Legislative Approaches to Combating Online Harms to Children

A guide to the approaches adopted or proposed by legislators to protect children from online harms.

Tim Bernard is an analyst and writer with an MBA from Cornell Tech, who focuses on content moderation and associated tech policy issues.
Table of contents

Legislative approaches to combating online harms to children
I. Introduction
   1. Do it for the children
II. The history and philosophy of combating online harms to children
   1. “The Internet Is for Porn”—but not for children
   2. Children’s rights
   3. Parental rights
III. Political dynamics
IV. Harms addressed by legislation
V. Harms created by legislation
VI. Review of approaches and their challenges
   Age assurance
      Varieties of age assurance
      1. Age-gating and content filtering
      2. Parental involvement
      3. Age-appropriate design codes
      4. Privacy
      5. Advertising
      6. Cyberbullying and sexting
      7. Perpetrators of sexual abuse and law enforcement
      8. Education
VII. Outlook
   1. Industry engagement
   2. Outcomes from enacted legislation
      a. Researcher access
      b. Risk assessments
      c. Regulator enforcement
   3. Lawsuits
   4. Technological advances
VIII. Conclusion
APPENDIX
   1. Mental health
   2. Pornography
   3. Dangerous behavior
   4. Sexual abuse
5. Sexting and non-consensual intimate imagery
6. Cyberbullying
7. Data privacy violations
8. Drugs and dangerous items
9. Advertising and financial harms

Legislative approaches to combating online harms to children

I. Introduction

Over the last few years, there has been a huge amount of legislative activity, particularly in the US, aiming to protect children who use the internet. This paper will lay out the relevant harms that children are thought to be subject to online and describe the major trends in policy responses—indeed, many identical and near-identical bills have been proposed in US State legislatures—as well as detailing outlier proposals. Current legal and political challenges to this legislation will also be discussed.

The related topic of tackling the sharing of child sexual abuse material (CSAM) amongst adult internet users is out of scope for this paper, and the complex issue of data privacy will not be discussed in great detail.

1. Do it for the children

It appears to be easier to forge consensus and build political capital to pass laws to protect children from online harms, even where similar concerns exist with respect to adults. (Indeed, in other contexts, some rules have been adopted for minors and later introduced for adults too, e.g. in 2016 the FTC announced that it was considering persistent identifiers to be personally identifiable information, a standard they had established for children in 2012.) Earlier drafts of the UK’s Online Safety Act, for example, contained strictures related to content that is legal but considered “harmful” to adults as well as to children, but the enacted version only retained this category for children.

For each harm and each proposed legislative remedy in this paper, one might also ask how it could apply to adults:

- Which, if any, of the harms are particular to children?
- Which remedies would alleviate identical or similar harms for adults?
- How do trade-offs (e.g. with burden on companies, privacy or speech rights) change if the remedies were applied to adults too? And what does this say about the way we see the place of children in society?
II. The history and philosophy of combating online harms to children

1. “The Internet Is for Porn”—but not for children

From the prehistory of the internet, on BBS servers and Usenet, and onwards there was a glut of material that many would deem unsuitable for children, chiefly pornography. In 1995, Time published a cover story about a study of online pornography, asking on the front cover, “Can we protect our kids—and free speech?” (The study was later widely critiqued, and the magazine story’s author, twenty years later, quoted his colleague as calling it, “one of the more shameful, fear-mongering and unscientific efforts that we ever gave attention to.”) These concerns prompted the passage of the 1996 Communications Decency Act (CDA), which may be the first major piece of legislation targeting children’s exposure to inappropriate content on the internet. The law reflected a widely-held attitude where society has the obligation to keep conventionally inappropriate material away from children. The following year, the Supreme Court ruled the CDA to be unconstitutional in Reno v. ACLU on First Amendment grounds, with the exception of the liability shield provision that is now widely known as Section 230.

Over the following few years, further laws in the US, both federal and in the states, were passed that attempted to restrict online pornography. Many were struck down after constitutional challenges, including because they created barriers to adults accessing this content.

2. Children’s rights

In contrast to the traditional assumptions underpinning anti-pornography laws, the other major trend informing new regulations on protecting children online emerges from an international framework of universal rights. The UN adopted the Declaration of the Rights of the Child in 1959 and a significantly expanded Convention on the Rights of the Child in 1989. (The United States is the only UN member country that has not ratified the Convention.) According to these documents, children are entitled to not only special protections from harm but also certain rights that mirror those of adults, such as freedom of thought, speech and assembly, and rights to the conditions they need to develop as independent human beings. The International Telecommunications Union was one of the earliest institutions to bring this approach to internet issues with the establishment of their Child Online Protection Initiative, which released its first set of guidelines in 2009. Updated versions were published in 2014 (in collaboration with UNICEF) and most recently in 2020, when they created a separate guide for policymakers.
One key advocate for this approach has been the 5Rights Foundation, which was tasked with gathering the views of children around the world to inform the drafting of a 2021 General Comment to the UN Convention on the Rights of the Child on “children’s rights in relation to the digital environment.” Three particular focuses of the General Comment (and of the 5Rights Foundation) are: protecting children’s online privacy, consideration of how platform design may impact children negatively, and the use of impact assessments as a tool for ensuring that online services consider how their services may impact the rights of children among their user base.

3. Parental rights

A third approach seeks to empower parents to control their children’s online experiences. The parental rights framing is common in US First Amendment case law, and in child-protection mechanisms in statutes around the world. Offline, a string of largely conservative movements has attempted to give parents the option to exempt their children from various aspects of public education and school activities, as well as from receiving the COVID-19 vaccine.

Without necessarily being rooted in the same worldview, data privacy laws often require consent for data collection or processing in certain cases, and establish ages of consent below which parents have authority over their children’s data. The 1998 Children Online Privacy Protection Act (COPPA) requires internet services to obtain parental consent to collect, use, or share the personal information of a user that it knows to be under the age of 13. As enforced, this has led to many services being officially unavailable to under-13s, while only requiring a simple affirmation of majority age to create an account without the protections of this law. (In the EU, under the General Data Protection Regulation (GDPR), 13, 14, 15 and 16 are all operative ages for digital consent in various countries.)

III. Political dynamics

While philosophical and political issues relating to combating online harms to children can at times be difficult to separate, a few standout political dynamics can be identified from recent years.

