American Journal of Law & Medicine (2023), 49: 2-3, 135–172
doi:10.1017/amj.2023.25

American Society of Law, Medicine & Ethics
Where Multidisciplinary Education Improves Practice

ARTICLE

ALGORITHMS, ADDICTION, AND ADOLESCENT MENTAL HEALTH: An Interdisciplinary Study to Inform State-level Policy Action to Protect Youth from the Dangers of Social Media

Nancy Costello[1], Rebecca Sutton[1], Madeline Jones[1], Mackenzie Almassian[1], Amanda Raffoul[2], Oluwadunni Ojumu[3], Meg Salvia[4], Monique Santoso[5], Jill R. Kavanaugh[4] and S. Bryn Austin[4]

[1]Michigan State University, East Lansing, MI, USA
[2]Harvard Medical School, Cambridge, MA, USA
[3]Harvard College, Cambridge, MA, USA
[4]Harvard T.H. Chan School of Public Health, Cambridge, MA, USA
[5]Stanford University, Stanford, CA, USA
Corresponding author: Nancy Costello; Email: costel29@msu.edu

Abstract

A recent Wall Street Journal investigation revealed that TikTok floods child and adolescent users with videos of rapid weight loss methods, including tips on how to consume less than 300 calories a day and promoting a "corpse bride diet," showing emaciated girls with protruding bones. The investigation involved the creation of a dozen automated accounts registered as 13-year-olds and revealed that TikTok algorithms fed adolescents tens of thousands of weight-loss videos within just a few weeks of joining the platform. Emerging research indicates that these practices extend well beyond TikTok to other social media platforms that engage millions of U.S. youth on a daily basis.

Social media algorithms that push extreme content to vulnerable youth are linked to an increase in mental health problems for adolescents, including poor body image, eating disorders, and suicidality. Policy measures must be taken to curb this harmful practice. The Strategic Training Initiative for the Prevention of Eating Disorders (STRIPED), a research program based at the Harvard T.H. Chan School of Public Health and Boston Children's Hospital, has assembled a diverse team of scholars, including experts in public health, neuroscience, health economics, and law with specialization in First Amendment law, to study the harmful effects of social media algorithms, identify the economic incentives that drive social media companies to use them, and develop strategies that can be pursued to regulate social media platforms' use of algorithms. For our study, we have examined a critical mass of public health and neuroscience research demonstrating mental health harms to youth. We have conducted a groundbreaking economic study showing nearly $11 billion in advertising revenue is generated annually by social media platforms through advertisements targeted at users 0 to 17 years old, thus incentivizing platforms to continue their harmful practices. We have also examined legal strategies to address the regulation of social media platforms by conducting reviews of federal and state legal precedent and consulting with stakeholders in business regulation, technology, and federal and state government.

While nationally the issue is being scrutinized by Congress and the Federal Trade Commission, quicker and more effective legal strategies that would survive constitutional scrutiny may be implemented by states, such as the Age Appropriate Design Code Act recently adopted in California, which sets standards that online services likely to be accessed by children must follow. Another avenue for regulation may be through states mandating that social media platforms submit to algorithm risk audits conducted by independent third parties and publicly disclose the results. Furthermore, Section 230 of the federal Communications Decency Act, which has long shielded social media platforms from liability for wrongful

acts, may be circumvented if it is proven that social media companies share advertising revenues with content providers posting illegal or harmful content.

Our research team's public health and economic findings combined with our legal analysis and resulting recommendations, provide innovative and viable policy actions that state lawmakers and attorneys general can take to protect youth from the harms of dangerous social media algorithms.

Keywords: social media; algorithms; mental health; consumer advocacy; freedom of speech; legislation as topic; child; adolescent

Introduction

In 2013, eleven-year-old Alexis Spence, a fifth grader, joined Instagram after her classmates made fun of her for not having a social media account.[1] She was two years under the platform's minimum age requirement to open an account, but other user content showed her how to obtain a parent's passcode to disable parental blocks to the social media platform.[2] On her tablet,[3] she made her Instagram app icon look like a calculator to hide it from her parents.[4] After joining the app, Alexis was confronted with algorithm-driven content portraying underweight models and links to extreme dieting websites that glorified anorexia nervosa, negative body image, and self-harm.[5] When she was twelve years old, Alexis drew a picture of herself crying on the floor next to her phone with words like "stupid," "ugly," and "fat" emanating from the screen, and "kill yourself" in a thought bubble.[6] She saved pictures of anorexic models as "motivation" to look at whenever she felt hungry.[7] Months after opening the Instagram account, Alexis started showing signs of depression and her parents sought mental health treatment, but she refused to continue to see a therapist after a handful of initial sessions.[8] In Instagram posts she shared in spring 2018, Alexis wrote: "I hate myself and my body….Please stop caring about me, I'm a waste of time and space."[9] Alerted by school counselors to the posts, Alexis's parents had her hospitalized.[10] Alexis was suffering from an eating disorder, anxiety, depression, and suicidal thoughts.[11]

As a result of Alexis's exposure to Instagram's toxic algorithm practices, she underwent years of professional counseling through in-patient and out-patient programs, participated in eating disorder treatment services, used a service dog, and required ongoing medical attention to ensure she did not relapse.[12] In June 2022, at the age of nineteen, the Social Media Victims Law Center filed a personal injury lawsuit on behalf of Alexis in California federal court alleging that Meta Platforms, Inc. (Meta), Instagram's parent company, purposely designed its social media platform to addict young users, and that Meta steered her down a years-long path of physical and psychological harm.[13]

Social media algorithms that push extreme content to vulnerable youth are linked to a pronounced increase in mental health problems for adolescents, including poor body image, eating disorders, and suicidality. A 2021 Wall Street Journal investigation revealed that TikTok floods child and adolescent users with videos of rapid weight loss methods, including tips on how to consume less than 300 calories a

[1]Complaint at para. 166, Spence v. Meta Platforms, Inc., (N.D. Cal. 2022) (No. 22CV03294), 2022 WL 2101825, at *135 [hereinafter Spence Complaint].
[2]*Id.* at para. 171.
[3]Alexis downloaded Instagram to an electronic tablet first and then later, in 2014, to her cell phone. *Id.* at para. 171, 189(g).
[4]*Id.* at para. 171.
[5]*Id.* at para. 192–195.
[6]*Id.* at para. 187.
[7]*Id.* at para. 153.
[8]*Id.* at para. 170.
[9]*Id.* at para. 204, 207.
[10]*Id.* at para. 204–205.
[11]*Id.* at para. 204–206.
[12]*Id.* at para. 216.
[13]*Id.* at para. 36-41.

day, and encourages a "corpse bride diet," showing emaciated girls with protruding bones.[14] The journalistic investigation involved the creation of a dozen automated accounts registered as thirteen-year-olds and revealed that TikTok's algorithm-driven "For You" page, a section of the platform that algorithmically recommends content to users, fed adolescent accounts tens of thousands of weight-loss videos within just a few weeks of joining the platform.[15]

Another report revealed the scale and intensity with which TikTok bombards vulnerable teen users with dangerous content that might encourage self-harm, suicide, and disordered eating. In 2022, researchers from the Center for Countering Digital Hate studied the TikTok algorithm by establishing new social media accounts posing as thirteen-year-old girls in the United States, United Kingdom, Australia, and Canada.[16] Researchers recorded the first thirty minutes of content automatically recommended by TikTok to these accounts in their "For You" page.[17] The study revealed that the volume of harmful content shown to vulnerable accounts (i.e., with the term "loseweight" in their username) was significantly higher than that shown to standard accounts.[18] For instance, vulnerable accounts were served twelve times more self-harm and suicide videos than standard accounts.[19]

Social media companies employ algorithms for a variety of reasons, with the primary purpose of keeping users engaged with constant feeds of information for extended periods of time; such engagement results in massive profits for the companies paid by advertisers targeting ads at a certain demographic.[20] A recent study by our research team, discussed later in this Article, found that in 2022, major social media platforms earned nearly $11 billion in advertising revenues from U.S. children ages zero to seventeen years.[21] Given these handsome profits, social media platforms have little incentive to moderate their own harmful practices. Policymakers must instead step forward and make changes to curb the harmful use of algorithms by social media platforms. Legal obstacles, however, stand in the way of such reform.

Social media platforms currently enjoy substantial First Amendment speech protection and are relatively insulated from culpability by the liability shield extended to website owners and operators under Section 230 of the Communications Decency Act (CDA).[22] Alexis Spence's lawsuit is part of a recent wave of litigation that attempts to circumvent Section 230 and other free speech protections, but the lawsuit is designed to help only one person, with relief made possible only after harm has been inflicted. Further, while studies have shown an association between exposure to social media algorithms and increased mental health harms in young users (detailed below), the task of demonstrating that social media has directly caused such harms has been difficult because platforms do not allow external researchers access to their algorithms. Stronger evidence of causation is needed to demonstrate that social media platforms are liable for harm.

This Article advocates for state and federal legislation requiring social media companies to conduct periodic algorithm risk audits that measure the incidence of harm inflicted on young users. Such risk audits should be conducted by independent third parties, and the results should be publicly disclosed. This policy measure is urgently needed to curb social media companies' pernicious use of relentless algorithms and to protect the millions of young users who are vulnerable to their harms.

---

[14]Tawnell D. Hobbs et al., *'The Corpse Bride Diet': How TikTok Inundates Teens with Eating-Disorder Videos*, Wall St. J. (Dec. 17, 2021, 10:45 AM), https://www.wsj.com/articles/how-tiktok-inundates-teens-with-eating-disorder-videos-11639754848 [https://perma.cc/9RAH-NXB5].

[15]*Id.*

[16]Press Release, *TikTok bombards teens with self harm and eating disorder content within minutes of joining the platform*, Center for Countering Digital Hate (Dec. 15, 2022), https://counterhate.com/blog/tiktok-bombards-teens-with-self-harm-and-eating-disorder-content-within-minutes-of-joining-the-platform/ [https://perma.cc/9RNP-PGYN0].

[17]*Id.*

[18]*Id.*

[19]*Id.*

[20]Spence Complaint at para. 33-41.

[21]Amanda Raffoul et al., *Estimated Social Media Platform Revenue from U.S. Children* [in preparation] (2023).

[22]Communications Decency Act, 47 U.S.C. §230 (2018).

The first section of this article examines the federal Children's Online Privacy Protection Act (COPPA), the age restriction for young users of social media, and the failure of platforms to verify the age of users. The second section discusses the results of public health and neuroscience studies that demonstrate evidence of mental health harms to adolescents resulting from social media use. This section also discusses how this evidence could help establish causation needed to prove unfair and deceptive business practice claims and products liability claims. The third section presents results of a new study that shows social media platforms are economically incentivized to keep young users actively engaged on their platforms. The fourth section discusses the legal obstacles to preventing harm to young people caused by social media algorithms and Supreme Court cases where the Court refused to rein in the blanket immunity currently granted to social media platforms under Section 230 of the CDA. The fifth section explores the strategies to circumvent First Amendment protection and immunity granted by Section 230 of the CDA, including bringing an unfair or deceptive business practice claim against social media platforms, filing a products liability lawsuit and bringing claims under public nuisance tort theory. The sixth section discusses recent state-level legislation, including the California Age Appropriate Design Code that is a promising step in addressing harms to young people, but faces a court challenge, and California's Data Protection Impact Assessment law and Utah's Social Media Amendments law, which are less effective because they rely on social media companies to evaluate themselves. It also discusses the Kids Online Safety Act, introduced in Congress in 2023, which would give parents and users under seventeen the ability to opt out of algorithmic recommendations, limit the time young people spend on a platform, and require platforms to do risk assessments conducted by independent third parties, but is uncertain whether the bill will become law. The final section advocates for states to require social media companies to conduct algorithm risk audits that would provide evidence for legal actions seeking to reform the harmful practices of social media platforms.

I. A Growing Number of Young People Have Easy Access to Social Media Platforms and Its Resulting Harms and Current Federal Law Restricting Young Users Is Ineffectual

Ease of access is a foundational issue in understanding how social media platforms affect adolescent mental health. Social media has become increasingly popular over the past two decades. The beginning of popular social media as we know it arguably dates to 2004, when MySpace became the first social media platform to reach one million monthly active users.[23] Throughout the next decade, social media became an integral part of many lives—especially those of adolescents. Popular platforms began to spring up, notably Facebook in 2004, Twitter in 2006, and Instagram in 2010.[24] The popularity of social media grew alongside the number of available platforms.. When TikTok launched in 2016, social media was so popular that the platform gained half a billion users worldwide in less than two years.[25] TikTok, a platform where young users create and share short videos often showing themselves singing, dancing, doing comedy, and lip-syncing, on average added twenty million new users each month during its first two years.[26] Large portions of social media memberships are populated by people under the age of eighteen.[27] And exposing minors to the harmful content on these platforms and the addictive design of these platforms has produced a generation plagued by the constant need to be online— only to be confronted by content that can be harmful to their mental health.

[23]Esteban Ortiz-Ospina, *The Rise of Social Media*, Our World in Data (Sept. 18, 2019), https://ourworldindata.org/rise-of-social-media [https://perma.cc/5PS5-9J29].

[24]Maryville University, *The Evolution of Social Media: How Did It Begin, and Where Could It Go Next?*, https://online.maryville.edu/blog/evolution-social-media/#:~:text=In%201987%2C%20the%20direct%20precursor,social%20media%20platform%20was%20launched [https://perma.cc/HQ9N-B6J9] (last visited July 8, 2022).

[25]Ortiz-Ospina, *supra* note 23.

[26]*Id.*

[27]Emily Vogels, et al., *Teens, Social Media and Technology 2022*, Pew Research Center (August 2022), https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022 [https://perma.cc/243K-8VY6].

Tammy Rodriguez is among the increasing number of parents who understand the toll social media platforms have on children's mental health. On January 20, 2022, she filed a wrongful death lawsuit against Meta on behalf of her eleven-year-old daughter, Selena, who took her own life as a result of being severely addicted to social media.[28] Due to Selena's use of Instagram and Snapchat, she was hospitalized for emergency psychiatric care and experienced "worsening depression, poor self-esteem, eating disorders, self-harm, and, ultimately, suicide."[29] In her complaint, Ms. Rodriguez claims that, due to the lack of parental controls on Instagram and Snapchat, the only way to effectively limit Selena's access to social media was to physically confiscate her phone, which caused her to run away to access her social media accounts on other devices.[30] Selena was solicited several times for sexually exploitative content and she once sent sexually explicit images, which were leaked to her classmates.[31] Ms. Rodriguez claims that Meta knew or should have known that its platform was harmful to a significant percentage of its minor users and still failed to redesign its products to ameliorate these harms.[32]

This is an ongoing lawsuit, and, unfortunately, one of many. Selena was exposed to social media platforms at a very young age and suffered severely because of it. Theoretically, Selena should have been protected online by COPPA, but loopholes in the law wrongly expose many children to the addictive design of and harmful content on social media platforms.

A. The Federal Children's Online Privacy Protection Act Does Not Adequately Protect Children Against the Mental Health and Addictive Harms of Social Media

Congress enacted COPPA in 1998 with the primary goal of placing parents in control of the information collected from their young children online.[33] COPPA prohibits social media platforms from collecting, using, or disclosing the personal information of children under the age of thirteen years without verifiable parental consent.[34] COPPA defines personal information as the child's first and last name; physical address; online contact information, including username; telephone number; social security number; persistent identifiers, such as IP address; photograph, video, or audio file that contain the child's image or voice; and geolocation.[35] COPPA applies to a social media platform where the platform either (1) is directed to children under thirteen or (2) has actual knowledge that they are collecting, using, or disclosing the personal information of someone under thirteen years old.[36]

As a result of the age restriction contained in COPPA, a vast majority of social media platforms require users to be at least thirteen years old to open an account.[37] These same platforms insist that

[28]Complaint at 11, Rodriguez v. Meta Platforms, Inc., No. 3:22-cv-00401 (N.D. Cal. Jan. 20, 2022), 2022 WL 190807, at *1 [hereinafter Rodriguez Complaint].

[29]*Id.* at 61.

[30]*Id.* at 54.

[31]*Id.* at 60..

[32]*Id.* at 7.

[33]*Complying with COPPA: Frequently Asked Questions*, Fed. Trade. Comm'n (July, 2020), https://www.ftc.gov/business-guidance/resources/complying-coppa-frequently-asked-questions [https://perma.cc/2UER-MSPU].

[34]15 U.S.C. § 6502(a)(1).

[35]15 U.S.C. §6501(8).

[36]15 U.S.C. § 6502(a)(1).

[37]Per the social media platforms terms of use, Facebook and Instagram, under the Meta umbrella, both require users to be at least thirteen years old to sign up for an account. *Terms of Service*, Facebook, https://www.facebook.com/terms.php [https://perma.cc/B9F5-J4PU] (last visited Apr. 17, 2023); *Terms of Use*, Instagram, https://help.instagram.com/581066165581870 [https://perma.cc/6DX3-BZMD] (last visited Apr. 17, 2023). To sign up for a Snapchat account or Twitter account, both platforms also require a user must be at least thirteen years old. *Snap Inc. Terms of Service*, Snap Inc. (Nov. 15, 2021), https://snap.com/en-US/terms [https://perma.cc/6NNQ-6VUD]; *Terms of Service*, Twitter (June 10, 2022), https://twitter.com/en/tos [https://perma.cc/H54J-UG5P]. TikTok requires a new user to pass through an age gate to guide that user into the right TikTok experience. *TikTok for Younger Users*, TikTok (Dec. 13, 2019), https://newsroom.tiktok.com/en-us/tiktok-for-younger-users [https://perma.cc/XHM3-WWKE] TikTok, in collaboration with the Digital Wellness Lab at Boston Children's

COPPA regulations do not apply to them, because the platforms are not directed at children under the age of thirteen.[38] These platforms' age-minimum "workarounds" result in mental health harm to adolescents who use them, because many do not enforce age verification, allowing young users to easily misrepresent their age to gain access to platforms.[39] Young users are thus left vulnerable to not only the harmful content present on the platforms but also to the exploitative business practices that manipulate people to stay on the platforms longer such as infinite scroll of content, encouraging posting content to obtain "likes," etc.[40] By failing to establish effective ways to verify users' ages, social media companies ultimately enable minors under the age of thirteen to set up accounts without verifiable parental consent—and place themselves squarely in direct violation of COPPA. These platforms' failures to verify user age circumvent COPPA's very purpose, which is to protect against the collection, use, and disclosure of the personal information of minors under the age of thirteen.

Many social media platforms also fail to comply with the advertising rules that COPPA sets forth. COPPA prohibits social media platforms from using behavioral or demographic advertising, due to the ban on collection of personal information from users under the age of thirteen absent verifiable parental consent.[41] (Behavioral advertising is curated based on the web-browsing behavior of the user, while demographic advertising is curated based on the personal demographic information of the user.[42]) Therefore, when adhering to COPPA regulations, social media platforms must deliver advertising on only a *contextual* basis—placing ads on webpages based on the context of those webpages—to those under thirteen.[43]

However, when users misrepresent their age to open an account, social media platforms that rely on that inaccurate data are essentially allowed to disregard COPPA's advertising restrictions, and instead expose their young users to behavioral and demographic advertising as well as contextual advertising. This issue is further compounded by the fact that many social media companies disregard COPPA because they blithely claim their platforms are not targeted to children;[44] some platforms do not even attempt to detect underage users who join the site with a falsified birthdate.

