cdt | Research

The Kids Are Online

Research-Driven Insights on Child Safety Policy

A Symposium Report

Michal Luria
Aliya Bhatia

February 2025

CENTER FOR
DEMOCRACY
& TECHNOLOGY

The **Center for Democracy & Technology** (CDT) is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

A report from

cdt | Research

The Kids Are Online

Research-Driven Insights on Child Safety Policy

A Symposium Report

Authors

Michal Luria and Aliya Bhatia

The findings and recommendations in this report don't reflect CDT's policy positions nor do they reflect the positions of all contributors of this report. This report is meant to be a starting point for discussion and collaboration and reflect areas of consensus amongst a set of diverse actors in the child safety space.

WITH CONTRIBUTIONS BY (ALPHABETICALLY)

Karla Badillo-Urquiola, Michael J. S. Beauvais, Scott Babwah Brennen, Leticia Bode, Becca Branum, Peter Chapman, Linda Charmaraman, Maddy Dwyer, Marissa Edmund, Sarah Forland, Timothy Hoagland, Samir Jain, Priya C. Kumar, Kostadin Kushlev, Amanda Lenhart, Jenna Leventoff, Alice Marwick, Nora Mcdonald, John Perrino, Sarah Philips, Elissa Redmiles, Shreya Sampath, Sydney Saubestre, Justice Shannon, Cori Stott, Dhanaraj Thakur, Kristin Woelfel.

ACKNOWLEDGEMENTS

This work is made possible through a grant from the John S. and James L. Knight Foundation.

SUGGESTED CITATION

Luria, M. & Bhatia, A. (2025). "The Kids are Online: Research-Driven Insights on Child Safety Policy." Center for Democracy & Technology. https://cdt.org/insights/the-kids-are-online-research-driven-insights-on-child-safety-policy/

References in this report include original links as well as links archived and shortened by the Perma.cc service. The Perma.cc links also contain information on the date of retrieval and archive.

Contents

Contents

Executive Summary

This report summarizes the key discussions and insights from an in-person symposium held in September 2024 on the topic of children's online safety policy. The event convened academic researchers, policy experts, and civil society representatives to explore research-driven approaches to addressing critical issues impacting young users in digital environments. During the symposium, we attempted to foster meaningful dialogue, identify areas of consensus and disagreement, and chart actionable paths forward. The symposium included a range of perspectives, and thus the report reflects a synthesis of ideas rather than unanimous agreement.

The symposium brought together 23 participants for a day-long event conducted under the Chatham House Rule. Attendees engaged in two rounds of thematic roundtables covering four key topics related to child safety on online platforms: **Connection**, **Content**, **Communication**, and **Characteristics**. The event concluded with an all-participant session that summarized some of the main discussions and identified strategies and opportunities to integrate research into policy.

We lay out some of the cross-cutting themes that we have identified in conversation; these highlight the interconnectedness of issues surrounding youth safety online, and emphasize the need for evidence-based and youth-centric approaches, particularly along the following lines:

- **No one-size-fits-all approach fixes current issues.** Researchers pointed to a range of ways for keeping young people safe online, yet most solutions raise thorny tradeoffs.
- **Experiences of all youth online should be examined, including those with different backgrounds.** Participants repeatedly raised that young users experience online environments differently based on factors like age, socioeconomic status, and identity. Tailored safety measures, they note, may be essential to address these varied experiences effectively. Some said that additional aspects like access and digital literacy require further consideration of tools that accommodate diverse user needs.
- **Consider the ecosystem of actors who are part of a young person's life holistically.** The discussions emphasized adopting a more holistic and collaborative approach to online child safety. Participants underscored the necessity of collective efforts that would involve parents,

educators, platform designers, and policymakers. Collaboration across these groups was identified as crucial for reaching feasible and balanced actionable steps.

- **Limited researcher access to data impedes evidence-informed solutions.** Researchers in the group agreed that a lack of access to comprehensive data impedes fully understanding online harms and prevents learning about the effectiveness of existing safety measures implemented by digital platforms. Most agreed that improved access to data is vital to develop evidence-informed policy.

Participants also proposed several practical steps with potential to enhance online safety for young people on digital platforms:

- **Establish default protections.** Participants agreed that implementing safety settings by default, such as private accounts, can potentially keep young users and all users safer.
- **Empower users with the ability to customize their online experiences.** According to participants, equipping youth — and all users — with features like customizable content filters and algorithm reset options could give them the reins to shape their own experiences online.
- **Provide researchers with privacy-preserving mechanisms to access data.** Participants emphasized the importance of providing researchers with access to platform data, especially data related to safety mechanisms (e.g., the rate of users who use safety tools or how these tools are being used). They noted that this would allow researchers to better study online experiences and evaluate the effectiveness of safety measures.
- **Support digital literacy and onboarding.** Participants recommended platforms to work towards supporting users' development of skills to navigate digital spaces responsible, as opposed to restricting access to young users altogether. Leveraging peer-to-peer education, more collaborative onboarding processes, and norm setting can all help acquaint young users with improving online norms and safety practices.

The conversation underscored the complexity of creating safer online environments and the importance of engaging researchers, who can share data-driven knowledge on approaches that have the potential to work. Participants emphasized the need for ongoing dialogue and actionable processes — safer digital spaces require sustained efforts to bridge gaps between research, policy, and platform design. This report serves as a step towards creating this shared space that would support the creation of safer digital environments for young users while respecting their rights and agency.

Introduction

On September 25, 2024, the Center for Democracy & Technology (CDT) hosted an in-person symposium that gathered academic researchers, policy experts, and civil society to discuss issues of platform governance and children's online safety. The goal of the gathering was to generate research-driven discussions, identify areas of consensus, and develop potential paths forward for online safety policy that would support young users.

While the symposium facilitated valuable discussions and brought together diverse perspectives, this report does not reflect unanimous agreement among participants. Instead, it represents a synthesis of the ideas shared during the event, capturing key issues and potential paths forward. In other words, not all contributors support every idea included in this report — the goal is to present a set of considerations that can guide policymakers in the process of addressing critical issues of child safety in online spaces.

Symposium Structure

The symposium was a day-long event with 23 attendees, conducted under the Chatham House Rule (allowing participants to share quotes and ideas from the event, but without attributing contributions to any particular participant). Participants were US-based academics who conduct research related to youth and platform governance, members of civil society with a focus on child safety policy, and U.S. government agency representatives.

After a limited set of presentations, participants split up into two rounds of roundtable discussions where they tackled research questions and findings related to one of four high-level topics defined based on a synthesis of categories and topics within current US child safety policy conversations. Participants were assigned to two of four roundtables based on preference (elicited in advance via survey) and prior expertise:

1. **Connection.** How do youth access platforms? How can we create safer entry points for young users?
2. **Content**. What content are youth exposed to, and how can policy reduce exposure to unwanted content across platforms?

3. **Communication.** How can we better support safe communication with others online, including in private messages?
4. **Characteristics**. How do platform design features influence youth online behavior, and how do they impact user safety and engagement?

Each roundtable covered the same two guiding questions:

1. What do we know about the topic based on youth-focused research? What works and what doesn't? What are current concerns and issues?
2. Looking forward, what are some ways in which platforms could improve youth safety while allowing them to enjoy the benefits of digital platforms?
   a. What are some limitations that should be set in place?
   b. What are some tools that should be offered?
   c. What additional research is needed?

The roundtables were facilitated by CDT staff members with expertise on the topic of child safety policy and research. Before each discussion, participants were asked to respond to the prompts in a shared document while adding citations. This allowed for time to consider the questions and recall prior research and discourse. It also helped us better document the conversation and ideas raised throughout the event.

The symposium concluded with a roundtable discussion with all participants about how to better facilitate researcher-policymaker collaborations to encourage evidence-based and rights-supporting courses of action. In addition, given the focus on research and the researchers conducting it, the roundtable considered opportunities to integrate this research into the policymaking process.

This report is organized according to the four predefined roundtable topics. For each thematic roundtable, we summarize the conversations, while identifying current concerns and paths forward raised by participants. We conclude with a summary of the concluding roundtable as well as a synthesis of cross-category themes.