One involves the influence of UK approaches to governing children’s experiences online. The British filmmaker and life peer, Baroness Beeban Kidron founded the 5Rights Foundation in 2014 to push for regulation that would protect the rights of children online, and has been instrumental in promoting Age-Appropriate Design Codes (AADCs) first in the UK and later in the US. Kidron has worked actively with US state lawmakers, in some cases leading to introduction or enactment of laws that closely follow the UK model. The public actions of technology companies in response to UK and EU regulation (such as GDPR, the Digital Services Act, and the UK’s AADC) may also be incentivizing action by US lawmakers.

Another involves the prominent role of US States—as opposed to federal lawmakers—in introducing and enacting laws protecting children from online harms. While lawmaking of this sort has been very active in
the states, nothing relevant has yet passed in Congress, although some draft federal laws such as the Kids Online Safety Act (KOSA), COPPA 2.0, Protecting Kids on Social Media Act, and the Strengthening Transparency and Obligations to Protect Children Suffering from Abuse and Mistreatment (STOP CSAM) Act have garnered attention and discussion. (Late in 2023, the Senate passed the Revising Existing Procedures On Reporting via Technology (REPORT) Act, so this narrower bill may end up becoming law.)

A final relevant political dynamic involves left/right politics. Support for bills regarding children and online harms in the abstract is bipartisan, and both red and blue states have enacted laws. Many of the bills that have been proposed federally (including Eliminating Abusive and Rampant Neglect of Interactive Technologies (EARN IT) Act, KOSA, REPORT Act, COPPA 2.0, Protecting Kids on Social Media Act) have had bipartisan sponsorship and support, but mistrust between the two parties may play a role in obstructing the passing of new laws, and expectations about how laws will be enforced may vary by party. A notable example: KOSA empowers State Attorneys-General, some of whom are perceived to have anti-LGBTQ+ political positions, to file suit under the law—though the February 2024 revisions substantially curtailed this power. The influential conservative think tank, the Heritage Foundation has openly endorsed using laws like KOSA to suppress content regarding transgender issues, and some LGBTQ+ groups had lobbied Democratic senators to oppose the bill. In the last few years, broadly speaking, different kinds of laws have passed in red and blue US states, most notably age-appropriate design codes being favored in Democratic strongholds, and straightforward age-gating (of pornography or social media) and parental consent laws passing in Republican-controlled legislatures.

IV. Harms addressed by legislation

While the three approaches outlined in section II may all be ideologically distinct, the legislation that has been proposed implies significant areas of overlap. Furthermore, even though laws emerging from a rights-based approach may not even need a justification rooted in known harms, the political will to create these laws typically emerges from awareness of (real or supposed) examples of children being harmed by their experiences online. In particular, the case of Molly Russell, a British teenager who died by suicide in 2017 and was known to have viewed self-harm content on social media, seems to have been a significant impetus behind the Online Safety Act’s development and eventual passage.

This section will briefly review some major harms that recent proposed and enacted legislation seeks to prevent. An appendix addresses each harm in more detail and with references.

- **Mental health**: Potential harms associated with social media use include, according to the US Surgeon General’s advisory report: low subjective well-being, depression, anxiety, poor body image, disordered eating, low self-esteem, self-harm, self-control challenges (related to addiction), sleep problems, attention problems and feelings of exclusion. Definitive evidence of causation, at least on a population level, has been elusive, with experts divided on the role played
by social media. Studies suggest, and the Surgeon General’s report recognized, particular benefits of social media usage for LGBTQ+ teenagers.

- **Pornography:** Traditionally assumed to be unsuitable for minors, pornography is now readily accessible to anyone online. There are ethical/methodological problems with researching this area, but the strongest connections have been drawn between pornography usage amongst adolescent males and sexual violence.

- **Dangerous behavior:** Young people have always been prone to engage in unwisely risky activities. Social media posts promoting such activities have included potentially fatal oxygen deprivation “games.”

- **Sexual abuse:** Abusers can use the affordances of online platforms, including anonymity, to approach minors and then abuse them online or offline. This can be facilitated by a grooming process or via “sextortion” after obtaining intimate imagery.

- **Sexting and non-consensual intimate imagery:** What may begin as willing sharing of intimate imagery with a partner can cause life-long harm if the content becomes public. Possession of images featuring underaged individuals may also pose serious legal risks to anyone, even to a minor that receives it from their partner.

- **Cyberbullying:** Bullying, a common harm offline for young people, can be exacerbated online, where bullies can reach into times and spaces that should be private and safe for their victim, occurring whenever and wherever they have internet access.

- **Data privacy:** Children are often understood to be below the age of consent for their data to be collected, used or shared. The collection of geolocation data that could provide critical information about a child’s whereabouts to someone who seeks to physically harm that child is a particular focus in some draft laws.

- **Drugs and dangerous items:** Minors have been able to obtain both illegal drugs and controlled items, such as weapons and graffiti supplies, online. This may be through ecommerce sites or informally via social media and messaging platforms.

- **Advertising and financial harms:** Children may be particularly susceptible to being persuaded by advertising and some seek to protect them from optimization-for-profit methods of social media platforms. Those with access to online payment methods have also been known to spend money on frivolous purchases and be victimized by online fraud.

V. Harms created by legislation

While proponents of child safety laws argue that they will redress harms, opponents often argue that they will only create new problems. In many cases, these are problems involving constitutional rights, meaning that the laws would be subject to legal challenge.
One major concern involves privacy. Whenever additional information about users, including their age, is required, or when platforms or vendors are required to track new information about anyone’s activity online, this requires the sacrifice of some degree of privacy. Privacy may be protected, whether by law, ethics, or the personal preference of the user. Legislation that requires platforms to verify or assess users’ ages may lead them to collect new sensitive information, ranging from government IDs to biometrics, from both adults and children. If platforms must withhold certain content (such as pornography) from certain users, this may entail tracking individual users’ online behavior in new ways. Some proposed legislation would also affect children’s privacy vis-à-vis parents, by requiring that parents approve or have opportunities to track children’s online actions, including messaging conversations with other minors.

