AI Insight Forum: Privacy & Liability

Tracy Pizzo Frey
Common Sense Media

Written Comments
Wednesday, November 8, 2023
Kennedy Caucus Room, Russell Senate Office Building
Washington, D.C.

Thank you Leader Schumer, Senator Rounds, Senator Heinrich, and Senator Young for the opportunity to participate in your AI Insight Forum on privacy and liability and to submit these written comments for your consideration. I am a senior advisor to Common Sense Media, the nation's leading nonprofit organization dedicated to improving the lives of kids and families by providing the trustworthy information, education, and independent voice they need to thrive in the 21st century. Common Sense's AI initiative is designed to ensure responsible AI for all children and families, and it includes an AI Literacy curriculum distributed to the more than 85,000 U.S. schools that use Common Sense's digital citizenship curriculum. This month, we will also launch the first-ever ratings and reviews system for AI products. I came to Common Sense Media after spending 11 years at Google, including three years at Google for Education. Relevant to this forum, in 2017 I created and subsequently ran all of Google Cloud's responsible AI work until March of 2022, including how we evaluated the products we were building, our customer engagements, and the various programs we built to support this work. I helped Google finalize its AI Principles, and the governance processes I built for Cloud served as a model for other areas of Google as they worked to align what they were doing with Google's AI Principles. I have been involved with a number of policy efforts in this space, including the EU AI Act and NIST's AI Risk Management Framework.

Artificial intelligence is evolving at an unprecedented pace without sufficient guardrails to protect human rights and our democracy. While AI presents many exciting opportunities for learning, discovery, innovation, and economic growth, the status quo presents significant challenges that puts the safety and privacy of all Americans – but especially children and teens – at risk, including the risk of undermining our basic social fabric. Congress should consider specific safeguards needed when data ingestion and analysis is inherently part of a product or service, including in ways that are opaque to many consumers.

There is no question that Congress and the states must do more to protect the data privacy of all users, but especially children and teens, and I encourage Congress in particular to pass legislation to establish strong data privacy protections for kids and teens.

Here are several reasons why AI has the potential to exacerbate data privacy and liability issues for children and teens:

1) **Lack of protections around high-risk categories of AI products and high-risk industries and sectors**

    a) *Generative AI*

        i) Generative AI (including those powered by large language models such as ChatGPT or Bard and text-to-image models such as DALL-E or Stable Diffusion) introduces new privacy concerns beyond commercial targeting, behavioral tracking, and third-party data sharing. Generative AI raises a special privacy concern regarding protecting personally identifiable information (PII) when products do not require collecting this information to provide their service.

        ii) Generative AI suffers from an unsolved technical challenge referred to as “memorization,” which essentially describes the well documented phenomena of language models “memorizing” long passages of text from their training data. Importantly, it does not take much for this to happen and has been known to occur after only one pass at training a model, and unfortunately “memorization” can be quite resistant to efforts such as de-duplication or filtering of generated outputs. When user inputs are leveraged to retrain the models, these inputs are also subject to “memorization”, making it possible for the model to “reuse” those inputs when generating new output to any user. The models do not, and cannot feasibly, distinguish whether this input includes PII, so this information is also subject to memorization. All firms should always provide user-visible warnings about input privacy, and at a minimum consumers would benefit from clear and required transparency standards.

        iii) Using inputs to retrain these models is an important tool for companies to combat other known challenges (such as generating responses that perpetuate harmful stereotypes or “hallucinations" — an informal term used to describe the false content or claims that are often output by generative AI tools). For this reason, use of inputs is often a default – as in the case with ChatGPT – and at times required – as in the case with Bard. Regardless, the same transparency standards should apply here, as users should be given clear information that is visible and persistent throughout their use of these products. While many generative AI products require users to be over age 18, or have parental permission to use them if they are over age 13, we know from research we have done at Common Sense that many children and teenagers are using these tools, regardless of these legal requirements. Adding these protections would not only benefit everyone, but would serve to help protect our most vulnerable populations even if they aren't supposed to be using these products.

    b) *Deep fakes, AI-generated porn, and AI generated CSAM*

        i) Text-to-image models such as DALL-E or Stable Diffusion present unique challenges for privacy. While it is technically impossible, given how diffusion models work, for these services to generate exact images containing public figures, they have the capacity to create photorealistic replicas that may increasingly become indistinguishable from actual photographs. There is a range of technical protections against generating images of public figures across these products. Importantly, however, any user can *upload* an image and then *modify* it using these tools, and the resulting images can be used in anything ranging from bullying and harassment to misinformation and disinformation campaigns. It is now well documented that AI is being used to generate pornographic images, primarily of women and teenage girls without their consent, and new CSAM material, at alarming rates, with no recourse by the victims to stop them from being generated and circulated. (Common Sense will have more information and guidance about the threats posed by these activities when we release our ratings and reviews later this year.)

c) *Education should be considered a high-risk category with regard to AI*

i) Education is now, and has been for years, viewed as a business opportunity by tech companies. But given the unique vulnerabilities of students and of schools themselves, the use of AI in education should be viewed as a high risk activity when it comes to transparency and governance of AI products. While not all uses of AI in education are high risk, it is because of the vulnerability of children and teens that we are recommending this higher scrutiny for all uses in education. Children, parents and other caregivers have limited or no ability to opt out of the use of AI when deployed in educational settings and they deserve the highest levels of transparency. Educators need sufficient information about products that use AI to determine what to procure for their schools. Firms that intend to have their products used in educational settings must take extra precautions for their intended users, and AI should only be used in cases where it equitably furthers educational goals with benefits that cannot be achieved with other technologies. In addition, many products not designed for use in education are still used widely in education, which makes it more important for AI products to be designed responsibly. Products built with ethics and responsibility 'by design' should include assessing use by children and teens and schools at the outset of the product's development. The result of this sort of responsible AI practice will serve to make these products both better for everyone, more beneficial to children, teens and education, and more likely to be successful for the organizations creating them. Given the high risk of kids and teens in school settings and working on school projects at home, it could be too late for a firm to try to adjust an AI product that was not originally intended for education to be sufficiently protective of student users.

