The Virginia Joint Commission

JCOTS

on Technology and Science

Protecting Youth Online: An Evaluation of Social Media Protection Tools

November 2024

JCOTS Membership

Delegate C.E. Cliff Hayes, Jr., Chair
Delegate Bonita G. Anthony*
Delegate Mike A. Cherry
Delegate Michelle Lopes Maldonado
Delegate David A. Reid
Delegate Anne Ferrell H. Tata
Delegate Michael J. Webert

Senator Adam P. Ebbin, Vice-Chair
Senator Lashrecse D. Aird
Senator Christie New Craig*
Senator Barbara A. Favola*
Senator Ghazala F. Hashmi*

*Indicates membership in the Online and Data Protections Subcommittee

Executive Director

Jodi Kuhn

Report Author

Zhiqiang Ji, Ph.D.
JCOTS Research Collaborator, Fall 2024

Table of Contents

Summary of Recommendations

This report aims to inform policymakers and stakeholders about the technical feasibility, user experience implications, and potential impact of social media protective measures while considering both their intended benefits and possible limitations.

1. **Warning Labels**

Warning labels on social media platforms, similar to those on tobacco products, would alert users that "social media is associated with significant mental health harms in adolescents."

- Develop context-specific warnings tailored to platform features, user age groups, and behavior patterns.
- Integrate warning labels with broader protective measures, including educational resources about healthy social media use and links to mental health support services.
- Regularly evaluate warning label effectiveness and adjust messaging based on user response data.

2. **Verifiable Parental Consent**

Verifiable parental consent (VPC) mechanisms, first implemented by the Children's Online Privacy Protection Act (COPPA) in 1998, require online services to obtain verifiable permission from parents or guardians before collecting, using, or disclosing personal information from minors.

- Expand beyond traditional verification methods to include privacy-preserving technologies that minimize data collection and retention.
- Implement exemption protocols that maintain safety while providing access to essential resources for at-risk youth, potentially through trusted intermediaries.
- Emphasize transparency and user control by offering granular controls that allow parents to customize access levels based on their child's needs and circumstances.
- Conduct regular effectiveness evaluations to assess the impact on different user groups, particularly vulnerable populations, and adjust approaches accordingly.

3. **Time-Based Restrictions**

Time-based restrictions on social media access aim to limit young users' exposure during specific periods when platform use may interfere with essential activities or pose heightened risks.

- Establish multi-layered verification frameworks that go beyond simple IP-based time zone detection to address technical circumvention challenges.
- Mandate graduated restriction models that include emergency override options and alternative access paths for crisis services to protect vulnerable youth.
- Integrate restrictions with educational components that explain the health and academic benefits of digital boundaries to improve effectiveness and user acceptance.

4. Cookie Opt-Out

Cookie opt-out mechanisms provide significant advantages in protecting user privacy and building trust between platforms and users.

- Mandate standardized, user-friendly consent interfaces that use clear, non-technical language, ensure equal prominence of accept/reject options, and include brief educational components.
- Establish guidelines that balance privacy protection with platform sustainability, including creating precise standards for distinguishing between necessary and optional cookies and allowing for privacy-preserving alternatives.
- Require regular evaluation of cookie consent mechanisms, including documentation of consent choices, periodic audits of cookie usage, and reporting on user engagement metrics.

5. Addictive Feed Restrictions

Addictive feed restrictions aim to regulate social media platforms' algorithmic content recommendation systems that may negatively impact young users' mental health and online behavior.

- Mandate regular, independent algorithm risk audits with public disclosure to evaluate the effects of platforms' recommendation systems on young users' mental health and well-being.
- Establish clear standards for alternative content delivery methods, requiring platforms to offer chronological feeds as the default for minors, implement transparent content curation criteria, and develop age-appropriate recommendation systems.
- Require platforms to maintain detailed documentation of their algorithmic practices and their impacts on minors, including regular reporting on key metrics and mechanisms for researcher access to anonymized data.

6. Beyond Platform Regulation

The protections noted thus far have been in the form of digital barriers; however, protection should also include equipping young people with the skills to navigate online spaces safely and meaningfully.

- Invest in digital literacy education programs for both youth and parents.
- Support collaboration between schools and platforms on educational initiatives.
- Promote family-based approaches to digital wellness.
- Strengthen parent-child relationships and communication about online activities.

1. Introduction

Social media has become an integral part of adolescent life, with nearly 95% of teenagers using these platforms and more than a third reporting being online almost constantly[1]. In a June 2024 New York Times opinion piece, U.S. Surgeon General Dr. Vivek Murthy raised alarm about this unprecedented level of engagement, comparing the current youth mental health crisis to the tobacco epidemic of the previous century[2]. While social media platforms offer unprecedented opportunities for connection and expression, mounting evidence suggests they may also contribute to a surge in youth anxiety, depression, and other mental health challenges. According to *Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory*, the design features of these platforms—from infinite scrolling to algorithmic content promotion—may be particularly harmful to developing minds, creating patterns of compulsive use that can interfere with sleep, healthy relationship development, and overall well-bein[1].