Connection: Age-appropriate Access to Online Services

“Connection refers to the means through which young users gain access to an online service and what interventions, if any, guide young users towards healthy, age-appropriate online experiences as they create an account or log into a service. Some methods outlined in proposed legislation or technical literature include age verification, which requires a user to provide age and proof of ID to an intermediary to be granted access to a service; age estimation, which requires an online intermediary to estimate a user’s age based on their characteristics or online behavior; or age declaration, which is the current status quo for many services, where a user has to attest their date or year of birth before being granted access to a service. Other methods to grant underage individuals access to a service include equipping parents with mechanisms to grant consent on their behalf.

The discussion in the “Connection” section opened with the question of how platforms currently grant young users access to their services and what concerns or opportunities these methods raise. During the discussion, a few participants questioned whether gating young users from online services or certain online spaces is effective or promotes healthy online behaviors at all. Some also observed that the lack of “child-only” spaces online — once more common but now less popular — is a key factor driving young users to mixed-audience sites (such as large social media platforms that cater to both younger and older individuals). Finally, even if effective gating methods were possible, participants noted that young users would always be adept at circumventing these methods, and potentially be directed to worse, “underground” online spaces.

Current Weaknesses in Facilitating Age-appropriate Access to Online Services

1. Age declaration methods are commonly used but easy to circumvent.

The most popular method currently used to enable or restrict access to online services based on age is platforms asking users about their ages, commonly known as “age declaration.” This method requires users to submit their year or date of birth when they create an account on the service or check a box attesting that they are above a certain age.

[According to some participants] the toll on privacy would not be not confined to children — age verification requires everyone to verify their age in order to use online content, and according to research can have not only significant privacy implications but also chilling effects on access to information.

Symposium attendees agreed that this method was easy to circumvent. They noted that all users were adept at circumventing these systems by lying or submitting any year of birth that would grant them access to the service. An Ofcom study found that a fifth of children aged 8 and 17 had a "false" adult age on online platforms as a consequence of giving the wrong birth year when prompted (Ofcom, 2022). Many parents and guardians also reported lying on their children's behalf to override the age declaration interstitial and grant their children access to a platform — this has been observed in previous generations of online services as well (boyd et al., 2011). Despite these shortcomings, online services continue to pursue this method because it is often the most straightforward, least privacy-invasive and least costly age assurance method.

2. More stringent methods of age verification or estimation raise privacy, accuracy, reliability, and equity concerns.

There was agreement in the room that more effective methods to collect accurate age-related data on users raise thorny tradeoffs. Citing reports produced by Open Technology Institute (Forland et al., 2024), University of North Carolina and Duke University (Marwick et al., 2024), and the Center for Growth & Opportunity at Utah State University (Brennen & Perault, 2025), researchers highlighted privacy concerns with requiring users to provide more data and sensitive information that would prove their age (e.g., government ID or biometric data like their facial image or voice patterns).

This approach to age verification may require users to provide government ID in order to gain access to an online service — according to some participants, this would be too privacy-invasive. They emphasized that the toll on privacy would not be not confined to children — age verification requires everyone to verify their age in order to use online content, and according to research can have not only significant privacy implications but also chilling effects on access to information (Marwick et al., 2024; Ruane et al., 2024; York, 2024).

Further, participants raised that some people do not have valid IDs. Without an ID, these individuals, more commonly undocumented, unhoused, or low-income individuals, would face barriers to accessing critical information provided on online services and would be limited in their rights to express themselves freely. They noted that children are also specifically less likely to possess an official ID or have access to theirs, making age verification methods akin to parental consent and control mechanisms (see the next subsection on "Parental control..."), whereby a child has to ask their parent to gain access to an online service. For some children, this would put them in harm's way.

Further, participants raised that not all young users are in supportive family dynamics, and in those cases, online spaces can be especially informative and even life-saving.

Some surveys found that parents and children have a slight preference for age estimation using biometric scanning (like scanning one's face or voice) or parental vouching — where a parent selects on their device whether to grant a child access to a service or not — as the mode to grant a child access to a service (Family Online Safety Institute, 2022). Researchers explained that these modes may be preferred because of convenience, ease, and familiarity, but that no method has gained unanimous support of either parents or young people.

Despite the convenience offered by approaches like face scanning, participants argued that they raise equity, reliability and accuracy concerns, and that they can also be easily circumvented. One example that was noted was an equity concern of using machine learning methods due to their error-prone nature for individuals who are outside the majority, such as users with disabilities who may look younger or older than their actual age, or users of color.

Participants concluded by saying that additional tools to estimate age are available to companies, who currently have strong financial incentives for these tools to be accurate (Raffoul et al., 2023), but that more transparency and research is needed to better understand these other methods, their effectiveness, and their potential impact on people's digital rights.

3. Parental control mechanisms may hurt more than they help.

Another common way to grant children access to age-appropriate experiences is by equipping parents with controls to verify children's age, as well as limiting what their children can and cannot access. However, participants said this can also backfire. Accordingly, recent research has supported a shift of paradigm from restrictive parental-based approaches to resilience-based approaches to increasing safety online (Park et al., 2024).

Further, participants raised that not all young users are in supportive family dynamics, and in those cases, online spaces can be especially informative and even life-saving (Redmiles, 2021). These spaces can provide young users with access to information that their parents may not provide, such as information about LGBTQ+ identity, reproductive healthcare, and violence prevention and support. In these cases, a participant said, requiring young people to ask permission of their parents for access can put young people in greater danger or lead to alienation.

4. Age-based restrictions often lead to an increase in deceptive practices due to low familiarity and trust.

One participant noted that increased use of age verification methods tends to increase lying or obfuscation among youth for many reasons, including low trust and familiarity with age gates — youth do not trust that these systems work in their best interest so they lie to age gates to protect themselves. In some cases, participants say, children are more aware of how age verification methods work than their parents (Family Online Safety Institute, 2022). In other cases parents have limited capacity and time to grant their children access to online services, particularly those with multiple jobs or immigrant parents who rely on their children to navigate English-language spaces such as government websites or parent-teacher letters.

Participants highlighted that deceptive practices and circumvention can also impact the types of experiences users will then have online. For example, if users lie about their age, the ads and feeds they will then see may also be mismatched or age-inappropriate. Participants noted that this tradeoff sets up a "vicious cycle" and dichotomy between privacy and safety where a user has to choose between receiving an age-appropriate feed or protecting sensitive information.

5. Device-based age declaration shows promise but could fail due to shared devices.

One method that was discussed to restrict users from accessing age-inappropriate services was to install device-based age limits or child controls, meaning that a user's age is set on a particular device through which they access the internet, as opposed to setting their age separately on every platform. One participant noted that this showed promise not only in working more smoothly, but in potentially alleviating concerns of notice fatigue by streamlining the number of times a parent or caregiver needs to input a child user's age and allowing parents to designate a device's age at the outset. Yet, as argued by another participant, a significant proportion of households rely on shared devices, making device-based age declaration difficult and ineffective. According to one survey, 35% of multi-person households shared a computer or a laptop and 10% of multi-person households shared a smartphone, with 58% of those sharing it at least once every day (Vogels, 2021). Sometimes a device isn't shared per-se, but an adult (e.g., a parent) may give their own device to a child and by doing so grant them access to the internet (Chiong & Shuler, 2010).

6. Age verification hinders the use of multiple accounts and identity exploration.

Another point that was raised is that young users tend to open multiple accounts to access and experience different communities, content, or identities on the same platform. Participants noted that online communities sometimes support essential identity formation by enabling young people to access different interests and "try on" identities in the safety of their own home (Luria & Foulds, 2021, Thakur et al., 2023). Sometimes users make use of multiple accounts to accommodate different parts of their identities and interests rather than having a single account that would amass all their likes and dislikes. Creating multiple accounts, one participant said, can also be a safety and content control tactic, as users can access content that they are curious about but do not want to access again or have it affect their content algorithms — this can include content about puberty, weight, sex education, etc.

At the same time, another participant noted that the phenomenon of multiple accounts, "fake accounts" or "finstas" (short for fake instagram account (Weaver & Issawi, 2021)), makes it challenging for an online service to create safe, age-appropriate guardrails as young users can easily circumvent them by creating additional accounts. Nevertheless, creating more stringent methods to verify a user's age, and by proxy their identity, may be a disproportionate response that could result in preventing their ability to create multiple accounts in the first place.

