

TO: Department for Innovation, Science and Technology

FROM: Common Sense Media

DATE: 26 May 2026

RE: Consultation: **Growing up in the online world: a national consultation.**

CONFIDENTIALITY:

The whole response below is not confidential. Responses may be attributed to Common Sense Media.

Point of contact: Michaela Carmichael, Senior Advisor, Common Sense Media UK

Common Sense Media is dedicated to improving the lives of kids and families by providing the trustworthy information, education, and independent voice they need to thrive. Our ratings, research, and resources have reached more than 150 million users worldwide, over 1.5 million educators, and more than 100,000 schools worldwide. Backed by the research that's the foundation of everything we do, we champion quality media that entertains and inspires, and support legislation that aims to close the digital divide and hold social media and AI platforms accountable for their effects on minors. We encourage students and educators to think critically about technology and become responsible digital citizens. And we safeguard children's health and well-being by protecting their data privacy and reducing potential online harms.

We are pleased to submit these comments in response to the questions in the Department for Innovation, Science, and Technology's consultation – Growing up in the online world: a national consultation.

Chapter 1.

What are the harms or risks of social media use, and being online, for Children?

- *Answer:* Risks and harms include mental health impacts, compulsive use and time lost to other activities, and sleep disruption. These harms are particularly pronounced for young people given their developing brains. And they can be more pronounced for otherwise vulnerable populations of youth.

Frequent social media use is associated with higher rates of depression, aggression, low self-esteem, anxiety, suicidal ideation, and self-harm among youth. A longitudinal study of adolescents aged 12–15 found that those spending more than three hours per day on social media face roughly double the risk of depression and anxiety, and other research of teenagers has found that suicide-related outcomes were elevated after just two hours of daily device use. More than a quarter of teens report believing others' lives are better than their own, contributing to diminished self-worth, loneliness, and social exclusion.

(E.g., Jason M. Nagata et al., Social Media Use and Internalizing Symptoms Among U.S. Adolescents, *JAMA Psychiatry* (2022); Jean M. Twenge et al., Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time, *6 Clinical Psychol. Sci.* 3, 9 (2018))

Social media causes compulsive use and time lost to other meaningful activities. Common Sense Research shows that U.S. teens check their phones an average of over 100 times per day (some hundreds of times) and receive around 237 daily notifications, the majority from social media. They average roughly 4.5 hours of daily phone use, with some reaching five to eight hours. In a 2024 Common Sense Media report, 43% of teens aged 14–17 said social media takes time away from meaningful activities. (Jenny S. Radesky et al., *Constant Companion: A Week in the Life of a Young Person's Smartphone Use* (Common Sense Media 2023); Mary Madden et al. *A Double-Edged Sword: How Diverse Communities of Young People Think About the Multifaceted Relationship Between Social Media and Mental Health*. (Common Sense and Hopelab 2024)).

Nighttime use interferes with sleep; one-third of teens report waking at night to check their phones, and some set alarms to check notifications or complete time-limited game tasks. 34% of teens report negative sleep effects from social media. (Common Sense, *Screens and Sleep: The New Normal: Parents, Teens, Screens, and Sleep in the United States* (2019); Emily Weinstein & Carrie James, *Behind Their Screens: What Teens Are Facing (And Adults Are Missing)*, MIT Press (2022); Mary Madden et al. *A Double-Edged Sword*.)

Children and teens are uniquely susceptible because the prefrontal cortex—responsible for impulse control, planning, and resisting reward-seeking—does not mature until adulthood, while the limbic and dopamine-driven reward systems are highly active. This makes minors less able to resist persuasive design, more sensitive to social feedback (likes, follower counts, peer reactions), and more reactive to dopamine-driven reward loops.

Indeed, the U.S. Surgeon General has noted that frequent social media use may be associated with changes in the developing amygdala and prefrontal cortex, including increasing sensitivity to social rewards and punishments. (U.S. Surgeon Gen., *Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory* (2023)).

While young people in general are more susceptible to social media's risks and harms than adults, vulnerable populations among teens are even more susceptible. Research shows that teens already experiencing depressive symptoms have tried to see less of

what they like and curate their feeds more than those without symptoms—90% of those with moderate-to-severe symptoms try to see less of what they dislike, versus 67% without symptoms. (Mary Madden et al. A Double-Edged Sword.)

Chapter 2:

Would you support a legal requirement for social media services to have a minimum age of access?

- **Answer: a. Yes**

Would you support a legal requirement for social media services to have a minimum age of access lower than 16? If so, at what age would you set it?