Hospital, introduced a 60-minute daily time limit for United States users between the ages thirteen to seventeen in March 2023. If the screen limit is reached, teen users are prompted to enter a passcode to extend their screen time on the app; however, this screen limit feature can be disabled entirely or continuously extended. If a user is under thirteen years old in the United States, they will be placed into the TikTok for Younger Users experience, which has additional privacy and safety protections designed specifically for this audience. For younger users, a 60-minute daily screen limit is applied but requires a parent or guardian to enter the passcode to enable an additional 30 minutes of watch time. Again, this additional screen time can be continuously extended. *See* Cormac Keenan, *New Features for Teens and Families on TikTok*, TikTok (Mar. 1, 2023), https://newsroom.tiktok.com/en-us/new-features-for-teens-and-families-on-tiktok-us [https://perma.cc/69VT-KXJM]. Other countries, like China, have stricter time restrictions for teen users.

[38] *See e.g.,* Joseph Marks, *App Makers Are Scooping Up Kids' Data With Few Real Checks*, Washington Post (June 9, 2022, 8:12 AM), https://www.washingtonpost.com/politics/2022/06/09/app-makers-are-scooping-up-kids-data-with-few-real-checks/ [https://perma.cc/Y4AF-TVRA]; Geoffrey A. Fowler, *Your Kids' Apps Are Spying on Them*, Washington Post (June 9, 2022, 8:00 AM), https://www.washingtonpost.com/technology/2022/06/09/apps-kids-privacy// [https://perma.cc/983Q-WVVA].

[39] *See* Jackie Snow, *Why Age Verification Is So Difficult for Websites*, Wall Street Journal (Feb. 27, 2022, 8:00 AM), https://www.wsj.com/articles/why-age-verification-is-difficult-for-websites-11645829728 [https://perma.cc/RE9R-AZM8].

[40] *See* Kaitlin Woolley & Marissa A. Sharif, *The Psychology of Your Scrolling Addiction*, Harv. Bus. Rev. Jan. 31, 2022, https://hbr.org/2022/01/the-psychology-of-your-scrolling-addiction [https://perma.cc/6GLB-PZTM].

[41] For example, on a webpage with a news article about running, behavioral advertisements would be based on the user's web-history. Perhaps the user has been frequently reading articles about how to lose weight and now receives a targeted ad on the article about running, detailing how many miles a day they need to run to lose a certain amount of weight. Under COPPA, this kind of targeted advertising is not allowed for children under thirteen without verifiable parental consent.

[42] *See* Jonathan Mayer, *Do Not Track Is No Threat To Ad-Supported Businesses*, The Center for Internet and Society: Blog (Jan 20, 2011, 2:12 AM), https://cyberlaw.stanford.edu/blog/2011/01/do-not-track-no-threat-ad-supported-businesses [https://perma.cc/3P3Z-5LAX].

[43] For example, in a news article about running, a contextual advertisement could be an ad for running shoes.

[44] Marks, *supra* note 38.

1. Disregard of COPPA Requirements by Social Media Platforms, Low Age Cut-off, and Inadequate Age Verification Procedures Result in Harm to Young Social Media Users

A 2022 study by Pixalate (a fraud protection, privacy, and compliance analytics platform) revealed that social media platforms' disregard of COPPA is likely exposing children to harmful advertising.[45] The study showed that while many social media platforms boast that advertisements are not targeted for a child's use, eight percent of Apple App Store apps and seven percent of Google Play Store apps actually *are* targeted to minors.[46] The study also found that child-targeted apps are forty-two percent more likely to share both GPS and IP address information with third-party digital advertisers than are non-child-targeted apps.[47] As noted earlier, geolocation and persistent identifiers, such as IP address, are considered personal information under COPPA; thus, COPPA forbids the collection, use, and disclosure this kind of information absent verifiable parental consent.[48] Lastly, the Pixalate study revealed that advertisers spend 3.1 times more money on child-targeted apps than they spend on apps directed to a general audience.[49] Thus, advertisers are focusing a larger sum of their financial resources on child-targeted apps, making exposure to potentially harmful advertising more likely for users of child-targeted apps, than users of general audience apps.

Congress has explicitly recognized the problem of minors' use of social media platforms without verifiable parental consent, and the reality of platforms' evasions of COPPA by claiming to not be child-targeted. In the 2021 Appropriations Act, Congress directed the Federal Trade Commission (FTC) to study and report on "whether and how artificial intelligence (AI) may be used to identify, remove, or take any other appropriate action necessary to address a wide variety of specified online harms."[50] AI is frequently used to address one such harm: age verification. AI can be used to determine whether users appear to be under the age of thirteen years.[51] Social media platforms can thus use AI to determine whether a user joined the site with a fake birthdate and whether COPPA regulations are applicable.

However, the FTC, in its responding report to Congress, advised against the usage of AI technology for this purpose, warning that using AI technology as a policy solution could lead to a myriad of unintended harms.[52] These harms derive from the inherent design flaws and inaccuracy of AI tools—including the potential bias built into the tool that reflects the biases of its developers—and the possibility of AI tools incentivizing and enabling invasive commercial surveillance and data extraction practices due to the vast amount of data required to develop, train, and use the tool.[53] The FTC advised that policies to alleviate online harms must not be rooted in the use of AI.[54] Rather, the FTC posits, it is imperative to understand the specific

[45] *Mobile Apps: Google vs. Apple COPPA Scorecards (Children's Privacy)*, Pixalate 1, 1 (2022), https://www.pixalate.com/hubfs/Reports_and_Documents/Mobile%20Reports/2022/App%20Reports/Active%20Apps/Child-Directed%20Apps/Q1%202022%20-%20Apple%20vs.%20Google%20COPPA%20Scorecard%20Report%20-%20Pixalate.pdf [hereinafter Pixalate].

[46] "Pixalate used automated processing derived from a combination of signals (which at times is coupled with human intervention) to determine if an app is likely to be child-directed, including the app's category, sub-category, content rating, and contextual signals (specifically, child-related keywords in app's title or the app's description)." *Id.*

[47] *Id.* at 3.

[48] 15 U.S.C. § 6501(8).

[49] "Pixalate calculates estimated programmatic ad spend through statistical models that incorporate programmatic monthly active users (MAU), the average session duration per user, the average CPM for the category of a given app, and ad density." Pixalate, *supra* note 45, at 16.

[50] Fed. Trade Comm'n, Combatting Online harms Through Innovation 1, 1 (2002), https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf.

[51] *See e.g.*, Ashley Johnson, *AI Could Make Age Verification More Accurate and Less Intrusive*, Info. Tech. & Innovation Found. (Apr. 5, 2023) https://itif.org/publications/2023/04/05/ai-could-make-age-verification-more-accurate-and-less-invasive/ [https://perma.cc/W2WA-D5U5].

[52] Press Release, Fed. Rad Comm'n, FTC Report Warns About Using Artificial Intelligence to Combat Online Problems (June 16, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems?utm_source=govdelivery [https://perma.cc/NMD8-GV92].

[53] *Id.*

[54] *Id.*

ways in which social media platforms are harmful to children and adolescents to enable policymakers to explore legal remedies and strategies that would hold the platforms accountable for the harm they create.

Congress is currently focused on the minimum age requirements that social media platforms impose upon users who wish to open accounts. By protecting only minors under the age of thirteen years, COPPA treats minors thirteen and older as adults, thus exposing them to the harms of social media without any age-related restrictions.[55] This issue has been recognized by bi-partisan members of the U.S. Senate who sponsored a bill in 2022, and reintroduced it in 2023, that would raise the cut-off age to seventeen years.[56] This bill is discussed later in this article.[57]

II. Public Health and Neuroscience Studies Point to Mounting Evidence of Mental Health Harms to Adolescents Resulting from Social Media Use

A. Rigorous Experimental and Longitudinal Public Health Studies of Social Media Effects Strongly Suggest Social Media Has a Harmful Impact on the Mental Health of Young Users

A growing body of evidence demonstrates that high amounts of social media use, and image-based social media in particular, is associated with poor mental health outcomes. Social media platforms are designed to provide content to retain viewers, using algorithms that populate individual feeds with material that entices users to stay engaged for longer periods of time.[58] Algorithm-driven features, such as limitless scrolling, social pressure and social reward (e.g., "likes" on posts), notifications, and individualized content feeds, are designed to maximize time spent on platforms.[59] Practices of social media platforms and apps designed to retain the attention of users are essential, albeit pernicious, features of platforms' business models that are predicated on monetizing users' time and attention.

These practices can be harmful to the mental health of users, particularly young users. To understand the associations between social media use and mental health outcomes, a plethora of research studies have been conducted, which have subsequently been summarized in several systematic reviews and meta-analyses.[60] A 2020 review summarizing the results of studies evaluating associations between social

[55]15 U.S.C. § 6501(1).

[56]Sen. Richard Blumenthal (D-CT) and Sen. Marsha Blackburn (R-TN) introduced The Kids Online Safety Act (KOSA), which would have given parents and users under seventeen the ability to opt out of algorithmic recommendations, prevent third parties from viewing a minor's data, and limit the time kids spend on a platform.

[57][ADD INFRA CITE]

[58]*See* Hilary Andersson, *Social Media Apps are 'Deliberately' Addictive to Users*, BBC (July 4, 2018) https://www.bbc.com/news/technology-44640959 [https://perma.cc/6SYB-QTHL].

[59]*See* Christian Montag et al., *Addictive Features of Social Media/Messenger Platforms and Freemium Games Against the Background of Psychological and Economic Theories*, 16 Int'l J. Env't Rsch. & Pub. Health 1, 4–6 (2019); Marco Zenone et al., *The Paradoxical Relationship Between Health Promotion and the Social Media Industry*, Health Promotion Prac. 1, 1–2 (2021); Thomas Mildner & Gian-Luca Savino, *Ethical User Interfaces: Exploring the Effects of Dark Patterns on Facebook*, CHI Conf. on Hum. Factors Computing Sys. Extended Abstracts 1, 2 (2021), https://dl.acm.org/doi/pdf/10.1145/3411763.3451659.

[60]Betul Keles et al., *A Systematic Review: The Influence of Social Media on Depression, Anxiety and Psychological Distress in Adolescents*, 25 Int'l J. Adolescence & Youth 79, 84–86 (2020); Amy Orben, *Teenagers, Screens and Social Media: A Narrative Review of Reviews and Key Studies*, 55 Soc. Psychiatry and Psychiatric Epidemiology 407, 408–11 (2020); Candice Odgers & Michaeline Jensen, *Annual Research Review: Adolescent Mental Health in the Digital Age: Facts, Fears, and Future Directions*, 61 J. Child Psych. & Psychiatry 336, 337–41 (2020); Laura Vandenbosch et al., *Social Media and Body Image: Recent Trends and Future Directions*, 45 Current Op. Psych. 1, 2–3 (2022); Elizabeth Ivie et al., *A Meta-Analysis of the Association Between Adolescent Social Media Use and Depressive Symptoms*, 275 J. Affective Disorders 165, 168–71 (2020); Alyssa N. Saiphoo & Zahra Vahedi, *A Meta-Analytic Review of the Relationship Between Social Media Use and Body Image Disturbance*, 101 Comput. Hum. Behav. 259, 264–67 (2019); Jenna Course-Choi & Linda Hammond, *Social Media Use and Adolescent Well-Being: A Narrative Review of Longitudinal Studies*, 24 Cyberpsychology, Behav., & Soc. Networking 223, 227–232 (2021); Samantha Tang et al., *The Relationship Between Screen Time and Mental Health in Young People: A Systematic Review of Longitudinal Studies*, 86 Clinical Psych. Rev. 1, 9 (2021); Sophia Choukas-Bradley et al., *The Perfect Storm: A Developmental-Sociocultural Framework for the Role of Social Media in Adolescent Girls' Body Image Concerns and Mental Health*, 25 Clinical Child & Fam. Psych. Rev. 681, 685–91 (2022); Ilaria Cataldo et al., *Social Media Usage and*

media use and indicators of mental health problems among adolescents published between 2011 and 2018 concluded there was a positive association, while also noting the complexity of the relationship.[61] The authors state that aspects of adolescents' personal and social identity formation may be vulnerable to the effects of social media use and described hypothesized mechanisms including limited self-regulation skills, displacement of sleep and/or physical activity, and negative social comparisons.[62] The study identified risk factors for mental health problems, including time spent on social media, personal investment, repeatedly checking for messages, and addictive use.[63]

While the body of research covering the early years of social media lays the groundwork for understanding the relationship between social media use and mental health outcomes, the platforms and their business practices have changed in profound ways, rendering more updated research necessary. Much of the research to date focuses on early platforms, such as Facebook and Twitter, which were created in 2004 and 2006, respectively, and does not adequately assess or account for the impact of currently popular platforms, such as Instagram and TikTok, which were created more recently in 2010 and 2016, respectively. Due to the rapid changes in the industry since its inception, the study designs used in the early years of social media research among youth provide limited insight into the effects of social media in its current form.

In recent years, scholars have more precisely assessed social media, and have employed more rigorous study designs, including experimental and longitudinal observational cohort studies with young people followed over years.[64] These enhancements strengthen the quality of the evidence generated by these studies and our ability to make causal conclusions about the relationship between social media use and mental health among youth. The most compelling studies in recent years have been those examining associations of social media use with body image and eating disorders and also those examining anxiety, depression, and suicidality.

Body image consists of the thoughts, feelings, and perceptions an individual has about the way they look,[65] and social media use has been associated with poor body image (i.e., body dissatisfaction).[66] Eating disorders are a serious public health concern,[67] and adolescence is a vulnerable window for the onset of disordered eating behaviors.[68] A number of studies have explored the relationship between social media use and body image using experimental designs, where participants are randomly assigned to different exposures or experiences of social media content to allow for comparisons between groups.[69] Random assignment helps researchers to best isolate the impact of the experiment, rather than external factors.[70] One European study found that an interaction between peer feedback and images of professional models contributed to adolescent girls' conceptualization of what an "ideal" body shape is, as well as differences in individual susceptibility to perceiving the ideal body as very thin.[71] Another experimental study randomly

Development of & Disorders in Childhood and Adolescence: A Review, 11 Frontiers Psychiatry eCollection: 1, 4-8 (2021); Patti M. Valkenburg et al., *Social Media Use and Its Impact on Adolescent Mental Health: An Umbrella Review of the Evidence*, 44 Curr. Opin. Psychol. 58, 59–60 (2022); Bohee So & Ki Han Kwon, *The Impact of Thin-Ideal Internalization, Appearance Comparison, Social Media Use on Body Image and Eating Disorders: A Literature Review*, 20 J. Evid.-Based Soc. Work. 55, 58–62 (2023).

[61] Keles, *supra* note 60, at 88.

[62] *Id.* at 80-81, 88.

[63] *Id.* at 88.

[64] *See e.g.,* Course-Choi & Hammond, *supra* note 60.

[65] Sarah Grogan, Body Image: Understanding Body Dissatisfaction in Men, Women, and Children 4 (2nd ed. 2008).

[66] Francesca Ryding & Daria Kuss, *The Use of Social Networking Sites, Body Image Dissatisfaction, and Body Dysmorphic Disorder: A Systematic Review of Psychological Research,* 9 Psych. Popular Media 412, 430 (2020).

[67] Janet Treasure et al., *Eating Disorders*, 395 Lancet 899, 899 (2020).

[68] Zachary J. Ward et al., *Estimation of Eating Disorders Prevalence by Age and Associations with Mortality in a Simulated Nationally Representative US Cohort.* 2 JAMA Network Open 1, 1 & 7 (2019).

[69] *See e.g.,* Ryding & Kuss, *supra* note 66.

[70] Christie N. Scollon, *Research Designs*, in R. Biswas-Diener & E. Diener, Noba Textbook Series: Psychology (2023), available at https://nobaproject.com/modules/research-designs.

[71] Jolanda Veldhuis et al., *Negotiated Media Effects. Peer Feedback Modifies Effects of Media's Thin-Body Ideal on Adolescent Girls,* 73 Appetite 172, 176–78 (2014).

assigned U.S. undergraduate college women to groups using either Facebook or Instagram, or a control group (the control group participants played a game rather than use social media).[72] The researchers found that Facebook and Instagram users reported engaging in more appearance comparisons than the control group.[73] Further, Instagram users also reported increased appearance comparison relative to Facebook users, and experienced decreased body satisfaction and increased negative affect.[74] In a third experimental study examining Instagram use, male and female college students viewed posts with two body-size conditions: a thin body type and a higher-weight body type.[75] Researchers measured attention to the Instagram posts and state body dissatisfaction[76] and found that exposure to images with thin-body portrayals resulted in both increased attention to the posts and increased body dissatisfaction compared to participants exposed to images of a higher-weight body type.[77] Female participants who perceived their own body type as higher-weight experienced increased body dissatisfaction in response to thin-image posts compared to higher-weight-image posts; this was not observed for females who perceived their body type as thin or average weight.[78] Another randomized controlled trial evaluated the impact of a break from social media (including Facebook, Instagram, TikTok, and Twitter) found individuals who were randomly assigned to a group that stopped using social media for a week saw improvements in measures of depression and anxiety[79] compared to participants who continued to use social media as usual.[80]

Ultimately, experimental study findings make clear that the kinds of social media platforms adolescents use have different effects on mental health outcomes, and that image-based or visual platforms are an important driver of the associations with worse mental health outcomes, particularly regarding body image and disordered eating.[81]

In addition to experimental study designs, longitudinal cohort study designs provide some of the most rigorous research findings to date linking social media use to eating disorders risk. For instance, a UK-based longitudinal observational study that enrolled youth (fifty-six percent male, mean age at time 1 = 14.3 years) and assessed social media use and body satisfaction at three times over one year found that adolescents with higher social media use engaged in more social comparison,[82] which was then associated with lower body satisfaction later in the year.[83] One U.S.-based study followed a cohort of adolescents beginning at age ten to thirteen years and measured social media use as a distinct component of media use.[84] This specificity in defining what to measure is key, as earlier research studies tended to assess "screen time" as a catch-all term that also included screen-based activities like television and

[72]Renee Engeln et al., *Compared to Facebook, Instagram Use Causes More Appearance Comparison and Lower Body Satisfaction in College Women*, 34 Body Image 38, 41 (2020).

[73]*Id.* 41 (2020).

[74]*Id.* at 41-42.

[75]Ciera Kirkpatrick & Sungkyoung Lee, *Effects of Instagram Body Portrayals on Attention, State Body Dissatisfaction, and Appearance Management Behavioral Intention*, Health Commuc'n 1, 5–6 (2021).

[76]State body dissatisfaction refers to a state of being or how someone feels in a particular moment, as opposed to trait body dissatisfaction, which is a more consistent and stable component of one's personality. Thomas F. Cash et al., *Beyond Body Image as a Trait: The Development and Validation of the Body Image States Scale*, 10 Eating Disorders 103, 103–04 (2002).

[77]Kirkpatrick & Lee, *supra* note 75.

[78]*Id.*

[79]The study used previously validated instruments including the Patient Health Questionnaire-8 (PHQ-8) to measure depressive symptoms and the General Anxiety Disorder Scale-7 (GAD-7) to measure anxiety symptoms.