Future Opportunities: Introduction of designated spaces, global standards, and more access to data

1. Introduction of more age-appropriate and child-specific spaces.

Participants argued that young people have always sought "older" content or communities (e.g., pre-teens reading Seventeen magazine), and researchers familiar with childhood education agree that younger users' interest in "older" or "age-inappropriate" content may be an essential part of development and identity formation. Still, in contrast to early years of the internet in which young people congregated on sites designed for youth, minor users now tend to "share" the web with adult users. As a result, participants believe that online platforms seeking to host content and facilitate an exchange of ideas amongst users of all ages bear the responsibility to facilitate access to age-appropriate experiences and potentially gate young users away from some information or interactions.

Participants asserted that one factor that guides young users to mixed audience sites is the dearth of child-only spaces online (both content platforms like YouTube Kids and interaction platforms like Minecraft). While more research is needed to understand the challenges and best practices in creating more child-focused spaces, participants suggested that it might be helpful to consider parallels from the physical world, like playgrounds, kids' museums and more. With more digital options for age appropriate and youth-focused platforms, young users could have safer online spaces to congregate, instead of being pushed to join mixed audience sites.

2. Global standards and independent third-party organizations to facilitate age verification.

Researchers noted the importance of standards bodies in developing standards for age-verification that would give effect to certain widely accepted privacy rules such as isolating age-verification data from other data. One legal researcher noted that there may be viable solutions to create a third-party, non-profit entity that authenticates a user's age without giving an online service access to the user's identity and sensitive information (Hanaoka et al., 2024; *EuConsent Making the Internet Age-Aware*, n.d.), as is being piloted in some places in Europe (Borak, 2024). At the same time, others argued that this would raise concern about new actors that could collect and sell sensitive data on users, including their government ID, to bad actors, in the absence of a federal privacy law in the U.S. and in many areas of the world. Participants noted that there are a few privacy-preserving ways to equip users with a digital ID to use online to prove their age, with the eID pilot in the EU being one example (*European Digital Identity (eID)*, n.d.) but that the viability and desirability of this hinges on additional legal privacy protections, which they said the US currently lacks federally and for many states.

3. Researcher access to data on the presence of young users across platforms.

Participants remarked on the lack of data regarding what platforms know about how users join and use their services based on age. In particular, one researcher noted that there was very little information on the experiences of users at the precipice of early adolescence (ages 11-13). Some researchers have conducted surveys to highlight the impact of early adolescent use of online services on digital behavior and literacy. However, in the absence of data directly from platforms on how many early adolescent users are actually there, researchers and policymakers are in the dark on the scope of the problem (Charmaraman et al., 2022).

Another researcher pointed to the passage of COPPA, the Children's Online Privacy Protection Act passed in 1996, as a driver for this opacity. COPPA introduced additional privacy and parental consent requirements for online services directed to children under 13. This research suggested that in response to COPPA, online service providers began modifying their terms of service to prohibit users under 13 from accessing their services. This, they argue, has led to under-monitoring users who were suspected of being under age — providers may fear liability under COPPA's privacy and consent requirements, which apply when they have "actual knowledge" that a user is under 13 (*Complying with COPPA*, 2020). Researchers argue that this has resulted in opacity for researchers who study users under the age of 13 due to inability to get data or an understanding of user behavior of very young users on major online platforms.

Complaints about the lack of access to information related to the ages of users on platforms were couched in a broader discussion of the lack of access to platform data, particularly as the few mechanisms to facilitate access to data have been either collapsed or made prohibitively expensive (CDT 2024; Stokel-Walker, 2023).

Content: Materials Online and Platforms that Moderate Them

"Content" refers to any material that users interact with on digital platforms, including videos, images, text posts and advertisements. Much of the policy conversation on this topic looks into how to design "age-appropriate" spaces with "age-appropriate" content (e.g., excluding violent, sexual, and other extreme content). Nevertheless, and as highlighted in the symposium roundtable, the boundaries of what is and is not age-appropriate remain subjective and difficult to agree on.

The topic of content is strongly tied to platform content moderation and recommendation systems, which act as the primary mechanisms for managing what content is shown and to whom. These systems — whether human, algorithmic, or more often a combination of both — play a critical role in determining the visibility of content based on age, preferences, and safety considerations. Managing content that is presented to young people on digital platforms is an ongoing challenge for platforms, trusted adults, and young people themselves.

Based on the discourse, participants suggested that improving content exposure and filtering on digital platforms requires both immediate enhancements to existing tools and a long-term shift in how platforms approach content management and moderation. The discussion centered on how prioritizing user agency and control and integrating transparency and human oversight into algorithmic systems can create safer and more positive content experiences for young people.

Current Weaknesses in Platform Content and Moderation Systems

1. Difficulty in defining age-appropriate content.

Platform moderation systems are processes designed to filter and curate the vast amount of user generated content shared on digital platforms. These processes encompass a combination of human moderators, automated content analysis, recommender systems, user reporting tools, and community guidelines to determine what content is visible, flagged, and removed. Broadly, moderation systems aim to enforce platform policies, comply with legal standards, and create a safe and easy to use user environment.

Despite their capabilities, participants assert that moderation systems have notable limits, which can have an effect on users, especially more vulnerable users like youth. One of the main challenges, participants noted, was in shaping automated content analysis systems to prioritize "age-appropriate" content when what exactly is "age-appropriate" is frequently disagreed upon amongst different communities. But as attendees pointed out, even if there is agreement, automated content analysis systems can misidentify or fail to catch all inappropriate or harmful content, especially when users deliberately bypass platform rules through codified language (Duarte & Llansó, 2017).

Moreover, participants pointed out that automated filtering technologies sometimes inadvertently limit users from accessing content that is appropriate (Kids Online Health and Safety Taskforce, 2024). Previous research that they mentioned showed that over-filtering can lead to excluding important content such as LGBTQ+ content (York, 2021). For example, some filters have previously blocked developmentally appropriate explorations of sexuality that were deemed inappropriate, reducing access to content for LGBTQ+ youth and questioning communities. As participants noted, this is in part due to the fact that algorithms cannot always accurately classify nuanced content, especially in cases where the difference between harmful and educational material (e.g., sexual health or LGBTQ+ content) is subtle. This becomes even more complicated and nuanced with the use of cultural and linguistic variations, leaving significant gaps in moderation efforts (Nicholas & Bhatia, 2023).

2. Engagement-based amplification of content can be harmful.

Participants repeatedly emphasized that most platforms use engagement-based algorithms that recommend content based on user interactions, not based on their actual stated preferences, and thus fail to distinguish between user interest and user engagement. In other words, just because a young user spends time on a video does not mean they want to see more of it, but as attendees noted, algorithms interpret engagement as a preference and tend to recommend similar content.

Just because a young user spends time on a video does not mean they want to see more of it, but as attendees noted, algorithms interpret engagement as a preference and tend to recommend similar content.

Participants explained that this is particularly problematic with questionable, extreme or harmful content — if a minor watches one inappropriate video (which could happen due to mere curiosity), the algorithm might suggest more similar content, potentially reaching a point where their feed is flooded with harmful material.

Research finds that many youth choose to share such content as a normal part of exploring their identity, engaging in relationships, or expressing themselves. On the other hand, participants noted, the associated risks are significant, including unauthorized redistribution, exploitation, and long-term harm to their reputations or mental health.

3. The challenging tension between wanting to share sensitive content (e.g., sexual imagery) and controlling its spread.

Participants agreed that the sharing of intimate content by youth presents a significant challenge. While there are differing opinions among adults regarding the behavior itself, research finds that many youth choose to share such content as a normal part of exploring their identity, engaging in relationships, or expressing themselves (Hasinoff, 2015). On the other hand, participants noted, the associated risks are significant, including unauthorized redistribution, exploitation, and long-term harm to their reputations or mental health. Research supports the claim that platform moderation systems, designed to limit the spread of sensitive or explicit material, often struggle to strike the right balance of protecting youth from harm while avoiding increasing stigma around sexual expression (Qin et al., 2024).

One participant pointed out that there are available tools such as blocking and flagging, but controls that can minimize unauthorized sharing of sensitive content are less available and may have limited efficacy (*SafeDigitalIntimacy.Org*, n.d.b.). Screen-capturing, for instance, further complicates content protection. According to participants, this tension underscores the need for thoughtful, nuanced approaches that protect teens from harm without unduly restricting their agency or ability to engage in self-expression.