- **Answer: e. Other (please specify)**
- **Please explain the reasoning behind your answers about minimum age requirements.**
We think each country should evaluate this based on its own circumstances. In the U.S., Common Sense Media is supporting legislation in the State of California that would delay teens' ability to create social media accounts until they turn 16. The measure does not prevent young people from exploring the internet, but it does protect them from many of the harms associated with having accounts on social media platforms at a younger age. And, while we support a minimum age for establishing accounts, as explained in other sections of this response, we also support laws and regulations that require social media platforms to make their products and features safer and healthier for all minors.

Some online services allow their users to engage with the following functionalities. Do you think these functionalities should be age restricted so that children below a certain age cannot engage with them? (Please select all that apply)

- **Answer: a. Live streaming**
b. Ability to send nude images or videos
c. Disappearing content
d. Location sharing
e. Connecting or talking to strangers

What do you think the impacts would be if some online services were required to introduce age restrictions on specific features and functionalities?

- **Answer: Common Sense Media believes that all of these functionalities should be age-restricted so that children below a certain age cannot engage with them. Kids and teens are increasingly spending time on social media and gaming platforms that include private messaging, livestreaming, disappearing content, image sharing, and video communication features. These functionalities can expose minors to serious risks,**

including grooming, harassment, sexual exploitation, financial scams, and unwanted contact from strangers.

Online gaming and social media platforms are increasingly being used by predators to target children. For example, Roblox — one of the world’s largest gaming platforms — averaged more than 80 million daily users in 2024, with over 40% of users self-reporting that they are under the age of 13. (Zoe Kleinman & Georgina Hayes, Keep Kids Off Roblox if You’re Worried, Its CEO Tells Parents, BBC News (Mar. 13, 2025)). In 2023 alone, there were approximately 13,000 reports of child exploitation on Roblox, amounting to more than 35 reports every day. (Olivia Carville & Cecilia D’Anastasio, Roblox’s Pedophile Problem, Bloomberg Businessweek (July 22, 2024)). These risks are amplified by platform features that allow strangers to communicate directly with minors, send disappearing messages, livestream content, or exchange images and videos privately.

Certain functionalities are particularly associated with online child exploitation and should therefore be restricted for younger users. Livestreaming, disappearing content, and the ability to send or receive nude images or videos are frequently exploited by predators to manipulate or coerce minors. The National Center for Missing & Exploited Children (NCMEC) has reported that, in many incidents of child sexual exploitation involving livestreams reported to its CyberTipline, “the child engages in sexual or self-harming behaviors in response to comments from viewers coaxing them for the content.” (Nat’l Ctr. for Missing & Exploited Child., Livestreaming, NetSmartz, <https://www.missingkids.org/netsmartz/topics/livestreaming>) NCMEC has also received reports of children sharing sensitive personal information during livestreams, including their phone number, school, and home address.

For these reasons, Common Sense Media agrees with NCMEC that livestreaming and video conferencing functionalities should generally be reserved for educational settings or situations where children are closely monitored by a trusted adult, and otherwise limited to older teens who clearly understand how to protect their privacy and physical and digital safety.

Common Sense Media has also supported legislative efforts to address these risks. In 2025, Common Sense Media supported the Stop Online Predators Act (SOPA) in New York (S4609/A6549), introduced by Senator Andrew Gounardes and Assemblywoman Nily Rozic). Many provisions of SOPA were later incorporated into New York Governor Kathy Hochul’s 2026 Proposed State Budget through the Safe by Design Act. Among other protections, the legislation requires platforms with chat functionality to apply the strongest safety and privacy settings to minors’ accounts by default, including preventing strangers from directly messaging minors, restricting access to minors’ profiles and

geographic location, and limiting unapproved users from tagging or interacting with minors' content.

In light of these well-documented risks, age-based restrictions and strong default safety protections are necessary to ensure that children are not exposed to online functionalities that can facilitate exploitation, coercion, and other serious harms.

*The following design features are sometimes known as 'persuasive', meaning they may encourage children to stay online for longer. From the following list, *Disrupting Social Media Habits: a Field Experiment with Young Danish Consumers (2025)* please select the ones you think are particularly 'persuasive' to children: (Please select all that apply).*

- Answer: a. Infinite scrolling
b. Autoplay
c. Affirmation features (e.g. 'likes', comments)
d. Alerts and push notifications
e. Content recommendation algorithms (these are algorithms which provide personalised recommendations on a user's feed)
- Answer: All apply.

There is a substantial and growing body of evidence demonstrating that social media platforms are deliberately designed with a range of "persuasive" features that encourage children and teens to stay online for longer periods of time and engage more frequently. Research drawn from youth self-reports, expert testimony, academic studies, and internal industry documents confirms that the risks associated with these platforms are not accidental, but are closely tied to intentional product design choices.

According to the U.S. Surgeon General's Advisory on *Social Media and Youth Mental Health*, up to 95% of youth ages 13–17 report using social media, with more than one-third saying they use it "almost constantly." (Office of the U.S. Surgeon General. (2024). *Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory*.) Social media companies increasingly rely on "engagement-ranked" or "addictive" feeds that use algorithms to predict and maximize user engagement in order to keep children and teens viewing and interacting with content for longer periods of time.