[80]Jeffrey Lambert et al. *Taking a One-Week Break from Social Media Improves Well-Being, Depression, and Anxiety: A Randomized Controlled Trial.* 25 Cyberpsychology Behav. Soc. 287, 290–291 (2022).

[81]Vandenbosch, *supra* note 47 at 186.

[82]The researchers assessed frequency of general comparisons, social comparisons, and appearance comparisons using nine survey questions with response options along a five-point Likert scale (e.g., answering "1=strongly disagree" to "5= strongly agree" in response to questions such as "I often compare myself with others on social media' and 'I often think that others are having a better life than me').

[83]Hannah K. Jarman et al., *Direct and Indirect Relationships Between Social Media Use and Body Satisfaction: A Prospective Study Among Adolescent Boys and Girls*, New Media & Soc'y 1, 11–12 (2021).

[84]Sarah M. Coyne et al., *Suicide Risk in Emerging Adulthood: Associations with Screen Time over 10 Years,* 50 J. Youth & Adolescence 2324, 2326-27 (2021).

internet browsing, making it difficult to ascertain the impact of social media as distinct from other activities.[85] Researchers found that adolescent girls with high social media use (two to three hours per day) early in adolescence who subsequently increased use over time had increased suicide risk ten years later.[86] Data from the longitudinal UK-based "Our Futures" study show that frequent social media use (two to three times per day) at the beginning of the study, when participants were age twelve to fourteen years, was associated with more psychological distress as measured by the General Health Questionnaire[87] at follow-up, when these same participants were ages fifteen to sixteen years.[88] Finally, in a longitudinal study conducted among U.S. high school students, appearance-related social media consciousness at the start of the study was associated with subsequent depressive symptoms one year later.[89]

Results of these experimental study designs and longitudinal cohort study designs provide rigorous evidence linking high amounts of social media use, and time spent on image-based social media in particular, to mental health harms in young users.

B . Neuroscience Pathways Directly Link Social Media Use to Mental Health Risks in Young Social Media Users

Evidence from the psychological literature highlights two psychological processes that are especially important in explaining how social media use negatively impacts adolescent mental health, particularly as related to eating disorders risk: upward comparison and thin-ideal internalization. Upward comparison occurs when an individual compares aspects of themselves (e.g., appearance) against more popular or esteemed others, such as social media influencers, professional models, or celebrities.[90] Research on Instagram suggests that among adolescents, use of the platform increases the tendency of users to do upward comparison, which is ultimately associated with body dissatisfaction.[91] Thin-ideal internalization[92] occurs when a society highly values being thin as a component of being attractive and can be especially harmful when an adolescent within that society adopts the cultural norm equating thinness with attractiveness as their own belief. Thin-ideal internalization is a risk factor for body image disturbances[93] and contributes to hyperconsciousness about one's appearance, including frequent body checking and shame (i.e., feeling like a bad person for how they look or weigh or feeling ashamed for not being smaller).[94] Evidence linking social media use to this type of harmful upward comparison and thin ideal internalization, and subsequently to eating disorders symptoms, is stronger in adolescent girls than

[85] *Id.* at 2334.

[86] *Id.* at 2328.

[87] The study team used the General Health Questionnaire (GHQ12) to measure mental health. It is a twelve-item scale where a score of three or higher signifies psychological distress. Russell Viner et al., *Roles of Cyberbullying, Sleep, and Physical Activity in Mediating the Effects of Social Media Use on Mental Health and Wellbeing Among Young People in England: A Secondary Analysis of Longitudinal Data*, 3 Lancet Child & Adolescent Health 685, 685 (2019).

[88] *Id.* at 691.

[89] Anne J. Maheux et al., *Longitudinal Associations Between Appearance-related Social Media Consciousness and Adolescents' Depressive Symptoms*, 94 J. Adolescence 264, 266 (2022).

[90] Federica Pedalino & Anne-Linda Camerini, *Instagram Use and Body Dissatisfaction: The Mediating Role of Upward Social Comparison with Peers and Influencers Among Young Females*, 19 Int'l J. Env't Rsch. Pub. Health 1, 7 (2022)

[91] *Id.* at 7.

[92] J. Kevin Thompson & Eric Stice, *Thin-Ideal Internalization: Mounting Evidence for a New Risk Factor for Body-Image Disturbance and Eating Pathology*, 10 Current Directions in Psych. Sci. 181, 181 (2001).

[93] Veldhuis, *supra* note 71 at 173, 176–79; Gemma López-Guimerà et al., *Influence of Mass Media on Body Image and Eating Disordered Attitudes and Behaviors in Females. A Review of Effects and Processes,* 13 Media Psych. 387, 401–02 (2010).

[94] Veldhuis, *supra* note 71. In the Veldhuis 2014 study, shame was assessed with the questions: "(1) I feel ashamed of myself when I haven't made an effort to look my best; (2) I feel like I must be a bad person when I don't look as good as I could; (3) I would be ashamed for people to know what I really weigh; (4) when I'm not exercising enough, I question whether I am a good person; and (5) when I'm not the size I think I should be, I feel ashamed." Sara M. Lindberg et al., *A Measure of Objectified Body Consciousness for Preadolescent and Adolescent Youth,* 30 Psych. Women Q. 65, 69 (2006).

boys.[95] This evidence highlighting a greater impact on adolescent girls is corroborated by a 2023 statement from the U.S. Surgeon General acknowledging that social media use "perpetuate[s] body dissatisfaction, disordered eating behaviors, social comparison, and low self-esteem, especially among adolescent girls."[96]

Evidence from the neuroscience literature has identified several features of the adolescent brain that lead to uniquely elevated risks of social media use in adolescents compared to adults or even to younger children. First, adolescence is a time of heightened sensitivity to peer feedback and social cues, which are processed by the brain's social cognition and emotional response circuitry, including brain regions such as the amygdala, striatum, and medial prefrontal cortex.[97] When sense of self-worth and identity is forming in adolescence, "brain development is especially susceptible to social pressures, peer opinions, and peer comparison."[98] As such, the adolescent brain is particularly tuned in to "rewarding" feedback on social media, such as "likes" on a post. Adolescents use this information to shape their understanding of social norms and values.[99] For example, if an adolescent posts an image of a thin model and subsequently receives hundreds of "likes," their brain interprets that they were "rewarded" for sharing the image and they are more likely than adults to use the "likes" to inform their concept of what images are socially desirable.

Additionally, the naturally uneven pace of development of different regions of the brain during adolescence exacerbates vulnerability to the harms posed by social media use. For instance, during adolescence, brain regions that process emotions (e.g., the amygdala) develop faster, while brain regions responsible for decision-making, reasoning, and impulse control (i.e., frontal regions) develop more slowly and continue to develop well into young adulthood.[100] This lopsided neural development is associated with heightened emotionality and sensitivity to emotion-inducing social media content, since adolescents' abilities to regulate emotional responses is hindered.[101] Lastly, adolescents are more sensitive than adults to social rewards (in contrast to monetary or other types of rewards).[102] The activation of reward processing regions in the brain when using social media platforms can make these platforms highly influential on teens. Features including getting "likes" on a post or comment, autoplay, infinite scroll and algorithms that leverage use data to serve content recommendations motivate continued engagement despite psychological harms and promote excessive use of social media.[103] As a common example, a teen who received several "likes" on a previously shared picture of themselves vaping is neurologically motivated to post a similar picture to receive the same stimulating reward.

In sum, neuroscience research has identified unique characteristics of the adolescent brain that place adolescents, rather than adults or younger children, at particular risk to negative mental health effects of social media use. These characteristics include: (1) the heightened sensitivity to social cues, (2) increased emotional responses as a product of underdeveloped judgment regions and more

[95]Ciara Mahon & David Hevey, *Processing Body Image on Social Media: Gender Differences in Adolescent Boys' and Girls' Agency and Active Coping,* 12 Frontiers Psych. eCollection 626763, 8 (2021); Marika Skowronski et al., *Links Between Exposure to Sexualized Instagram Images and Body Image Concerns in Girls and Boys,* 34 J Media Psych. eCollection 55, 59 (2022); Illyssa Salomon & Christia Spears Brown, *The Selfie Generation: Examining the Relationship Between Social Media Use and Early Adolescent Body Image,* 39 J Early Adolescence 539, 548–52 (2022).

[96]*Social Media and Youth Mental Health: The US Surgeon General's Advisory*, (May 2023) https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf [hereinafter *US Surgeon General's Advisory*].

[97]Leah H. Somerville, *The Teenage Brain: Sensitivity to Social Evaluation*, 22 Current Directions Psych. Sci. 121, 122 (2013).

[98]*US Surgeon General's Advisory*, *supra* note 73.

[99]*Id.*

[100]B.J. Casey et al., *The Adolescent Brain*, 1124 Annals N.Y. Acad. Scis. 111, 116 (2008).

[101]*Id.* at 117; Somerville, *supra* note 97, at 122.

[102]Paige Ethridge et al., *Neural Responses to Social and Monetary Reward in Early Adolescence and Emerging Adulthood*, 54 Psychophysiology 1786, 1792–93 (2017).

[103]Lauren E. Sherman et al., *The Power of the Like in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media*, 27 Psych. Sci. 1027, 1031 (2016); *see also US Surgeon General's Advisory*, *supra* note 73.

mature emotion processing regions, and (3) social media's ability to activate reward processing regions in the brain to motivate continued engagement. Social media platforms that are highly visual or image-based, where digitally altered and unrealistic images of body shape and thinness are common, compound the links between social media use and subsequent mental health harms.[104] Adolescents are especially sensitive to peer feedback that communicates social preferences and have exaggerated emotional responses because of the brain's reduced ability to regulate emotional responses. Persistent exposure to social media content is driven by algorithms and platform practices that engage sensitive reward processing structures to motivate teens to stay on platforms longer. Altogether, the interaction of normative adolescent neurodevelopment with features of social media platforms, particularly those that are image-based, increases mental health risks to young people. This developmental-stage-based vulnerability must be accounted for when (1) assessing the harm inflicted on young users by social media platform practices and (2) creating legislation and regulations to curb such harms.

III. Social Media Platforms Are Economically Incentivized to Use Relentless Algorithms that Push Harmful Content to Young Online Users

The immense advertising revenue social media platforms generate from young users discourages efforts by the platforms to self-regulate and curb the online harms caused to young people. The economic benefit social media companies enjoy from exploiting young social media users is assumed to be considerable but has not been well-documented. Social media platforms have no obligation to release data surrounding the types of content to which young users are exposed, nor the impacts of such content.[105] And platforms are highly incentivized to keep youth online; children's online experiences are heavily monetized through advertising revenue on social media platforms and mobile applications.[106]

Since platforms are not held accountable to children nor regulatory agencies,[107] they are not required to report advertising revenue nor the age breakdown of users. To fill gaps in the information on how much revenue social media platforms generate from minors, authors of this article obtained data from a business marketing source and from public survey data[108] to conduct a novel simulation analysis that would provide the first known estimates of the number of users and the annual advertising revenue generated from U.S.-based users aged zero to twelve and thirteen to seventeen years for six major social media platforms. We found, across the major platforms, that annual advertising revenue from U.S. children ages zero to twelve years is estimated to be over $2 billion in U.S. dollars, and from all children ages zero to seventeen years is nearly $11 billion.[109] For several social media platforms, thirty to forty percent of their annual advertising revenue is generated from users ages zero to seventeen years. The massive revenue generated from young users discourages social media platforms from self-regulation and further demonstrates the need for government policy and legislative intervention to curb harms.

---

[104]Mara van der Meulen et al., *Brain Activation Upon Ideal-body Media Exposure and Peer Feedback in Late Adolescent Girls*, 17 Cognitive, Affective, & Behav. Neurosci. 712, 720 (2017).

[105]*See* Brooke Auxier & Monica Anderson. *Social Media Use in 2021.* Pew Research Center (Apr. 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/; Orben, *supra* note 47 at 411.

[106]Yolanda Reid Chassiakos et al., *Children and Adolescents and Digital Media*, 138(5) Am. Acad. Pediatrics (Nov. 1, 2016); Marisa Meyer et al., *Advertising in Young Children's Apps: A Content Analysis*, 40(1) J. Dev.'l & Behav. Pediatrics 32, 38 (2019).

[107]Caitlin R. Costello et al., *Adolescents and Social Media: Privacy, Brain Development, and the Law*, 44(3) J. Am. Acad. Psychiatry & L. 313, 313 (2016).

[108]Auxier & Anderson, *supra* note 105; Victoria Rideout et al., *The Common Sense Census: Media Use by Tweens and Teens, 2021*, Common Sense Media (Mar. 9, 2022), https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2021 [https://perma.cc/588T-XFCE].

[109]Amanda Raffoul et al., *Social Media Platforms Generate Billions of Dollars in Revenue from U.S. Youth: Findings from a Simulated Revenue Model*, 18 PLoSONE 12 (2023), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0295337 [https://perma.cc/VLL3-DLZK].

IV. Legal Obstacles to Preventing Harm to Young People Caused by Social Media Algorithms and the Strategies to Circumvent Them: How to Grapple with First Amendment Speech Protections and Section 230 of the Communications Decency Act

A. First Amendment Protection for Content on Social Media Platforms Allows Harm to Be Inflicted on Young Users Through the Platforms' Use of Algorithms

Those attempting to regulate harms to children and teens resulting from time spent on social media platforms face the daunting legal obstacles of the First Amendment and Section 230 of the federal CDA. As technology becomes more and more entangled with the everyday life and communication of most Americans, social media platforms like Facebook, Instagram, Twitter, Snapchat, and TikTok have become forums where individuals can exercise their right to free speech. The First Amendment protects a wide swath of speech, ranging from highly protected political speech to lesser protected commercial speech and sexually explicit speech.[110] Certain categories of speech are bluntly illegal and will not enjoy First Amendment protection, such as defamation, bribery, incitement, fighting words, conspiracy to commit a crime, etc.[111] Generally, however, laws trying to regulate the specific content of speech, for instance, hate speech, will be found unconstitutional, while content-neutral laws that instead regulate the time, place, and manner of speech no matter its content will not be deemed violative the First Amendment.[112]

Those trying to regulate harms of social media platforms risk violating First Amendment free speech rights because of restrictions they seek to impose on content, specifically algorithms used by social media platforms.[113] Thus, the more specific issue is whether algorithms—in this case, computer programming that can sort and recommend content for users of social media—is protected speech.[114] Though the Supreme Court has not ruled on the issue of whether algorithms

[110]*See, e.g.,* Jenkins v. Georgia, 418 U.S. 153 (1974) (discussing high standards needed for sexually explicit content to reach levels of obscenity unprotected by the First Amendment.).

[111]Victoria L. Killion, *The First Amendment: Categories of Speech*, Congressional Rsch. Serv., IF11072, https://crsreports.congress.gov/product/pdf/IF/IF11072.

[112]An example of a constitutional content-neutral law would be a law disallowing anyone to use a bullhorn to say anything in a public square after 8 p.m. because it could disrupt sleep and the quiet evening solitude for those nearby. The law does not restrict speech based on its content, rather it restricts any speech based on the disruption it could cause to those trying to enjoy quiet and peaceful late evening hours. In contrast, a law that would restrict someone from using a bullhorn in the town square to announce the strengths of a political candidate running for town council but allow someone to use a bullhorn to announce an upcoming performance of a play at a local theater would be a content-based law and would be unconstitutional. *Id.*

[113]Some states have attempted blanket crackdowns on social media platforms and have faced immediate pushback. For instance, Montana Governor Greg Gianforte, signed a bill on May 17, 2023, prohibiting individuals from using or downloading TikTok in the state of Montana. Any entity, defined as an app store or TikTok, faces a $10,000 penalty for each time a user downloads, accesses, or is able to access TikTok. An additional $10,000 penalty is added for each day the violation continues. The law does not impose fines on individual Tik Tok users. The ban will be void if TikTok is acquired by a country that is not incorporated in a country "designated as foreign adversary." *See* S.B. 419, Gen. Sess. (Mont. 2023). It is unclear how Montana would enforce the law. While many members of Congress expressed their wariness regarding TikTok and the mental health of teen users at the March 2023 Congressional Hearings, Montana's law is directed at privacy and security concerns involving the Chinese Communist Party. *Id.* This ban is currently the most extreme prohibition of the app in the United States and faced immediate legal challenges regarding its feasibility and constitutionality. TikTok filed suit just days after the Montana law was adopted, alleging that the ban is "extreme" and violates the First Amendment, as well as other federal laws. The social media company claims concerns that the Chinese government could access the data of U.S. TikTok users — which are a key motivation behind the ban — are "unfounded." *See* Clare Duffy, *Tik Tok sues Montana over new law banning the app*, CNN Business, (May 23, 2023, 5:31 AM), https://www.cnn.com/2023/05/22/tech/tiktok-montana-lawsuit/index.html [https://perma.cc/VLB9-W2D8]. Further, NetChoice, a tech trade group that includes Google, Meta and TikTok, sued the state of Arkansas in June 2023 claiming the state's newly passed Social Media Safety Act is unconstitutional. Netchoice asserted the law allegedly treads on First Amendment free speech rights by making users hand over private data to access social networks. It also asserts the Act hurts privacy and safety by making internet companies rely on a third-party service to store and track kids' data. See John Fingas, *Tech Firms Sue Arkansas Over Social Media Age Verification Law.* (June 30, 2023), https://www.engadget.com/tech-firms-sue-arkansas-over-social-media-age-verification-law-180002953.html [https://perma.cc/5DE5-AUWH].

[114]Alexander S. Gillis, *Definition: Algorithm*, Tech Target (May 2022), https://www.techtarget.com/whatis/definition/algorithm [https://perma.cc/R6YA-4CLE] ("An algorithm is a procedure used for solving a problem or performing a computation. Algorithms act as an exact list of instructions that conduct specified actions step-by-step in either hardware-

are protected under the First Amendment, algorithms are likely protected speech under the Free Speech Clause because algorithms are, in essence, a computer code,[115] and federal courts have repeatedly found that a computer code is speech protected under the First Amendment.[116] Similarly, federal courts have found that search engine results are protected speech under the First Amendment.[117] Algorithms and search engine outputs function similarly in that both algorithms and search engines are edited compilations of speech that are generated from other individuals, such as engineers, and are arranged to appear in a specific order on a user's social media feed.[118] Due to the similarity, courts would likely find that a social media algorithm is a type of computer code or output code, and is consequently protected under the First Amendment.[119]

Thus, the content circulated by a social media platform's algorithm—the information that shows up in the feed of social media users—likely cannot not be specifically targeted by any legislation due to First Amendment protections.[120] Neither can the actual computer code of the algorithm that selects and directs the content, even though some of it is harmful to young people.[121] Legislation aiming to regulate harmful algorithms must surmount this high bar of First Amendment speech protection.