ii) The most dangerous uses for AI in education are:

(1) Surveillance can create privacy, safety and security risks and limit children's freedom of choice, self-determination, and willingness to express themselves. Concerns regarding technologies that can be used for surveillance in any industry are heightened in the context of children and education.

(2) This is also the same for risk prediction use cases, verbal aggression detectors, remote proctoring, automated abstract work grading (e.g. essays, creative writing), bot/toy interactions where bot is not declared, bots/toys claiming to have feelings, automated admissions decisions, and use of "AI detectors" which are currently extremely unreliable.

(3) While not specific to privacy, "Emotion detection" is growing in popularity as a use for AI. Critically, this – along with all forms of affective computing – are not sufficiently supported by science and present significant risk to all who are subject to it. Alongside the need for far greater research and scientific support, any ability for AI to be trustworthy in this space would need to be able to account for the myriad and different ways in which facial cues, for example, are associated with different feelings across cultures. Given the individual variation in how emotions are visibly expressed as well, any claim that today's AI is able to accurately detect emotions should be met with the utmost scrutiny.

(4) For the reasons discussed above, generative AI calls out for greater privacy protections and greater liability protections than exist today to protect all consumers, especially children and teens.

We recommend Congress focus on the following solutions.

1) **Strengthen data privacy for kids and all consumers and establish guardrails for kids on social media platforms**
    a) A baseline and bipartisan step Congress should take to maximize AI's benefit while protecting the American people is passing strong privacy and platform accountability legislation. This includes updating existing laws, such as COPPA, with the bipartisan COPPA 2.0 that would strengthen protections for young children and extend protections to teens, and adopting the Kids Online Safety Act (KOSA). It also includes passing comprehensive privacy legislation that would protect everyone. These baseline protections are important across all types of technology and with respect to all uses of data, and should include protections against commercial targeting, behavioral tracking, and third-party data sharing in addition to requirements for data minimization and transparency. Whatever new technology we face, basic privacy and accountability protections are key to protecting individuals.
2) **The ideal: By design, all AI is built with ethics and responsibility**
    a) Many of the concerns that exist today for privacy, liability, and the ways in which AI continues to change our social landscape are directly connected to the data used to train the models that power them and what–if any–ethical and responsible AI considerations were taken into account early in development. Generative AI chatbots serve as a good example here. These chatbots are able to generate responses to a wide range of questions and prompts because they are trained on massive amounts of information scraped from the internet. To date, organizations that have released these chatbots work to limit generation of harmful content with *post-training*

techniques. Should, however, companies take a different approach and thoughtfully curate what data to include at the beginning, many of the concerns raised by these tools would be significantly lessened.

a) *It is not impossible for AI products to protect privacy and afford liability protection*
    i) It is not impossible for AI products to protect privacy, but in order to do so they need to be designed with that goal from the beginning. Unfortunately, the vast majority of existing AI tools were not designed from the start with privacy in mind. And, as in all fast moving areas of technology, it is critical that any federal legislation not foreclose states' ability to innovate in the future.
    ii) It's important that federal efforts to regulate AI include strong and clear language around liability. This provides firms with clear guidance on the liability that attaches to their activities. If the government defines liability clearly at the outset, then the government will be better equipped to induce firms to participate in a robust "pre-market" regulatory framework.
b) *Recognize vulnerabilities of children and teens on AI platforms and do more to protect them.*
    i) Young children especially should be kept off inappropriate products, and companies must be held accountable when kids are on those products.
    ii) Many AI products have age requirements in their terms of service. Often these are obscure documents, and even if an acknowledgement of terms gate exists, this serves more to provide legal protection for the companies than clarity for consumers about who is allowed to use the product and in what ways. In addition, some companies allow use for teenagers 13+, and at times legally require parental / guardian permission to do so. While age gates sometimes exist, they do not always provide the needed transparency, nor do they always block a user younger than the terms require from signing up. Companies should not be allowed to hide behind TOS and ineffective age gates.
    iii) Companies should not be able to pretend that children are not on their sites when they know they are. In order to avoid the unnecessary and potentially hazardous collection of additional data on kids, Congress should update the "knowledge standard" in COPPA and any other legislation under consideration with regard to social media platform regulations.
    iv) Age assurance should be done appropriate to the level of risk. While complex, evaluating AI products along a continuum of risk can help determine the appropriate age assurance actions to take.
c) *EU AI Act and a call for transparency*
    i) Our upcoming AI ratings and review systems fill a need to provide the public with a trusted source of information that thoroughly assesses the safety, transparency, ethical use, and impact of AI products. Our hope is that this will also serve as an example of auditing and meaningful stakeholder-relevant transparency.

Thank you again for the opportunity to submit these written comments on behalf of Common Sense Media. We look forward to the forum and to continuing this conversation with you to ensure the greatest protections and opportunities for children and families with regard to artificial intelligence and other existing and emerging technologies.