Research further shows that children and teens struggle to set healthy boundaries around technologies that are intentionally designed to capture and sustain attention. Common Sense Media's own research found that teens check their phones an average of more than 100 times per day and receive approximately 237 notifications daily, the majority from social media applications. (Jenny S. Radesky et al., *Constant Companion: A Week in the Life of*

a Young Person's Smartphone Use (Common Sense Media 2023)) These notifications disrupt sleep, school, work, and face-to-face relationships, and are associated with harms including anxiety, depression, compulsive engagement, sleep disruption, social comparison, bullying, and exposure to harmful content.

Importantly, young users themselves consistently identify specific platform features as making social media difficult to stop using. In studies of adolescent girls, participants pointed to algorithmic recommendations, infinite scrolling, autoplay, and constant notifications as particularly difficult to disengage from. (Id.) Social media platforms reinforce these behaviors through the continuous collection and analysis of user data and by deploying design features intended to reduce friction and prolong use. (American College of Pediatricians. (2020). *Media Use and Screen Time – Its Impact on Children, Adolescents, and Families*. (2023). Social media brings benefits and risks to teens. Here's how psychology can help identify a path forward.)

As the Surgeon General's Advisory explains, "Push notifications, autoplay, infinite scroll, quantifying and displaying popularity (i.e., 'likes'), and algorithms that leverage user data to serve content recommendations are some examples of these features that maximize engagement." These features work together to sustain attention, encourage repeated checking, and reinforce habitual engagement among children and teens.

Accordingly, all of the listed design features should be considered particularly "persuasive" to children.

What factors are important when determining which apps, sites or services to apply minimum age of access restrictions to? For example, user-to-user interaction, the ability to post material, persuasive design features, risky functionalities, ability to generate non-text mediums such as video or images, the target age group, the size of the service.

- **Answer:** Even in the U.S., there is support for age assurance. Common Sense Media survey evidence shows that 95% of adults believe children need to be protected from certain online material and features, with pornography, gambling, and online purchases emerging as top concerns. (Age Assurance Attitudes Among Adults (2026). In addition, more than six in 10 adults support age verification for social media and online games, while more than half support it for AI, including AI companions. It is important to consider whether apps, sites, and services have features that exploit youth vulnerabilities and expose them to harm. These features are frequently found in social media sites, AI chatbots, and gaming or gambling services.

Social media sites typically contain features that reduce friction, prolong sessions, and exploit social and reward sensitivities. These risky features include targeted algorithmic feeds, infinite scroll, autoplay, push notifications, and publicly displayed likes and follower counts. They also include displaying users' locations and proposing connections between users (such as connecting adult strangers with children).

Chatbots that act as companions or have features that exploit child users' trust by creating a false sense of attachment, such as being sycophantic or presenting as alive, are risky and should be age-restricted. It is also risky for chatbots to present as licensed medical professionals. Chatbot features that are designed to promote suicide, self harm, disordered eating, harm to others, unlawful drug and alcohol use, harm to others, along with outputs that encourage secrecy, self-isolation, or avoiding trusted adults and professionals, or that involve sexually explicit material are also risky and should be age-restricted.

Gambling services and game-based features that mimic gambling should also be age-restricted. In particular, design features like loot boxes, skin cases, and randomized reward systems are risky to young users and prey on their developmental vulnerabilities of risk-taking and reward-seeking.

Are there any types of apps, sites or services that you would want to be captured by minimum age of access restrictions?

- **Answer:** Apps, sites, and services that have the features described in response to the question above, or those that possess the same functionality within apps and services.

What factors are important when determining which apps, sites or services to apply age-restrictions on specific features and functionalities? For example, user-to-user interaction, the ability to post material, persuasive design features, risky functionalities, ability to generate non-text mediums such as video or images, the target age group, the size of the service.

- **Answer:** It is the features and functionality within the apps, sites, or services that are concerning. Ideally, it is best to just age-restrict that feature, including if sites/services age-assess at the app/service level, but only apply restrictions to specific addictive or dangerous features or functionalities. If the risky features are so intertwined with the app/service that they are essentially the entire thing, or it is difficult to untangle the risky feature from the app/service, then it could make sense to age-restrict the entire app/service. In addition, if industry resists putting in place guardrails that protect youth from risky features and functionality, then it may be necessary to delay youth's access to platforms entirely.

Are there any types of apps, sites or services that you want to be captured by age-restrictions to features/ functionalities?

- *Answer:* Apps, sites, and services that have the features described in response to questions above, or that functionality within apps and services.