Efforts to protect children and adolescents online date back to the early days of the Internet, as exemplified by the Child Online Protection Act (COPA) of 1998[3]. However, this and similar well-intentioned legislation have consistently faced challenges in implementation due to vague standards, unclear definitions, and potential conflicts with First Amendment rights[7]. The rise of mobile Internet and social media platforms, particularly their widespread adoption among youth, has introduced new complexities to protective policymaking. In recent years, states have taken increasingly active roles in regulating social media companies to protect minors. Arkansas, Utah, Texas, California, and Louisiana have enacted legislation addressing various aspects of youth protection online. At the federal level, bipartisan efforts like the Kids Online Safety Act (KOSA)[5] and the Children and Teens' Online Privacy Protection Act (COPPA 2.0)[6] have emerged, though they face scrutiny from tech industry and human rights groups regarding their effectiveness and constitutionality[8][16]. These legislative efforts, while well-intentioned, raise concerns about potential unintended consequences for vulnerable youth groups, such as LGBTQ+ teens or those in abusive households who rely on social media for support and community connection[15][41][44].

Virginia legislators have also initiated efforts to protect young users through various legislative measures. Rather than examining specific legislation, this report provides a focused analysis of five protective mechanisms that have emerged from recent legislative proposals: **warning labels**, **verifiable parental consent (VPC) requirements**, **time-based access restrictions**, **enhanced cookie control options**, and **regulations addressing addictive feed designs**. For each of these tools, we evaluate their practical implementation challenges, assess their potential effectiveness, and provide recommendations for improvement. This analysis aims to inform policymakers and stakeholders about the technical feasibility, user experience implications, and potential impact of

these protective measures while considering both their intended benefits and possible limitations.[1]

2. Analysis of Five Protection Tools

2.1 Warning Labels

Warning labels on social media platforms, like those on tobacco products, would alert users that "social media is associated with significant mental health harms in adolescents." This approach, championed by U.S. Surgeon General Dr. Vivek Murthy, aims to raise awareness about the potential risks of social media use, particularly for young users[2]. The concept has gained significant traction, with 42 state attorneys general supporting the implementation of such warning labels at the federal level[9].

Benefits

Warning labels can serve as awareness tools and behavior modifiers. Evidence from tobacco warning labels demonstrates their ability to increase public awareness and influence user behavior. Research from Harvard shows that the visibility and design of labels significantly affect their impact—for example, changing label colors from light gray to blue increased user awareness by 15 percentage points[10][11].

Studies also show that warning labels can effectively reduce the spread of misinformation; even people who are more hesitant about the warnings' source are still likely to behave in ways that treat the warning as credible[12]. Additionally, warning labels respect user autonomy by allowing individuals to make their own choices while providing important safety information.

Limitations

Several significant challenges limit the effectiveness of warning labels. First, unlike tobacco or alcohol, social media's effects vary significantly among users and usage patterns, making universal warnings potentially oversimplified. Research indicates that warning labels can become ubiquitous and ignored over time, similar to California's chemical warning labels that appear on numerous products[13]. Furthermore, labeling all digital tech use as 'dangerous' could destroy credibility with teens because most report their social media experience as positive, even as they recognize the problems[14].

[1] Four of these mechanisms correspond to specific Senate Bills in the Virginia General Assembly: verifiable parental consent (SB 432), restricted hours (SB 532), cookie opt-out (SB 252), and addictive feeds (SB 359).

Implementation Recommendations

To maximize the effectiveness of warning labels, we recommend developing state-mandated warnings that are context-specific and consider platform-specific features and risks, user age groups, and/or behavior patterns. These warnings should be integrated with broader protective measures, including educational resources about healthy social media use and links to mental health support services. Regular evaluation systems should monitor warning label effectiveness and adjust messaging based on user response data. Most importantly, warning labels should be viewed as one component of a more comprehensive approach to social media safety rather than a standalone solution.

2.2 Verifiable Parental Consent (VPC)

Verifiable parental consent (VPC) mechanisms, first established by the Children's Online Privacy Protection Act (COPPA) in 1998[4], require online services to obtain verifiable permission from parents or guardians before collecting, using, or disclosing personal information from minors. While COPPA initially focused on protecting children under 13, recent state legislation has expanded both the age thresholds and the scope of protection, creating a complex landscape of requirements across different jurisdictions[8][15].

The implementation of VPC varies significantly across states and platforms. Some states, like Florida, require one-time parental consent only for account creation for users aged 14-15, while others, such as Utah, mandate ongoing parental supervision with continuous access to minors' activities[16]. The scope of consent requirements ranges from basic account creation to more comprehensive oversight of specific features, including data collection practices, privacy settings modifications, messaging capabilities, and time limitations.

To facilitate these requirements, the Federal Trade Commission (FTC) has approved various verification methods that balance security with accessibility[17]. Traditional approaches include signed consent forms and financial verification through credit card transactions, while more modern methods incorporate technological solutions such as facial recognition matching and knowledge-based authentication. For internal-use data collection, platforms may employ the simplified "email plus" method, though this is not sufficient when information will be shared with third parties.

Benefits

VPC mechanisms serve multiple protective functions in the digital space. When properly implemented, these systems enable parents to make informed decisions about their children's online engagement and maintain oversight of data collection practices. VPC systems can help platforms maintain compliance with privacy regulations while offering parents granular control

over specific features, including privacy settings, data collection permissions, and interaction capabilities[15].