B. Immunity Granted to Social Media Platforms by Section 230 of the Communications Decency Act Protects Social Media Companies Against Injury Claims

The second major obstacle to regulating harm caused by social media platforms is Section 230 of the CDA. Section 230 grants immunity to online services, meaning that online services are not liable for the speech of third parties published on their platforms.[122] Enacted in 1996, Section 230 was considered

---

or software-based routines. Algorithms are widely used throughout all areas of IT. They are the building blocks for programming, and they allow things like computers, smartphones, and websites to function and make decisions. In mathematics and computer science, an algorithm usually refers to a small procedure that solves a recurrent problem. Algorithms are also used as specifications for performing data processing and play a major role in automated systems. An algorithm could be used for sorting sets of numbers or for more complicated tasks, like recommending user content on social media. Algorithms typically start with initial input and instructions that describe a specific computation. When the computation is executed, the process produces an output … There are several types of algorithms, all designed to perform different tasks, including a search engine algorithm, encryption algorithm, greedy algorithm, recursive algorithm, backtracking algorithm, divide-and-conquer algorithm, divide and conquer algorithm, dynamic programming algorithm, brute-force algorithm, sorting algorithm, hashing algorithm, randomized algorithm, etc.")

[115]Veronica Balbuzanova, *First Amendment Considerations in the Federal Regulation of Social Media Networks' Algorithmic Speech, Part I*, Am. Bar Ass'n (Jan. 29, 2021), https://www.americanbar.org/groups/litigation/committees/privacy-data-security/articles/2021/first-amendment-social-media-algorithmic-speech-part-1/ [https://perma.cc/Q8W5-6B5L].

[116]*Id. See, e.g.,* Universal City Studios, Inc. v. Corley, 273 F.3d 429, 449 (2d Cir. 2001) (holding that "computer code, and computer programs constructed from code can merit First Amendment protection"); Johnson Controls v. Phoenix Control Sys., 886 F.2d 1173, 1175 (9th Cir. 1989) (holding that "[s]ource and object code, the literal components of a program, are consistently held protected by a copyright on the program… Whether the non-literal components of a program, including the structure, sequence and organization and user interface, are protected depends on whether on the particular facts of each case, the component in question qualifies as an expression of an idea, or an idea itself"); Green v. United States DOJ, 392 F. Supp. 3d 68 (D.D.C. 2019); Bernstein v. U.S. Dep't of State, 922 F. Supp. 1426, 1436 (N.D. Cal. 1996) (holding that "copyright law also supports the 'expressiveness' of computer programs").

[117]Balbuzanova, *supra* note 110. S*ee, e.g.,* e-ventures Worldwide LLC v. Google, Inc., 188 F. Supp. 3d 1265 (M.D. Fla. 2016); Zhang v. Baidu.Com, Inc., 10 F. Supp. 3d 433 (S.D.N.Y. 2014); Langdon v. Google, Inc., 474 F. Supp. 2d 622 (D. Del. 2007); Kinderstart.Com, LLC v. Google, Inc., No. CO6-2057KF(RS), 2007 U.S. Dist. LEXIS 22637 (N.D. Cal. Mar. 16, 2007); Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457-M, 2003 U.S. Dist. LEXIS 27193 (W.D. Okla. May 27, 2003).

[118]Balbuzanova, *supra* note 110.

[119]*Id.*

[120]Veronica Balbuzanova, *First Amendment Considerations in the Federal Regulation of Social Media Networks' Algorithmic Speech, Part II*, Am. Bar Ass'n (Feb. 8, 2021), https://www.americanbar.org/groups/litigation/committees/privacy-data-security/articles/2021/first-amendment-social-media-algorithmic-speech-part-11/ [https://perma.cc/2PD7-PS8B].

[121]*Id.*

[122]Ashley Johnson & Daniel Castro, *Overview of Section 230: What It Is, Why it Was Created, and What It Has Achieved*, Info. Tech. & Innovation Found. (Feb. 22, 2021), https://itif.org/publications/2021/02/22/overview-section-230-what-it-why-it-was-created-and-what-it-has-achieved/ [https://perma.cc/N3X4-Z3TH].

foundational in supporting the internet as a free speech medium. At that time, however, social media did not exist, and was not a primary source of communication and information like it is today

Nevertheless, courts in recent years have applied Section 230's protections to social media platforms, including Facebook and Twitter.[123] As a result, Section 230 has become a major roadblock to legislation aimed at protecting children and teens from online harms. Certainly, Section 230 has had a positive impact, notably keeping the social media companies that have become so central to everyday life in business.[124] However, many legislators attempting to regulate harmful content in digital spaces argue that Section 230 is overbroad and has granted immunity to platforms that knowingly profit from harmful content on their social media platforms.[125]

When drafting legislation to regulate social media harms, legislators might achieve in circumventing Section 230's rigorous protections in the three following circumstances:

> First, if the defendant may have induced or contributed to the development of the illegal content in question, then Section 230 does not apply. Second, if the plaintiff's claim does not arise from the defendant's publishing or content moderation decisions, then Section 230 does not apply because Section 230 does not protect providers from all liability (only liability from its role as a publisher). Third, if the case relates to a content-removal decision and the defendant fails to meet Section 230 (c)(2)'s "good faith" requirement, then Section 230 does not apply because the defendant does not qualify for its protection.[126]

1. U.S. Supreme Court Declines to Limit Section 230 Blanket Immunity for Social Media Companies, but a Revenue Sharing Claim that Could Remove Section 230 Protection Potentially Exists

Of note are the legal challenges to the broad immunity granted to social media platforms by Section 230 that have arisen in recent years, including a case heard by the U.S. Supreme Court in 2022. In two recent decisions, *Twitter, Inc. v. Taamneh* and *Gonzalez v. Google*, the U.S. Supreme Court declined to limit the broad immunity Section 230 offers to social media companies for promoting inappropriate content that is published by third parties on their platforms. (The decisions did, however, appear to leave intact a revenue sharing theory where a plaintiff may allege a platform that commercially profits from an algorithm that pushes illegal content could be considered an information content-provider, thus removing the immunity protection of Section 230 and opening the platform up to liability.) These rulings were celebrated by Tech companies and their allies as a win for free expression on the internet,[127] while critics of Section 230 viewed the decisions with disappointment.

In *Twitter, Inc. v. Taamneh*, the family of Nawras Alassaf, who was killed in an ISIS terrorist attack on the Reina nightclub in Istanbul, Turkey, alleged that social media companies knowingly aided ISIS in violation of the Anti-Terrorism Act by allowing ISIS content on their platforms, failing to remove such

---

[123] *Id.*; *see also* Doe v. MySpace, 528 F.3d 413 (5th Cir. 2008).

[124] Michael D. Smith & Marshall Van. Alstyne, *It's Time to Update Section 230*, Harvard Bus. Rev. (Aug. 12, 2021), https://hbr.org/2021/08/its-time-to-update-section-230 [https://perma.cc/YX7E-K95N].

[125] Johnson, supra note 116. Numerous bills have been proposed aiming to amend Section 230 of the CDA including The Biased Algorithm Deterrence Act of 2019; Protecting Americans from Dangerous Algorithms Act; Justice Against Malicious Algorithms Act of 2021; Federal Bich Tech Tort Act, Safeguarding Against Fraud, Exploitation, Threats, Extremism, and Consumer Harms Act (SAFE TECH Act); Stop Shielding Culpable Platforms Act; and Social Media NUDGE Act. H.R. 492, 116th Cong. (2019-2020); H.R. 2154, 117th Cong. (2021-2022); H.R. 5596, 117th Cong. (2021-2022); H.R.3421, 117th Cong. (2021-2022); H.R. 2000, 117th Cong. (2021-2022); S.3608, 117th Cong. (2021-2022).

[126] *Id.*

[127] *See* Robert Barnes & Cat Zakrzewski, *Supreme court Rules for Google, Twitter, on terror-related content*, Wash. Post (May 18, 2023, 11:04 AM), https://www.washingtonpost.com/politics/2023/05/18/gonzalez-v-google-twitter-section-230-supreme-court/ [https://perma.cc/53Q3-XMF4]. ("Tech companies and their surrogates celebrated the ruling, which followed extensive lobbying and advocacy campaigns to defend Section 230 in Washington. Changes to the law, they said, could open a floodgate of litigation that would quash innovation and have wide-ranging effects on the technology that underlies almost every interaction people have online, from innocuous song suggestions on Spotify to prompts to watch videos about conspiracy theories on YouTube.").

content, and recommending ISIS content using algorithms.[128] The Supreme Court unanimously held that social media companies, including Twitter, did not "aid and abet" ISIS simply because their algorithms recommended ISIS content, failed to remove such content, or knew such content existed on the platform.[129] The Court explained that these actions did not rise to the level of substantial assistance as required to seek damages under the Anti-Terrorism Act for a secondary-liability claim.[130] Although the *Taamneh* decision did not forthrightly address Section 230, it declined to impose any third-party liability on Twitter because it did not "knowingly" provide substantial assistance and thus could not have aided and abetted ISIS in the terrorist attack on the Reina nightclub.

In *Gonzalez v. Google*, the Court left Section 230 fully intact and declined to rule definitively on whether it protects a platform's recommendation algorithms because the plaintiffs in *Gonzalez* failed to state a claim.[131] Nohemi Gonzalez was killed in 2015 during an ISIS terrorist attack in Paris while studying abroad.[132] Nohemi's family alleged that Google, Twitter, and Facebook aided and abetted ISIS with algorithms and recommended video content.[133] Specifically, the family asserted a revenue sharing theory, alleging that the platforms placed paid advertisements in proximity to ISIS-related content and shared in resulting ad revenue; therefore, the three social media companies should be liable for the ISIS-related content it generated revenue from.[134]

The revenue sharing theory articulated in *Gonzalez* asserts that, if a platform is commercially profiting off of an algorithm, it should be considered an information content-provider under Section 230, thus barring immunity from liability. A recent California case, *In re Apple Inc. Litigation*, analyzed this theory. This case involved social media casino apps, including the purchase of virtual "chips" to wager for gambling purposes.[135] Here, the plaintiffs asserted two distinct theories for revenue sharing that would arguably bar the platforms from immunity under Section 230. Under one theory, plaintiffs alleged that the platforms operated as the payment processor for all purchases of virtual chips and thus aided in the "exercise of illegal gambling by selling chips that [were] substantially certain to be used to wager on a slot machine."[136] The court found that since this theory was grounded in the platforms' own bad *acts*, and not in the *content* of the social media casino app, the platforms could not rely on Section 230 to escape liability.[137] The second revenue sharing theory asserted that platforms were liable for "offering, categorizing, and promoting" social casino applications in their App Stores, which helped the platforms generate a profit by targeting advertisements at specific users.[138] The plaintiffs alleged the platforms not only recommended content but helped develop advertisements to attract users to the social casino apps, making the illegal product "more appealing and addicting."[139] But the court noted that the platforms' contribution of data, which aided in the creation of advertisements, did not

[128] *See* Gonzalez v. Google LLC, 2 F.4th 871, 883 (9th Cir. 2021).

[129] The fact that these algorithms matched some ISIS content with some users thus does not convert defendants' passive assistance into active abetting. Twitter, Inc. v. Taamneh, 143 S.Ct. 1206, 1209 (2023).

[130] *See id.* at 1226 (explaining that recommendation algorithms do not go "beyond passive aid and constitute active, substantial assistance"). Plaintiffs also alleged that the Defendants' knowledge of ISIS content and failure to screen the publication of such content rose to the level of "aiding and abetting" ISIS; however, the Court disagreed. *See id.* at 1222–24.

[131] *See* Gonzalez v. Google LLC, 143 S. Ct. 1191 (2023).

[132] *See* Patrcik Garrity et al., *American Student Nohemi Gonzalez Identified As Victim in Paris Massacre*, NBC News (Nov. 14, 2015, 1:50 PM,), https://www.nbcnews.com/storyline/paris-terror-attacks/american-student-nohemi-gonzalez-idd-victim-paris-massacre-n463566 [https://perma.cc/525N-NJA7].

[133] *See Gonzalez*, 2 F.4th at 882.

[134] *See id.* at 880. Plaintiffs further allege that defendants should be directly liable for committing acts of international terrorism, and for conspiring with, and aiding and abetting ISIS's acts of international terrorism because the platform's algorithm directed ISIS videos to users and recommended ISIS content to users. *Id.* at 881. Ultimately, the Ninth Circuit held in favor of the defendants because the plaintiffs "did not plausibly allege" that Google, Twitter, and Facebook's actions qualified as an act of international terrorism and conspiracy or aiding and abetting. *Id.* at 913.

[135] *See* In re Apple Inc. Litig., 2022 U.S. Dist. LEXIS 159613 *1, *20 (N.D. Cal. 2022).

[136] *Id.* at 72–73.

[137] *See id.* at 74–76.

[138] *Id.* at 72.

[139] *Id.* at 77.

create and *develop* the casino apps, rather, the contribution of data was akin to offering publishing advice.[140] Thus the platforms behaved like editors, rather than content providers, and were shielded from liability by Section 230.[141]

The Ninth Circuit found in *In re Apple* that Section 230 did not bar the ad revenue sharing claims, because such allegations did not seek to hold the social media platforms liable for any content provided by a third party.[142] Under the facts of *Gonzalez*, the Ninth Circuit held that the plaintiffs "failed to state a claim for aiding-and-abetting liability" because the allegations were devoid of any statements "about how much assistance Google provided" and therefore did not plausibly allege "that Google's assistance was substantial."[143] By failing to demonstrate that Google provided substantial assistance to ISIS, the plaintiffs did not have a viable claim under the Anti-Terrorism Act and thus, Google could assert a Section 230 immunity defense.

The Supreme Court agreed with this holding, because the *Gonzalez* complaint "allege[d] nothing about the amount of money that Google supposedly shared with ISIS, the number of accounts approved for revenue sharing, or the content of the videos that were approved."[144] The Court thus explained that there was nothing in the complaint "to view Google's revenue sharing as substantial assistance" and that without more the plaintiffs failed to demonstrate "that Google knowingly provided substantial assistance" to the Reina attack, the Paris attack, or any other ISIS terrorist attack.[145] Because Google did not violate any law, it could still benefit from Section 230 immunity.

While the revenue sharing claims did not succeed in this particular case, the Ninth Circuit acknowledged that, in a different scenario, ad revenue sharing by a social media platform would not be immune to liability under Section 230.[146] The Supreme Court did not reject this idea, but explained that without a viable claim, in this case "aiding and abetting" under the Anti-Terrorism Act, it could not address Section 230.[147] Rather, to bar a platform from asserting Section 230 immunity, a plaintiff would first need to raise a viable claim for the platforms to be held liable for their conduct.

Unfortunately, when considering how a revenue sharing liability claim could challenge the harm platforms inflict on adolescents through relentless algorithms, the second revenue sharing liability theory in *In re Apple* seems to apply best. Creating algorithms that offer, categorize, and promote harmful content would likely not be considered a content-providing act, but rather an editorial function protected from liability, even if a social media platform earns profits from ad revenue. Platforms may be aiding in, and profiting from, targeting harmful content to minor users, but they are *not creating the content* itself and thus may be immunized under Section 230. Further, both *Gonzalez* and *In re Apple* involved online sites engaged in illegal activity—terrorism and gambling—which provided more reason for the Ninth Circuit and district court to hold the social media platforms liable for their conduct. In contrast, while platforms feeding harmful content to minors through the use of algorithms—and earning tidy sums from ad revenues—may be injurious to young users, the platforms are not promoting an illegal activity.[148] These differences,

[140] *See id.* at 76–78.

[141] *See id.*

[142] *See Gonzalez*, 2 F.4th at 909–913.

[143] *Id.* at 907.

[144] *Twitter*, 143 S. Ct. at 1209.

[145] *Id.*

[146] The court held the *Gonzalez* Plaintiffs' revenue-sharing allegations were not directed to the publication of third-party information. The revenue sharing did not depend on the particular content ISIS places on Youtube; the theory is solely directed to Google's unlawful payments of money to ISIS. Therefore, the alleged violation could be remedied without changing any of the content posted by Youtube's users. The allegations of revenue sharing do not seek to hold Google liable for any content provided by a third-party. *See* Gonzalez, 2 F.4th at 913. The Supreme Court did not reject this reasoning, suggesting that a potential revenue sharing liability claim may be used in the future. *See* Twitter, 143 S. Ct. at 1209–1210.

[147] Additionally, the plaintiffs in the Gonzalez case "did not seek review of the Ninth Circuit's holdings regarding their revenue-sharing claims," so the Supreme Court did not address this issue in its opinion of Gonzalez. *See* Gonzalez, 143 S. Ct. at 1192.

[148] With the exception of demographic and behavioral advertising targeted at minors under the age of thirteen which is in violation of COPPA, the broad use of algorithms to feed content to minors, including content that can result in harm, is not illegal.

and the Supreme Court's decision to not review the revenue sharing theory of liability claim in *Gonzalez* and *Taamneh* cases, suggest at the claim's potential—but it is difficult to predict how exactly it could be used in the future to bar social media platforms from immunity under Section 230 of the CDA.

V. Legal Strategies to Circumvent First Amendment Protection and Immunity Granted by Section 230 of the Communications Decency Act to Social Media Platforms

There are a few legal strategies that can be used to regulate harms created in the online world that surmount the obstacles created by the First Amendment and Section 230 of the CDA.[149] First, the FTC or states' attorneys general could bring claims against social media companies for unfair or deceptive business practices. Second, products liability lawsuits could be brought against social media platforms, though such suits may benefit only a few people and only after harm has occurred. Lastly, states could pass legislation based on products liability theory that would require a study of social media design functions and the reform of those functions to prevent harm to users.[150]

A. Unfair or Deceptive Business Practice Claims Brought Against Social Media Companies by the FTC or States' Attorneys General Could Withstand a First Amendment Free Speech Defense and Circumvent Section 230 of the Communications Decency Act

One approach to regulating the harms that children and teens experience due to social media use is applying Section 5 of the FTC Act, which declares unlawful "unfair or deceptive acts or practices in or affecting commerce."[151] The FTC finds that an act or business practice is unfair where "(1) the act or practice causes or is likely to cause substantial injury to consumers which (2) is not reasonably avoidable by consumers themselves and (3) not outweighed by countervailing benefits to consumers or to competition."[152] In addition, the FTC finds that an act or business practice is deceptive where (1) a representation, omission, or practice misleads or is likely to mislead the consumer; (2) a consumer's interpretation of the representation, omission, or practice is considered reasonable under the circumstances; and (3) the misleading representation, omission, or practice is material.[153]

The FTC Act does not grant a private right of action; enforcement of the FTC Act can only be achieved through the FTC itself.[154] An action could be brought under Section 5 of the FTC Act if it can be proven that the persistent algorithmic pushing of harmful content, such as eating disorder content shown to a user through the social media platform's algorithm-driven feed, meets the definition of an "unfair or deceptive

[149]This article does not explore the legal remedies that provide a less examined solution to preventing harm inflicted on minors by social media platforms, but they are mentioned here. One remedy is taxation. The Maryland Digital Advertising Gross Revenues Tax is the nation's first tax on the revenue from digital advertisements that are sold by social media platforms displayed inside the state of Maryland. The tax went into effect on January 1, 2022 and was projected to gain $250 million dollars in its first year of implementation. This tax has been challenged on constitutional grounds in federal and state court. Similar tax legislation has been introduced in five other states. *See* David McCabe, *Maryland Approves Country's First Tax on Big Tech's Ad Revenue*, New York Times (Feb. 12, 2021), https://www.nytimes.com/2021/02/12/technology/maryland-digital-ads-tax.html [https://perma.cc/THE8-WKVA].