A second major concern involves freedom of expression and information. When content or communications are impeded in any way, this may be an obstacle to free expression (including that of children), which is also protected by law, perhaps most radically by the First Amendment in the United States. We may be concerned not only about (adult or minor) users’ ability to make themselves heard online, but also to access the expression of others. As courts noted in striking down earlier child protection laws in the U.S., laws that effectively prevent adults from accessing lawful material, including pornography, may violate the First Amendment even if lawmakers might legitimately limit children’s access to that same content. Separately, some may identify concerns about online services’ own freedom to choose what content they host and to whom they make it available.

Both privacy and information access concerns can be particularly acute with regard to certain minority groups, in particular LGBTQ+ minors, who may require privacy—even from their parents—while accessing supportive community and important information online.

One final concern, perhaps less specific to this area of regulation than the first two, is that new regulatory requirements can be expensive for companies, and, as such, favor large incumbents who can afford the cost of compliance. (This has been suggested as an actual outcome of the process requirement of the GDPR regime.) In some cases, the more mature platforms may have already implemented many of the newly legislated features and practices, whereas newer startups, or established but smaller and less well-resourced companies may find compliance a struggle that sets back their ability to compete in their verticals, thus entrenching the advantage of current tech giants.

VI. Review of approaches and their challenges

This section will review a number of approaches taken by legislators to combat harms to children online. Some proposed and enacted laws take one approach to one harm; a few combine approaches in the attempt to tackle all online harms for both adults and children; many fall somewhere in between on one or more of these dimensions. Only those measures aimed specifically at minors are included here, and the section is organized by approach, rather than by legislation or by harm. (Examples of laws and bills are in
no way comprehensive!) Challenges, including legal, practical and political aspects, are also summarized for each approach.

Age assurance

Many legislative approaches require internet services to change their operations and offer different versions of their services when the user is a minor. The child-oriented version might, for example, display different content or offer different privacy protections. Some laws would require platforms to actively seek to ascertain users’ ages; others impose obligations that are triggered once a platform knows or "should know" that a user is a child.

The task of identifying the age of the user is broadly often referred to as “age assurance.” Age assurance technologies themselves can create new risks, including in relation to users’ privacy, as authorities including France’s data protection regulator have noted. One principle relevant to deciding on the level of certainty for age assurance is proportionality, or a “risk-based approach,” i.e., balancing the possible harm to minors from the content against the privacy compromises involved and the value of the content to adults. Laws vary widely in their specificity about what form of assurance is required for compliance, and some direct agencies to do the tricky work of defining the standards more concretely.

<table>
<thead>
<tr>
<th>Examples of language regarding age assurance for online services</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSA: “aware with reasonable certainty”</td>
</tr>
<tr>
<td>COPPA: “has actual knowledge”</td>
</tr>
<tr>
<td>COPPA 2.0: “actual knowledge or knowledge fairly implied on the basis of objective circumstances”</td>
</tr>
<tr>
<td>KOSA: “knows or reasonably should know”</td>
</tr>
<tr>
<td>Utah SB 152: “shall verify the age of an existing or new Utah account holder”</td>
</tr>
<tr>
<td>Connecticut SB 3: “consumers whom such controller has actual knowledge, or wilfully disregards, are minors”</td>
</tr>
<tr>
<td>California AADC: “Estimate the age of child users with a reasonable level of certainty appropriate to the risks that arise from the data management practices of the business”</td>
</tr>
</tbody>
</table>

Varieties of age assurance

This is an area of active research by NGOs, government agencies, academics and third-party vendors to identify how to optimize for certainty / accuracy, privacy, equity, ease-of-use and cost. Vendors are of
course incentivized to portray their products in the best light to policymakers, while privacy activists often stress the weaknesses of these technologies, much of which is very new. Online services may be most concerned with cost and barriers for users, and promote solutions that shift the burden of age assurance to other players. In some models, the services would collect identity data themselves and verify it or transmit it to a vendor for verification. Alternatively, government agencies or vendors could confirm the user’s information and provide a digital token to the services that need to assess ages without the service itself having access to the personal data involved. Some primary varieties of age assurance are:

- **Age verification:** requiring a government ID (this can be in hard copy in-person, an uploaded image or entered details, or using a digital token provided by a third party including an official agency) or transactional data, such as an account with a credit card, bank, utility provider or voting records.
- **Age estimation:** users’ ages can be probabilistically estimated through biometrics (such as facial features or keystroke dynamics), capacity testing (such as solving puzzles), or through data profiling similar to that currently used for “programmatic” or targeted digital advertising.
- **Account confirmation:** using an existing online account of the user or their responsible adult to vouch for the age of a new one.
- **Device-level control:** during device setup or manufacture, the age / date of birth of the user or a binary adult/minor variable is set, and can only be changed through a passcode or by an authorized third-party, such as a retailer.
- **Self-declaration:** users are asked to enter or confirm their age, or give a date of birth, when accessing a site or setting up an account. This method is especially vulnerable to user misrepresentation, but can be somewhat enhanced by not permitting changes or asking for subsequent confirmation of matching details.

1. **Age-gating and content filtering**

Many proposals to protect children from harms involve “age-gating” or preventing children from accessing particular kinds of content or services. Pornography sites are a particular focus of these efforts, but social media platforms have also been targets in some legislation. Content filtering at the device, local network or ISP level can be directed to not only block specific lists of sites, but also objectionable content by the webpage.

In order to block services or content from being accessible to children, the user’s age must be made known to the technology layer that is performing the filtering. At the broadest level of restriction are laws that may require device retailers or manufacturers, or alternatively ISPs, to filter the sanctioned services unless an adult user opts out, which may require showing a physical ID and paying a fee. Others might make the restrictions at this level opt-in. In approaches that focus on schools, vendors to schools, and libraries, the assumption can be made that any user on the network or device is a minor. Some laws may also prohibit adults from giving minors access to these services by providing them with unfiltered devices.
or passcodes to bypass the filters. Others can require the services themselves to use age assurance, often requiring a strong form of verification.

Perhaps the most salient critique of age-gating is the argument that existing technology to identify child users poses threats to the privacy of adults as well as children, because services must confirm the ages of all users in order to comply. Many laws attempt to mitigate this concern by prohibiting the retention of identity data after age assurance has been confirmed. However, users may not want to rely on this stipulation being strictly followed when faced with significant privacy risks such as their identities appearing on a stolen database of visitors to a pornographic website, along with the search terms they entered whilst there.