Public support for VPC requirements is notably strong. Recent Pew Research Center surveys indicate that 81% of U.S. adults support the requirement for parental consent for minors to create social media accounts, with strong bipartisan backing across different age groups and parental statuses. Even among teenagers, 46% support parental consent requirements, with only 25% opposing such measures[18]. This broad public support suggests that VPC mechanisms align with societal expectations for protecting young users online while maintaining parental oversight of their digital activities.

Limitations

Technical and implementation barriers significantly hinder VPC's effectiveness. The verification process itself presents numerous challenges, from managing multiple verification methods (credit cards, government IDs, facial recognition) to maintaining consistent user experiences across platforms. These systems can also be circumvented by tech-savvy youth using fake documentation, borrowed credentials, or stored payment information[44].

Access and equity issues pose another significant concern. The verification requirements create substantial barriers for non-traditional families, including those where children have different last names than guardians, those in foster care, or those under the care of relatives. Furthermore, some verification methods rely on devices or documentation that not all families may have, potentially excluding children from disadvantaged backgrounds.

Most critically, VPC mechanisms can cause unintended harm to vulnerable youth populations. In Utah alone, where parental consent legislation was recently passed, 9,695 children were confirmed as victims of abuse and neglect in 2022, with the majority of perpetrators being parents[44]. For LGBTQ+ youth or those in abusive households, social media often serves as a crucial lifeline for accessing support resources and communities. Requiring parental consent could effectively cut off these vital support networks, potentially putting already vulnerable youth at greater risk[8][19][41][44].

Privacy concerns also emerge from the verification process itself. Parents must share sensitive personal information with platforms or third-party verification services to prove their guardian status. And children must sometimes share biometric data, such as face scans, with platforms or third parties, which creates new and additional data protection issues. These aspects of the verification process create additional privacy risks, potentially exposing both parents and children to privacy breaches.

Implementation Recommendations

The implementation of VPC should prioritize flexible verification methods that protect user privacy while maintaining effectiveness. Platforms should expand beyond traditional verification methods to include privacy-preserving technologies that minimize data collection and retention. For example, a zero-knowledge proof system could allow parents to verify their identity and relationship with the minor without storing sensitive personal information. In this approach, a trusted third-party verifier could validate the parent-child relationship using existing records (such as school or medical records) and issue a cryptographic proof that platforms can verify without accessing the underlying personal data.

To protect vulnerable youth populations, platforms should implement thoughtful exemption protocols that maintain safety while providing access to essential resources. This could include establishing alternative verification pathways for at-risk youth through trusted intermediaries such as school counselors, social workers, or licensed mental health professionals. These intermediaries could verify the need for access to support resources while maintaining confidentiality and protecting vulnerable users from potential harm.

A balanced implementation framework should emphasize transparency and user control while maintaining protective measures. Platforms should provide clear, granular controls that allow parents to customize access levels based on their child's specific needs and circumstances. This includes the ability to enable access to educational and support resources while maintaining restrictions on potentially harmful features. Regular effectiveness evaluations should be conducted to assess the impact on different user groups, with particular attention to vulnerable populations, and platforms should adjust their approaches based on these assessments to ensure that protective measures don't inadvertently cause harm[20]. These evaluations should be published publicly and/or the data should be made accessible to independent researchers at state agencies and academic institutions.

2.3 Time-Based Restrictions

Time-based restrictions on social media access aim to limit young users' exposure during specific periods when platform use may interfere with essential activities or pose heightened risks. The implementation of these time-based restrictions relies on technical measures such as IP address verification to determine the user's time zone, as specified in SB532, ensuring that restrictions align with local time standards.

Recent legislative efforts across different jurisdictions have established comprehensive frameworks for these restrictions. California's SB 976[22], for instance, prohibits notifications between 12 a.m. and 6 a.m. and during school hours (8 a.m. to 3 p.m. on school days) while also implementing a default one-hour daily time limit on "addictive feeds" for minors.

Time restrictions typically operate on three interconnected levels. First, platform access restrictions may completely block usage during specified hours, particularly late-night periods. Second, feature-specific limitations disable certain functions like notifications, messaging, or infinite scrolling during designated times. Third, usage duration controls implement daily time limits and require parental override for extensions. Virginia's approach through SB532 focuses primarily on the first level, establishing nighttime access restrictions while providing flexibility through parental consent options.

Benefits

Time-based restrictions serve as an important protective mechanism by addressing one of the most direct impacts of social media use on youth well-being: sleep disruption. Research indicates that teenagers need 8-10 hours of sleep for healthy development, yet nighttime social media use frequently interferes with this essential requirement. By implementing technical controls to limit platform access during late-night hours, these restrictions help establish and maintain healthy sleep patterns[24][25][26]. But while controlled trials demonstrate that limiting social media use to 30-60 minutes daily led to significant improvements in depression, anxiety, and sleep quality—there is currently a lack of direct experimental evidence demonstrating the long-term effectiveness of such restrictions[40].

These restrictions also support academic success and daily routines[26]. By preventing social media access during school hours and establishing clear boundaries between online and offline time, time-based restrictions help students maintain focus on their studies and develop balanced daily schedules. This structured approach is particularly beneficial during critical learning periods and helps establish healthy digital habits.

Rather than completely blocking access, these measures create natural breaks in platform engagement while still allowing beneficial social connections during appropriate hours. This balanced approach helps young users develop self-regulation skills while maintaining access to the positive aspects of social media use.