[150]Another potential legal remedy is the withholding of government contracts from platforms which fail to uphold and implement standards to keep their platforms safe for children and teens users. The San Francisco Green Building Code is an example. Under the Code, developers of buildings who do not comply with the standards that ensure that buildings are healthy and sustainable places to live and work will not be afforded government contracts. Ord. 3-20, File No. 190974 (2020). This same practice could be adopted to withhold government contracts from social media companies which fail to provide a safe platform to teens and children.

[151]15 U.S.C. § 45(a)(1).

[152]15 U.S.C. § 45(n).

[153]Chairman James C. Miller III, *FTC Policy Statement on Deception*, Federal Trade Commission 1, 1–2, (Oct. 14, 1983), https://www.ftc.gov/system/files/documents/public_statements/410531/831014deceptionstmt.pdf [hereinafter *FTC Policy Statement on Deception*].

[154]*What the FTC Does*, Fed. Trade Comm'n, https://www.ftc.gov/news-events/media-resources/what-ftc-does [https://perma.cc/4JNA-B6ES] (last visited Apr. 17, 2023).

business practice," regardless of the platform's intent to harm the user. When an action is brought under Section 5 alleging unfair or deceptive business practices, the defendant may not use good faith as a defense because intent to deceive the consumer is not an element of the claim.[155]

States' attorneys general offices can also bring claims of unfair or deceptive business practices against social media companies, because the FTC assigns certain enforcement authority to states in this area.[156] State consumer protection laws also grant attorneys general significant authority to bring such claims.[157] The FTC Act has prohibited unfair or deceptive acts and practices since 1938, and states followed suit in the 1970s and 1980s when they began to adopt their own forms of consumer protection statutes, largely modeled after the FTC Act.[158]

A multi-state investigation against TikTok and Meta was launched in 2022 by attorneys general in eight states; it focused on the methods and techniques used by social media companies to boost engagement among young users.[159] Specifically, attorneys general are examining the methods used to increase the duration of time spent on the platforms as a means to uncover the harm such usage may cause young people and what social media companies know about those harms.[160] In the investigation, attorneys general will likely seek disclosure and reform from social media companies related to the effect algorithmic operations have on adolescent users,[161] as well as gathering information on a growing number of public health studies that examine mental health harms suffered by young users of social media.[162]

States' attorneys general, however, may struggle to bring a successful claim against social media platforms because of the difficulty in proving that harms suffered by young social media users are caused by unfair or deceptive business practices such as algorithms to push harmful content. Most difficult to prove would be the first element of an unfair practice claim, which requires "the act or practice causes or is likely to cause substantial injury to consumers,"[163] and the third element of a

[155]Federal Trade Commission v. LeadClick Media, LLC, 838 F.3d 158, 168 (2d Cir. 2016) (quoting F.T.C. v. Verity Intern., Ltd., 443 F.3d 48, 63 (2d Cir. 2006)).

[156]*See generally* Ryan Strasser et al., *State AGs Lead the Way in False Advertising Enforcement*, Troutman Pepper Law Firm (Feb. 2, 2022), https://www.troutman.com/insights/state-ags-lead-the-way-in-false-advertising-enforcement.html [https://perma.cc/BKZ6-5R3T].

[157]*Id.*

[158]*Id.* Today, each of the fifty states and the District of Columbia has some form of a consumer protection law, often referred to as the state's "Unfair and Deceptive Acts and Practices Act" (UDAP) or "Consumer Protection Act" (CPA). Generally, these state consumer protection laws prohibit deceptive practices in consumer transactions, and although the substance of the statutes varies widely from state to state, many also prohibit unfair or unconscionable practices. State UDAPs and CPAs are primarily civil statutes, but others also create criminal penalties for severe violations. *Id.*

[159]Matthew Lewis, *The Role of the Attorney General in Reforming Social Media for Children*, N.Y. J. of Legis. & Pub. Pol'y (Oct. 10, 2022), https://nyujlpp.org/quorum/lewis-how-state-attorneys-general-can/ [https://perma.cc/MTG8-MM8N]. The multi-state investigation includes attorneys general offices in Massachusetts, California, Florida, Kentucky, Nebraska, New Jersey, Tennessee, and Vermont. *Id.* Attorneys general in forty-two states filed legal actions against Meta in October 2023 alleging it violated consumer protection laws by unfairly ensnaring children and deceiving users about the safety of its platforms. *See* Cecilia Kang & Natasha Singer, *Meta Accused by States of Using Features to Lure Children to Instagram and Facebook*, New York Times (Oct. 23, 2023), https://www.nytimes.com/2023/10/24/technology/states-lawsuit-children-instagram-facebook.html [https://perma.cc/3GH9-YXQQ].

[160]Lewis, *supra* note 159. Congress has introduced legislation aimed at curbing alleged harms inflicted on youth by social media platforms, but it has met considerable opposition. Legal analysts suggest it will be more effective for state attorneys general to pursue claims social media companies to alleviate harms. *Id.*

[161]*Id.* "State attorneys general are equipped in three ways to serve the public interest and address the harms of social media against children: (A) investigation and litigation against social media platforms; (B) advocating for policy reform in their state legislatures, Congress, and directly to platforms; and (C) educating the public. Attorneys general have broad power to subpoena documents and compel testimony by social media company executives to force disclosure on all information related to the operation of algorithms and their effect on adolescent users from social media platforms." *Id.*

[162]Members of the Strategic Training Initiative for the Prevention of Eating Disorders (STRIPED), who are the authors of this article, met with attorneys general offices in more than 12 states in 2022, including several involved in the multi-state investigation, to discuss the harmful effects of social media algorithms on youth, identify the economic incentives that drive social media companies to use them, and look at possible legal strategies to regulate social media platforms' use of algorithms.

[163]15 U.S.C. § 45(n).

deceptive practice claim, which requires "the misleading representation, omission, or practice is material."[164]

Numerous public health studies (explored earlier in this article) can undoubtedly establish an association, and these studies strongly suggest causation between social media use by young people and the mental health harms they suffer. However, risk audits of social media platforms—and specifically legislation that would require those audits—are necessary to show that the algorithmic function of social media platforms is directly linked to substantial harms to young users, many of whom suffer from body dissatisfaction, eating disorders, substance abuse, anxiety, depression, self-harm, and suicidality.

1. Product Liability Claims Focused on Harmful Design of Social Media Platforms Could Withstand a First Amendment Challenge and Circumvent Section 230 of the Communications Decency Act

A second legal remedy that could surmount the obstacles imposed by the First Amendment and Section 230 immunity would be a products liability claim, brought under a negligence theory. A plaintiff bringing a products liability lawsuit against a social media platform would need to allege that the social media platform is harmful and that the platform knew or should have known that the design of the product, the social media app or website, would cause harm.[165] The plaintiff would need to show that the defendant is not immune from liability under Section 230 of the CDA in order to be successful.[166] For example, in *Lemmon v. Snap, Inc.*, plaintiffs alleged that their two sons died in a high-speed car crash due to the negligent design of the Snapchat app, which, through the use of a Speed Filter, encouraged their sons to drive at excessively high speeds while the app measured and displayed their speed in real time.[167] The court held that Snap Inc., as a products manufacturer, had a duty to design a reasonably safe product.[168] Moreover, the court found that Snapchat was not immune from liability under Section 230 of the CDA because its duty to design a reasonably safe product was separate from its role in monitoring and publishing third-party content.[169] Another product liability case, discussed earlier, is being pursued by Tammy Rodriguez, the mother of an eleven-year-old suicide victim, alleging that Meta and Snap must be held liable for the wrongful death of her daughter, Selena Rodriguez.[170] Ms. Rodriguez alleges that Meta and Snap "knowingly" and "purposefully" designed their platforms to be addictive, making these platforms unreasonably dangerous to minor users, such as Selena.[171] Under a products liability strategy, companies such as Meta and Snap may be held liable for the physical and mental harm to users of their platforms if the court finds that the company knew or should have known that the design of the platform posed unreasonable dangers.[172] A products liability strategy circumvents both the First Amendment free speech protections and Section 230 of the CDA because it does not challenge the content and speech found on a platform but, instead, cites fault with the harmful design of the platform itself.[173]

---

[164]*FTC Policy Statement on Deception*, *supra* note 138. A material misrepresentation or practice is defined as a misrepresentation or practice "which is likely to affect a consumer's choice of or conduct regarding a product." *Id.*

[165]Allison Zakon, *Optimized for Addiction: Extending Product Liability Concepts to Defectively Designed Social Media Algorithms and Overcoming the Communications Decency Act*, 2020 Wis. L. Rev. 1107, 1118–19 (2020).

[166]*Id.* at 1119-21.

[167]Lemmon v. Snap, 995 F.3d 1085, 1087 (9th Cir. 2021).

[168]*Id.* at 1093.

[169]*Id.* at 1087.

[170]Jason Ysais, *Meta Platforms, Inc. and Snap, Inc. Face Wrongful Death Lawsuit for Causing the Suicide of 11-year-old Selena Rodriguez*, Social Media Victims Law Center (Jan. 20, 2021), https://socialmediavictims.org/press-releases/rodriguez-vs-meta-platforms-snap-lawsuit [https://perma.cc/78TN-8QX9].

[171]*Id.*

[172]Zakon, *supra* note 165, at 1118–19.

[173]The U.S. Surgeon General's Advisory in 2023 suggested using a multifaceted approach, including a products liability strategy, to curb the harms caused to young people by social media. "The U.S. has often adopted a safety-first approach to mitigate the risk of harm to consumers. According to this principle, a basic threshold for safety must be met, and until safety is demonstrated with rigorous evidence and independent evaluation, protections are put in place to minimize the risk of harm from products, services, or goods. For example, the Consumer Product Safety Commission requires toy manufacturers to undergo third-party testing and be certified through a Children's Product Certificate as compliant with the federal toy safety standard for toys intended for use by children….Given the mounting evidence for the risk of harm to some children and

Relying on a products liability claim to combat the harms inflicted on children by social media has its limitations, however. Product liability cases typically involve only a single plaintiff, or a specific class of plaintiffs in a class action suit. At best, product liability cases are a reactive legal strategy that address harms occurring in the online world only *after* the harms have occurred. A product liability claim does not address the continuing or future harm social media platforms inflict on the general public of users.

In addition, proving that a faulty design of a product was the direct cause of an injury can be difficult. Generally, for a products liability case to be successful, a plaintiff must prove: (1) the product caused them to be injured; (2) the product that injured them was defective; (3) the defect of the product is what caused their injury; and (4) the product was being used the way it was intended.[174] It is not enough to argue that one was injured while using the defective product correctly; the plaintiff must also demonstrate specifically that their injury was *caused* by the defect itself.[175] In some cases, linking the defect in the product to the injury is fairly straightforward, but in other cases, it is not. That is specifically the problem in cases where young persons experience harm by engaging with social media platforms. Public health studies can conclusively demonstrate an *association* between social media use and the mental health harms suffered by young social media users, but direct *causation* of harm is harder to prove, but a groundswell of reputable public health, psychology, and neuroscience studies in recent years go a long way in directly linking social media use to severe harms suffered by young social media users. Mandatory algorithm risk audits of social media platforms, discussed in detail later in this article, would delineate that faulty and intentional design of social media platforms cause many harms experienced by youth.[176]

2. Public Nuisance Theory Brought Against Social Media Giants by School Districts Could Restrain Platforms From Targeting Addictive Social Media Algorithms at Minors.

A third legal remedy that could circumvent First Amendment speech protections and Section 230 immunity is the tort theory of public nuisance. In January 2023, Seattle School District No. 1 (Seattle Schools) brought a case against Meta, Snapchat, TikTok, and YouTube, alleging that the social media platforms "intentionally marketed and designed their social media platforms for youth users, substantially contributing to a student mental health crisis."[177] Seattle Schools specifically allege that the four social media platforms intentionally design their platforms to maximize the time youth users spend on the platform with the use of harmful algorithms.[178] Furthermore, Seattle Schools allege the harm to adolescent mental health is reasonably foreseeable.[179]

Seattle Schools brought this complaint because the school district is a primary provider for mental health services to children and teenagers who are specifically targeted by social media platforms.[180] In

adolescents from social media use, a safety-first approach should be applied in the context of social media products." *US Surgeon General's Advisory*, *supra* note 73.

[174]David Goguen, *Proving a Defective Liability Claim*, NOLO, https://www.nolo.com/legal-encyclopedia/proving-defective-product-liability-claim-29531.html [https://perma.cc/ZM33-H4DX].

[175]*Id.*

[176]The U.S. Surgeon General recommends that social media companies "[c]onduct and facilitate transparent and independent assessments of the impact of social media products and services on children and adolescents [and] assume responsibility for the impact of products on different subgroups and ages of children and adolescents, regardless of the intent behind them." The Surgeon General also recommends that results of independent assessments "be transparent" and that social media companies share assessment findings and underlying data with independent researchers and the public. The Surgeon General urges that platform design and algorithms should prioritize health and safety as the first principle, seek to maximize the potential benefits, and avoid design features that attempt to maximize time, attention, and engagement. Issued periodically, a US Surgeon General's Advisory is a public statement that calls the American people's attention to an urgent public health issue and provides recommendations for how it should be addressed. Advisories are reserved for significant public health challenges that require the nation's immediate awareness and action. *US Surgeon General's Advisory*, *supra* note 73.

[177]Complaint at 23, Seattle School District No. 1 v. META (Case 2:23-cv-00032).

[178]*Id.* at 1.

[179]*Id.* at 87.

[180]*Id.* at 73.

2023, there were 109 schools within the Seattle Schools district with a population of 53,873 students.[181] Seattle Schools claims harmful social media algorithms have created a mental health crisis for children and teens and Seattle Schools have struggled to provide adequate mental health services to adolescents in their schools to meet the growing need.[182] In a similar suit, in April 2023, Dexter Community Schools in Washtenaw County, Michigan joined a lawsuit with at least eleven other Michigan schools against major social media platforms, including Meta, Snapchat, TikTok, and YouTube.[183] Plaintiffs are seeking damages for past and future harm resulting from social media addiction and for funding for school counselors to address the mental health crisis resulting from high social media use.[184]

Seattle Schools brings its complaint under a public nuisance theory. Under the relevant code, public nuisance is defined as "whatever is injurious to health or indecent or offensive to the senses, or an obstruction to the free use of property, so as to essentially interfere with the comfortable enjoyment of life and property."[185] A public nuisance occurs when someone commits an act or performs a duty that "annoys, injures, or endangers the comfort, repose, health or safety of others, offends decency…or in any way renders other persons insecure in life, or in the use of property."[186] It impacts an entire community.[187] In the past, public nuisance claims have been used to address pollution, road obstructions, and operating houses for prostitution.[188] More recently, public nuisance has been used to litigate claims regarding climate change, gun violence, and teen vaping.[189]

Because a public nuisance is experienced by the entire community, a plaintiff has standing to sue under this theory if they are one of the following: (1) a public authority charged with the responsibility of protecting the public; or (2) an individual who has suffered harm from the specific nuisance.[190] In this case, Seattle School District No. 1 has standing to bring this claim under the first category of plaintiffs.

When defending against a public nuisance allegation, a defendant can assert various defenses: contributory negligence, assumption of the risk, coming to the nuisance, or statutory compliance.[191] If these defenses fail and the plaintiff prevails, the typical remedy the court awards is damages.[192] In some cases, an injunction may be appropriate, wherein the defendant would be restrained for continuing the wrongful conduct.[193] Defendants may be fined for committing a public nuisance in addition to the court issuing an injunctive order.[194] While a public nuisance claim may help many students, especially if an injunction is ordered, it addresses harm in only the specific school districts in which the suit is brought. Thus, the number of young people who are protected is limited.

Juul, a company that sells electronic cigarettes, has faced ongoing litigation from school districts, cities, and counties across the nation under the public nuisance theory for contributing to nicotine addiction

[181] *Seattle School District No. 1*, Pub. Sch. Rev., https://www.publicschoolreview.com/washington/seattle-school-district-no-1/5307710-school-district [https://perma.cc/94NQ-5LRR] (last accessed Apr. 19, 2023).

[182] Complaint at 74, Seattle School District No. 1 v. META (Case 2:23-cv-00032).

[183] Isabel Lochman, *Dexter schools sue social media giants, citing child mental health crisis*, Bridge Michigan (Apr. 14, 2023), https://www.bridgemi.com/talent-education/dexter-schools-sue-social-media-giants-citing-child-mental-health-crisis [https://perma.cc/Q7E4-TZWF].

[184] *Id.*

[185] Complaint at 85, Seattle School District No. 1 v. META (Case 2:23-cv-00032).

[186] RCW 7.48.120.

[187] RCW 7.48.130.

[188] *Public nuisance*, Britannica, https://www.britannica.com/topic/nuisance [https://perma.cc/6Y8X-5XXV] (last accessed Apr. 19, 2023).

[189] Gene Johnson, *Schools sue social media companies for targeting children*, Kare (Jan. 11, 2023, 4:46 AM), https://www.kare11.com/article/news/nation-world/schools-sue-social-media-companies/507-3fcdc58b-deaa-4f84-8594-57d1cda667b0 [https://perma.cc/VM8E-WV4S].

[190] *Tort Law: The Rules of Public Nuisance*, Law Shelf, https://lawshelf.com/shortvideoscontentview/tort-law-the-rules-of-public-nuisance [https://perma.cc/EVD3-NJD9] (last accessed Mar. 1, 2023).

[191] *Nuisance*, Cornell Law School, https://www.law.cornell.edu/wex/nuisance#:~:text=A%20public%20nuisance%20is%20when,through%20a%20thing%20or%20activity https://perma.cc/XP73-38PB (last accessed Feb. 27, 2023).

[192] *Id.*

[193] *Nuisance*, JR Rank, https://law.jrank.org/pages/8871/Nuisance-Remedies.html [https://perma.cc/4LXR-C6P3] (last accessed Apr. 19, 2023).

[194] *Id.*

among adolescents.[195] The first of these was brought in Massachusetts in 2018.[196] Plaintiffs in these lawsuits alleged that Juul marketed e-cigarettes to youth, using false and misleading language describing e-cigarettes as fun and safe for adolescents.[197] The Connecticut Attorney General claimed that Juul "relentlessly marketed vaping products to underage youth, manipulated their chemical composition to be palatable to inexperienced users, employed an inadequate age verification process and misled consumers about the nicotine content and addictiveness of its products."[198] In December 2022, Juul agreed to settle the claims coming from 10,000 lawsuits for a sum of $1.2 billion.[199] Similarly, in a lawsuit involving six states, in April 2023, Juul agreed to settle claims that it unlawfully marketed addictive products to minors for $462 million.[200] In March 2023, Juul also agreed to settle a complaint brought under the public nuisance theory by Minnesota in December 2019, however the terms of the settlement have not yet been released.[201]

While public nuisance claims against Juul have been successful, allegations that Juul is unlawfully marketing an addictive product—electronic cigarettes—to teens is vitally different from allegations that social media giants are targeting an addictive product—social media— to teens. It is illegal to encourage minors to use electronic cigarettes and consume nicotine; it is completely legal to encourage minors to use social media. Seattle Schools, and any other school districts bringing a public nuisance claim, will have difficulty proving that the social media platforms are involved in illegal conduct. Furthermore, Seattle Schools will need to prove that the social media companies *caused* the mental health crisis among youth in the school district, not merely that there is a correlation between negative mental health and increased social media use.[202]

VI. The California Age Appropriate Design Code Leads the Nation in Attempts to Address Harms to Young People Inflicted by Social Media Platforms and Should Have Ripple Effects Nationally

The passage of the California Age Appropriate Design Code (California Code) on September 15, 2022, constituted a significant step forward in the United States to combat online harms to children and adolescent users, including online content that contributes to eating disorders, depression, anxiety, social media addiction, and other mental health harms.[203] The law thwarts First Amendment challenges

[195]Joe Toppe, *Juul Labs will pay $1.2B for role in youth-vaping epidemic*, Fox Bus. (Dec. 9, 2022, 1:39 PM), https://www.foxbusiness.com/markets/juul-labs-pay-one-point-two-billion-role-youth-vaping-epidemic; https://www.kare11.com/article/news/nation-world/schools-sue-social-media-companies/507-3fcdc58b-deaa-4f84-8594-57d1cda667b0 [https://perma.cc/B35S-TL92].