In the US, the precedents of the Communications Decency Act and subsequent laws suggest that age assurance can constitute an unacceptable barrier to legal content for adults under the First Amendment. Policymakers and age assurance vendors argue that the realities of new technology have changed the calculus, making this barrier much lower than it was a decade or two ago. Even if this claim is valid, privacy concerns and the issue of blocking content that minors themselves have a First Amendment right to access may still remain. Another legal pitfall for these laws regards how the legislation describes the content that needs to be blocked. If the language of the bill is too ambiguous in this regard, then the laws may not withstand First Amendment scrutiny.

Filtering of content by third-party vendors at the website or webpage page level, as in schools, is liable to lead to acceptable or beneficial content also being made inaccessible. (The Supreme Court in 2003 rejected a First Amendment challenge to a law that raised somewhat analogous issues by conditioning library funding on use of potentially over-inclusive pornography filters for children who used computers in libraries, however.) When there is a substantial risk of under-filtering due to legislative demands, the vendors are incentivized to err on the side of being excessively restrictive.

Lastly, the narrow definitions of regulated services pose problems for justifying the laws as fulfilling policy goals. Legislation has been passed (and since blocked by a Federal judge after a challenge by NetChoice) in Arkansas that age-gates social media, but excludes from its mandates any service whose “predominant” function is “[d]irect messaging consisting of messages, photos, or videos,” perhaps excluding Snapchat, one of the most widely-used—and criticized—social media platforms amongst children and adolescents. The text of several recent laws in US States define covered sites as those where more than one-third of their content is pornographic, thus excluding a site like Reddit, which also contains vast amounts of easily-accessible pornography.

Examples

- Iowa HF 526 (2023): This Bill would have prohibited social media platforms opening accounts for those under 18.
1. Minors and Parental Control

Examples

- **Louisiana Act No. 440** (2022): Requires websites with more than one-third pornography to use age verification, either using State digital IDs, or third-party verification. Prohibits retention of identity information. ([Legal challenge by Free Speech Coalition](https://example.com))
- **Mississippi HB 1315** (2023): Requires vendors to K-12 schools to block content that is pornographic or relates to children and sexual activity in a variety of ways (including otherwise undefined “inappropriate materials”).
- **Oklahoma HB 1050** (2023): This bill would require device retailers and manufacturers to have filters blocking pornographic sites (and some others) that are enabled at activation and can only be disabled by adults paying a fee.
- **UK Online Safety Act, Part 5** (2023): Requires the use of “highly effective” age assurance to prevent children accessing pornographic content, with the acceptable age assurance technologies to be further defined by regulator guidance, based on principles of ease-of-use, equitable effectiveness and interoperability.

2. Parental involvement

Parental permission is a variation on age-gating, where a minor’s access to certain services is only permitted with the consent of their parent or guardian, most often during creation of a social media account. Other legislation in this category gives parents access to aspects of their children’s online experience, for example notifying them when they download a new app, or allowing them to see the minor’s viewing, posting or messaging histories. Parents may also be given the authority to control safety, privacy and payment settings on their children’s accounts, including setting curfews, and to submit abuse reports on behalf of their child.

A practical obstacle to these approaches is in the verification of the relationship between a child user and their parent or guardian—will birth certificates or similar documentation be needed, and how will they be reviewed for authenticity when these differ widely between regions? What happens when parents (with or without custody of the child) disagree? Regardless of the method adopted, this will surely extend the privacy concerns of age assurance as both parent and child will need to submit details.

A further concern regards allowing older minors the freedom to seek information and connection with a degree of privacy from their parents. This is especially problematic when parents seek to exert coercive control over their children’s lives, or when they would disapprove of their developing interests or identity. Minors who are unable to obtain parental consent for logistical reasons, such as those with parents frequently away on military service, could also be negatively impacted.

**Examples**

- **Utah SB 152** (2023): Requires parental consent for minors to open social media accounts; parents to be granted access to posts by their children and given ability to set time restrictions for their
accounts. (Following challenges by NetChoice and FIRE, the legislature is considering a new bill, SB 194 (2024), amending the Act.)

- **Protecting Kids on Social Media Act**: Includes a provision requiring parental consent for under-18s to use social media.
- **Kansas SB 222** (2023): This bill included a requirement for mobile devices to have a default setting notifying parents of app downloads.
- **Kids Online Safety Act** (2023): Section 4(b) requires that parents are given the ability to view and change their children’s accounts; to see metrics of how much time they spent on the platform; and to submit reports regarding content that is harmful to minors.

3. **Age-appropriate design codes**

Laws and regulations that can be described as age-appropriate design codes tend to contain a range of related but distinct sets of requirements. Some of these measures are also pursued separately in smaller pieces of legislation or integrated into bills with other types of laws:

- **Privacy**: AADC laws are often framed as data protection bills, and include a range of protections for children’s data. These will be briefly discussed in the next section.
- **Generalized responsibility**: A duty to to act in, or to consider, the “best interests” of users who are children, or a “duty of care” to protect children from coming to harm—either narrowly defined or left vague—as a result of their use of the platform, is a typical feature of AADCs and associated legislation. Some bills define specific carve-outs from intermediary immunity for user-generated content platforms, effectively creating similar duties. While the legal ramifications of these formulations may vary quite significantly, they all impose some degree of responsibility on service operators to design and/or operate their platforms so as to shield young users from harm.
- **Risk assessments**: Platforms in scope are required to publish risk assessments with regard to harms to minors, including plans for mitigating the risks that are identified. Specific age ranges for the assessments are sometimes specified. These may be public or submitted to a government agency. Some legislation spells out the specific harms that must be included in the assessments. Suitably conducting these assessments can be the “price” of liability immunity. Regulators in Australia and the UK government have promoted a Safety by Design approach, which is encouraged through risk assessment frameworks and generalized responsibilities.
- **Design features**: A range of design features and settings that are thought to impact harm to minors may be required or prohibited, for example: addictive features, account discoverability by unknown users, opt-outs for algorithmic recommendations, and default safety and privacy settings. (The design code recently published by the USC Marshall’s Neely Center for Ethical Leadership and Decision Making, while not exclusively focused on children, is instructive for the kind of features that may be pertinent for harm reduction on social media platforms.) “Dark patterns” which nudge young users into sharing more data may be prohibited, and information about privacy and safety may have to be presented to minors in age-appropriate language.
One notable common requirement in AADCs is that, when parents have access to their children’s activity on a platform, the child user must be clearly alerted to this. While the required parental settings may otherwise be similar to some parental control laws, this reveals the distinct ideological genealogy as described in Part II. The age assurance standard for these laws tends to be age estimation.