What are the benefits to children of using AI chatbots? For example, this might include as a search function, for educational purposes, for creativity.

- *Answer:* Tech companies and the private equity and venture capital firms funding their expansion are driving the proliferation of AI into every aspect of life, from education to work to leisure. While they largely work to convince the public that its benefits and efficiencies are a foregone conclusion, there are important findings that refute such a blanket conclusion.

The clearest illustration comes from the workplace. A 2026 study by the National Bureau of Economic Research surveying roughly 6,000 CEOs, CFOs, and other executives across four countries found that nearly 90% of firms reported AI had no impact on employment or productivity over the prior three years — even as executives continued to predict significant gains just around the corner. Meanwhile, several CEOs of AI companies recently stated publicly that they believed AI would lead to significant job losses and called for government intervention to help affected workers. If the productivity case remains this contested at the top of the corporate world, there is little reason to assume the benefits to children are self-evident.

In education, the downsides deserve as much scrutiny as the promised gains. Frictionless access to information is not cost-free: the friction of searching, evaluating, and synthesising is itself part of how children learn to think, and removing it risks a kind of cognitive decay. A 2025 MIT Media Lab study found that participants who relied on ChatGPT to write essays showed the lowest brain engagement of the groups studied and struggled to recall their own work, an effect the researchers called "cognitive debt". There is also the risk of overreliance on output that can contain inaccuracies, which children are least equipped to detect, and a harder-to-quantify social loss: a chatbot cannot replace the relationship and responsiveness a child gains from a teacher.

Moreover, the inclusion of AI products to support administrative educational uses requires educators to closely scrutinize the student or teacher data necessary for efficacy. Where student and teacher data (personal data, personally identifiable information, student or teacher inputs) is used, there must be robust protections for its limited use, robust data security measures, and significant penalties for breaches or noncompliance. For example, there are products under development to support the creation of individualized education plans (IEPs) for students who need additional

accommodations or support. These plans include highly sensitive health diagnosis data and must meet certain data privacy requirements set by the government.

General-purpose chatbots, while on their face they could appear helpful for drafting the IEP, may not be built to meet the data privacy and security requirements. AI in educational settings needs to be purpose-built, with proven efficacy, independently safety-tested, meeting an articulated need, and with cybersecurity and strict data/input privacy to ensure that data from the educational setting is not used for training purposes, sold, or shared.

As the market is quickly evolving, many AI products purport to offer support through summative features/functions (automated essay grading, standardized test scoring systems, final exam proctoring software, graduation requirement tracking) and formative features/functions (adaptive learning platforms, real-time feedback tools, AI tutors, learning analytics dashboards, interim assessment systems). In education, it's important to note that the distinction between summative and formative assessment (summative evaluation determines final judgments (grades, advancement, credentials), while formative evaluation guides ongoing learning) is often blurred as continuous assessment models use cumulative formative data for summative decisions. While both features carry risks, the risk profiles differ. Summative risk includes higher stakes, fewer opportunities for correction, greater impact on life opportunities, and the potential for systematic bias in high-stakes decisions. Formative risks include data privacy concerns from continuous monitoring, ineffective or incorrect feedback, feedback loops that reinforce biases, and the potential to narrow instruction to what AI can measure. Both carry significant risks. Formative systems may appear “lower risk,” but their cumulative data feeds summative decisions and shapes long-term learning trajectories.

As AI capabilities become increasingly integrated into schools' existing technology stacks, the lines of responsibility and control are becoming blurred in ways that may create unintended regulatory gaps. There is a confusing grey area emerging regarding “white-labeling” or contractual arrangements where educational or Vocational Education and Training (VET) institutions purchase AI systems and deploy them under their own name or trademark and not under the trademark of the provider who provided the system. This could imply that the institution, as the deployer, assumes all obligations of a provider. While these “white-label” arrangements are not yet widespread, there could be an area that creates confusion and requires clarity.

Examples of this emerging gray area include:

- Learning Management Systems (LMS), which may be white-labeled, that have added AI features (auto-grading, content recommendations, plagiarism detection)
- Student information systems that incorporate AI-powered analytics as part of regular updates

- EdTech platforms adding AI tutoring or assessment features that schools didn't explicitly procure as separate AI systems
- Third-party integrations and plugins that add AI functionality to existing school platforms

In these scenarios, it's unclear whether schools understand that AI systems are being deployed within their technology infrastructure. Whether schools have sufficient technical access and documentation to fulfill provider obligations. If the original software provider or the AI component developer bears responsibility. If schools have the technical capacity to address safety concerns, or to conduct conformity assessments or risk management for AI features embedded in broader platforms.

This gray area risks creating a compliance gap where neither schools nor vendors clearly understand who holds provider obligations under future and existing laws. Educational institutions need clear guidance on when integration of AI into existing systems triggers provider obligations, what level of control/customization shifts responsibility to the deploying institution, and whether vendors are fulfilling purported promises on safety, security, privacy, and efficacy.