4. Extensive exposure to ads, including ads misaligned with interests or age-inappropriate.

Young users, like all users, are exposed to advertisements online. Sometimes, these ads are not aligned with their interests, or are inappropriate for their age group, but participants claimed that there are very limited ways of changing that. Prior research showed that ad controls are largely ineffective, and generally fail to provide users with a way to express their ad preferences (Ali et al., 2023). With few semantic controls existing (e.g., "less eating disorder content," a setting on YouTube), attendees noted, users need to use vague tools like "show me less" and "snooze" to try to manipulate their algorithm to match their actual preferences.

One participant pointed out that features that do allow more specific forms of control, like turning off ads based on a specific interest, are somewhat ineffective, as research shows that, for example, over 25% of advertisers on Facebook do not make use of user interests to target their ads (Ali et al., 2023). Researchers have flagged in past reports that all users including youth would benefit from more control over the content that appears on their feeds (Marwick et al., 2024).

A broader discussion on the topic was whether ads are at all appropriate for younger users, given that they may be more vulnerable to manipulation, and may lack the literacy skills to differentiate between ads and content. This is a years-long argument predating the internet-era, with orgs like Fairplay advocating for a 'commercial-free childhood' (*The Facts about Marketing to Kids*, n.d.). Thus, some participants suggested that platforms should be prohibited from presenting ads to minors altogether.

Participants advocated that platforms should provide user-friendly tools to enable users to proactively establish preferences about items to be recommended and items to be blocked to ensure the content is aligned with what they actually want to see.

Future Opportunities: More preference-based and control-based tools to manage content

1. Replace engagement-based with preference-based.

Some participants suggested that platforms should consider moving away from engagement-based algorithms toward models that focus on user preferences, as previous advocacy efforts have suggested (Szymielewicz, 2024). For instance, recommender systems can allow users to select their interests, or to flag content they are not interested in, even if their behavior (i.e., engaging with content) signals otherwise to platforms. Participants said that research into non-engagement-based content curation models offers a promising alternative to current recommendation systems (Cunningham et al., 2024; Stray et al., 2024). These models could present content based on a user's stated preferences rather than relying solely on their past interactions, potentially leading to more accurate and safer recommendations. Participants advocated that platforms should provide user-friendly tools to enable users to proactively establish preferences about items to be recommended and items to be blocked to ensure the content is aligned with what they actually want to see.

2. Human oversight in moderation when possible.

Participants proposed that human oversight in content moderation should be prioritized, especially in cases where algorithms are likely to miss the nuance of what is appropriate for different age groups, and where there may be cultural differences in what people consider to be suitable for young users. Further, participants noted, moderation teams with expertise in child safety and digital literacy should play a larger role in curating content for young users.

For example, some suggested encouraging platforms to implement just-in-time nudging features that warn users before they view inappropriate content, like nudity detectors or content warnings. This proactive approach is said to give young users more control over unwanted exposure to content without relying entirely on reactive blocking and reporting tools.

3. Proactive content protections.

Participants shared that some platforms offer parental control tools that would mediate content, but that they are often ineffective and difficult to use. According to participants, parental control tools often offer limited options and lack granularity, leaving significant gaps in the types of content that can be filtered. Some mentioned that parental controls also do not keep up with emerging forms of content or the slang that youth use to bypass restrictions.

Instead, participants say that previous research and design efforts have proposed a range of platform-side features that could scaffold content for young people in a manner that leaves control in users' hands and supports their sense of agency, potentially dissuading them from using circumvention tactics. For example, some suggested encouraging platforms to implement just-in-time nudging features that warn users before they view inappropriate content, like nudity detectors or content warnings. This proactive approach is said to give young users more control over unwanted exposure to content without relying entirely on reactive blocking and reporting tools.

Several participants mentioned that in the context of sensitive content sharing, such as intimate content, more robust protections should be implemented, for example, ones that would prevent screenshots of intimate content. That said, participants noted that users also need to be able keep records of abusive behavior for reporting purposes. One way that was suggested to address this gap is to apply an e-discovery type approach in which platforms retain records of unsent or expired content for a disclosed period of time to equip users to send or refer to a report regarding that content later (*SafeDigitalIntimacy.Org*, n.d.a).

4. Greater transparency and awareness of safety features.

In addition to providing users with greater control, participants noted that platforms should offer more transparency about content presentation and content moderation. They suggested that a better understanding of these processes alongside increased visibility of what content-related tools and settings exist or are turned on by default can support user agency in shaping their own experience and staying safe online. Although platforms have implemented new safety features like the ability to block specific hashtags or types of content, young users do not always know their feeds are able to be changed.

Transparency should also include transparency around ads, participants noted, as well as user-friendly and semantic-based ad control systems that would allow youth and their parents to clearly indicate the kinds of ads they do or do not want to see. This would help limit exposure to inappropriate ads that research shows are currently slipping through the cracks of automated systems (Gak et al., 2022). Some symposium participants pointed to prior research that explored ways of making ad control more accessible to users (Im et al., 2023); others suggested that banning advertising to youth altogether should also be considered, although this may raise First Amendment implications.

Communication: Connecting to Others on Digital Platforms

More than anything else, participants say youth use social media and other online platforms as a means of communication with others, including school friends and trusted adults, as well as individuals they do not know (Luria & Scott, 2023). Researchers and advocates asserted that young people's ability to communicate freely is vital for their development, and overly restricting this capacity can have significant downstream effects on their social and emotional well-being (Ybarra et al., 2015; Gray et al., 2023). Researchers in the symposium pointed out several current drawbacks in digital communication and how they can be addressed.

Current Weaknesses in Communication Channels with Others Online

1. Under-moderated digital environments

Platforms with high interaction rates and unique interaction modalities (e.g., avatars, in-game chat) are often under-moderated, according to participants. Although these can be spaces for peer connection and community, they can also expose youth to hate speech and harassment (Breuer, 2017). Some participants noted that unfiltered or live chat features, especially in multiplayer gaming and streaming platforms, are challenging to moderate effectively (Gorwa & Thakur, 2024). One participant noted the possibility of direct communication between users on these platforms that tend to have a greater volume of unknown users, which can pose safety risks to younger users.

According to researchers, many teens express concern that reporting someone's behavior can trigger retaliation, is likely to have negligible impact, or will simply be ignored.

2. Limitations of blocking and reporting mechanisms

Prior research has shown that youth generally understand blocking and reporting functions but tend not to use them due to perceived ineffectiveness and limited trust in platforms' response actions (Vilk & Lo, 2023); according to researchers, many teens express concern that reporting someone's behavior can trigger retaliation, is likely to have negligible impact, or will simply be ignored. One participant shared that reporting inappropriate behavior and conversation in gaming environments is even more challenging, as it includes multiple steps and is more disruptive to the gameplay.

Participants pointed out another aspect of reporting that is not as frequently talked about but that is especially relevant in the context of gaming platforms — the weaponization of safety features such as reporting and blocking to exclude and harass others. According to participants, safety tools, while essential, are sometimes exploited; recent reports suggest that young people use reporting and blocking tools maliciously to remove others' content or to prevent others from accessing platforms (i.e., "mass reporting") (Han et al., 2023; Elswah, 2024). They suggest that such abuse often goes unchecked on gaming platforms and raises equity concerns, particularly for vulnerable and marginalized youth.

Participants in the symposium highlighted the substantial impact of platform design choices on youth behavior, particularly the reliance on default interaction and communication settings. These defaults often dictate how users initially engage with the platform, shaping their experiences and setting the tone for their online behavior.

3. Impact of social media norms and default settings on youth behavior

As social media content shapes norms and expectations among youth, participants in the symposium highlighted the substantial impact of platform design choices on youth behavior, particularly the reliance on default interaction and communication settings. These defaults often dictate how users initially engage with the platform, shaping their experiences and setting the tone for their online behavior. Some researchers argue that existing defaults tend to prioritize engagement or ease of access, but that they tend to perpetuate risks of addictive online behavior and potentially increase exposure to harmful conduct (Flayelle et al, 2023).

Additionally, participants underscored the variability in moderation frameworks and default protections across different platforms. Some expressed concern that when teens gravitate toward platforms with weaker moderation and safeguards, they may be exposed to more risk. In such spaces, participants emphasized that the need for robust defaults designed with user safety in mind is even more crucial.