Limitations

Technical circumvention poses a significant challenge to the effectiveness of time-based restrictions. Research shows that tech-savvy youth can easily bypass these controls by using VPNs, manipulating device clock settings, or accessing platforms through alternative devices. For example, South Korea's implementation of a nightly gaming curfew for adolescents in 2011 proved largely ineffective, reducing internet use by only two minutes on average with no measurable impact on sleep patterns[27].

Implementation challenges create practical barriers to effectiveness. Time zones, different school schedules, and varying family routines make it difficult to establish appropriate restriction periods that work for all users. Additionally, platforms face technical challenges in accurately determining local time through IP addresses, as specified in legislation like SB532, which can be circumvented through location-masking technologies.

Perhaps most concerning is the potential harm to vulnerable youth populations. Time-based restrictions may inadvertently cut off access to vital support networks and resources during critical hours. This is particularly problematic for teens who rely on social media for emotional support, especially those in difficult home situations or those seeking connection with supportive communities during late hours when in-person support may be unavailable. Tracking IP addresses may also introduce new data privacy risks for children if there is a data breach or data on the locations of young users is sold to third parties.

Evidence suggests that rigid time restrictions may even be counterproductive. Rather than promoting healthy digital habits, they can push young users toward less regulated platforms or riskier behaviors to maintain their social connections. Research indicates that technological interventions alone have shown weak effects in achieving their intended outcomes, suggesting the need for more nuanced approaches that consider individual circumstances and needs[26][45].

Implementation Recommendations

To address technical circumvention challenges while maintaining protective benefits, legislation should establish flexible verification frameworks that go beyond simple IP-based time zone detection. These frameworks should require platforms to implement multi-layered verification systems that combine device settings, network information, and user-provided data to accurately enforce time restrictions while including mechanisms to detect and prevent common circumvention methods.

To protect vulnerable youth while maintaining the protective intent of time-based restrictions, legislation should mandate graduated restriction models that include emergency override options and alternative access paths for crisis services. This approach should require platforms to enable customization of restriction periods based on individual circumstances, such as different school schedules or family situations, and include provisions for trusted adults like counselors or mental health professionals to authorize specific exemptions.

To improve effectiveness and user acceptance, legislation should require platforms to implement restrictions alongside educational components that explain the health and academic benefits of digital boundaries. This should include requirements for platforms to provide tools for users to track their own usage patterns, conduct regular effectiveness assessments to identify unintended consequences and integrate with school and family-based digital wellness programs to create

consistent, supportive frameworks for healthy social media use. Enhanced data protections should be put in place for geolocation and other data used in multi-layered time restriction verification systems that restrict the storage and sale of this data.

2.4 Cookie Opt-Out

Cookie opt-out mechanisms provide users control over how their personal data is collected and used through website cookies. These tools distinguish between "strictly necessary" cookies essential for basic website functionality and non-essential cookies used for tracking, advertising, or analytics. This approach aligns with the COPPA requirements for protecting minors' data privacy, as it helps prevent unauthorized collection and use of children's personal information online.

Under Virginia's SB252, controllers must provide clear opt-out methods, explicitly disclose cookie purposes, and obtain prior express consent for any non-essential cookies. They are also prohibited from denying service to users who decline cookie placement. The legislation marks a significant shift toward enhanced user privacy protection.

Benefits

Cookie opt-out mechanisms provide significant advantages in protecting user privacy and building trust between platforms and users. For young users specifically, cookie opt-out mechanisms offer three key benefits, particularly on social media platforms. First, they prevent platforms from building detailed behavioral profiles through tracking cookies, limiting the collection and processing of minors' personal data without appropriate consent. Second, these mechanisms help protect young users from targeted advertising and algorithmic content recommendations that might affect their online experiences and mental health. Third, by requiring explicit consent for non-essential cookies, these protections give parents greater control over their children's digital footprint.

From a compliance perspective, proper implementation of cookie opt-out mechanisms helps platforms avoid significant penalties. Under current regulations, violations of children's privacy protections can result in fines of up to $43,280 per violation from the FTC, while the California Consumer Privacy Act violations can result in civil penalties of up to $2,500 per violation, or $7,500 for each intentional violation or violations involving minors' personal information[29]. By providing clear opt-out mechanisms and obtaining proper consent, platforms can protect users and avoid these substantial financial risks.

Limitations

Cookie opt-out mechanisms face several significant challenges that limit their effectiveness as a privacy protection tool. Research from Carnegie Mellon University reveals that only 0.1% of users actively engage with cookie consent interfaces, indicating a critical failure in achieving meaningful user participation[29]. This extremely low engagement rate is particularly concerning for young users who may lack the understanding or attention span to make informed privacy decisions.

User comprehension and experience issues create additional barriers to effectiveness. Studies from the IT University of Copenhagen demonstrate that users often ignore cookie disclaimers not due to privacy apathy but rather from resignation to cookie usage, viewing the disclaimers as nuisances rather than meaningful controls[31]. Many users harbor misconceptions about cookie functionality, incorrectly assuming cookies won't be used unless explicitly accepted, even in opt-out systems. This confusion is exacerbated by complex technical terminology and frequent notifications, leading to "consent fatigue," where users make hasty decisions without proper consideration.