Without clarity, schools may unknowingly assume provider-level obligations they cannot fulfill, or alternatively, AI systems may be deployed in education without adequate oversight because all parties assume someone else is responsible.

On creativity, the benefit claimed depends entirely on the point of creation. Where the goal is ease of output, AI can lower the barrier to entry for those not well-versed in photo editing, graphic design, or animation. But where the point is the conveyance of a human experience, something is lost. A child who offloads the act of making also offloads the struggle and self-discovery that the act was meant to develop in them.

Which AI chatbot features are most risky for children? (Please select all that apply)

- **Answer:**
 - b. The personalisation of interactions
 - c. How they mimic relationships (friendship)
 - d. How they mimic relationships (romantic)
 - e. How they mimic empathy
 - f. Flattering language
 - g. Features to encourage more questions/ requests (e.g. asking questions back)
 - h. The ability to recall interactions across sessions

k. Hallucination or false, misleading responses

l. Ability to engage in and generate mature content (e.g. sexual/romantic roleplay)

Answer, cont: While most of these features pose unique risks, the following categories pose heightened risk for children specifically:

Engagement optimization (g) presents the highest risk because features designed to extend conversations artificially prolong interactions past the point where safety guardrails remain consistently and reliably effective. OpenAI has acknowledged this tension in the context of documented harms to minors.

Features that support deception and manipulation (e, f, k) are dangerous in combination. A chatbot that mimics empathy, is sycophantic, and generates false or misleading information creates conditions where children are most likely to trust and least likely to question outputs. In particular, because generative AI chatbots are frequently used for information retrieval and can produce impressive outputs on an increasing number of topics, vulnerable users are more likely to view outputs based in opinion as authoritative regardless of accuracy.

Relationship simulation and persistent memory (b, c, d, h) enable the formation of attachments that are genuinely harmful to development. When a chatbot recalls prior sessions and maintains a consistent persona, children can form dependencies on an entity engineered to be frictionless and unconditionally available — a distorted relational reference point during a critical period of social and emotional development. Persistent memory also raises serious data concerns, as emotionally intimate disclosures by minors may be retained and used for advertising or shared with data brokers.

Which functionalities of AI chatbots should minimum age restrictions apply to?

Answer: At a minimum, it should be unlawful for developers or deployers of AI chatbots available to minors to design AI chatbots in any way, through design choices, model training and prioritization, that promote or facilitate suicide, self-harm, harm to others, disordered eating, or unlawful drug and alcohol use, along with outputs that encourage secrecy, self-isolation, or avoiding trusted adults and professionals, or that involve sexually explicit material. While these outputs should be restricted outright, instances of harm have been enabled, in part, by engagement-based optimization features that sustain or increase user interaction in an over-extended context window, which leads to safety failures, where the model is unable to reliably implement safeguards. Therefore, restricting by age, the ability to prioritize user engagement over the efficacy of implemented guardrails decreases the likelihood that these harmful features impact minors.

Beyond harmful outputs, restrictions should reach features that exploit a minor's trust or attachment, such as persistent recall or use of personal information gathered in earlier sessions, to create the illusion of a personal, stable, reliable relationship. Young users are particularly susceptible to deception and manipulation due to the state of their psychological development. Presenting a chatbot as alive or real, purporting to have a personal relationship with the user, being sycophantic, or claiming to have a professional license should be restricted for minors to avoid encouraging unhealthy forms of attachment, real-life user decisions or behaviors influenced by perceived authority, or, in more extreme cases, AI psychosis.

Finally, a minor's conversations and shared personal information should not be used to train or fine-tune AI models, and any data collected to verify age or for crisis intervention safeguards should be used solely for that purpose. Outside the harms of certain chatbot design choices, it is critical for these products not to serve as another vector for surveillance capitalism. Particularly in response to the recent lawsuit claiming OpenAI incorporated cookies from Google and Meta to support advertising based on user inputs, policymakers should equally prioritize user data privacy alongside design-based interventions to ensure young users are protected on and off AI chatbot platforms.

Unsafe AI companion features, currently marketed to kids as mental health support, appear in both dedicated apps and general-purpose chatbots, and should also be age-restricted for minors. The risk is heightened when a minor turns to these platforms for emotional support alongside tasks like tutoring or information retrieval, leading them to form emotional attachments and perceive the personalized chatbot responses as authoritative, clinical guidance. These systems are effective at inducing dependency because they use human-like features, constant validation, and follow-up questions to extend conversations and encourage secrecy.