Future Opportunities: Scaffolding safe communication for youth

1. Privacy-first, opt-in public communication.

Participants suggested that platforms should implement the most privacy-preserving settings as default for youth, and potentially all, accounts, requiring opt-in for expanded communication options or public profile settings.

Participants suggested that platforms should implement the most privacy-preserving settings as default for youth, and potentially all, accounts, requiring opt-in for expanded communication options or public profile settings.

2. Simplified and accessible user tools.

There was also agreement that platforms should streamline reporting features to be accessible and easy to use, with an emphasis on gaming platforms that tend to have fewer user safety tools. That could include, for example, a tool that informs users of basic response actions (e.g., muting, blocking, reporting) as well as any additional responses, such as temporary removal from chat groups or restricted private communication from unknown users. Participants emphasized that transparency on both feature availability and platform response processes (such as follow-up notifications of reported users), could encourage more consistent engagement with reporting tools (Luria & Scott, 2023).

3. Research on diverse user needs and feature efficacy.

Participants highlighted the need for inclusive research focusing on how different demographic groups experience platform communication managing tools, especially their engagement with privacy settings and reporting tools. Additional studies are also needed on the efficacy of new tools that have been introduced in various platforms, as well as on the trade-offs between safety and user connection, and how they manifest among different demographic groups.

4. Digital literacy, skill building and community support.

Some participants suggested investing in integration of onboarding modules that educate users on platform tools, privacy settings, and safe interaction with others as an approach to raise awareness to risks and mitigation strategies, especially if designed in an engaging and interactive manner. Such tutorials, they said, can help teens establish healthy online behaviors and identify appropriate responses to negative interactions. Previous research has similarly suggested that helping kids gain digital literacy may be more effective in the long run than restricting access (Livingstone et al., 2019).

Characteristics: Design Choices that Impact Interaction

Much of the current policy discussion on child safety focuses on the design features and affordances of online services that guide users towards either "negative" or "positive" experiences online. Some features that gained extensive attention from both researchers and policymakers include content autoplay, recommendation to add friends, and notifications 'nudges' that attempt to encourage users to interact when they have gone idle online. While there are many specific features that may impact behavior, participants' conversation centered mostly on features that attempt to increase time online and the tendency towards parental control. Participants also extensively discussed the dire need for more data and research that examines the impact of design patterns and safety features.

Current Weaknesses in the Design Choices that impact Youth Behavior Online

1. Design elements encouraging extended use.

Participants noted that infinite scroll, autoplay, and variable rewards exemplify "persuasive design" or "dark patterns," user interface elements designed specifically to retain users, including youth, on a platform for extended periods of time. They suggested that these features often prioritize platform engagement over user well-being and may contribute to the overuse of digital platforms. While there have been some efforts to monitor these behaviors, like screen-time overviews, little is known about the effectiveness of these features and how they interact with session-prolonging designs on digital platforms.

2. Limited data on the impact on wellbeing and mental health.

Participants identified that part of the reason there is little known about the outcomes of a range of design choices and platform features is that most data collected by companies lacks standardized, validated measures of mental health, loneliness, and overall user wellbeing, and any such data that does exist is not made available to independent researchers. Instead, datasets often prioritize usability metrics or opinion polls, offering a narrow view of user experiences and overlooking impact on user wellbeing.

Researchers said that this creates a critical knowledge gap about how specific design features affect users, both on and off the platform. They noted that it is also very difficult to measure the impact of a single change to a user interface, especially as many changes are not optional for users. For instance, the impact of platform content warnings on users' mental health and well-being remains largely unexplored and empirically hard to measure. Without more data that aims to answer these complex and sensitive questions, researchers noted, it is very difficult to assess which safety features and designs can foster healthier behavior.

3. Parental controls as a default protection measure.

Many platforms provide parental controls to improve and oversee child safety online, yet participants explained that these are frequently difficult to navigate or are inadequate, and that many parents lack the digital literacy to use them effectively. One participant noted that parents with multiple jobs, those who do not speak a dominant language, or those from different socioeconomic or ethno-cultural backgrounds may have different exposure or willingness to make use of parental controls (Milosevic et al., 2022); a study conducted by Data & Society finds that parents and caregivers with higher incomes and more education are more likely to help their children navigate privacy settings and other online spaces than parents with lower incomes and educational attainment levels (Redmiles, 2018).

Moreover, participants noted that parental intervention has been shown to be limited in its effectiveness and can even lead to adverse outcomes (Stoilova et al., 2023). Thus, they argued that while a focus on parental controls and features may meet the expectations of parents, it tends to overlook the perspectives and needs of young users themselves. Participants particularly raised concerns related to parental control and supervision tools being weaponized against young LGBTQ+ users or any users who find themselves in unsupportive family dynamics or have parents who are unable to navigate digital environments.

Future Opportunities: More customized experiences alongside research and data sharing

1. Introducing features for greater user control and flexibility.

Participants suggested that platforms could enhance user agency by offering tools that provide more control and more choice regarding how platforms and features are presented to them. One example that was discussed was an "algorithm reset" button would allow users to "clear" their algorithmic preferences when they notice the platform pushing unwanted content, be that extreme content or simply content they are not interested in.

2. Research on the effectiveness of screen time reduction and other safety features.

Attendees stated that in recent years platforms have introduced many kinds of safety features, but there is very limited understanding on how such features (e.g., screen time limits, content filters, etc.) truly impact user wellbeing on digital platforms. They said that there is a need for more research to be conducted, specifically research that evaluates real-world impacts of platform safety tools and features — understanding how effective these measures are across different platforms would be essential to determine what safety approaches work best. One participant elaborated that different methods may prove more suitable for certain types of platforms or user demographics.

Attendees stated that in recent years platforms have introduced many kinds of safety features, but there is very limited understanding on how such features (e.g., screen time limits, content filters, etc.) truly impact user wellbeing on digital platforms. They said that there is a need for more research to be conducted, specifically research that evaluates real-world impacts of platform safety tools and features.

3. Increased platform data and knowledge sharing.

Part of the reason for the minimal knowledge on the effectiveness of safety features, according to participants, is the limited access to research findings conducted within companies. Some participants noted that platform data could also be beneficial as it would allow independent or academic researchers to do the research themselves, but that too is restricted. Thus, participants agreed that encouraging greater transparency and fostering privacy-protective data sharing between companies and researchers can help shed light on the effectiveness of current and future safety features. This kind of knowledge would be crucial for stakeholders and policymakers to develop more informed and impactful safety tools for all users.

Overarching Themes and Areas of Agreement

The discussions during the symposium revealed that while each roundtable focused on distinct aspects of child safety online — Connection, Content, Communication, and Characteristics — many themes and challenges transcended individual categories. These overarching themes highlight the interconnectedness of the issues and the need for holistic approaches to improve youth safety in digital environments. Below, we synthesize some of those areas of agreement that emerged across discussions, such as the need for more research and access to data on current interventions, and the value of youth-led and youth-centric perspectives throughout.

1. Not all children are the same.

Across all themes, researchers repeatedly noted that young users experience online services and harms differently. Thus, it is important to tailor safety measures to specific user groups in addition to specific harms. Some researchers who had expertise working with marginalized children raised the importance of understanding how harms manifested for different youth communities and how to tailor solutions to their unique challenges. Overall researchers agreed that more research on different youth populations and their experiences online was essential to better inform future policy (*LGBTQ Young People of Color in Online Spaces*, 2023).

2. Not all children and parents use technology the same way.

Additionally, researchers flagged that different children and caregivers interact with technology differently. One researcher noted that the way users of all ages interact with online services and how they experience harmful outcomes differ by age, education level, socioeconomic status, and more. They argued that some young users face barriers when attempting to access safety tools that were not designed with their needs in mind. Some common assumptions that participants raised include: that all households are the same; that digital skills and literacy are uniformly available across households; that all users have similar socioeconomic status and cultural backgrounds; that parents are involved in scaffolding digital communication and are involved in good faith; and that everyone has the same cognitive and physical ability to access online safety tools. One tool developed by researchers at the University of Notre Dame and Vanderbilt University that was shared with the group is a tool that collects user experiences on social media to aid in understanding diverse experiences (Badillo-Urquiola et al., 2022).