[196]Ty Roush, *Juul To Pay $1.2 Billion To Settle Youth-Vaping Lawsuits*, Forbes (Dec. 9, 2022, 1:28 PM) https://www.forbes.com/sites/tylerroush/2022/12/09/juul-to-pay-12-billion-to-settle-youth-vaping-lawsuits/?sh=6d8cad09345c [https://perma.cc/W529-JUYR].

[197]Angelica LaVito, *Lawmaker accuses Juul of illegally advertising vaping as a way to quit smoking*, CNBC (Sept. 5, 2019, 3:23 PM), https://www.cnbc.com/2019/09/05/juul-accused-of-illegally-advertising-vaping-as-a-way-to-quit-smoking.html [https://perma.cc/8ZGH-57PM].

[198]Roush, *supra* note 196.

[199]Reuters, *Juul agrees to pay $1.2bln in youth-vaping settlement-Bloomberg News*, Reuters (Dec. 9, 2022, 11:47 AM), https://www.reuters.com/legal/juul-agrees-pay-12-bln-youth-vaping-settlement-bloomberg-news-2022-12-09/ [https://perma.cc/WCQ2-Q9NM].

[200]Reuters, *Juul to pay $462 million to California, New York, and other states over claims it marketed vapes to minors*, NBC News (Apr. 12, 2023, 1:09 PM), https://www.nbcnews.com/health/health-news/juul-to-pay-462-million-claims-marketed-vapes-minors-rcna79375 [https://perma.cc/FY9Y-XK4X].

[201]*See* Ananya Bhattacharya, *Minnesota is trying to prove Juul got teens addicted on vaping in a first-of-its-kind trial*, Quartz (Mar. 24, 2023), https://qz.com/minnesota-jull-altria-trial-public-nuisance-vaping-1850260880; *Juul's Trial in Minnesota End With a Settlement*, CS News (Apr. 18, 2023), https://www.csnews.com/juuls-trial-minnesota-ends-settlement [https://perma.cc/8847-EKZQ].

[202]Julian Shen-Berro, *As Seattle schools sue social media companies, legal experts split on potential impact*, Chalk Beat (Jan. 17, 2023, 6:00 AM), https://www.chalkbeat.org/2023/1/17/23554378/seattle-schools-lawsuit-social-media-meta-instagram-tiktok-youtube-google-mental-health [https://perma.cc/3NUE-JTNG].

[203]Samantha Gross, *The California Age-Appropriate Design Code Act Places New Obligations on Companies Collecting Information About Children Online*, JD Supra (Feb. 24, 2023), https://www.jdsupra.com/legalnews/the-california-age-appropriate-design-1510066/ [https://perma.cc/35FZ-U3VQ].

because it does not regulate the content of speech on social media platforms, but instead focuses on the functional design of the platforms that cause harm, essentially applying a products liability theory, but more broadly.[204]

The California Code sets forth certain standards that social media platforms must comply with. For example, it mandates that a social media company must conduct a Data Protection Impact Assessment for services or platforms likely to be accessed by consumers younger than the age of 18, [205] establish the age of consumers using the platform with a level of certainty,[206] and ensure that minor users' platform websites and apps are set to the highest level of privacy possible.[207]

Furthermore, the Code prohibits social media platforms from using private information of a child user in a way that is harmful to the physical and mental health of the child,[208] collecting the geolocation information of a user,[209] and using deceptive design functions, such as targeted advertising, and exposing children to harmful content and contacts that pressure children to provide personal, private information beyond what is necessary.[210] These are only a few of the many important standards social media companies must comply with under the California Code. To implement and enforce these standards, the California Code requires the establishment of the California Children's Data Protection Task Force.[211] The California Code was signed into law in September 2022[212] and goes into effect on July 1, 2024.[213]

Of the many pieces of legislation filed in the United States aimed at addressing online harms, specifically those harms that impact minor users on a platform, the California Code will be perhaps the most influential. Currently, social media platforms must follow COPPA, which imposes requirements on online service operators to protect the privacy of users under the age of thirteen years.[214] Where COPPA places the burden on parents to take control of their child's privacy online, in contrast, the California Code places the burden on the social media platform to create services and devices that are safe for the physical and mental wellbeing of children.[215] California is a leader in U.S. law regarding technology and privacy rights and is a technology hub in itself; thus, it is likely that the California Code will have a ripple effect throughout the nation in changing the structure and design of social media platforms for the better.

The California Code is modeled after the United Kingdom Age Appropriate Design Code (UK Code), which has inspired significant change by social media companies. The UK Code came into full force on

[204]*See generally*, Meg Crowley, *California's New Age-Appropriate Design Code Act: Violation of Free Speech?*, Hastings Commc'ns & Ent. J. (Jan. 26, 2023), https://www.hastingscomment.org/online-content/californias-new-age-appropriate-design-code-violation-of-free-speechnbsp [https://perma.cc/BA7C-HXCK].

[205]The California Age Appropriate Design Code Act, AB 2273, State Assemb. 2021-2022 Sess. (Ca. 2022) (1).

[206]*Id.* at §1798.99.31(a)(5).

[207]*Id.* at §1798.99.31(a)(6).

[208]*Id.* at §1798.99.31(b)(1).

[209]*Id.* at §1798.99.31(b)(5).

[210]*Id.* at §1798.99.31(b)(7).

[211]*Id.* at §1798.99.32(a). The California Data Protection Working Group will be assembled by April 1, 2023, and will sanction regulations under the Age Appropriate Design Code by April 1, 2024. Members of the Taskforce will be appointed by the California Privacy Protection Agency (CPPA). The taskforce will be "Californians with expertise in the areas of privacy, physical health, mental health, and well-being, technology, and children's rights." *See, e.g.*, Chloe Altieri & Kewa Jiang, *California Age-Appropriate Design Code Aims to Address Growing Concern About Children's Online Privacy and Safety*, Future of Privacy Reform (June 28, 2022), https://fpf.org/blog/california-age-appropriate-design-code-aims-to-address-growing-concern-about-childrens-online-privacy-and-safety/ [https://perma.cc/PA2F-N9KH].

[212]Megan Brown et al., *California Age-Appropriate Design Code Act to Impose Significant New Requirements on Businesses Providing Online Services, Products, or Features*, JD Supra (last updated Sept. 19 2022). https://www.jdsupra.com/legalnews/california-age-appropriate-design-code-8105166/ [https://perma.cc/MS6B-VS8R].

[213]*Id.* The California Code has been challenged by tech companies in court so implementation of the law may be forestalled.

[214]Katie Terell Hanna, *COPPA (Children's Online Privacy Protection Act)*, Tech Target (Mar. 2022), https://www.techtarget.com/searchcio/definition/COPPA-Childrens-Online-Privacy-Protection-Act [https://perma.cc/ZJ3B-AP5B].

[215]Musadiq Bidar, *California lawmakers push ahead with sweeping children's online privacy bill*, CBS News (May 12, 2022, 1:50 PM), https://www.cbsnews.com/news/online-privacy-california-age-appropriate-design-code-teens-children/ [https://perma.cc/L828-XQA2].

September 2, 2021, after a twelve-month transition period.[216] Since the UK Code came into effect, many social media platforms have made changes to their services and devices to comply with its requirements. For example, TikTok has turned off notifications past bedtime for children less than thirteen years old and has provided safe search mechanisms as a default, Instagram has disabled targeted advertisements for minor users, YouTube has disabled autoplay for minor users, and Google has stopped targeted advertising for minor users.[217] Given the success of the UK Code, passage of the California Code should produce similar instrumental changes in the United States.

A . California's Data Protection Impact Assessment Requirement to Measure Social Media Harms on Young Users Is Well Intentioned but Very Limited Because It Relies on Social Media Companies to Assess Themselves

The California Code requires businesses to complete a Data Protection Impact Assessment before a new online service, product, or feature is offered to the public, and to maintain documentation of this assessment.[218] A Data Protection Impact Assessment is a "systematic survey" that assesses and mitigates risks that "arise from data management practices of the business to children who are reasonably likely to access the online service, product, or feature at issue."[219] Specifically, the assessment will address whether a product, service, or feature could harm children or expose children to harmful content, could lead children to experiencing harmful contacts, could permit children to witness or participate in harmful conduct, or whether algorithms used and whether targeted advertisements could harm the child.[220]

However, the Data Protection Impact Assessments are confidential and will not be publicly disclosed to anyone other than the California Attorney General's Office.[221] Moreover, the Data Protection Impact Assessments will be conducted internally by a social media company, instead of by a third party.[222] For example, Facebook would be in charge of running a Data Protection Impact Assessment of its own design on its own platform. These two factors drastically weaken the impact the Data Protection Impact Assessments could have, particularly because the social media platform would not be subject to outside scrutiny. California should replace the assessments with algorithm risk audits that are conducted by independent third parties and are required to be publicly disclosed, therefore providing for greater accountability and enforceability of the California Code objectives.

The California Code is arguably the strongest state law in the United States that addresses the mental health of children and teens and the role social media plays. The First Amendment and Section 230 of the CDA have been barriers to legislation aimed at regulating online harms caused by social media platforms.[223] However, the California Code circumvents the First Amendment and Section 230 of the CDA by regulating the design and function of social media platforms, rather than the content or speech posted on the social media platforms. In this way, the California Code employs a products liability

[216]Natasha Lomas, *UK now expects compliance with children's privacy design code*, Tech Crunch (Sept. 1, 2021, 7:01 PM), https://techcrunch.com/2021/09/01/uk-now-expects-compliance-with-its-child-privacy-design-code/ [https://perma.cc/8R5N-UKLV].

[217]*Id.*; Alex Hern, *Social media giants increase global child safety after UK regulations induced*, The Guardian (Sept. 5, 2021, 10:14 AM), https://www.theguardian.com/media/2021/sep/05/social-media-giants-increase-global-child-safety-after-uk-regulations-introduced [https://perma.cc/7KVJ-FC84]; *Google Announcement Shows Impact of Children's Code*, 5Rights Found. (Aug. 10, 2021), https://5rightsfoundation.com/in-action/google-announcement-shows-impact-of-childrens-code.html [https://perma.cc/XX28-UZPD].

[218]The California Age Appropriate Design Code Act §1798.99.30(a)(1)(A) (2022), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB2273 [https://perma.cc/2U4Y-W472].

[219]*Id.* at §1798.99.30(b)(2).

[220]*Id.* at §1798.99.30(b)(4).

[221]*Id.* at §1798.99.31(c). The Design Code appoints the California Attorney General's Office as the main enforcer of the state law. Similarly, the Attorney General may bring actions against businesses for unfair or deceptive practices, mirroring the claims the FTC can bring under Section 5 of the FTC Act. *See* Strasser et al., *supra* note 141.

[222]Ca. AB 2273 §1798.99.30(b)(2).

[223]Balbuzanova, *supra* note 110; Johnson, *supra* note 119.

theory, but has the potential to be more effective in curbing harms caused by social media platforms than a single products liability lawsuit such as *Lemmon v. Snap, Inc.* or *Google v. Rodriguez.*

The California Code is preventive rather than reactive in nature. Unlike products liability cases, which involve a singular plaintiff and are brought after a harm has already occurred, the California Code attempts to prevent online harms before they occur by requiring social media platforms to comply with certain standards. The law places a burden on social media platforms to create services and products that are safe for the mental and physical wellbeing of users, with specific attention to the vulnerabilities of children using their platforms. Unlike products liability suits that benefit only a handful of users or just one person, the California Code will likely have a broad impact on all young users of social media in California.

B . Social Media Platforms Must Institute a Reliable Age Verification Method to Protect Minors and Enact Laws to Assess the Injuries that Platforms Inflict on Young Users

While the California Code leads the nation in legislatively contemplating the harms caused to adolescents by social media, it falls short of identifying actual injury teens experience and does not go far enough to prevent those harms. For the California Code to be effective, platforms' age verification processes must be appropriately addressed.[224] COPPA commands that operators of online services restrict their platforms to users ages thirteen years or older, absent verifiable parental consent.[225] For example, both Facebook and Instagram require users to be thirteen years old or older to create an account, but implement this requirement only by asking for the user's birthdate during account creation.[226]

This COPPA regulation is not easy to enforce; many child users are able to evade the age requirements on social media platforms by simply misrepresenting their birth date when registering for an account.[227] Social media platforms must administer a mechanism to verify the age of minor users to a degree of certainty, including those minor users who have lied about their age. For example, Instagram is currently testing the following three options to verify the age of Instagram users: (1) the user must upload an image of their ID, (2) the user must record a video of themselves, or (3) the user must ask friends to verify their age.[228] While these options are currently being explored only in instances where an Instagram user attempts to change their age from under eighteen to eighteen years or older, they may be implemented by other social media companies to ensure all users are age thirteen years or older.[229]

---

[224] *What is the Age-Appropriate Design Code - and How is it Changing the Internet?*, Parent Zone (July 27, 2022), https://parentzone.org.uk/article/what-is-the-age-appropriate-design-code [https://perma.cc/7W82-3U5U].

[225] Paul Harper & Catherine Micallef, *How Old Do You Have to be to Have Facebook and Instagram Account? Social Media Age Restrictions Explained*, The U.S. Sun (June 8, 2022, 12:33 PM), https://www.the-sun.com/tech/289567/age-restrictions-facebook-snapchat-twitter-instagram/ [perma.cc/PY3Z-LE4M].

[226] *Id.*

[227] Ariel Fox Johnson, *13 Going on 30: An Exploration of Expanding COPPA's Privacy Protections to Everyone*, 44 Seton Hall Legis. J. 419, 448–49 (2020). Congress is currently considering legislation that would amend COPPA to strengthen protections related to the online collection, use, and disclosure of personal information of minors under age 17. Co-authored by U.S. Senators Edward Markey (D-MA) and Cassidy (D-LA), COPPA 2.0 would: (1) expand protections to teens age 13–16 by requiring their opt-in consent before data collection; (2) ban targeted advertising to all covered minors; (3) close a loophole in COPPA that allows sites and apps to turn a blind eye to young people using their services and evade compliance; (4) create an "eraser button" for parents and kids requiring companies to delete personal information of minors; 5) establish a "Digital Marketing Bill of Rights" that minimizes the amount of data collected and used on minors; and (6) enhance enforcement by establishing a Youth Privacy and Marketing Division at the Federal Trade Commission. *Children and Teens Online Privacy Protection Act: Legislation to Strengthen Privacy Protections for Minors*, Common Sense, https://www.commonsensemedia.org/sites/default/files/featured-content/files/coppa-2.0-one-pager-2023.pdf (last accessed on Aug. 2, 2023).

[228] *Introducing New Ways to Verify Age on Instagram*, Instagram (June 23, 2022), https://about.instagram.com/blog/announcements/new-ways-to-verify-age-on-instagram [https://perma.cc/YHQ8-P63M].

[229] *Id.*

Congress is currently focusing on the low cut-off age requirement that social media platforms impose upon users wishing to open accounts. On May 2, 2023, Senators Richard Blumenthal (D-CT) and Marsha Blackburn (R-TN) reintroduced to Congress the Kids Online Safety Act (KOSA). This bill seeks to give parents and users under seventeen the ability to opt out of algorithmic recommendations, prevent third parties from viewing a minor's data, and limit the time young people spend on a platform.[230] KOSA has received bipartisan support from U.S. Senators across the country and is endorsed by several mental health organizations and associations.[231]

KOSA lists a set of harms that social media companies must mitigate, including preventing the spread of content that promotes suicidal behaviors, eating disorders, substance use disorders, sexual exploitation, advertisements for certain illegal products (e.g. tobacco and alcohol), and other matters.[232] Mitigation efforts could include removing rewards given to young users for time spent on the platform or other features that result in compulsive usage.[233] KOSA also requires social media companies "to perform an annual independent, third-party audit that assesses the risks to minors."[234] This audit must be made public and must evaluate the risks to minors who use the platform.[235] The bill further requires platforms "to enable the strongest privacy settings by default" for kids.[236]

Unlike an earlier version of the bill proposed in 2022, "KOSA 2.0" addresses concerns that it could inadvertently cause harm to young people. Opponents of the earlier version of the bill expressed concerns that KOSA would create pressure to over moderate content and allow political agendas to influence what information was accessible to young people.[237] For example, if a young teenager has an eating disorder and is looking for resources to receive counseling or health care resources, such content, though beneficial, might be censored by the online platform to avoid liability. However, KOSA 2.0 includes protections for beneficial support services like the National Suicide Hotline, substance use disorder resources, and LGBTQ youth centers.[238] These safeguards ensure young people's access to such groups is not hindered by the bill's requirements.

Despite these changes, Big Tech groups and some civil liberty organizations, including the American Civil Liberties Association, oppose the legislation, raising concerns about young people's privacy and First Amendment rights.[239] While KOSA 2.0 has addressed many of the concerns of children's mental

[230]KOSA was previously introduced by Senators Blumenthal and Blackburn in February 2022. *See Blackburn, Blumenthal Introduce Bipartisan Kids Online Safety Act*, Marsha Blackburn (May 2, 2023), https://www.blackburn.senate.gov/2023/5/blackburn-blumenthal-introduce-bipartisan-kids-online-safety-act [https://perma.cc/U6AJ-CKVN]. Despite a unanimous, 28-0 vote, by the Commerce Committee, the bill failed to continue in the legislative process. *See id.*

[231]The latest version of KOSA has thirty-nine bipartisan co-sponsors and has endorsements from Common Sense Media, American Psychological Association, American Academy of Pediatrics, American Compass, Eating Disorders Coalition, Fairplay, Mental Health America, and Digital Progress Institute. *See id.*

[232]S.B. 1409, Gen. Sess. (2023-2024).

[233]*See id.* (requiring platforms to allow minor users the ability to access safeguards to "limit features that increase, sustain, or extend use of the covered platform by the minor, such as automatic playing of media, rewards for time spent on the platform, notifications, and other features that result in compulsive usage of the covered platform by the minor").

[234]*See id.*

[235]*See id.*

[236]*See id.*

[237]LGBTQ advocates, who viewed KOSA's language as too restrictive, voiced concern that such limitations would ultimately harm marginalized young people's ability to learn about important information that they otherwise could not gain access to. *See* Lauren Feiner, *Kids Online Safety Act may harm minors, civil society groups warn lawmakers*, CNBC News (Nov. 28, 2022, 12:01 AM), https://www.cnbc.com/2023/05/02/updated-kids-online-safety-act-aims-to-fix-unintended-consequences.html [https://perma.cc/AZP4-XQJF].