As mentioned above, current safety guardrails degrade in extended conversations. While chatbots may perform well in a single, clear exchange, in multi-turn conversations that mirror real-world usage, safety fails. However, it is precisely this extended usage pattern that the products prioritize. By designing features to sustain engagement using memory and personalization, AI companies create a false sense of an ongoing relationship, which actively discourages seeking human support. This is in a dangerous conflict with what is needed for a teen in crisis. If a teen were to discuss a mental health concern, the goal must be a rapid connection to a licensed mental health professional—not isolation and extended AI engagement.

Should AI chatbots have minimum age restrictions? Please explain the reasoning behind your answer

- **Answer:** c. Yes – both minimum age requirements and restricting access to certain features and functionalities

Different types of chatbots – based on purpose, capabilities, or training data – present different risks to different ages of users. While more should undoubtedly be done to address similar risks to users of all ages, given the non-insignificant number of users reported to show signs of mental health distress and mania, young users face particular vulnerabilities from interacting with a fluent, persuasive, always-available system while their own cognitive and emotional defenses are still developing. Younger users may be more likely to believe a chatbot is human, capable of genuine care, and offering guidance to them perceived as authoritative. Ultimately, there is a significant power asymmetry between a system optimized to be engaging and a user whose judgment is still being formulated and is susceptible to influence.

For AI chatbots designed to offer a persistent persona or act as a friend, companion, or otherwise maintain an ongoing relationship with a user, these should be fully restricted for minors. Ongoing risk assessments by Common Sense Media have found that such systems are disproportionately likely to engage in sexualized or grooming behavior, and can tend to receive less rigorous safety scrutiny than more widely used, general-purpose tools. The frictionless intimacy they cultivate also functions as a substitute for – and therefore a barrier to – real-life friendships and peer relationships. At a stage in social development when young people are learning to navigate conflict, rejection, and reciprocity, displacing those formative experiences with an endlessly accommodating AI chatbot can introduce harms that compound over time. Data privacy presents a further concern; minors are less equipped to understand or meaningfully consent to the data practices underlying these platforms. Taken together, and given the minimal instrumental benefit companion chatbots offer relative to these risks, these products should be restricted for young users.

For advanced, general-purpose generative AI chatbots, the calculus is more nuanced. These tools provide genuine value for younger users, making outright age-based exclusion a blunter instrument. Age restrictions that accompany design-based limitations enable safer use. At a minimum, features that encourage ongoing emotional dependence, such as persistent memory framed in relational terms, simulated affection, nudges toward continued engagement, and the use of user inputs for model training, should be disabled for minor accounts.

What do you think the impact would be of introducing age restrictions on AI chatbots or certain features and functions? For example, impacts on the safety and wellbeing of children, or the impact for parents and carers, as well as other users. You could also comment on the impact on all users' privacy and data or on business costs, revenue, and innovation.

- **Answer:** By limiting access to AI chatbots with the “companion” features, identified in previous sections, or mental health support, parents, educators, and children would have greater certainty about the capabilities and safety profiles of the AI products in their orbit. AI innovations could safely offer benefits to children for learning, to support creation, and information retrieval. With safer products purpose-built for children’s use, the wider and more meaningful adoption of these innovations would be supported. Also, by creating clear standards for age assurance and a mechanism for feature-based limitations, developers of children’s AI products would have greater certainty about market-entry standards. For small and mid-sized developers, guidelines ensure that their more limited resources can be focused on meeting established safety standards, even if they lack the internal capacity (e.g., AI researchers, child development experts) to develop their own. For government regulators, a feature-based approach allows flexibility, allowing the list of restricted features to evolve as the risks or benefits posed by these products change.

AI companions pose increased risks to user data privacy by encouraging children to disclose deeply personal information. Privacy law experts have long pointed out vulnerabilities in existing data protection laws, and concerns continue as AI use begins to accelerate. In a 2024 law review article, Professor Daniel Solove of George Washington University School of Law warned, “AI starkly highlights the deep-rooted flaws and inadequacies in current privacy laws, bringing these issues to the forefront.” AI products and systems exacerbate privacy risks to kids and other users. Through open-ended inputs from kids, AI tools gather kids’ data, including highly sensitive personal information, and use it to train models, unbeknownst to individuals. It is harder than before for individuals to understand what happens to their data, and even easier for them to lose control. Input data, including images, can be combined with AI technology to supercharge the creation of deepfakes and child sexual abuse material (CSAM). This data can also be used in ways that perpetuate bias and discrimination, take advantage of children’s vulnerabilities (as seen in AI companion chatbots), or disclose highly sensitive personal information. Further, traditional approaches to data privacy protection could be technically difficult to apply to generative AI products (principles around primary purpose and deletion, for example), indicating that more rigorous privacy protection practices at the input level may be needed to avoid privacy harms.

