



## Insight

# Primer: AI Impact on Kids Social Media Safety Legislation Discussion

ANGELA LUNA | JULY 25, 2024

## Executive Summary

- A wide range of legislation has been introduced this Congress to safeguard minors on social media, yet these proposals often fail to fully consider the effect that legislation would have on artificial intelligence (AI) tools.
- Advances in AI allow social media companies to enhance automated content moderation, implement sophisticated age-verification techniques, and improve user experiences, but these tools could also amplify potential risks to minors' safety and well-being.
- As Congress debates kids' safety social media legislation, it should ensure that proposals target specific harms and do not limit the potential benefits of AI applications.

## Introduction

The 118th Congress has considered a variety of bipartisan bills aimed at safeguarding minors on social media, including the [Kids Online Safety Act](#), the [Social Media Child Protection Act](#), the [Safe Social Media Act](#), and the [Protecting Kids on Social Media Act](#). These proposals offer a wide spectrum of policy recommendations, including age assurance requests, funding research on the effects of social media on minors' mental health, data protections, and establishing a duty of care for social media platforms. Yet these proposals often fail to fully consider the effect that legislation would have on advancements in artificial intelligence (AI) tools or mitigate the specific harm that these tools can amplify.

All social media platforms [utilize AI](#) in different ways. With AI-driven capabilities that allow social media companies to enhance automated content moderation, implement sophisticated age verification techniques, improve user experiences, and allow image and video editing, AI is transforming social media experiences, including those of kids and teens. As AI tools evolve, platforms can improve safety by removing spam and fake accounts and provide an experience that aligns with personal interests. Yet the advantages brought by these new features also come with trade-offs. For example, techniques designed to enhance user experiences, such as automatically loading additional posts to allow for endless scrolling, could lead to overuse, while image-editing features might negatively impact minors' self-esteem by advancing unrealistic beauty standards.

As Congress debates kids' safety social media legislation, it should consider the trade-offs of AI-driven features and ensure that proposals target specific harms without limiting the potential benefits of AI applications to minors and users more generally. This primer discusses the applications of AI in social media and the benefits and challenges of these applications for young users, and reviews current minors' safety bills and other initiatives that attempt to mitigate the harms of AI applications on social media.

## How AI Shapes Social Media Experiences for Minors

AI-driven improvements to social media may offer broad, significant benefits for younger users. For example, the Surgeon General Advisory [highlights](#) that social media has the potential to cultivate a supportive community where people with shared identities, abilities, and interests connect, access important information, and express themselves. The American Academy of Pediatrics [affirmed](#) that benefits include exposure to new ideas and knowledge acquisition, increased opportunities for social contact and support, and new opportunities for community engagement. The technology also comes with trade-offs, however, that could negatively impact children's mental health and well-being.

AI tools can be designed to protect children and provide a safer environment, such as through the use of [AI content moderation](#) that identifies, sorts, and removes content that doesn't meet a platform's standards. For instance, AI can identify and block comments that include hate speech and violence or remove otherwise inappropriate content to foster a safe and positive online environment for kids and teens. At the same time, AI content moderation is not immune to error and bias and may require access to user data, raising concerns about data privacy.

Additionally, content curation and recommendation are mechanisms that employ AI algorithms to provide users with engaging and interactive content and features. Platforms utilize AI to select and organize (content curation) and suggest (content recommendation) what to show users based on their preferences and behavior. For minors, this could involve presenting entertaining images, videos, or educational material on their feeds that match their interests and preferences. For example, [Meta](#) uses AI models to predict how important a particular piece of content will be to a user of Facebook or Instagram to make experiences more relevant. Yet by creating personalized experiences, AI algorithms make users' experience more engaging, which could potentially lead to minors' overuse or minors ending up in a harmful [filter bubble](#), a unique digital space tailored to each user.

Similarly, AI age-assurance techniques employ AI to verify the age of social media users through facial characterization, biometric verification, and other [estimation](#) methods. AI age verification enables social media to control access and safeguard children from inappropriate content, but this process involves analyzing personal and biometric data, which may threaten minors' data privacy.

Finally, AI-powered [image editing](#) is another trending application that uses AI-powered filters for users to modify photos and videos, such as by smoothing skin or altering facial features. While making the platform more interactive, such technology can also exacerbate unrealistic beauty standards, potentially leading minors to feelings of inadequacy. More research is needed in this field, but many studies suggest that social media can expose children to content that makes them feel worse about themselves and could lead to depression and anxiety: A JAMA Psychiatry study [indicates](#) that adolescents who spend more than three hours per day on social media face double the risk of experiencing poor mental health outcomes. Similarly, in a meta-analysis conducted by the Journal of Abnormal Child Psychology, researchers [showed](#) that more social media use is significantly associated with elevated depression symptoms.

## Initiatives to Make Social Media a Healthy Environment

Given the potential harm AI-driven social media can inflict on young users, the 118th Congress has introduced several bills to bolster online protections. These initiatives take a wide range of approaches to protecting children online, and most would have an impact on how social media companies deploy AI tools.

Most notably, the [Kids Online Safety Act](#) would create a duty of care for social media platforms, requiring them to take reasonable measures in the design and operation of their services if they know the service is used by

minors. Specifically, the bill would mandate safeguards for minors' data and parental supervision tools on covered platforms and require large online platforms to notify users about algorithm use and offer an alternative version that doesn't prioritize user-specific data. This approach would directly impact many of the AI features of social media because it would broadly apply to almost all aspects of social media services, meaning some harm could be mitigated. At the same time, fear of liability could prevent social media companies from offering AI-driven tools and features that attempt to provide users with helpful or relevant content, negatively impacting the experience for all users without directly tackling the specific online harms faced by minors.

Other bills, such as the [Social Media Child Protection Act](#) and the [Protecting Kids on Social Media Act](#), would set specific ages at which a user can create a social media account, and require platforms to verify the age of users to ensure compliance. Social media companies have been implementing various methods to comply with these requests, but these practices come with trade-offs. Traditional [age gating](#) techniques involved asking users if they are the appropriate age via a checkbox system, an unreliable practice as minors could easily bypass the age gating systems by providing false information. Other practices included requesting government IDs to verify ages, but they can potentially exclude many young users who lack the necessary documentation, and [courts have been skeptical](#) of such approaches under the First Amendment. Current age-assurance techniques are more sophisticated. Many include AI-driven methods, including facial age estimation, biometric verification, and inference models (such as analyzing browsing history) that use AI to evaluate facial geometry and other biometrics, and online activities like social media engagement and screen time to determine users' age or age range. While these methods are more effective in determining age, and thus protecting minors online, they also struggle with accuracy and bias and can intrude on privacy, allowing for the potential collection, use, or sale of user activity data. Congress mandating age assurance, rather than allowing companies the flexibility to determine if age assurance is necessary, could cause significant harms to users or limit new developments in AI-driven age assurance out of fear that the technique would not sufficiently comply with the law.

A more prudent approach could be something like that proposed in the [Safe Social Media Act](#), which would mandate a study on social media use among minors, focusing on frequency, mental health effects, and policy recommendations. Such a study could better identify the specific harms that need mitigation through legislation, without necessarily limiting the development of new services and offerings.

## **Conclusion**

Congress has made kids' safety on social media services a priority, yet much of the proposed legislation could impact the development and deployment of AI-driven features without adequately addressing the specific harms they could cause or exacerbate. As Congress debates kids' online safety legislation, it should fully consider the interplay of these factors to inform legislation to mitigate harm while preserving AI's benefits.