[238]The earlier version of the bill did not include these safeguards.

[239]The ACLU, which was opposed to the earlier version of the bill, expressed its continued opposition to KOSA 2.0, stating that "[KOSA] would ironically expose the very children it seeks to protect to increased harm and increased surveillance." Lauren Feiner, *Lawmakers update Kids Online Safety Act to address potential harms, but fail to appease some activists, industry groups*, CNBC News (May 2, 2023, 1:12 PM), https://www.cnbc.com/2023/05/02/updated-kids-online-safety-act-aims-to-fix-unintended-consequences.html [https://perma.cc/LDY7-MSPV]. (quoting ACLU Senior Policy Counsel Cody Venzke)). Additionally, NetChoice, a lobbying group for multinational technology companies including Google, Meta, TikTok and

health advocates, and would force social media companies to be transparent in its potentially harmful business practices, it is unclear whether the bill will garner the necessary legislative support (especially from the U.S. House of Representatives) to become a federal law.[240]

Laws must also be enacted to examine the calculated addictive design of social media platforms and to prevent platforms from targeting vulnerable adolescent users. Social media platforms typically have a three-step method that draws users in and makes it psychologically more difficult to set down the phone: (1) a trigger, such as a notification, which pushes the user to check their device; (2) an action, where the user "clicks" to open and use an application on their device; and (3) a reward, like a favorite or "like" on a post, that motivates continued engagement on the platform.[241] Minor users, such as eleven-year-old Selena Rodriguez, who took her own life in July 2021, are most vulnerable to social media addiction and the resulting mental and physical harms.[242] Tammy Rodriguez, Selena's mother, alleges in her suit against Instagram and Snapchat that the platforms were purposefully designed to "exploit human psychology" and addict users to their platforms; therefore, Instagram and Snapchat should be liable for the harm that resulted from Selena's addiction to their platforms.[243]

To protect young users like Selena, another piece of legislation, the Social Media Duty to Protect Children Act, was considered in California in 2022. The bill attempted to impose a duty upon social media companies to not addict children to their platform, but the bill did not pass.[244] With this bill, social media platforms would have been prohibited from addicting a child to a social media platform through "using a design, feature, or affordance that the platform knew, or by the exercise of reasonable care should have known, causes a child user, as defined, to become addicted to the platform."[245] The Duty to Protect Children Act took a direct and controversial path in holding social media accountable and, not surprisingly, big tech lobbyists worked hard to make sure the measure was defeated.[246] It was voted down by the California Senate in August 2022.[247]

A law similar to California's failed legislation, however, found success in Utah.[248] In March 2023, the Utah legislature adopted the Social Media Usage Amendments law that prohibits social media companies

---

Amazon, has continued to express concern regarding "how this bill would work in practice …" as it "still requires an age verification mechanism and data collection on Americans of all ages." *See id.* NetChoice has also sued California challenging its Age-Appropriate Design Code Act.

[240]When this article was published in 2023, KOSA and COPPA 2.0 were passed in the U.S. Senate Committee on Commerce, Sciente and Transportation. The bills could be moved to a Senate floor vote later in 2023.

[241]Larissa Sapone, *Moving Fast and Breaking Things: An Analysis of Social Media's Revolutionary Effects on Culture and its Impending Regulation*, 59 Duq. L. Rev. 362, 367–69 (2021).

[242]*Addiction*, Social Media Victims Law Center, https://socialmediavictims.org/social-media-addiction/ [https://perma.cc/A6EF-SWUS] (last accessed July 27, 2022).

[243]Megan Cerullo, *Mom sues Meta and Snap over her daughter's suicide*, CBS News (Jan. 21, 2022), https://www.cbsnews.com/news/meta-instagram-snap-mom-sues-after-daughter-suicide/ [https://perma.cc/BES8-3F2U]; Rodriguez Complaint, *supra* note 29.

[244]The California Social Media Duty to Protect Children Act, AB 2408, State Assemb. 2021-2022 Sess. (Ca. 2022).

[245]*Id.* (emphasis omitted).

[246]Evan Symon, *Bill to Punish Social Media Companies for Addictive Features for Minor Users Killed in Senate*, California Globe (Aug. 12, 2022) https://californiaglobe.com/articles/bill-to-punish-social-media-companies-for-addictive-features-for-minor-users-killed-in-senate/ [https://perma.cc/7RCU-9PJB].

[247]*Id.*

[248]In March 2023, Utah passed two new laws to protect minors from perceived harms caused by social media. One law, entitled the *Social Media Regulation Amendments*, requires social media platforms to verify the ages of account holders and enforces a digital curfew, from 10:30pm to 6:30am, for teen users. This law also requires social media companies to verify the age of users. Verification procedures will be determined by the Utah Division of Consumer Protection and may not be limited to government issued identification cards. Parental consent is also required for teen users to have a social media account, and parents or guardians are granted full access to their teen's account. *See* S.B. 152, Gen. Sess. (Utah 2023). Arkansas enacted a similar law, the *Social Media Safety Act*, in April 2023 which also requires age verification and parental consent. This law will be enforced by the Arkansas Attorney General's Office and prohibits teen users from having a social media account without the express permission of a parent or guardian. To verify the ages of users, the *Social Media Safety Act* requires social media companies to use a third- party vendor to "perform reasonable age verification," which includes checking a user's government issued identification card or other "commercially reasonable age verification method[s]." *See* S.B. 396, Gen. Sess. (Arkansas

from using practices, designs, or features that the company knows or should know would cause a young person to form an addiction to that social media platform.[249] To enforce this, the law gives the Utah Division of Consumer Protection the ability to audit the records of social media companies to determine compliance with the law and to investigate a complaint alleging a violation.[250] If a social media company is found to be in violation, the company is subject to civil penalties of "$250,000 for each practice, design, or feature of its platform shown to have caused addiction."[251] The social media company can also face penalties up to $2,500 for each teen user who is shown to have been exposed to the addictive practice, design, or feature. The court may also issue an injunction or award actual damages to the injured young person.

The law also creates a private right of action allowing individuals to social media companies for "any addiction, financial, physical, or emotional harm suffered by a Utah young person as a consequence of using or having an account on the social media company's platform."[252] Any minor who suffers such harms is entitled to an award of "$2,500 per each incident of harm" in addition to other relief the court deems necessary.[253] If a young user or account holder is under the age of sixteen, it is assumed that a harm is caused as a result of having or using a social media account unless it can be proven otherwise.[254]

In response to an alleged violation, the social media company can assert an affirmative defense to such penalties if a quarterly audit of its practices, designs, and features is conducted to detect potential addiction of a young user and the company corrects, within thirty days of the completion of an audit, any violation. The law does not *require* social media companies to conduct audits, but rather allows social media companies to *use* quarterly audits as a means to assert an affirmative defense.[255] The law did not specify how these audits would be conducted, but does suggest that social media companies would audit themselves.[256]

The Utah law, however, faces a significant legal battle because social media companies have filed suit in December of 2023 claiming free speech violations. Tech advocacy groups, established and funded by members of the Big Tech industry, including NetChoice, publicly opposed the passage of the law.[257] NetChoice has already sued to challenge California's Age-Appropriate Design Code Act for restricting young users' social media usage and may file a similar claim against the Utah legislation.[258]

2023). While well-intentioned, these laws face criticism for invading teen privacy and freedom of speech rights. Social media companies have yet to announce any plans to challenge these new laws, but it is anticipated the laws will face future legal battles.

[249]*See* H.B. 311, Gen. Sess. (Utah 2023). In this context, a "young person" refers to minors, anyone younger than eighteen years old.

[250]*Id.*

[251]*Id.*

[252]*Id.* Similarly, a law passed recently in Arkansas entitled, the *Social Media Safety Act* creates a private right of action for teen users to sue social media companies for any damages incurred by their access to social media platforms without the consent of their parent or guardian. The social media companies face a penalty of $2,500 per violation, in addition to other fees and damages ordered by a court. *See* S.B. 396, Gen. Sess. (Arkansas 2023).

[253]*See* H.B. 311, Gen. Sess. (Utah 2023).

[254]*See id.* (explaining that for users under the age of 16, "there shall be a rebuttable presumption that the harm actually occurred and that the harm was caused as a consequence of using or having an account").

[255]*Id.*

[256]*Id.* The Utah law states that "[a] social media company shall not be subject to a civil penalty for violating this section if the social media company, as an affirmative defense, demonstrates that the social media company: (i) instituted and maintained a program of at least quarterly audits of the social media company's practices, designs, and features to detect practices, designs, or features that have the potential to cause or contribute to the addiction of a minor user; and (ii) corrected, within 30 days of the completion of an audit described in Subsection (3)(b)(i), any practice, design, or feature discovered by the audit to present more than a de minimus risk of violating this section."

[257]*See* Bryan Scott, *Utah faces new lawsuit over social media restrictions for minors*, Salt Lake City Tribune (Dec. 18, 2023, 9:43 p.m.), https://www.sltrib.com/news/politics/2023/12/18/utah-faces-new-lawsuit-over-social/ [https://perma.cc/Q8AJ-TLZ9] (In the suit, Netchoice claims Utah's regulations unconstitutionally restrict the ability of minors and adults to access content that otherwise would be legal).

[258]NetChoice alleges that the California Age-Appropriate Design Code Act (CAADCA) infringes on users' privacy rights and the First Amendment. It also argues the CAADCA violates the Commerce Clause and is preempted by COPPA and Section 230. *See* Mot. for Prelim. Inj. at 1-7, NetChoice v. Bonta, No. 5:22-cv-08861-BLF (N.D. Cali. 2022).

The Utah law, and the bills that were defeated or watered down in California, serve as models of viable legal remedies to curb social media harm that could be employed elsewhere. The California Code is a significant step forward in attempts to reduce harm to minors using social media, but the defeat of the Social Media Duty to Protect Children Act demonstrated that bluntly identifying social media use as addictive and dangerous to children may be politically difficult to advance. Further, the Data Protection Impact Assessment requirement under the California Code, which had the potential to directly identify the social media functions that harm minors, was rendered toothless because the assessments will not be conducted by independent third-party auditors nor publicly disclosed. Similarly, the Utah Social Media Amendments law, while well-intentioned, is weakened by its apparent reliance on social media companies to conduct their own internal audits.

Essentially, social media companies are entrusted to police themselves, which will undoubtedly result in superficial auditing and ineffectual enforcement. To ensure laws impose the right regulations to alleviate harm to minors, and that social media companies comply by taking the best corrective action, the social media functions that pose the greatest risk to minors must be accurately assessed and the results made publicly available. Algorithm risk audits conducted by independent third parties that continuously measure harmful algorithmic practices directed toward minors who use social media should be required by legislation that is passed in tandem with a law similar to the California Code.

VII. Laws Requiring Algorithm Risk Audits Will Provide Compelling Evidence Linking Social Media's Use of Algorithms to Harm to Children and Thereby Enhance Enforcement of Laws Mandating Reform of Social Media Platforms

"The real problem of humanity is the following: We have Paleolithic emotions, medieval institutions and godlike technology."[259] *– E.O. Wilson*

This ponderance by the late E.O. Wilson, a pioneer of evolutionary biology, captures the uneven balance between human vulnerability and the technologically advanced spaces in which we spend so much time. In a world where social media technology is constantly developing and ultimately outpacing the ability for humans to navigate its effects, legislation must be implemented to protect minor users from the harms this technology can cause. To best draft such legislation, policymakers must fully understand what harms social media causes and the effects of these harms on young people. The most pernicious practice is arguably the use of algorithms that relentlessly direct targeted content to minors on their social media feeds.

Public health, psychology, and neuroscience studies clearly demonstrate an alarming rise in depression, anxiety, suicidality, and other mental illnesses among adolescents in the last decade[260] coinciding with the introduction of social media platforms[261] such as Instagram (2010), Snapchat (2011), and TikTok (2016), which are all heavily used by young people.[262] Therefore, any law aimed at protecting minors online must address how social media platforms employ algorithms in the function and design of their products. To be effective, the laws must incorporate enhanced means of enforcement, rather than mere prohibitions on particular acts. This objective could best be accomplished through the use of algorithm risk audits.

[259]Edward O. Wilson, Debate at the Harvard Museum of Natural History, Cambridge, Mass., (Sept. 9, 2009) https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00016553;jsessionid=0CDC082C53C019ACD4F203281506A378 [https://perma.cc/55PX-LCAV].

[260]Office of the Surgeon General (OSG). Protecting Youth Mental Health: The U.S. Surgeon General's Advisory 9 (2021).

[261]Jean 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 Psych. Sci. 3, 8–9 (2018).

[262]Emily Vogels et al., *Teens, Social Media and Technology 2022*, Pew Rsch. Center (Aug. 2022), https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/ [https://perma.cc/H7VK-TPXP].

Legally mandating algorithm risk audits is a relatively new strategy that is gaining traction nationally and globally.[263] New York City was among the first jurisdictions to mandate these types of audits, passing a law on December 11, 2021, that requires annual audits assessing bias in automated employment decision tools, which use algorithms to screen applicants for employment positions.[264] By requiring these audits, known as "bias audits," the law helps identify when an algorithm might intentionally or unintentionally weed out applicants based on certain demographics, such as race and gender.[265] The New York City law, which took effect on April 15, 2023, requires an impartial evaluation by an independent auditor, the results of which must be made publicly available.[266]

The bias audits will measure the disparate impact the use of algorithms has on a specific demographic by comparing the number of applicants from a specific demographic selected to move forward in the hiring process to the number of those in the most highly selected demographic.[267] For example, the bias audit might compare the number of applicants who are women selected to move forward in the hiring process to the number of applicants who are men, who were the most selected demographic. This comparison will allow the independent auditors to assess whether the use of algorithms in the hiring process disproportionately impacts a certain demographic, such as women.[268] The demographic categories the bias audits assess are gender, race/ethnicity, and intersectional (i.e., overlapping demographics, such as an applicant who is a woman of a minoritized race).[269]

Another real-world example to look to for guidance on how an algorithm risk audit might work is the recent settlement between Meta Platforms, Inc. (Meta) and the U.S. Department of Justice (DOJ). On June 21, 2022, the DOJ announced its entrance into a settlement agreement that resolved allegations that Meta engaged in discriminatory advertising in violation of the Fair Housing Act (FHA).[270] The agreement also resolved a lawsuit filed against Meta by the United States, which alleged that "Meta's housing advertising system discriminated against Facebook users based on their race, color, religion, [gender], disability, familial status, and national origin."[271] Meta was charged with unevenly displaying housing ads to Facebook users of certain FHA-protected demographics, such as gender and race.[272] The

[263]The European Union has legally mandated algorithm risk audits under the Digital Services Act passed in July 2022. It will take effect no later than January 1, 2024. The DSA imposes obligations on very large online platforms, with users in the European Union, to manage systemic risks through various means, including independent audits. It requires platforms to conduct internal annual risk assessments and implement reasonable, proportionate, and effective mitigation measures. The independent audits the DSA requires will result in publicly disclosed reports. Regulation on a Single Market for Digital Services and amending the Directive 2000/31/EC (Digital Serv. Act), Oct. 19, 2022, Eur. Parl. Doc. 2022/2065.

[264]20 NYCRR 871.

[265]*Id.* 871(a)(1), (b)(2).

[266]*Id.* 871(b)(1)-(2).

[267]This method, known as the selection rate, describes one way the impact ratio for a particular demographic category can be measured. Alternatively, the impact ratio can be measured using a scoring rate, which is applicable when an automated employment decision tool scores applicants. The scoring rate is determined by the rate at which individuals in a demographic category receive a score above the sample's median score. Under this model, the bias audits measure the disparate impact the use of algorithms has on a specific demographic category by comparing the scoring rate of applicants from that specific demographic to those in the demographic category with the highest scoring rate. City N.Y. Rules, tit 6, § 5-300 (2023).

[268]*Id.*

[269]*Id.* at § 5-301.

[270]The United States Attorney's Office for the Southern District of New York, *United States Attorney Resolves Groundbreaking Suit Against Meta Platforms, Inc., Formerly Known As Facebook, To Address Discriminatory Advertising For Housing*, Dep't of Just. (June 21, 2022), https://www.justice.gov/usao-sdny/pr/united-states-attorney-resolves-groundbreaking-suit-against-meta-platforms-inc-formerly [https://perma.cc/DG5N-569C].

[271]*Id.*

[272]*Id.* This lawsuit was based on an investigation and charge of discrimination by the Department of Housing and Urban Development, which found that all three aspects of Facebook's ad delivery system delivered housing ads based on FHA-protected characteristics. The complaint for the case against Meta challenged three key aspects of Meta's ad targeting and delivery system. First, the complaint alleged that "Meta enabled and encouraged advertisers to target their housing ads by relying on race, color, religion, sex, disability, familial status, and national origin to decide which Facebook users [would] be eligible, and ineligible, to receive housing ads." Second, the complaint alleged that Meta created an ad targeting tool—the Special

settlement between Meta and the DOJ required Meta to develop a new system to make housing ads more evenly displayed across race and gender groups, and therefore address the discrimination caused by its algorithms.[273]

The settlement set forth a three-step approach: (1) identify the specific harm, (2) determine how to measure the extent of harm, and (3) agree on reporting periods and benchmarks to reduce harm.[274] The first step was to identify the specific harm, which was the discrimination caused by housing ads being unevenly displayed to Meta users of certain demographics, namely gender and race, in violation of the Fair Housing Act. The second step—to determine how to measure the extent of harm—required Meta and tech experts to figure out how to measure Meta's data to assess the extent of the discriminatory harm. The discriminatory harm is shown through variances between the eligible and actual audiences for housing ads.

The eligible audience includes all users who (1) fit the targeting options selected by an advertiser for an ad, and (2) were shown one or more ads on a Meta platform over the past 30 days.[275] The actual audience includes all users in the eligible audience who actually viewed the specific ad.[276] Once these audiences are identified, a measurement is taken to determine the variance between them using a measurement method called the Earth Mover's Distance.[277] To conceptualize this measurement, think of side-by-side pie charts. One pie chart shows the eligible audience for a housing ad—suppose it is split fifty percent for male users and fifty percent for female users.[278] The other pie chart shows the actual audience for a housing ad—suppose this is split forty percent for male users and sixty percent for female users. To determine the variance, compare the differences between corresponding pieces of the pie charts.[279] The total variance is the sum of the differences between corresponding slices of the pie charts. Here, there is a ten percent difference for male users (fifty percent in the eligible audience chart and forty percent in the actual audience chart) and a ten percent difference for female users (fifty percent in the eligible audience chart and 60% in the actual audience chart).[280] Once the variance for each demographic is found, add the variances together and divide by two (since any decrease in one slice becomes an equivalent increase in another slice, so it is double-counted) to determine the total variance. In this case, the total variance is (10% + 10%) / 2 = 10%.[281]

As a word of caution, this calculation of the Earth Mover's Distance is quite simple, and works best in a context where demographic groups are of relatively equal size in the eligible population.[282] The metric, however, will be less useful under scenarios where demographic groups are of widely varying sizes, as is the case when comparing across racial/ethnic groups in the United States.[283] For this reason, it would be prudent for the Earth Mover's Distance metric to be supplemented with an additional metric to flag when any particular group, for instance, a small demographic group, experiences a

Ad Audience—which used an algorithm "to find Facebook users who share[d] similarities with groups of individuals selected by an advertiser using several options provided by Facebook."# In doing this, Meta "allowed its algorithm to consider FHA-protected characteristics—including race, religion, and sex—in finding Facebook users who 'look like' the advertiser's source audience." Third, the complaint alleged that Meta's ad delivery system used algorithms that relied, in part, "on FHA-protected characteristics—such as race, national origin, and sex—to help determine which subset of an advertiser's targeted audience [would] actually receive a housing ad."# In total, the complaint alleged that Meta "used these three aspects of its advertising system to target and deliver housing-related ads to some Facebook users while excluding other users based on FHA-protected characteristics." *Id.*

[273] *Id.*

[274] *Id.*

[275] *Id.*

[276] *Id.* at 7.