Despite the focus on design rather than content, where the imposed design features impact what content is or is not presented to users, or where the generalized responsibility includes the consequences of exposure to content, the law may be vulnerable to First Amendment challenges in the US. Design features that do not impact content are less likely to raise First Amendment concerns. (These issues have been playing out in recent years, outside the First Amendment context, in two civil suits regarding Snapchat’s Speed Filter, *Lemmon v. Snap* in federal court and *Maynard et al. v Snap* in Georgia court.) Some US legislation attempts to make explicit that the generalized responsibility pertains only to design, and not to content, though this distinction can be unclear.

The design-focused generalized responsibility at the heart of AADCs can also be critiqued for relying on technological fixes for broad societal problems. Just as the tech industry is sometimes accused of making overly grandiose claims for the power of technology to help society, proponents of this legislation may affirm, explicitly or implicitly, that it is within the power of social media platforms to seriously ameliorate complex issues like adolescent mental illness.

This provision is also subject to political misuse, as mentioned in section II.4. Even outside the US, where potentially harmful speech has less stringent protections, a generalized responsibility could be understood as a controversial mandate for platforms to scan all content (a “general monitoring obligation”) to determine whether or not it could be harmful to children. Such obligations are limited, in part as a matter of free expression protection, under EU law. If applied to private messaging, this monitoring could also fall foul of communications privacy laws in various jurisdictions or it may effectively outlaw end-to-end encryption.

**Examples**

- [UK Age appropriate design: a code of practice for online services](“The Children's Code”, 2020): The first AADC, a statutory code of practice prepared under the Data Protection Act, consists of “flexible standards” to guide compliance.
- The [California Age-Appropriate Design Code Act](AB-2273, 2022): The first big test of AADC methods in the US, currently blocked under legal challenge by NetChoice, and pending appeal.
- [Kids Online Safety Act](2024): Perhaps the federal bill in this area with most enthusiasm behind it, it attempts to respond to the concerns of critics by specifying what aspects of a site are subject to the duty of care, while still allowing some enforcement via State AGs as well as the Federal Trade Commission.
• **Singapore Code of Practice for Online Safety**, part (Aii) (2023): Following the 2022 passage of the Online Safety (Miscellaneous Amendments) Bill, the regulator released this code with the obligations of designated social media services, which include protecting children from harmful content in specific categories and making a range of safety features available for children and their parents.

• **Utah HB 311 / A 5069** (2023): Prohibits social media platforms from using design that is addictive to minors, while exempting from liability those that conduct quarterly assessments and mitigations; creates liability for varied harms to minors. (Following challenges by NetChoice and FIRE, the legislature is considering a new bill, **SB 194** (2024), amending the Act.)

• **Draft EU Regulation to Prevent and Combat Child Sexual Abuse** (2022): Would require platforms to assess and mitigate risks of child sexual abuse, including grooming, and establish the possibility for detection orders requiring the use of automated technology to scan for this activity. (The most extreme powers of detection orders are opposed by some member states as well as by the European Parliament’s Civil Liberties Committee.)

4. Privacy

Privacy is a complex topic and a full treatment of the legislation about it is not feasible in this paper. In brief, there have been attempts to control the collection, retention, use, storage and sharing of minors’ data. One of the more tangible provisions under this rubric gives minors—and their parents—better ability to delete public online content that the minor has shared (see Massachusetts **H.80**, as well as **COPPA 2.0** with respect to personal information). Several broad data protection bills in US states categorize all personal data about minors as “sensitive” giving it a higher level of protection that pertains only to some data relating to adults. Some laws focus on schools and school vendors, restricting student data practices and requiring data security measures. The existing federal privacy law, COPPA, gives an age of consent of 13, and some new privacy laws extend extra protections to those under 16 or 18.

While popular with the public and civil society, some of these laws have a direct negative impact on the business models for certain online services.

**Examples**

- **COPPA 2.0** (2024): This would strengthen protections relating to the online collection, use, and disclosure of personal information of minors, including raising the age of consent to 17.

- **India Digital Personal Data Protection Act, 2023**: Includes a section requiring parental consent to process children’s data, prohibiting some forms of tracking, and requiring government approval of data security measures in order to process children’s personal data.

- **Florida SB 662** (2023): Significantly restricts school tech services’ collection, use, and retention of student data, including prohibiting targeted advertising; requires their adoption of acceptable data security practices.
5. Advertising

Selling products like offensive weapons, tobacco, lottery tickets, pornography and fireworks to minors is already illegal in many jurisdictions; advertising may be as well. A few recent bills would have added new restrictions on advertising these products to minors. These are unlikely to encounter much public opposition.

A more common feature in recent legislation is outlawing the use of personal information to target minors with advertising, though in some cases this appears to be more motivated by privacy concerns rather than concerns about the advertising problematically influencing the child’s purchasing activity, as in COPPA 2.0.

Examples

- **Massachusetts H 80** (2023): Prohibits internet services from advertising or marketing harmful products to minors.
- **EU Digital Services Act (Article 28.2)**: Outlaws using minors’ data to target them with advertising.
- **UK Age-Appropriate Design Code, Section 5**: Recommends caution with respect to advertising to children, referencing the use of behavioral advertising, exploitation of young people’s vulnerabilities, and harmful products (including unhealthy food).

6. Cyberbullying and sexting

While potentially different topics, these may be treated together as perhaps the most salient harm of sexting is that the intimate media is shared beyond the intended recipient resulting in social ostracization. Bills and laws may target the platform, the user, or both to disincentivize and put a stop to cyberbullying. They would also enable law enforcement to require platforms to reveal what they know of the identities of anonymous perpetrators. Regarding sexting, proactive and reactive education may be implemented, potentially as part of an alternative to standard criminal justice procedures under CSAM laws, which may not be well-suited for minors sharing this intimate imagery with other minors.

One obstacle for some of these measures regards defining cyberbullying: the Australian Online Safety Act refers to the likely effect on the victim, whereas a Massachusetts bill is based on the intent of the perpetrator. Intent and likely effect are very difficult to ascertain at scale and without situational knowledge about the perpetrator, the victim, and their relationship, which can make platform obligations unclear.