Economic considerations present significant implementation challenges, particularly for publishers and advertising-supported platforms. According to recent industry reports, when third-party cookies are disabled in Chrome without enabling privacy-preserving alternatives, publishers experience a 34% drop in programmatic revenue through Google Ad Manager. Even with privacy-preserving technologies like Privacy Sandbox in place, revenue losses still average around 20%[32]. This substantial revenue reduction could threaten the sustainability of free services and educational content for platforms serving youth audiences, potentially limiting access to valuable online resources that rely on advertising revenue[32][42].

These limitations suggest that while cookie opt-out mechanisms serve an important protective function as outlined in SB252, their implementation requires a careful balance between privacy protection, service functionality, and economic sustainability. The challenge lies in making these controls both effective and accessible while maintaining the benefits of necessary data collection for service improvement.

Implementation Recommendations

To address the critically low user engagement with cookie consent interfaces and widespread comprehension issues, legislation should mandate standardized, user-friendly consent interfaces. These interfaces should use clear, non-technical language to explain cookie purposes, ensure equal prominence of accept/reject options, and include brief educational components that explain data collection implications in simple terms. Most importantly, the interfaces should avoid “dark patterns” that nudge users toward less privacy-protective options. And cookie consent options

should be coupled with broader protective measures, including educational resources and digital literacy support for families and children.

Given the significant economic impact on platforms, legislation should establish clear guidelines that balance privacy protection with platform sustainability. This includes creating precise standards for distinguishing between "strictly necessary" cookies and optional ones, allowing platforms to implement privacy-preserving alternatives for analytics, and establishing safe harbor provisions for platforms that adopt approved privacy-preserving technologies. Such guidelines would help maintain essential services while protecting user privacy.

To enhance transparency and effectiveness, SB252 should require regular evaluation of cookie consent mechanisms. Platforms should be required to maintain clear documentation of user consent choices, conduct periodic audits of cookie usage and effectiveness, and report on user engagement metrics. This ongoing assessment would help identify areas where consent mechanisms need improvement and ensure that privacy protections remain effective as technology evolves.

2.5 Addictive Feed Restrictions

Addictive feed restrictions aim to regulate social media platforms' algorithmic content recommendation systems that may negatively impact young users' mental health and online behavior. According to Virginia's SB359, an "addictive feed" is defined as a website, online service, or application that recommends, selects, or prioritizes content for display based on information associated with the user or the user's device, unless specific exemptions apply (such as user-selected privacy settings or direct communications).

Recent legislative efforts across different jurisdictions have established comprehensive frameworks for these restrictions. New York's Stop Addictive Feeds Exploitation (SAFE) for Kids Act[33] prohibits platforms from delivering algorithm-based content to users under 18 without parental consent and mandates alternative options like reverse-chronological feeds. Similarly, California has passed legislation mirroring these protections, while Virginia's SB359 proposes to prohibit social media platforms from using addictive feeds for users under 18 without obtaining verifiable parental consent.

These restrictions specifically target algorithmic content delivery mechanisms such as:

- Personalized "For You" content sections based on user behavior
- Algorithmic content recommendations
- Auto-playing videos
- Infinite scrolling features that continuously load new content
- Content prioritization based on user engagement patterns

Instead, platforms are required to offer alternative, non-algorithmic options such as chronological feeds or content organized by specific topics or sources selected by the user.

Benefits

Research demonstrates the urgent need to restrict algorithmic content feeds for young users. Recent Gallup studies show that teens now spend an average of 4.8 hours daily on social media, with concerning mental health implications. Among teens with the highest social media use, 41% rate their mental health as poor or very poor, compared to 23% of those with the lowest use[34][35][36].

From a neurodevelopmental perspective, research findings highlight that normal processes of brain development in mid-adolescence may heighten vulnerability to exaggerated emotional responses to platform algorithmic practices. A Wall Street Journal investigation revealed that platforms' algorithmic feeds could rapidly expose young users to harmful content. For instance, TikTok's "For You" page fed teens tens of thousands of weight-loss videos within just weeks of joining the platform, including dangerous content promoting extreme dieting[37]. The Center for Countering Digital Hate found that vulnerable teen accounts were served twelve times more self-harm and suicide-related content than standard accounts within just the first thirty minutes of platform use[38].

Beyond mental health protection, these restrictions also provide data privacy benefits. By limiting platforms' ability to collect and analyze user behavior data for content recommendation, the restrictions help prevent the extensive tracking and profiling of minors' online activities. This addresses what New York State Attorney General Letitia James identifies as the "grave risk" of having children's location and personal data tracked, shared, and potentially accessed by malicious actors[39].

While the evidence strongly suggests the need for algorithmic feed restrictions, it is important to acknowledge current research limitations. Although recent short-term experiments have shown promising results, most existing research has focused on documenting the harms of algorithmic content delivery rather than evaluating the impact of specific interventions to limit it. Long-term studies are needed to assess whether and how restrictions on addictive feeds improve youth mental health outcomes and online experiences.

Limitations

Implementation challenges pose significant barriers to the effectiveness of addictive feed restrictions. While SB359 defines an "addictive feed" as content that is "recommended, selected, or prioritized for display to a user based on information associated with the user or the user's device," the distinction between different types of content delivery mechanisms remains complex, especially when platforms offer multiple features and functionalities.

