will provide space and opportunity for children to benefit from online spaces where they can not only be safe, but also be empowered to learn, explore, play, grow, and evolve.

AUTHORSHIP AND ACKNOWLEDGEMENTS

This annex reflects contributions from the following members of the Task Force for a Trustworthy Future Web: Lauren Buitta, Girl Security; Emma Day, DFRLab; Riana Pfefferkorn, Stanford Internet Observatory; and Leah Plunkett, Harvard Law School. It also reflects contributions from contributing experts: Sara Grimes, University of Toronto, and Anne Collier, the Net Safety Collaborative. This report does not represent the individual opinion of any contributor, member of the task force, or contributing organization. Rather, it serves to consolidate collective research, feedback, and contributions gathered over a five-month period. The contributors are grateful to additional members of the task force and outside experts for their review and feedback, as well as to Lauren Quittman of Duco Experts. Finally, the contributors would like to thank John Perrino, policy analyst at the Stanford Internet Observatory, for providing a briefing note on common policy approaches to children's online-safety issues and for allowing the task force to incorporate portions of the briefing note into this annex.