Examples
**Australian Online Safety Act, Part 5** (2021; updated from the 2015 *Enhancing Online Safety For Children Act*, which was largely focussed on cyberbullying of children): Requires online services to promptly remove content cyberbullying a child after receiving a notice from the eSafety Commissioner. Notices may also be sent to the posting user, requiring them to take down the content, desist from similar actions in the future and apologize to the victim.

**Massachusetts H 1745 (2023):** This bill would create an educational diversion program for minors being charged with CSAM offenses regarding the dangers of sexting and cyberbullying. Creates a criminal offense for sharing non-consensual intimate imagery (of an adult or child) with intent to harass (etc.) the subject.


7. Perpetrators of sexual abuse and law enforcement

Whereas the legislation in the last section dealt primarily with avenues whereby minors harm their peers online, a range of measures have been proposed to impact adults who may seek to harm minors, specifically with respect to sexual abuse. Some of these are aimed at formerly convicted sexual offenders and attempt to restrict their ability to identify or groom new victims online. In 2016, The Supreme Court upheld a First Amendment challenge to a particularly broad North Carolina law of this sort, which restricted offenders’ access to an unduly wide array of sites.

Others laws create new offenses—or extend existing offline ones—to make the enticement of a child online either illegal or an enhanced offense. A related category of legislation funds law enforcement efforts to address child abuse that is facilitated online.

It is unclear if measures that require sexual offenders to register their internet identifiers will actually prevent abuse, as these identifiers are trivial to bypass, and it is not clear that there are widely-used systems that check the registries and attach restrictions or warnings to their public online identities. Rather, they may be intended to make it easier to reconvict an offender if they are even attempting to seek out a new victim online.

**Examples**

- **Nebraska LB 107** (2023): Would have created “digital grooming” offense: using online communication by adults with under 16s to entice sexual activity.
- **Massachusetts H 1582** (2023): Adds online approaches to the established offense of enticement.
- **New York S 3569** (2023): This bill includes provisions for adding internet identifiers to the sex offenders register.
- **Project Safe Childhood Act** (2023): Renews funding for and revises a nationwide DOJ initiative to make improvements to the investigation and prosecution of online child exploitation.
- **Vermont H 441** (2023): Appropriates funds for the Vermont Internet Crimes Against Children Task Force and creates staff positions.
Related

- The **REPORT Act** (2023): Late in 2023, the Senate passed this bill, which broadens the reporting requirements for online services: platforms would have to report the sexual coercion and enticement of minors to the National Center for Missing and Exploited Children, which works with law enforcement.

8. Education

A number of US State bills focus on education, typically in K-12 school, as a lever to improve children’s safety online. Some bills place internet safety in the context of digital / media literacy, digital citizenship, or sexual abuse prevention curriculums, rather than as a discrete area of study. At least one piece of proposed legislation would mandate the education of parents, rather than children, about internet safety, and another would require instructors, as well as students, to be trained on the safe use of technology. The directives vary with some establishing task forces to develop material, and some permitting, requiring, or funding the teaching of the topics, or some combination of these.

Few would complain about teaching children how to safely navigate the internet, and this approach is welcomed by industry (Google has created a [curriculum for 2nd-6th grade and interactive resources](https://www.google.com) around internet safety). Some critics worry that this diverts the responsibility for keeping children safe from well-resourced platforms who facilitate the dangerous activity to the young victims and their teachers.

**Examples**

- **Indiana SB 142** (2023): Would permit schools to teach internet safety and instruct the state education department to create curricular materials at, minimally, elementary and high school levels.
- **Georgia HB 338** (2023): Would require instructor and student education on safe use of technology.
- **Texas HB 2673** (2023): Includes provision requiring schools to educate parents about internet safety.
- **New York A 35** (2023): This bill establishes an advisory committee to make recommendations regarding teaching media literacy, including internet safety, in schools.

VII. Outlook

This section will highlight some key developments that will impact the future of legislation in this area.
1. Industry engagement

Recently, some of the largest companies impacted by the laws under discussion here (Google / YouTube, Meta, and Microsoft) have released proactive statements indicating the kind of policies that they would support. Previous efforts by the tech industry have largely focussed on critiquing proposals they dislike and challenging them in court once they are enacted. Predictably, these proposals protect some of the companies’ own interests. For example, Meta suggests putting the burden of age assurance on app stores, and Google recommends strict age verification only for “high-risk” content that the company typically steers clear of.

However this positive engagement is a sign that the industry acknowledges that regulation is inevitable and so it is seeking to ensure that it is created and implemented in ways that are easiest for their compliance. This includes: specific approaches that are amenable for their products; consistency across jurisdictions; and ensuring that requirements are applied across the industry so that their competitive advantages are not negatively impacted. Competitive factors like growth can constitute a conflict for platforms when safety features are proposed internally; this gives another reason for industry players to support certain legislative efforts: so that they can be the good actors they would perhaps like to be without sacrificing market share.

2. Outcomes from enacted legislation

The recently enacted EU Digital Services Act (DSA) and UK Online Safety Act (OSA), both broad new regimes that are not solely focused on children’s safety, will be significant tests of a range of approaches regarding regulation of internet services. Of particular interest will be:

    a. Researcher access

As discussed above, there is significant uncertainty around much of the evidence around internet harms on children, especially social media usage. There can also be a high constitutional bar for US regulation impacting speech platforms. The DSA’s Article 40 includes access to data for vetted researchers that has the potential to be of significant value in enhancing the academic and public understanding of the impact of the covered platforms. (The Platform Accountability and Transparency Act in the US shares some features.) More clarity around the harms of these platforms on minors can inform the direction of future legislation and improve its defensibility in US courts.