Age verification presents another significant challenge. Although SB359 requires platforms to use "commercially reasonable methods" to determine if users are minors, research shows that existing age verification methods can be easily circumvented. According to a Brookings report, current approaches to age verification vary significantly across platforms and jurisdictions, creating inconsistencies in implementation and enforcement[8]. The report highlights how some states mandate third-party verification while others leave the methods largely undefined, leading to potential gaps in protection.

Economic considerations create additional barriers to effective implementation. Research reveals that social media platforms earn substantial advertising revenue from users aged 0-17 ($11 billion annually), creating strong financial disincentives for platforms to restrict addictive feeds fully. Features like algorithmic content recommendation and targeted advertising are central to their revenue models[42][43]. This economic reality suggests that without strong enforcement mechanisms and penalties, platforms may prioritize revenue over protection, potentially implementing superficial restrictions that fail to address the core issues of algorithmic content delivery to minors[42].

These limitations suggest that while addictive feed restrictions represent an important step toward protecting young users, their effectiveness depends heavily on robust implementation frameworks, reliable age verification systems, and strong enforcement mechanisms that can overcome platforms' economic incentives to maintain engagement-driven features.

Implementation Recommendations

To strengthen the effectiveness of algorithmic content restrictions, legislation should mandate regular algorithm risk audits conducted by independent third parties. These audits should evaluate how platforms' recommendation systems affect young users' mental health and well-being, with particular attention to vulnerable populations. The results should be publicly disclosed to ensure transparency and enable evidence-based policy adjustments. This aligns with SB359's goal of protecting minors from potentially harmful algorithmic content delivery while providing concrete mechanisms for oversight.

To enhance the effectiveness of current restrictions while maintaining platform functionality, legislation should establish clear standards for alternative content delivery methods. This includes requiring platforms to offer chronological feeds as the default option for minor users, implementing transparent content curation criteria, and developing age-appropriate recommendation systems that prioritize educational and positive content. These alternatives should be designed to maintain user engagement while avoiding the potentially harmful effects of current algorithmic practices, as identified in research showing how algorithmic feeds can rapidly expose young users to dangerous content.

To improve compliance monitoring and enforcement, platforms should be required to maintain detailed documentation of their algorithmic practices and their impacts on minor users. This documentation should include regular reporting on key metrics such as content exposure patterns, user engagement data, and mental health indicators. Additionally, platforms should be required to establish clear mechanisms for researchers to access anonymized data about algorithmic content delivery to minors, enabling ongoing independent assessment of these systems' effects on youth well-being. This approach would help ensure that restrictions on addictive feeds achieve their intended protective goals while providing data needed to refine and improve these protections over time.

3. Beyond Platform Regulation: Supporting Digital Literacy and Family Relationships

Unlike restricting access to drugs or explicit content, protecting minors' privacy, safety, and mental health on social media requires a more nuanced approach. As noted in the Surgeon General's Advisory, while social media poses risks to youth mental health, it also provides crucial educational, social, and psychological support resources. For example, LGBTQ youth often find vital emotional support and community connections through social media that may not be available in their immediate environment. Studies also indicate that strong family relationships and supportive home environments significantly reduce the likelihood of youth experiencing mental health issues when using social media[35][36]. The goal should not be to create a generation of youth who avoid social media entirely but rather to develop digitally literate young people who can identify risks, resist addiction, and leverage these platforms for positive social interaction, knowledge acquisition, and entertainment[14].

Research indicates that restrictive approaches alone may be counterproductive. Studies show that youth with less restrictive parents tend to use the Internet for a broader range of informational and creative activities, while those with more restrictive parents lean toward entertainment-only activities[15]. Instead of focusing solely on restrictions, legislation should promote:

- Digital literacy education programs for both youth and parents
- Resources for schools and platforms to collaborate on educational initiatives
- Support for family-based approaches to digital wellness
- Programs that strengthen parent-child relationships and communication about online activities

Rather than creating digital barriers, legislation should support building digital bridges—equipping young people with the skills to navigate online spaces safely and meaningfully. The most effective protection comes not from external controls but from developing internal wisdom

and resilience. By fostering open dialogue between parents and children, providing comprehensive digital literacy education, and supporting positive online engagement, we can help young people harness the benefits of social media while managing its risks.

4. Conclusion

This report has examined several key protective mechanisms for youth social media safety, including verifiable parental consent, time-based restrictions, cookie opt-out controls, and restrictions on addictive feeds. While these tools show promise in protecting young users, significant work remains to be done in two critical areas of research and development.

First, further legal analysis is needed to examine potential conflicts between these protective measures and both First Amendment protections and Section 230 immunity. While California's approach to regulating content delivery methods rather than the content itself offers a promising direction, the constitutionality and practical implementation of various protective mechanisms require careful consideration.

Second, technical challenges remain in developing accurate, robust age verification systems that are both difficult for youth to circumvent and protective of user privacy. The ideal solution would be decentralized, preserving user anonymity while ensuring reliable verification—a complex technical challenge that deserves dedicated research and development efforts.

Protecting youth online safety requires coordinated efforts from families, schools, platforms, and government entities working together to establish appropriate frameworks. Given the inherently interstate nature of internet usage, state legislation must consider alignment with federal youth privacy and protection laws to ensure consistent and effective protection across jurisdictions.