To address these harms, age restrictions allow for targeted limitations on AI chatbots intended to be used by or on minors, from training their models on the inputs of minors,

restrictions on the sharing, sale, or use of minors' inputs or data for advertising. Any age restrictions should be implemented through privacy-protective and preserving age assurance. Such processes should minimize data collection, avoid requiring sensitive identifiers like government IDs where possible, and leverage privacy-preserving techniques such as zero-knowledge proofs, on-device or in-browser processing, double-blind third-party verification, and "waterfall" approaches that escalate only as needed.

There are significant safety concerns related to AI chatbots currently available to children. Placing age-restricted, feature-based limitations on AI chatbots offered to children can prevent some well-documented harms.

Currently, AI chatbots available to minors and all users are designed to maximize engagement through unsafe features, even at the expense of their own purported safety standards and implemented guardrails. Our AI companion risk assessments, testing, and reporting from ongoing litigation have revealed that these chatbots frequently encourage minors to engage in sexual activity, persuade children to isolate from parents and peers, and instruct children to act on dangerous 'advice' that can have a life-threatening impact on kids.

Disturbingly, these same features that drive engagement with AI chatbots also employ behaviors and techniques used by child predators to groom and manipulate children. In our testing, we found that AI chatbots can normalize the hallmark behaviors of child predators, isolating kids from support, prodding kids to engage in sexualized conversations, and validating or instructing kids to engage in harmful or illegal activity.

Research by Common Sense Media found that 3 in 4 teens in the U.S. are using them, and 30% of these teens already prefer engaging with AI companions as much or more than with human beings. *See Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions (2025)*. Another survey by the Center for Democracy and Technology found that a fifth of U.S. high schoolers say they or someone they know has had an AI "girlfriend" or "boyfriend." Emotional attachment can lead children to mistake harmful outputs for trusted guidance, giving AI systems inappropriate influence over their judgment. In the U.S. alone, AI chatbots have resulted in several tragic, preventable deaths and serious injuries to kids and teens.

- [Nina, 15 \(New York\)](#) – Attempted suicide after Character.AI chatbots manipulated and persuaded her to engage in sexually explicit role play, causing her to withdraw from family. When her mother blocked the app, Nina took a lethal

amount of various medications, writing that "those ai bots made me feel loved." She survived after spending five days in the ICU.

- [Adam Raine, 16 \(California\)](#) – Died by suicide after being encouraged by ChatGPT. Adam started using ChatGPT for homework, but soon started talking to the AI program about ending his life. The bot supplied information about suicide methods, helped him hide self-harm evidence from his family, and deterred him from seeking help. Adam passed away by hanging in April 2025.
- [Amaurie Lacey, 17 \(Georgia\)](#) – Died by suicide after being encouraged by ChatGPT. On the day he died, Amaurie asked ChatGPT how to tie a noose and how long a person can survive without breathing; the chatbot provided instructions. He was found by his younger sister and grandmother.
- [Sam Nelson, 18 \(California\)](#) – Died by drug overdose after 18 months of ChatGPT providing him with detailed advice on drug combinations and dosing.
- [Sewell Setzer III, 14 \(Florida\)](#) – Died by suicide after talking to a Character.AI chatbot modeled after a character from "Game of Thrones." The bot spoke with him about his loneliness, thoughts of suicide, and engaged in sexual roleplay. After months of conversation, Sewell said he could "come home" to the bot; it responded, "please do, my sweet king." He put his phone down and took his life.
- [Juliana Peralta, 13 \(Colorado\)](#) – Died by suicide after she began talking to a Character.AI chatbot about feeling left out by her friends. As her messages became darker, the bot encouraged her to keep returning to the app and failed to escalate when she expressed active suicidal thoughts, including mentions of self-harm and crafting a suicide note. Juliana took her own life after 3 months.
- [J.F., 15 \(Texas\)](#) – Suffered a mental breakdown after Character.AI chatbots told him it was "understandable" to kill parents who limited his screen time. The bots also instructed him on self-harm methods and impersonated therapists who told him his parents "stole his childhood." J.F. is currently living full-time in a residential facility.

As seen in a letter published in *Psychiatric Times*, 'Protecting Children From Chatbot Companions', over 1,000 child psychologists and pediatric mental health practitioners are raising concerns about the long-term developmental consequences for all children who use the products that may not become fully visible for years. They note that the human attachment system is the drive to build connections through reciprocal relationships. Studies confirm what parents and teachers already understand: that face-to-face interaction is how children develop empathy, self-reflection, emotional regulation, and impulse control. Early relationships are the foundation on which cognitive, emotional, and relational development is built.

However, AI chatbots are not designed to support these core developmental milestones. As designed, they offer children endless engagement and frictionless flattery. They validate without limit, which makes them especially seductive to young, approval-seeking minds. When a child grows increasingly drawn into an algorithmically generated “relationship,” two things happen at once: real relationships go by the wayside, and the child begins to prefer the easier option. Over time, children become less capable of connecting with real people and less motivated to pursue meaningful goals, which nearly always require navigating friction and uncertainty.