3. Limited data on current interventions and their effect on wellbeing.

Participants agreed that limited data access is part of what impedes a full and comprehensive understanding of the harms young people experience and the best ways to support them. Participants echoed that without access to significant and meaningful usage data from online services, online experiences and harms young people face online would be very difficult to capture, and as a result, researchers and policymakers would stay in the dark about recommending future safety measures. The conversation surfaced a few key issues on limited data.

Participants agreed that limited data access is part of what impedes a full and comprehensive understanding of the harms young people experience and the best ways to support them. Participants echoed that without access to significant and meaningful usage data from online services, online experiences and harms young people face online would be very difficult to capture, and as a result, researchers and policymakers would stay in the dark.

First, participants explained that the data online services make available to researchers is often of limited perspective or value and is increasingly difficult to access. For instance, X and Meta have historically made data available to researchers studying how their services fit into the larger information environment. But in the last few years, X has increased the price of API access, making it prohibitively expensive to most researchers and institutions; Meta has shut down its researcher-access platform, CrowdTangle; one participant noted that more niche services commonly used by young users offer even less consistent access to data.

Second, participants observed that very limited information is available about how young marginalized users benefit from and are harmed by digital communication, and what safety measures are effective in supporting them. This includes LGBTQ+ youth, youth of color, youth across socioeconomic strata, youth with disabilities, foster youth, and immigrant youth. The sensitivity of researching these communities already makes it a difficult area of research, and this is exacerbated, as participants noted, as academia tends to not hold this type of research in high regard, compared to studies that capture larger and more "general" user groups. In turn, it disincentivizes academics from doing research on more nuanced experiences, and instead encourages more generalizable research (i.e., research on a more representative sample or a sample of the "majority of users").

Finally, participants noted that data is even more limited on the youngest of online users, especially those under 13, as most platforms are required by law to restrict their services to users aged 13 and older. Nevertheless, some youth under 13 are using digital platforms, and are especially vulnerable to harm online.

4. The need for specificity in defining concerns.

Participants observed that concern frequently expressed about screen time and digital platform use often lacks specificity. They suggested that one way of moving forward toward addressing harms is to talk more specifically about people's fears; answering questions about concrete scenarios and interactions that youth experience, and specific harms that should be addressed can make discussions more productive.

One way of moving forward toward addressing harms is to talk more specifically about people's fears; answering questions about concrete scenarios and interactions that youth experience, and specific harms that should be addressed can make discussions more productive.

5. Recognition of the connection between offline and online harms.

Several participants argued that recognizing and addressing the interplay between offline environments and online behaviors is crucial to create more effective safety strategies. Some examples they noted were that children from homes with violence or low parental involvement may spend more time online, gravitate towards certain types of content, or engage on specific platforms in ways shaped by their offline circumstances. These patterns are examples that are often less influenced by parental controls or safety measures and more by lived experiences. By addressing these connections holistically, participants noted that platforms and policymakers can develop solutions that account for the broader context of kids' lives.

From Insights to Action: Practical Steps towards Safer Digital Spaces

The symposium underscored a critical need for design, moderation, and policy strategies that not only protect youth but also respect their agency. Participants agreed that this combined approach encourages youth empowerment while providing safety tools and improving digital literacy. Future efforts should extend to fostering safe, supportive environments that align with youth developmental needs and social realities, participants say, as a way of preparing them to engage responsibly and healthfully in a digitally interconnected world. Below, we share actionable strategies that participants identified as ways to potentially address current concerns about interaction and communication on digital platforms, and moving towards safer environments for young users and all users.

1. User defaults as a step towards protecting children.

Due to disparate levels of familiarity and literacy with online safety tools and practices, researchers agreed that setting high defaults is one step in the right direction, especially for young users. These defaults can include setting accounts to private upon creation, limiting recommendations of users' content to friends or people who follow them, or limiting messaging capabilities to people they know.

Future efforts should extend to fostering safe, supportive environments that align with youth developmental needs and social realities, participants say, as a way of preparing them to engage responsibly and healthfully in a digitally interconnected world.

2. Enhanced user control and customization.

Participants suggested that platforms should offer customizable tools that empower youth to shape their own experiences and to mitigate some of the risks they encounter themselves. Features that participants suggested include a reset option for recommendation algorithms, permanent filters to block certain types of content, and clear settings to opt out of data collection and personalization when using AI platforms.

3. Nuanced, research-driven and context-sensitive design choices.

Participants said that platforms should move beyond screen-time limits and adopt a more holistic perspective, focusing on use cases and contextual factors (e.g., the type of content viewed, purpose of use). Further research, including internal studies, could inform how design features affect youth based on variables such as mental health, cultural background, and personal experiences. This kind of research could help identify support mechanisms tailored to individuals' unique needs without compromising their autonomy.

4. Transparency and data access to individual researchers.

Participants identified a pressing need for platforms to provide researchers with controlled access to data, as mandated by policies like the Digital Services Act (DSA) in the EU. This transparency would help researchers evaluate the effectiveness of design changes and features that set out to promote youth wellbeing.

5. Develop resources for young people to acquire and share digital safety skills.

One researcher noted that young users needed to be taught digital skills and online safety "just like we teach skills in driver's education, health and sex education" and that it required repetitive training and frequent reminders.

One researcher noted that young users needed to be taught digital skills and online safety "just like we teach skills in driver's education, health and sex education" and that it required repetitive training and frequent reminders. Others agreed that kids need to be prepared to use digital tools and platforms over the course of their lives — shielding them completely until they turn eighteen is unlikely to be the answer. Participants suggested that a better approach may be to gradually prepare them as they grow, providing increasing autonomy along with the necessary scaffolding (Park et al., 2024).

Researchers further noted that not all users, particularly young users, were aware of the possible harms they may be exposed to online. For example, a poll conducted by Pew Research found that LGBTQ+ youth were more likely than their heterosexual peers to think that social media may expose them to harm online (Gelles-Watnick & Vogels, 2023). The finding held with generative AI systems, with 34% of LGBTQ+ youth asserting that they didn't use generative AI tools due to concerns about inaccuracy and bias, as compared to only 14% of their straight and cis-gendered peers (Odgers & Jensen, 2020). This highlights the opportunity for greater investment into digital literacy and more tailored literacy resources for different communities.

One promising opportunity to make sure online safety principles were taught and retained that was mentioned in conversation is enabling peer-to-peer education and discussion. A participant shared prior research that found that young users are the most effective in setting good practices and modeling social norms for other young users on online platforms. In other words, young users are more likely to listen to other young users when it comes to developing an understanding of norms online (James et al., 2017).

6. Consider pathways to shape norms on online platforms.

Some researchers noted that developing robust onboarding processes on platforms commonly used by young users may be one way to familiarize them with best practices and norms that would keep them and others safe (Digital Wellness Lab at Boston Children's Hospital, 2023). Researchers noted that some platforms already hosted versions of this — for example, gaming spaces and online groups sometimes require users to agree to rules or restate them before being granted access to an online service. Limitations were also raised, as some participants argued this can quickly become performative, as in click-through end-user license agreements.

Conclusion

The symposium underscored the complexity of creating safer online environments for young users. With a range of backgrounds and expertise, participants explored the nuanced challenges and opportunities in addressing child safety online, primarily on four broad and predefined topics: connection, content, communication and characteristics.

A recurring theme throughout the event was the need for collaboration between researchers, policymakers, and platform designers. Bridging the gap between research and policy is essential to ensuring evidence-based, rights-respecting digital environments that protect and empower young users. Fostering partnerships and developing frameworks would strengthen the integration of research insights into policymaking processes.

The symposium marked an important step forward by highlighting a range of challenges and opportunities for action. Nevertheless, participants acknowledged that achieving safer online spaces for youth will demand continued dialogue informed by rigorous research and inclusive perspectives, as well as follow-up action, such as building coalitions and piloting some of the suggested interventions while measuring their effectiveness.

By sustaining momentum and expanding on the strategies discussed, researchers, policymakers and platform developers can work toward a shared goal: creating a digital ecosystem where young people and all people thrive, are safe, and have their rights and agency upheld.