[277] *Id.*

[278] Interview with Jacob Appel, Chief Strategist, Oneill Risk Consulting & Algorithm Auditing (June 1, 2022) (on file with Strategic Training Initiative for the Prevention of Eating Disorders legal team) [hereinafter Interview with Jacob Appel].

[279] *Id.*

[280] *Id.*

[281] *Id.*

[282] *Id.*

[283] *Id.*

large relative variance, such as exceeding fifty percent, when comparing eligible to actual audience sizes.[284]

Under the final step of the approach described settlement, Meta and the DOJ must agree on reporting periods and benchmarks to reduce harm.[285] Meta must meet "certain [benchmarks] within a specific period of time" to reduce the variance between the eligible and actual audience for housing ads.[286] These benchmarks call for Meta, by December 31, 2023, to reduce variances to "less than or equal to 10% for 91.7% of those ads for [gender] and less than or equal to 10% for 81.0% of those ads for […] race/ethnicity."[287] By the end of 2023, Meta must ensure that for 91.7% of all housing ads on its platform, the variance between the eligible and actual audience for gender is 10% or less. Additionally, Meta must ensure that for 81% of housing ads on its platform, the variance between the eligible and actual audience for race/ethnicity is 10% or less.

To meet these benchmarks, Meta has developed a system called the Variance Reduction System (VRS), which helps reduce variances between the eligible and actual audiences for housing ads.[288] Once a variance is detected between the eligible and actual audiences using the Earth Mover's Distance measurement, Meta can use the VRS to help reduce that variance. Think of the two working in tandem with one another, similar to how a radar and auto-pilot work with a plane.[289] The radar identifies when there is a hazard ahead and the autopilot shifts the plane's speed or altitude to avoid the hazard. Likewise, the Earth Mover's Distance identifies the variance between the audiences and the VRS works to shrink that variance.[290]

Additionally, under the settlement, Meta must prepare a report every four months confirming that it has met the benchmarks for the previous four-month period.[291] Importantly, Meta and the DOJ selected an *independent, third-party reviewer* "to investigate and verify on an ongoing basis" whether the benchmarks are being met.[292] The third-party reviewer, therefore, serves as an objective check on Meta's compliance with the DOJ agreement.

This settlement agreement marks the first time Meta will be subject to court oversight for its ad targeting and delivery system.[293] The settlement requires Meta to alter the way its algorithms target and deliver housing ads to ensure compliance with the Fair Housing Act. We believe that this three-step approach to monitoring and measuring harm caused by algorithms can be adapted to assess harm caused by social media platforms to adolescent users in the form of an algorithm risk audit.

A. Legislation Based on the Meta/DOJ Settlement Should Require Social Media Companies to Conduct Algorithm Risk Audits to Reduce Harm to Children

New legislation that would legally mandate algorithm risk audits would mirror the three-step approach used in the Meta/DOJ settlement: (1) identify the specific harm(s), (2) determine how to measure the extent of each harm, and (3) agree on reporting periods and benchmarks to reduce harm. Our model legislation does not provide a specific set of harms to be measured, but lawmakers could customize it to

[284] *Id.*

[285] *Id.*

[286] Emma Roth, *Meta's new ad system addresses allegations that it enabled housing discrimination*, The Verge (Jan. 9, 2023, 6:03PM), https://www.theverge.com/2023/1/9/23547191/meta-equitable-ads-system-settlement [https://perma.cc/Z4G7-JP89].

[287] The United States Attorney's Office for the Southern District of New York, *United States Attorney Implements Groundbreaking Settlement With Meta Platforms, Inc., Formerly Known As Facebook, To Address Discrimination In The Delivery Of Housing Ads*, Dep't of Just. (Jan. 9, 2023), https://www.justice.gov/usao-sdny/pr/united-states-attorney-implements-groundbreaking-settlement-meta-platforms-inc-formerly [https://perma.cc/LJR8-UGZE] [hereinafter *Jan. US Attorney's Office*].

[288] Settlement Agreement at 6, United States v. Meta Platforms, Inc., No. 1:22-cv-05187 (S.D.N.Y. June 21, 2022).

[289] Interview with Jacob Appel, *supra* note 247.

[290] *Id.*

[291] Settlement Agreement at 9, United States v. Meta Platforms, Inc., No. 1:22-cv-05187 (S.D.N.Y. June 21, 2022).

[292] *Jan. US Attorney's Office*, *supra* note 255.

[293] *Id.*

determine what kind of harms they want to address.[294] To conceptualize how an algorithm risk audit would work, consider the specific harm adolescent users experience when confronted with pro-eating disorder content. Pro-eating disorder content may include very restrictive dieting plans, extreme exercise regimens, and images of very thin bodies that intend to serve as "inspiration" for users who are seeing the content.[295] An algorithm risk audit could be used to measure the extent of this harm, which could lead to social media platforms being pressured, and possibly required, for instance by attorneys general to comply with existing prohibitions on unfair or deceptive business practices, to alter the way their algorithms function to reduce the harm.

Using the audit's three-step approach, the first step would be to identify the specific harm. The specific harm might be described as "eating disorder rabbit holes,"[296] such as when adolescent social media users begin searching for and engaging with content related to mental health and body image and then are progressively shown more and more pro-eating disorder related content.[297]

The second step would be to determine how to measure eating disorder rabbit holes.[298] For this step, a social media platform might be required to measure the number of users who have made the transition from mental health and body image-related content to pro-eating disorder related content (e.g., an extremely restrictive dieting plan) within a certain number of minutes, hours, or days. The social media platform could measure the users who plunge into eating disorder rabbit holes and compare the demographics of these users. If the specific concern is adolescent users, the social media platform could compare the number of *all users* who enter eating disorder rabbit holes to that of *adolescent users* who do. Comparing the difference between these numbers would show whether adolescent users are disproportionately likely to tumble down eating disorder rabbit holes.

The social media platform and the governmental body that enacted a law requiring an algorithm risk audit would then move to the third step—agreeing on reporting periods and benchmarks to reduce harm. Determinations for reporting periods and benchmarks could be made in collaboration with the social media platforms and some governmental entity serving as an enforcement group for the law, including the enacting legislative body, state attorneys general offices, or state administrative agencies. The enforcement group could determine that the social media platform needs to implement a new system, similar to Meta's development and implementation of the VRS, to alter its current algorithm to address the disparate impact it has on adolescent users. In implementing such a change, the parties would need to determine benchmarks for improvement and reporting periods to ensure compliance with those benchmarks. Reporting periods could be required at any reasonable rate, such as on a quarterly, monthly, or even weekly basis. Similar to the Meta/DOJ settlement, a law requiring algorithm risk audits would require that the reports be evaluated by a *third-party, independent reviewer* to ensure compliance with the benchmarks agreed upon by the parties.

Our proposed legislation for algorithm risk audits, however, would move beyond the requirements dictated by the Meta/DOJ agreement. In addition to the three-step approach, a law mandating algorithm risk audits would require public disclosure of a social media platform's compliance with the agreed upon benchmarks. Indeed, the compliance reports developed by the platform, and reviewed by a third-party, should be made publicly available. These reports could be required on a quarterly,

[294]Interview with Jacob Appel, *supra* note 247.

[295]*See* Suku Sukunesan, *Examining the Pro-Eating Disorders Community on Twitter Via the Hashtag #proana: Statistical Modeling Approach*, 8 JMIR Mental Health 1, 2 (2021).

[296]*See, e.g.*, Jennifer A. Harriger, *The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms*, 41 Body Image 292 (2022).

[297]There are sources that raise this specific concern. *E.g.,* Sapna Maheshwari, *Young TikTok Users Quickly Encounter Problematic Posts, Researchers Say*, N.Y. Times (Dec. 14, 2022), https://www.nytimes.com/2022/12/14/business/tiktok-safety-teens-eating-disorders-self-harm.html [https://perma.cc/FU7G-HD4G] ("[TikTok] starts recommending content tied to eating disorders and self-harm to 13-year-olds within 30 minutes of their joining the platform, and sometimes in as little as three minutes … .").

[298]WSJ Staff, *Inside TikTok's Algorithm: A WSJ Video Investigation*, Wall St. J. (July 21, 2021, 10:26 AM), https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477 [https://perma.cc/4CFJ-UK2S].

monthly, or even weekly basis, allowing policymakers to determine the necessary frequency. This level of transparency would encourage social media platforms to be diligent in their prevention of harms caused by algorithms. Additionally, public disclosure would provide users with information about a platform's algorithmic practices, including its benefits and harms, which would allow users to choose whether or not to use a platform that employs such an algorithm.[299] Further, required public disclosure would provide data to researchers examining the potential harms caused to adolescents by social media and could also inform policymakers as to the actual risks of harm and inspire concrete legislative solutions to remedy it.[300]

Significantly, while harms caused to adolescents by social media platforms are currently criticized as theoretical in nature, algorithm risk audits would curate evidence of instances of harm that could significantly add to the mounting evidence that demonstrates a causal link between social media platforms' business practices and harm to adolescents.[301] Indeed, if social media platforms are able to alter their practices to comply with benchmarks required under a law of this kind, it might indicate that these platforms have at least some control over the harms their algorithms cause. Policymakers, state attorneys general offices, and state administrative agencies could therefore pursue lawsuits aimed at holding social media platforms accountable for the harm caused to adolescent users that they negligently create and ignore.

Notably, to create an algorithm risk audit, a social media platform would need to share data with the independent third-party assessing compliance with the agreed upon metrics.[302] Social media platforms may object to this, fearing that trade secrets or proprietary information would be exposed, which might allow competitors to gain a business advantage. However, a law requiring algorithm risk audits could require that only the measured harms to adolescents be publicly disclosed and not the company's proprietary data.[303]

Finally, and significantly, a law requiring algorithm risk audits would survive a constitutional challenge under the First Amendment due to its content neutral nature. An algorithm risk audit would not regulate content; or speech, on social media platforms, nor prohibit the use of particular algorithms; rather, it would measure the effects an algorithm has on its users.[304] Such evidence could be used to help establish *causation*, which is the most difficult element to prove in FTC claims against businesses for unfair or deceptive practices and in products liability claims. Such findings could be a catalyst for attorneys general to enforce state laws aimed at preventing harm caused by social media platforms, including the California Age Appropriate Design Code. Publicly disclosed algorithm risk audits would, therefore, provide vital new evidence needed to compel social media companies to change their harmful practices.

---

[299]*See Auditing Algorithms: The Existing Landscape, Role of Regulators and Future Outlook*, Digit. Coop. F. (Sept. 23, 2022), https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook#introduction-and-purpose ("Consumers or those affected by algorithmic systems who have a better understanding of these systems can then take informed decisions about how or when they engage with different products and services.").

[300]*See id.* ("Where the outputs of an algorithmic processing system have impacts on individuals, the system will be subject to regulatory expectations such as ensuring consumers or citizens are treated fairly, not discriminated against, and have their rights to privacy respected.").

[301]*Id.* ("Auditing can indicate to individuals that they have been harmed … . It can provide them with evidence that they could use to seek redress.").

[302]*See* Miles Brundage et al., Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims 11 (2020) ("Third party auditors can be given privileged and secured access to … private information, and they can be tasked with assessing whether safety, security, privacy, and fairness-related claims made by the AI developer are accurate.").

[303]*See* James Kobielus, *How We'll Conduct Algorithmic Audits in the New Economy*, InformationWeek (Mar. 4, 2021), https://www.informationweek.com/ai-or-machine-learning/how-we-ll-conduct-algorithmic-audits-in-the-new-economy (arguing that audit scopes should be clearly and comprehensively stated in order to make clear what aspects of audited algorithms may have been excluded and why they were not addressed in a public report (e.g., to protect sensitive corporate intellectual property)).

[304]*See e.g*, Balbuzanova, *supra* note 110; *Universal City Studios, Inc.*, 273 F.3d at 429; *Johnson Controls*, 886 F.2d at 1173; *Bernstein*, 922 F.Supp. at 1426 (N.D. Cal. 1996); *e-ventures Worldwide, LLC*, 188 F. Supp. 3d at 1265; *Zhang*, 10 F. Supp. 3d at 433 (finding that computer codes and search engine outputs are protected speech under the First Amendment).

VIII. Conclusion

Mental and physical health injuries to children and adolescents caused by harmful algorithm feeds on Instagram, TikTok, and other social media platforms are far-reaching and must be confronted as a public health crisis. Social media companies employ relentless feeds of algorithm-driven content to keep young users engaged on platforms, which results in billions of dollars in annual revenue for the platforms paid by advertisers targeting ads at children. With such economic incentives, platforms will not take it upon themselves to cure practices that harm young social media users. That task must fall to policymakers in Congress and state legislatures. Any new law must be careful not to run afoul of First Amendment free speech protection for social media platforms and must circumvent the immunity currently granted to social media platforms under Section 230 of the CDA. The Supreme Court, in the cases of *Twitter, Inc. v. Taamneh* and *Gonzalez v. Google*, recently declined to diminish the immunity from liability social media companies currently enjoy under Section 230, but appeared to leave intact a revenue sharing theory where a plaintiff may allege a platform that commercially profits from an algorithm that pushes illegal content could be considered an information content-provider, thus removing the immunity protection of Section 230 and opening the platform up to liability. Further, laws such as the California Age Appropriate Design Code, which requires Data Protection Impact Assessments, while positive, do not provide enough enforcement to be truly effective in curbing social media harms. Claims lodged by state attorneys general against platforms for unfair or deceptive business practices will also fail if the causal link between social media practices and harm to minors cannot be established.

To best accomplish this, social media companies should be required to conduct algorithm risk audits that identify specific sections of computer code as deceptive design elements. The U.S. Senate is contemplating a law, KOSA 2.0, that would mandate risk audits of social media algorithms by independent third parties. The results of those audits would be made public. But the law faces opposition by respected civil liberty groups and Big Tech raising concerns about young people's privacy and First Amendment rights. The bill may also fail to garner necessary legislative support in the U.S. House of Representatives to become law. Thus, state legislatures must be urged to craft legislation that requires algorithm risk audits, but the neutrality and transparency of the audits is imperative. Any algorithm risk audit that a social media company conducts must be administered by an independent, third-party auditor, and the results should be publicly disclosed. Such disclosure will allow law enforcement organizations, such as attorneys general, and researchers examining the risks and benefits of social media practices to access the audit's findings. Requiring algorithm risk audits is a crucial step to protecting children who risk their mental and physical well-being when they delve into the relentless algorithmic information feeds of social media.

Funding. This study was supported by the Becca Schmill Foundation and the Strategic Training Initiative for the Prevention of Eating Disorders. A Raffoul is supported by the Canadian Institutes of Health Research Institute of Population and Public Health grant MFE-171217. SB Austin is supported by the US Maternal and Child Health Bureau training grant T76-MC00001. The funders were not involved in the conduct of the study. The authors do not have financial conflicts of interest with this study.

Nancy A. Costello supervises the legal research team for the social media study conducted by the Strategic Training Initiative for the Prevention of Eating Disorders. She serves as a Clinical Professor of Law and Director of the First Amendment Law Clinic at Michigan State University College of Law. She formerly worked as a journalist for The Associated Press and Detroit Free Press.

Rebecca Sutton is a legal research assistant with the Strategic Training Initiative for the Prevention of Eating Disorders and a 2023 graduate of Michigan State University College of Law, where she graduated Magna Cum Laude with her Juris Doctorate. Sutton is also a graduate of Virginia Tech, where she majored in Multimedia Journalism and earned a B.A. in Communication in 2019. Rebecca is licensed to practice law in the State of Michigan and currently works as a research attorney for the Michigan Court of Appeals.

Madeline Jones is a 2023 graduate from Michigan State University College of Law, where she obtained her Juris Doctorate. She worked with the Strategic Training Initiative for the Prevention of Eating Disorders during law school from January 2022 until May 2023. Madeline is now a barred attorney in Minnesota and practices family law.

Mackenzie Almassian is a Juris Doctorate candidate, 2024, at Michigan State University College of Law. She earned her B.A. in 2021 from Michigan State University. She is a Notes Editor for the Michigan State Law Review and a past member of the Michigan State University College of Law First Amendment Clinic. After graduation she will be working in private practice as a municipal law attorney.

Amanda Raffoul, PhD is an Instructor in Pediatrics at Boston Children's Hospital and Harvard Medical School. Her mixed methods research explores policies for the prevention of eating disorders, ranging from bolstering clinician education to strengthening evidence for public health legislation. She is actively involved in legislative advocacy with the Strategic Training Initiative for the Prevention of Eating Disorders.

Oluwadunni Ojumu is an A.B. Candidate in Neuroscience at Harvard College '25 and a research assistant at the Strategic Training Initiative for the Prevention of Eating Disorders at the Harvard T.H. Chan School of Public Health. With a passion for global health and African development, Dunni plans to continue a career in global health and medicine upon graduation.

Meg Salvia is a doctoral candidate in Population Health Sciences at Harvard University and a Registered Dietitian Nutritionist specializing in the treatment of eating disorders. She has worked as a graduate research assistant with the Strategic Training Initiative for the Prevention of Eating Disorders for the past several years and will continue research in public health and nutrition, particularly with respect eating disorder treatment and prevention, after graduation.

Monique Santoso was the Program Coordinator of the Strategic Training Initiative for the Prevention of Eating Disorders based in the Division of Adolescent and Young Adult Medicine at Boston Children's Hospital in Boston, MA, at the time of this study. She is currently a PhD Student at Stanford University.

Jill R. Kavanaugh holds a B.A. in Media, Information, and Technoculture and a Master of Library and Information Science, along with a Certificate in Writing, all from Western University in Ontario, Canada. Jill's research centers on social media's positive and negative effects on adolescent health, particularly in the context of body image and media literacy. Additionally, Jill is an advocate for responsible regulation of social media companies, aiming to address concerns related to their influence on public health.

S. Bryn Austin, ScD, is an award-winning researcher, teacher, and mentor. She is Professor at Harvard Chan School of Public Health and Boston Children's Hospital and is director of the Strategic Training Initiative for the Prevention of Eating Disorders: A Public Health Incubator. She is a social epidemiologist and behavioral scientist with a research focus on public health approaches to eating disorders prevention with an emphasis on policy translation research and advocacy. Her research also focuses on determinants of sexual orientation and gender identity health inequities. She is Past President of the Academy for Eating Disorders and Eating Disorders Coalition.

Cite this article: Costello, N., Sutton, R., Jones, M., Almassian, M., Raffoul, A., Ojumu, O., Salvia, M., Santoso, M., Kavanaugh, J.R. and 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**, pp. 135–172. https://doi.org/10.1017/amj.2023.25