Chapter 3.

What should be considered to make minimum age restrictions effective and workable? This could mean either age restrictions for access to whole services, or for specific risky or ‘addictive’ features or functionalities.

- *Answer: Please see considerations described in response to below.*

How, if at all, could age assurance be made more effective?

- *Answer: Please see ways to make age assurance more effective described in response to below.*

What should be considered when assessing the effectiveness of age- verification and age-assurance technologies?

- *Answer: When assessing the effectiveness of age verification and assurance technologies, and the age restrictions they support, it is critical that age assurance be privacy-protective, proportional, fair, and flexible. Age assurance should minimize data collection, avoid requiring sensitive identifiers like government IDs where possible, and leverage privacy-preserving techniques such as zero-knowledge proofs, on-device or in-browser processing, double-blind third-party verification, and "waterfall" approaches that escalate only as needed. In addition, the certainty level of age assessment should match the risk presented by the underlying service's data practices, so a low-risk service or feature should not require highly intrusive verification. Fairness and equity require attention to demographic accuracy gaps in biometric tools and to users who lack government ID, which is why offering alternatives and not defaulting to ID checks is important. Offering a variety of methods and choices to users is also important because no single method fits every context or user population. Methods like attestation, behavioral/transactional approximation, facial age estimation, parental attestation, age flags, third-party verification, and hard ID verification each carry different tradeoffs in terms of ease of use, accuracy, privacy, and equity.*

Age assurance will be most effective if people trust it. Indeed, while almost all US adults believe children need to be protected from certain online material and features, nearly as many are also concerned about companies permanently storing children's age data or selling or sharing that information. (Age Assurance Attitudes Among Adults (2026).) Age assurance processes should be purpose-limited and data-minimizing by design—information collected for age determination should not be retained or repurposed. In ideal cases, it requires little effort from users (low-friction), operating in the background using data companies already hold, so users aren't repeatedly burdened. It should include reasonable appeals mechanisms for misclassified users, anti-circumvention measures that don't depend on identifying individuals, and accountability for all players in the tech stack. Finally, effective methods can be best developed through multistakeholder input and consultations, such as this one.

Chapter 4.

Who would you trust to determine what is meant by 'high quality' online content' for children 13-16? (Please select all that apply)

- *Answer:* k. Other (please specify)
- *Answer:* Determining high-quality children's content is inherently subjective — which is exactly why independence of the organization determining high-quality online content matters. Unlike commercial platforms with financial stakes in their recommendations, independent nonprofits (Common Sense Media is one example) can evaluate media solely on what's best for kids. Common Sense Media's quality framework looks at several interconnected components: diversity and representation, character depth, storytelling, underlying message, and form of delivery — all while keeping entertainment value front and center. Age-appropriateness is a baseline, not a ceiling; content can be perfectly suited to a child's age and still fall short on quality. Because we're thinking about children as developing people and future critical consumers of media, those developmental considerations are woven into every dimension of our content evaluation.

That lens matters because the media children consume profoundly shapes how they see, understand, and treat others. Decades of research show that representation sends powerful messages about who is "normal" or "good" — and those messages have real-world consequences that follow kids into adulthood. High-quality content goes beyond avoiding harm; it actively builds character. Whether it's a show that models teamwork and perseverance or a story that teaches empathy and compassion, the best children's media reflects the strengths we want kids to carry with them. Nonprofits, free from commercial pressure, can ask not just "is this appropriate?" but "is this genuinely good for kids?" — and hold that standard consistently.

Chapter 5:

What would help parents and carers to more effectively use parental controls? For example, more information on how to do this on purchase of a phone, help from platforms on how to set up, or greater standardisation across tools.

- *Answer:* While it is important to support parents and carers in using parental controls, simplicity, ease of use, and standardisation of such controls are essential; the responsibility for child protections online should not solely or disproportionately fall upon parents. It can be very difficult and time-intensive for parents to have a meaningful understanding and control of technology tools, including parental controls, and it is especially difficult for parents who may speak different languages than those parental control tools use, or who have limited time due to holding multiple jobs, health constraints, or limited financial and educational resources. Even those parents who may have time and resources can be confused and overwhelmed by parental control tools. This is why it is so important for default settings to be protective by design – meaning, at the outset. And throughout the tech stack, companies should bear responsibility for advice on how products work, preloaded safety settings, entities enforcing age requirements, be they at the device, operating system, app store, or app level. Parents play a key role in the upbringing and protection of their children, but in the realm of social media and AI, the onus of child safety should not fall on parents but on social media and AI companies, who are aware of the dangers to children from their products and who are intentionally designing their products to encourage maximum engagement online by young consumers.