■

References

Ali, M., Goetzen, A., Mislove, A., Redmiles, E. M., & Sapiezynski, P. (2023). *Problematic Advertising and its Disparate Exposure on Facebook*. 5665–5682. https://www.usenix.org/conference/usenixsecurity23/presentation/ali [perma.cc/E9UZ-JQKY]

Badillo-Urquiola, K., Antoine, C., Nisenbaum, A., Daniel Shea, Z., & J. Wisniewski, P. (2022). "30 Days:" An EMA Diary Mobile App & Web Tool. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–5. https://doi.org/10.1145/3491101.3519888 [perma.cc/XFY3-268M]

Borak, M. (2024). *Belgium launches national digital identity wallet | Biometric Update*. https://www.biometricupdate.com/202405/belgium-launches-national-digital-identity-wallet [perma.cc/7EHB-22XC]

boyd, d., Hargittai, E., Schultz, J., & Palfrey, J. (2011). Why parents help their children lie to Facebook about age: Unintended consequences of the "Children's Online Privacy Protection Act." *First Monday*. https://doi.org/10.5210/FM.V16I11.3850 [perma.cc/WEH8-8SFM]

Brennen, S., & Perault, M. (2025). Keeping Kids Safe Online: How Should Policymakers Approach Age Verification? *The Center for Growth and Opportunity*. https://www.thecgo.org/research/keeping-kids-safe-online-how-should-policymakers-approach-age-verification/ [perma.cc/A29T-FSTD]

Charmaraman, L., Lynch, A. D., Richer, A. M., & Grossman, J. M. (2022). Associations of early social media initiation on digital behaviors and the moderating role of limiting use. *Computers in Human Behavior, 127*, 107053. https://doi.org/10.1016/j.chb.2021.107053 [perma.cc/J5GR-2UUA]

Chiong, C., & Shuler, C. (2010). Learning: Is there an app for that? *Joan Ganz Cooney Center*. https://joanganzcooneycenter.org/publication/learning-is-there-an-app-for-that/ [perma.cc/5APF-E6VF]

Complying with COPPA: Frequently Asked Questions. (2020, July 20). Federal Trade Commission. https://www.ftc.gov/business-guidance/resources/complying-coppa-frequently-asked-questions [perma.cc/5GVM-TMUU]

Cunningham, T., Pandey, S., Sigerson, L., Stray, J., Allen, J., Barrilleaux, B., Iyer, R., Milli, S., Kothari, M., & Rezaei, B. (2024). *What We Know About Using Non-Engagement Signals in Content Ranking* (arXiv:2402.06831). arXiv. https://doi.org/10.48550/arXiv.2402.06831 [perma.cc/C3ZC-SMA7]

Digital Wellness Lab at Boston Children's Hospital. (2023). *Creating a Positive Foundation for Greater Civility in Online Spaces*. https://digitalwellnesslab.org/wp-content/uploads/Digital-Wellness-Lab-White-Paper-Civility-Online.pdf [perma.cc/KRC3-9UKY]

Duarte, N., & Llansó, E. (2017, November 28). *Mixed Messages? The Limits of Automated Social Media Content Analysis.* Center for Democracy and Technology. https://cdt.org/insights/mixed-messages-the-limits-of-automated-social-media-content-analysis/ [perma.cc/9NUR-Y8T9]

Elswah, M. (2024). *Moderating Maghrebi Arabic Content on Social Media*. Center For Democracy And Technology. https://cdt.org/insights/moderating-maghrebi-arabic-content-on-social-media/ [perma.cc/2698-HV5D]

EuConsent Making the internet Age-Aware. (n.d.). EuConsent. Retrieved December 13, 2024, from https://euconsent.eu/ [perma.cc/EAT7-FMM7]

European digital identity (eID): Council adopts legal framework on a secure and trustworthy digital wallet for all Europeans. (n.d.). Consilium. Retrieved January 14, 2025, from https://www.consilium.europa.eu/en/press/press-releases/2024/03/26/european-digital-identity-eid-council-adopts-legal-framework-on-a-secure-and-trustworthy-digital-wallet-for-all-europeans/ [https://perma.cc/YZ6L-SMXY]

Family Online Safety Institute. (2022). *Making Sense of Age Assurance: Enabling Safer Online Experiences*. https://cdn.prod.website-files.com/5f47b99bcd1b0e76b7a78b88/636d1325723267567 2619f45_MAKING%20SENSE%20OF%20AGE%20ASSURANCE%20FULL%20REPORT%20-%20FOSI%202022_compressed.pdf [perma.cc/S4Q8-FNQE]

Forland, S., Meysenburg, N., & Solis, E. 2024. *Age Verification: The Complicated Effort to Protect Youth Online*. (n.d.). New America. Retrieved January 14, 2025, from http://newamerica.org/oti/reports/age-verification-the-complicated-effort-to-protect-youth-online/ [perma.cc/W63B-LW7X]

Gak, L., Olojo, S., & Salehi, N. (2022). The Distressing Ads That Persist: Uncovering The Harms of Targeted Weight-Loss Ads Among Users with Histories of Disordered Eating. *Proc. ACM Hum.-Comput. Interact., 6*(CSCW2), 377:1-377:23. https://doi.org/10.1145/3555102 [perma.cc/7YJE-Z8RV]

Gelles-Watnick, R., & Vogels, E. A. (2023, April 24). Teens and social media: Key findings from Pew Research Center surveys. *Pew Research Center*. https://www.pewresearch.org/short-reads/2023/04/24/teens-and-social-media-key-findings-from-pew-research-center-surveys/ [perma.cc/SVH2-7AX5]

Gorwa, R., & Thakur, D. (2024, November 21). *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/ [perma.cc/2YS4-MWTC]

Gray, P., Lancy, D. F., & Bjorklund, D. F. (2023). Decline in Independent Activity as a Cause of Decline in Children's Mental Well-being: Summary of the Evidence. *The Journal of Pediatrics, 260*. https://doi.org/10.1016/j.jpeds.2023.02.004 [perma.cc/T34V-33AV]

Han, C., Seering, J., Kumar, D., Hancock, J. T., & Durumeric, Z. (2023). Hate Raids on Twitch: Echoes of the Past, New Modalities, and Implications for Platform Governance. *Proc. ACM Hum.-Comput. Interact., 7*(CSCW1), 133:1-133:28. https://doi.org/10.1145/3579609 [perma.cc/AM5P-9VLF]

Hanaoka, K., Ngan, M. L., Yang, J., Quinn, G. W., Hom, A., & Grother, P. J. (2024). *Face analysis technology evaluation: Age estimation and verification* (NIST IR 8525; p. NIST IR 8525). National Institute of Standards and Technology (U.S.). https://doi.org/10.6028/NIST.IR.8525 [perma.cc/D9KB-NVBE]

Hasinoff, A. A. (2015). *Sexting panic: Rethinking criminalization, privacy, and consent*. University of Illinois Press. https://books.google.com/books?hl=en&lr=&id=lD7oBgAAQBAJ&oi=fnd&pg=PP1&dq=sexting+panic&ots=Tw3hT1_zej&sig=-pfaaSRBO_VKCyKM8gEGp7itwZU [perma.cc/DF9A-LJLV]

Im, J., Wang, R., Lyu, W., Cook, N., Habib, H., Cranor, L. F., Banovic, N., & Schaub, F. (2023). Less is Not More: Improving Findability and Actionability of Privacy Controls for Online Behavioral Advertising. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–33. https://doi.org/10.1145/3544548.3580773 [perma.cc/4HRJ-CHTY]

James, C., Davis, K., Charmaraman, L., Konrath, S., Slovak, P., Weinstein, E., & Yarosh, L. (2017). Digital life and youth well-being, social connectedness, empathy, and narcissism. *Pediatrics, 140*(Supplement_2), S71–S75. https://publications.aap.org/pediatrics/article/140/Supplement_2/S71/34171/Digital-Life-and-Youth-Well-being-Social [https://perma.cc/WE5C-8Q6S]

Kids Online Health and Safety Taskforce. (2024). *Online Health and Safety for Children and Youth: Best Practices for Families and Guidance for Industry*. https://www.ntia.gov/sites/default/files/reports/kids-online-health-safety/2024-kohs-report.pdf [perma.cc/EL6H-AX4E]

LGBTQ Young People of Color in Online Spaces. (2023, July 19). The Trevor Project. https://www.thetrevorproject.org/wp-content/uploads/2023/07/The-Trevor-Project_LGBTQ-Young-People-of-Color-in-Online-Spaces.pdf [https://perma.cc/TGV8-STCQ]