As we navigate this complex digital landscape, our ultimate goal must extend beyond mere restrictions to fostering digital resilience. Success should be measured not just by the effectiveness of protective mechanisms, but by our ability to nurture a generation of confident, capable digital citizens who can harness social media's benefits while managing its risks. This balanced approach—combining technical safeguards with youth empowerment—offers the most promising path toward ensuring both the safety and digital flourishing of our youth.

References

[1] Murthy, V. H. (2023). Social media and youth mental health: The U.S. Surgeon General's Advisory. https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf

[2] Murthy, V. H. (2024, June 17). Opinion | Surgeon General: Why I'm calling for a warning label on social media platforms. *The New York Times*. https://www.nytimes.com/2024/06/17/opinion/social-media-health-warning.html

[3] Child Online Protection Act of 1998, Pub. L. No. 105-277, 112 Stat. 2681-736 (1998), invalidated by Ashcroft v. ACLU, 542 U.S. 656 (2004).

[4] Children's Online Privacy Protection Act of 1998, Pub. L. No. 105-277, 112 Stat. 2681 (1998) (codified at 15 U.S.C. §§ 6501-6506).

[5] Kids Online Safety Act, S. 1409, 118th Cong. (reported to Senate, Dec. 13, 2023).

[6] Children and Teens' Online Privacy Protection Act, S. 1418, 118th Cong. (reported to Senate, Dec. 13, 2023).

[7] Purdy, E. R. (2024, July 5). Child Online Protection Act of 1998 (1998). The Free Speech Center. https://firstamendment.mtsu.edu/article/child-online-protection-act-of-1998/

[8] Jang, K., Pan, L., & Lee, N. T. (2023, August 14). The fragmentation of online child safety regulations. Brookings. https://www.brookings.edu/articles/patchwork-protection-of-minors/

[9] Attorney General James calls on Congress to require social media warning labels. (2024, September 10). New York State Office of the Attorney General. https://ag.ny.gov/press-release/2024/attorney-general-james-calls-congress-require-social-media-warning-labels

[10] Salahi, L. (2024, August 30). Warning labels could help regulate social media. But will it make us healthier? Association of Health Care Journalists. https://healthjournalism.org/blog/2024/08/warning-labels-could-help-regulate-social-media-but-will-it-make-us-healthier/

[11] Nassetta, J., & Gross, K. (2020). State media warning labels can counteract the effects of foreign misinformation. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-45

[12] MIT Sloan Office of Communications. (2024, September 2). Warning labels from fact checkers work — even if you don't trust them. MIT Sloan. https://mitsloan.mit.edu/press/warning-labels-fact-checkers-work-even-if-you-dont-trust-them

[13] Robinson, L. A., Viscusi W. K., & Zeckhauser R. (2016, November 30). Consumer warning labels aren't working. *Harvard Business Review*. https://hbr.org/2016/11/consumer-warning-labels-arent-working

[14] Rutledge, P. B. (2024, June 24). Why warning labels on social media miss the mark. Psychology Today. https://www.psychologytoday.com/us/blog/positively-media/202406/why-warning-labels-on-social-media-miss-the-mark

[15] Luria, M., & Bhatia, A. (2024, May 15). Opinion: Restricting and monitoring social media won't protect kids — here's what will. CNN. https://www.cnn.com/2024/05/15/opinions/social-media-monitoring-restriction-legislation-mediation-luria-bhatia/index.html

[16] Lee, L. T., Jenna Zhang, Diana. (2024, August 9). State and federal developments in minors' privacy in 2024. Inside Privacy. https://www.insideprivacy.com/childrens-privacy/state-and-federal-developments-in-minors-privacy-in-2024/

[17] Federal Trade Commission. (2013, June 30). Children's Online Privacy Protection Rule: A six-step compliance plan for your business [Guidance]. https://www.ftc.gov/business-guidance/resources/childrens-online-privacy-protection-rule-six-step-compliance-plan-your-business

[18] Anderson, M., & Faverio, M. (2023, October 31). 81% of U.S. adults – versus 46% of teens – favor parental consent for minors to use social media. Pew Research Center. https://www.pewresearch.org/short-reads/2023/10/31/81-of-us-adults-versus-46-of-teens-favor-parental-consent-for-minors-to-use-social-media/