Details of how the access will work are still to be determined but there are at least two notable caveats. Firstly, Article 40 only applies to very large online platforms and search engines (VLOPs and VLOSEs), and so study of smaller, but still important, services are not included. Secondly, some argue that causal inference requires the ability to conduct experiments in collaboration with platforms, which is harder to require with legislation. Moreover, some suggest platforms themselves may be deterred from conducting such research precisely because of compulsory transparency for the results.
b. Risk assessments

Both the DSA and the OSA require risk assessments, a key feature of age-appropriate design codes. As these assessments are released (and approved or rejected by regulators) the public may get a sense of how successful they are at promoting the companies to make safety improvements in their products, and where they have little success, how future legislation or regulator guidance might be adjusted to increase their impact.

c. Regulator enforcement

As with risk assessments, much of the impact of these two laws will depend on how they are enforced. The OSA regulator, Ofcom, is planning to release its draft guidance on compliance with aspects of the act regarding harms to children in Spring 2024, including a full code of practice. The EU, UK, and US have different traditions of relationship between regulators and industry, and the relative merits of the EU and UK systems in this context may be interesting to observe. While these laws rely on relatively well-resourced dedicated agencies, many of the proposed State laws in the US rely on the departments of State attorneys general who will not have the same ability to fine-tune guidance and work together with regulated companies, and in some cases allowing private civil suits to be brought.

3. Lawsuits

Rather than waiting for legislation to prohibit specific practices, some in the US have decided to sue platforms based on the harms that they are purportedly causing to minors. This tactic is sometimes compared to actions that were taken against tobacco companies in the 1990s, which led to major settlements. The cases against Snap regarding the Speed Filter and purchase of fentanyl are cited earlier and in the appendix. Other key suits include:

- **A.M. v. Omegle.com**: The plaintiff, as an 11-year-old girl, was matched with a man who subsequently sexually abused her. An Oregon district court judge ruled that Section 230 protections did not apply, and Omegle shut down in November 2023, days after settling with the plaintiff.
- **Social Media Adolescent Addiction/Personal Injury Products Liability Litigation**: A consolidation of suits brought by school districts, local and state governments and individuals against Meta, Snap, ByteDance and YouTube regarding a slew of alleged harms to young people.
- **33 State AGs v. Meta**: This suit has several claims, but focuses on the addictive features of Facebook and Instagram and alleges breaches of COPPA due to inaction despite knowledge of underage users.

Some have warned that the success of suits like these will discourage platforms from conducting internal studies into possible harms from their products, or from collaborating with external researchers, as the findings, whether accurately represented or not, may open them up to legal risk. A similar concern is that risk assessments performed for DSA and OSA compliance, especially if they are performed with full rigor and honesty, may be mined for ammunition for future lawsuits.
4. Technological advances

As technology develops, so will online risks to children and the tools available to mitigate them. Proposed and enacted regulation can also spur this development, require its usage, and even directly fund research. Recent advances in artificial intelligence may lead to more accurate age assurance methods without the need for entering personal details (though this can raise new privacy concerns instead). Classifiers are also improving rapidly, which might impact the feasibility of granular content controls. Finally, a combination of psychological research and tech solutions could enable a more individualized approach to mental health and online activity that may be more effective than broad requirements for all individuals below a certain age.

VIII. Conclusion

While the extraordinarily high volume of proposals and new laws in 2023 may not be repeated in the future, there is sure to be prolonged interest in developing regulation that impacts the safety of children online. It is far from clear whether we will see further consolidation or diversification of approaches. If lessons can be learned from continuing research into harms, outcomes from the implementation of existing legislation, and legal challenges, the next wave of policymaking could have increased potential to enhance the protection of children online.

Acknowledgments

Daphne Keller, Director, Program on Platform Regulation, Stanford Cyber Policy Center Tech Policy Press, which commissioned and published my initial study of US State bills in this area.
APPENDIX

This appendix will explore in more detail the harms to children that policymakers often seek to address. Some of the harms described below are self-evident, at least on an individual basis, while others may require more evidence. However, it is almost impossible to prove that children’s experiences online, overall, are causing more harm than would have resulted if these children were not exposed to the internet, in part due to the difficulties of creating suitable control groups.

Some commentators have concluded that the public attention to harms to minors from social media and other online services is merely a moral panic, comparable to previous panics over new media forms such as popular print media, comic books, film, television, and video games. Haphazard and ineffective public policy that has sometimes been proposed to counter online harms to children, often on a flimsy evidentiary basis, can lend support to this perspective. This appendix flags some weaknesses in evidence and policy, but does not attempt to resolve the central disagreement between those convinced that the harms warrant legislative action and those who are deeply skeptical.

1. Mental health

The issue of social media’s impact on mental health is perhaps, currently, the most well-known and most studied harm to minors. The Facebook/Meta whistleblower, Frances Haugen, famously revealed that the company’s internal studies showed some evidence of body image-related harms for a subset of adolescent girls who use Instagram. More recently, the US Surgeon General released an advisory report, “Social Media and Youth Mental Health.” The introductory portion of the report contains the following summary statement:

More research is needed to fully understand the impact of social media; however, the current body of evidence indicates that while social media may have benefits for some children and adolescents, there are ample indicators that social media can also have a profound risk of harm to the mental health and well-being of children and adolescents. At this time, we do not yet have enough evidence to determine if social media is sufficiently safe for children and adolescents. We must acknowledge the growing body of research about potential harms, increase our collective understanding of the risks associated with social media use, and urgently take action to create safe and healthy digital environments that minimize harm and safeguard children’s and adolescents’ mental health and well-being during critical stages of development.

Mental health is not just one discrete harm, nor is social media one discrete product, nor are all children the same. Indications of positive mental health outcomes are particularly strong for LGBTQ+ children, for example. The report cites studies that find evidence for: low subjective well-being, depression, anxiety, poor body image, disordered eating, low self-esteem, self-harm, self-control challenges (related to addiction), sleep problems, attention problems and feelings of exclusion.
Jonathan Haidt and Zachary Rausch of NYU maintain a series of collaborative review documents that collect and categorize the many studies that pertain to social media and the mental health of minors, amongst other reviews, including the impact of video games, smartphone usage and pornography on minors. While Haidt is generally convinced by the evidence that social media has a strong negative effect on adolescent mental health, the review documents include notes suggesting weaknesses in the studies, studies showing small or non-existent effects, and one review document is devoted to alternative explanations of the seeming decline in mental health amongst young people in some countries.

Identifying clear causal links between internet usage and mental health outcomes, especially when dealing with children, is intrinsically difficult. The Surgeon General’s report, amongst others, recognizes some positive outcomes of social media for young people and identifies significant gaps in the evidence for harms. “Social Media and Adult Health,” a recent consensus report by the National Academies of Sciences, Engineering, and Medicine concluded that a review of the literature “did not support the conclusion that social media causes changes in adolescent health at the population level,” while still making recommendations for reducing harm and increasing benefit. A health advisory by the American Psychological Association came to roughly similar conclusions.