Livingstone, S., Stoilova, I., & Nandagiri, R. (2019). Children's data and privacy online: Growing up in a digital age. *London: London School of Economics and Political Science, 24*(1), 153–173. https://eprints.lse.ac.uk/101283/1/Livingstone_childrens_data_and_privacy_online_evidence_review_published.pdf [https://perma.cc/2VS7-JUWP]

Luria, M., & Foulds, N. (2021). Hashtag-Forget: Using Social Media Ephemerality to Support Evolving Identities. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems,* 1–5. https://doi.org/10.1145/3411763.3451734 [perma.cc/Y4VC-8KYJ]

Luria, M., & Scott, C. F. (2023, November 9). *More Tools, More Control: Lessons from Young Users on Handling Unwanted Messages Online*. Center for Democracy & Technology. https://cdt.org/insights/more-tools-more-control-lessons-from-young-users-on-handling-unwanted-messages-online/ [perma.cc/AMK4-RBZB]

Marwick, A., Smith, J., Caplan, R., & Wadhawan, M. (2024). Child Online Safety Legislation (COSL)—A Primer. *The Bulletin of Technology & Public Life*. https://doi.org/10.21428/bfcb0bff.de78f444 [perma.cc/B4KY-8367]

Milosevic, T., Kuldas, S., Sargioti, A., Laffan, D. A., & O'Higgins Norman, J. (2022). Children's Internet Use, Self-Reported Life Satisfaction, and Parental Mediation in Europe: An Analysis of the EU Kids Online Dataset. *Frontiers in Psychology, 12,* 698176. https://doi.org/10.3389/fpsyg.2021.698176 [perma.cc/2ZQ6-2KTL]

Nicholas, G., & Bhatia, A. (2023). *Lost in Translation: Large Language Models in Non-English Content Analysis* (arXiv:2306.07377). arXiv. https://doi.org/10.48550/arXiv.2306.07377 [perma.cc/TY4T-S5C7]

Odgers, C. L., & Jensen, M. R. (2020). Adolescent mental health in the digital age: Facts, fears, and future directions [Annual research review]. *Journal of Child Psychology and Psychiatry, 61*(3), 336. https://doi.org/10.1111/jcpp.13190 [https://perma.cc/B64N-JJPQ]

Ofcom. (2022, October 11). *A third of children have false social media age of 18+*. Www.Ofcom.Org.Uk. https://www.ofcom.org.uk/online-safety/protecting-children/a-third-of-children-have-false-social-media-age-of-18/ [perma.cc/F63L-DLQU]

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. IEEE Security & Privacy. https://doi.org/10.1109/MSEC.2024.3417804 [perma.cc/US6M-8SSY]

Qin, L., Hamilton, V., Wang, S., Aydinalp, Y., Scarlett, M., & Redmiles, E. M. (2024). *"Did They {F***ing} Consent to That?": Safer Digital Intimacy via Proactive Protection Against {Image-Based} Sexual Abuse*. 55–72. https://www.usenix.org/conference/usenixsecurity24/presentation/qin [perma.cc/9X47-C2E3]

Raffoul, A., Ward, Z. J., Santoso, M., Kavanaugh, J. R., & Austin, S. B. (2023). Social media platforms generate billions of dollars in revenue from U.S. youth: Findings from a simulated revenue model. *PLOS ONE, 18*(12), e0295337. https://doi.org/10.1371/journal.pone.0295337 [perma.cc/9JDC-PL6M]

Redmiles, E. (2018). Net Benefits: Digital Inequities in Social Capital, Privacy Preservation, and Digital Parenting Practices of U.S. Social Media Users. *Proceedings of the International AAAI Conference on Web and Social Media, 12*(1). https://doi.org/10.1609/icwsm.v12i1.14997 [perma.cc/C9XW-WW9U]

Redmiles, E. (2021). *Apple's New Child Safety Technology Might Harm More Kids Than It Helps*. Scientific American. https://www.scientificamerican.com/article/apples-new-child-safety-technology-might-harm-more-kids-than-it-helps/ [perma.cc/2Q6K-X86H]

Ruane, K., Branum, B., Doty, N., & Jain, S. (2024, September 23). CDT Files Amicus Brief in Free Speech Coalition v. Paxton, Challenging TX Age Verification Law. *Center for Democracy and Technology*. https://cdt.org/insights/cdt-files-amicus-brief-in-free-speech-coalition-v-paxton-challenging-tx-age-verification-law/ [perma.cc/M8NZ-ZBHB]

SafeDigitalIntimacy.org. (n.d.)a. Safer Digital Intimacy. Retrieved January 23, 2025, from https://www.safedigitalintimacy.org/recommended-features [perma.cc/UX3T-KNNP]

SafeDigitalIntimacy.org. (n.d.)b. State of the Industry. Retrieved January 23, 2025, from https://www.safedigitalintimacy.org/state-of-the-industry [https://perma.cc/RE4T-HSAQ]

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 [https://perma.cc/Q4P3-XFW8]

Stray, J., Halevy, A., Assar, P., Hadfield-Menell, D., Boutilier, C., Ashar, A., Beattie, L., Ekstrand, M., Leibowicz, C., Sehat, C. M., Johansen, S., Kerlin, L., Vickrey, D., Singh, S., Vrijenhoek, S., Zhang, A., Andrus, M., Helberger, N., Proutskova, P., ... Vasan, N. (2024). Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *ACM Transactions on Recommender Systems, 2*(3), 1–57. https://doi.org/10.1145/3632297 [perma.cc/V6AG-CF5B]

Szymielewicz, K. (2024). *Safe by Default: Moving away from engagement-based rankings towards safe, rights-respecting, and human centric recommender systems*. Panoptykon Foundation. https://panoptykon.org/sites/default/files/2024-03/panoptykon_peoplevsbigtech_safe-by-default_briefing_03032024.pdf [perma.cc/UES7-997B]

Thakur, D., Grant-Chapman, H., & Laird, E. (2023, July 31). *Beyond the Screen: Parents' Experiences with Student Activity Monitoring in K-12 Schools*. Center for Democracy and Technology. https://cdt.org/insights/report-beyond-the-screen-parents-experiences-with-student-activity-monitoring-in-k-12-schools/ [perma.cc/B5K5-CJRH]

The Facts About Marketing to Kids. (n.d.). Campaign for a Commercial-free Childhood. Retrieved January 23, 2025, from https://nepc.colorado.edu/sites/default/files/CERU-0502-115-OWI.pdf [perma.cc/3PWC-KMUT]

Vilk, V., & Lo, K. (2023). *Shouting into the Void: Why Reporting Abuse to Social Media Platforms Is So Hard and How to Fix It—PEN America*. https://pen.org/report/shouting-into-the-void/ [perma.cc/9GVU-CKKC]

Vogels, E. A. (2021, June 22). Digital divide persists even as Americans with lower incomes make gains in tech adoption. *Pew Research Center*. https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/ [perma.cc/QP6Y-BNXT]

Weaver, C., & Issawi, D. (2021, September 30). 'Finsta,' Explained. *The New York Times*. https://www.nytimes.com/2021/09/30/style/finsta-instagram-accounts-senate.html [https://perma.cc/68ZY-VWMC]

Ybarra, M. L., Mitchell, K. J., Palmer, N. A., & Reisner, S. L. (2015). Online social support as a buffer against online and offline peer and sexual victimization among U.S. LGBT and non-LGBT youth. *Child Abuse & Neglect, 39*, 123–136. https://doi.org/10.1016/j.chiabu.2014.08.006 [perma.cc/F9VJ-YP4A]

York, J. C. (2021, August 18). *How LGBTQ+ Content is Censored Under the Guise of "Sexually Explicit."* Electronic Frontier Foundation. https://www.eff.org/deeplinks/2021/08/how-lgbtq-content-censored-under-guise-sexually-explicit [perma.cc/VXH6-4SG8]

York, C. R., Dan. (2024, September 23). Age Verification Law Weakens Internet Privacy and Security. *Internet Society*. https://www.internetsociety.org/blog/2024/09/texas-mandatory-age-verification-law-will-weaken-privacy-and-security-on-the-internet/ [perma.cc/XL3C-5XQ4]

cdt.org

cdt.org/contact

Center for Democracy & Technology
1401 K Street NW, Suite 200
Washington, D.C. 20005

202-637-9800

@CenDemTech

@cendemtech.bsky.social

CENTER FOR
DEMOCRACY
& TECHNOLOGY