[19] Craig, S. (2023, June 29). Social media can hurt kids. For many 2SLGBTQ+ youth, it's a lifeline. Maclean's. https://macleans.ca/society/social-media-can-hurt-kids-for-many-2slgbtq-youth-its-a-lifeline/
[20] Future of Privacy Forum. (2023, June). Verifiable parental consent: The state of play. https://fpf.org/verifiable-parental-consent-the-state-of-play/
[21] S.B. 532, 2024 Sess. (Va. 2024)
[22] S.B. 976, Ch. 321, 2024 Sess. (Cal. 2024)
[23] Online Safety Act 2023, c. 50 (UK)
[24] van den Eijnden, R. J. J. M., Geurts, S. M., ter Bogt, T. F. M., van der Rijst, V. G., & Koning, I. M. (2021). Social media use and adolescents' sleep: A longitudinal study on the protective role of parental rules regarding Internet use before sleep. *International Journal of Environmental Research and Public Health*, *18*(3), 1346. https://doi.org/10.3390/ijerph18031346
[25] Yu, D. J., Wing, Y. K., Li, T. M. H., & Chan, N. Y. (2024). The impact of social media use on sleep and mental health in youth: A scoping review. *Current Psychiatry Reports*, *26*(3), 104–119. https://doi.org/10.1007/s11920-024-01481-9
[26] Perrault, A. A., Bayer, L., Peuvrier, M., Afyouni, A., Ghisletta, P., Brockmann, C., Spiridon, M., Hulo Vesely, S., Haller, D. M., Pichon, S., Perrig, S., Schwartz, S., & Sterpenich, V. (2019). Reducing the use of screen electronic devices in the evening is associated with improved sleep and daytime vigilance in adolescents. *Sleep*, 42(9), zsz125. https://doi.org/10.1093/sleep/zsz125
[27] Schneiders, P., & Gilbert, A. (2024, March 13). Banning children's social media use: A wave of symbolic regulations, but at what cost? Internet Policy Review. https://policyreview.info/articles/news/banning-childrens-social-media-use/1744
[28] S.B. 252, 2024 Sess. (Va. 2024)
[29] California Consumer Privacy Act of 2018, Cal. Civ. Code § 1798.100 et seq.
[30] Habib, H., Li, M., Young, E., & Cranor, L. (2022). "Okay, whatever": An evaluation of cookie consent interfaces. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–27. https://doi.org/10.1145/3491102.3501985
[31] Kulyk, O., Gerber, N., Hilt, A., & Volkamer, M. (2020). Has the GDPR hype affected users' reaction to cookie disclaimers? *Journal of Cybersecurity*, *6*(1), tyaa022. https://doi.org/10.1093/cybsec/tyaa022
[32] Secure Privacy. (2024, August 22). What it means for the future of digital advertising. https://secureprivacy.ai/blog/googles-third-party-cookie-deprecation
[33] S.7694A, 2023-2024 Leg. Sess. (N.Y. 2023)
[34] DeAngelis, T. (2024, April 1). Teens are spending nearly 5 hours daily on social media. Here are the mental health outcomes. *Monitor on Psychology*, *55*(3), 80. https://www.apa.org/monitor/2024/04/teen-social-use-mental-health
[35] Rothwell, J. (October 27, 2023). Parenting mitigates social media-linked mental health issues. Gallup. https://news.gallup.com/poll/513248/parenting-mitigates-social-media-linked-mental-health-issues.aspx.
[36] Rothwell, J. (2023). How parenting and self-control mediate the link between social media use and mental health. https://ifstudies.org/ifs-admin/resources/briefs/ifs-gallup-parentingsocialmediascreentime-october2023-1.pdf.
[37] Hobbs, T. D., Barry, R., & Koh, Y. (2021, December 17). 'The corpse bride diet': How TikTok inundates teens with eating-disorder videos. *Wall Street Journal*. https://www.wsj.com/articles/how-tiktok-inundates-teens-with-eating-disorder-videos-11639754848
[38] Lavelle, T. (2022, December 15). *TikTok bombards teens with self harm and eating disorder content within minutes of joining the platform*. Center for Countering Digital Hate. https://counterhate.com/blog/tiktok-bombards-teens-with-self-harm-and-eating-disorder-content-within-minutes-of-joining-the-platform/
[39] Attorney General James champions legislation to protect kids from addictive social media feeds in national USA Today op-ed [Press release]. (2024, March 14). New York State Office of the Attorney

General. https://ag.ny.gov/press-release/2024/attorney-general-james-champions-legislation-protect-kids-addictive-social-media
[40] Dolan, E. W. (2024, May 28). Experiment finds limiting social media use can reduce mental health issues in distressed youth. PsyPost. https://www.psypost.org/experiment-finds-limiting-social-media-use-can-reduce-mental-health-issues-in-distressed-youth/
[41] Berger, M. N., Taba, M., Marino, J. L., Lim, M. S. C., Cooper, S. C., Lewis, L., Albury, K., Chung, K. S. K., Bateson, D., & Skinner, S. R. (2021). Social media's role in support networks among LGBTQ adolescents: A qualitative study. *Sexual Health*, *18*(5), 421–431. https://doi.org/10.1071/SH21110
[42] Costello, N., Sutton, R., Jones, M., Almassian, M., Raffoul, A., Ojumu, O., Salvia, M., Santoso, M., Kavanaugh, J. R., & Austin, S. B. (2023). Algorithms, addiction, and adolescent mental health: An interdisciplinary study to inform state-level policy action to protect youth from the dangers of social media. *American Journal of Law & Medicine*, *49*(2–3), 135–172. https://doi.org/10.1017/amj.2023.25
[43] Raffoul, A., Ward, Z. J., Santoso, M., Kavanaugh, J. R., & Austin, S. B. (2023). Social media platforms generate billions of dollars in revenue from U.S. youth: Findings from a simulated revenue model. *PLOS ONE*, *18*(12), e0295337. https://doi.org/10.1371/journal.pone.0295337
[44] Kelley, J. (2023, May 12). The law should not require parental consent for all minors to access social media. Electronic Frontier Foundation. https://www.eff.org/deeplinks/2023/05/law-should-not-require-parental-consent-all-minors-access-social-media
[45] McAlister, K. L., Beatty, C. C., Smith-Caswell, J. E., Yourell, J. L., & Huberty, J. L. (2024). Social media use in adolescents: Bans, benefits, and emotion regulation behaviors. *JMIR Mental Health*, *11*(1), e64626. https://doi.org/10.2196/64626