There is some consensus that it is often particular subsets of minors who are vulnerable to mental health harms from social media usage. Though not focussing on young people, Stanford’s Human Screenome Project has identified connections between individuals’ device usage patterns and their mental wellbeing and this may turn out to be a more promising approach than pinpointing effects on broad swathes of users.

2. Pornography

While the legal term “harmful to minors,” as used in reference to pornography, originates in a combination of the psychological development theory of the time and traditional morality, more recent research has identified some connection between minors’ consumption of pornography and sexual violence, particularly for males viewing violent pornography. However, as the cited report notes, “[r]esearch into the impact of pornography exposure on harmful sexual behaviour has ethical and methodological challenges,” especially when concerning children, that are an obstacle to obtaining clear data. Here again, more details of studies can be found in Haidt and Rausch’s collaborative review document on the topic.

3. Dangerous behavior

Young people have always engaged in diverse and inventive risky behaviors, often at the encouragement of their peers. Social media has been identified as a vector of such encouragement in recent years. The preamble to Montana’s SB 419 banning TikTok from the state (for residents of all ages; currently blocked by federal court on First Amendment grounds and pending appeal) lists a number of these:
“WHEREAS, TikTok fails to remove, and may even promote, dangerous content that directs minors to engage in dangerous activities, including but not limited to throwing objects at moving automobiles, taking excessive amounts of medication, lighting a mirror on fire and then attempting to extinguish it using only one's body parts, inducing unconsciousness through oxygen deprivation, cooking chicken in NyQuil, pouring hot wax on a user’s face, attempting to break an unsuspecting passerby's skull by tripping him or her into landing face first into a hard surface, placing metal objects in electrical outlets, swerving cars at high rates of speed, smearing human feces on toddlers, licking doorknobs and toilet seats to place oneself at risk of contracting coronavirus, attempting to climb stacks of milkcrates, shooting passersby with air rifles, loosening lug nuts on vehicles, and stealing utilities from public places…”

There have been multiple cases of death attributed to imitating an oxygen-deprivation “blackout challenge” on TikTok. Similar activities were of course also observed before the social media era, following trends that were not technologically mediated. It is possible, while perhaps difficult to prove, that the ability to reach outside one's social circle using online services and to “go viral” magnifies the impact of these unfortunate trends by accelerating them and increasing their reach.

Distinguishing between genuine encouragement to perform an activity and some shade of making a mostly-harmless joke about it can also be very difficult, as James Grimmelmann lays out in his analysis of the so-called “Tide Pod Challenge,” and some fears over online content of this nature leading to harmful actions have turned out to be unwarranted on several occasions.

4. Sexual abuse

Minors with social media accounts have been sought out by predators for the purpose of online or in-person sexual abuse at all levels of severity. One way this can play out is similar to in-person grooming, where the minor is befriended and the abuser slowly gains control over them, and eventually is able to coerce them into submitting to abuse. This can be exacerbated by the abuser being able to hide their identity, including their age and gender, online.

Anonymity is also key to sextortion. Sextortion cases typically start with an abuser procuring intimate media from a minor while pretending to be of a similar age. Once at least one image or video has been obtained, the abuser then threatens to make it public unless the victim does as they are instructed. The demands can be to provide further intimate imagery, to perform sexual acts, or, increasingly often, to pay the culprit (financial sextortion). Young people, especially girls, are also known to be victims of cyberflashing, the sending of unwanted intimate imagery via messaging apps or Apple Airdrop, either from people they know or from strangers.

The distribution of child sexual abuse material (CSAM) is a significant topic for platform policy, law enforcement, and legislation, and it is often conceived of as perpetrating a new harm to the victim (and as
a new crime) every time it is spread. Given the highly specialized legal, policy, and moral issues relating to CSAM, including self-generated and deepfake CSAM, it is largely out of scope for this paper.

5. Sexting and non-consensual intimate imagery
The sharing of intimate imagery with a partner, even when entirely voluntary, poses serious risks (sexting can also be used to refer to intimate text conversation, which is much lower-risk). If the files are shared or stolen—at any later point in the subject’s life—they can lead to embarrassment, bullying, sanction from peers, workplaces, school or family, and even sexploitation. Content depicting minors can also legally be considered CSAM, and this can, depending on the jurisdiction, pose significant legal risks to the sender and recipient.

6. Cyberbullying
As the internet is a major locus of young people’s social life, it is no surprise that bullying takes place online as well as offline. Bullies can take advantage of any affordances that online platforms provide, such as tagging, blocking, anonymity, commenting, and disappearing messages to optimize the tormenting of their victims. Cyberbullying is also distinct from more traditional bullying in that it can reach into times and spaces that should be private and safe, occurring whenever and wherever the victim has internet access.

7. Data privacy violations
The contemporary internet, especially advertising-dependent social networks, collects a lot of data about users. As noted earlier, children are understood to be below the appropriate age to give consent for their data to be collected, used or shared. The specific details of harms related to a lack of data privacy constitute a whole field of study and policy and cannot be fully discussed here. One particular concern is geolocation data, which could provide critical information about a child’s whereabouts to someone who seeks to physically harm that child.

8. Drugs and dangerous items
Many products that are illegal or strictly controlled, or are prohibited for sale to minors, are available online. This includes recreational drugs and their paraphernalia, alcohol and tobacco, as well as offensive weapons, fireworks, and graffiti supplies. Those products that are not outright illegal for most adults may sometimes be commercially advertised to minors on mainstream platforms, and all of these could be bought by minors online. The purchases can be through major retailers, lesser-known ecommerce sites, or informally over chat platforms, as in the well-publicized cases of drugs containing fentanyl being obtained via Snapchat.
9. Advertising and financial harms

As discussed above, there are products that ought not be advertised to children as the products they advertise are harmful. However, some express concern that children may be particularly susceptible to being persuaded by advertising, even for intrinsically unproblematic products (this has been studied with regards to television advertising). A related issue can be expressed as a desire to protect children from the “profit maximization” methods of social media platforms, which may in turn incentivise potentially harmful techniques.

Young people with access to payment methods, whether through credit/debit cards, payment apps or via App Stores and in-app purchases, have been known to spend money on frivolous purchases (such as in-game items, including controversial elements like loot boxes that are akin to gambling) and have been victimized by online scams like impersonation scams.