A Case for Protecting Youth from the Harmful Mental Effects of Social Media

Kaidyn McClure∗

INTRODUCTION

For years, businesses have executed strategies to engage viewers with their products or services.[1] Since engagement strategies can be manipulative, marketers must consider whether and when certain marketing tactics are unethical.[2] Today, social media companies may have the same basic objective to engage their audiences, but their engagement strategies utilize artificial intelligence.[3] To keep users engaged on social media, these platforms deploy algorithms that manipulate what the user views based on the user's predicted interests.[4] But the algorithm doesn't just dictate what a user sees.[5] It *amplifies* the user-generated content, meaning that, while the underlying content may be created by a human, the user's experience of the content, or of reality, is mediated by the algorithm.[6]

This amplification is harmful because it enables the platform to show an unprecedented amount of personalized content to the viewer,[7] ultimately promoting a message to the viewer that targets and preys on the viewer's vulnerabilities and insecurities.[8] This harm is evidenced by social media's strong association with a rise in mental health challenges, primarily among teenagers.[9] The

---

[1] *See, e.g.*, Dr. Sydney Ceruto, *The Psychological Concept That Can Make You a More Effective Marketer*, FORBES: LEADERSHIP (Feb. 19, 2020, 8:45 AM), http://www.forbes.com/sites/forbescoachescouncil/2020/02/19/the-psychological-concept-that-can-make-you-a-more-effective-marketer/?sh=56f413c821a4 [http://perma.cc/E42B-8G7V] (describing how brands use classical conditioning to "train" customers to think about and turn to their brand).

[2] *See id.*

[3] *See* Hidden Forces, *The Age of A.I. and our Human Future*, APPLE PODCASTS, at 39:45 (Nov. 11, 2021) http://podcasts.apple.com/tj/podcast/the-age-of-a-i-and-our-human-future/id1205359334?i=1000541470640 (speaking with Eric Schmidt, former CEO of Google, and Daniel Huttenlocher, inaugural dean of MIT Schwarzman College of Computing).

[4] *See* Elizabeth D. Levin, *Theoretical Justifications for Government Regulation of Social Media Platforms*, 24 VA. J.L. & TECH. 1, 10 (2021).

[5] *See id.* at 6–7.

[6] *See id.* at 7, 11–12; *see also* Swathi Sadagopan, *Feedback Loops and Echo Chambers: How Algorithms Amplify Viewpoints*, THE CONVERSATION (Feb. 4, 2019, 4:18 PM), http://theconversation.com/feedback-loops-and-echo-chambers-how-algorithms-amplify-viewpoints-107935 [http://perma.cc/YEV8-XXCZ] (describing that algorithmic amplification is "when some online content becomes popular at the expense of other viewpoints" and experience shows that users viewing "a lighter version of a topic" are then recommended "more hardcore content").

[7] *See* Elizabeth D. Levin, *Theoretical Justifications for Government Regulation of Social Media Platforms*, 24 VA. J.L. & TECH. 1, 14 (2021).

[8] *See* Georgia Wells et al., *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show*, THE WALL ST. J. (Sept. 14, 2021, 7:59 AM), http://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739 (pointing to Facebook's internal studies reporting that teens blame Instagram for increases in the rate of anxiety and depression).

[9] Recent research demonstrates that increasing social media use is an important factor affecting adolescents' mental health, and it particularly adversely impacts girls. *See* Lennart Raudsepp & Kristjan Kais, *Longitudinal Associations Between Problematic*

existing scientific research shows the strong association between social media use and a decline in teen mental health.[10] Teens are devoting so much time and effort to social media use that it limits other social activities, which researchers have coined "problematic social media use" or "PSMU."[11] In 2021, the U.S. Surgeon General squarely addressed the impact of harmful social media messages on teen mental health in a public advisory, stating that "too often, young people are bombarded with messages . . . that erode their sense of self-worth—telling them they are not good looking enough, popular enough, smart enough, or rich enough."[12] Evidence of the connection between mental harm and social media is further represented by lawsuits brought by parents against social media platforms, such as one against Instagram, alleging that the addictive algorithm caused their daughters' poor self-esteem and depression, ultimately leading to suicide.[13]

Not only do independent studies and public voices emphasize this strong association, but internal research performed by one of the social media platforms itself—Facebook (also the owner of Instagram)—exemplified that Facebook use caused mental harm to teens.[14] In September 2021, Frances Haugen, a former

Social Media Use and Depressive Symptoms in Adolescent Girls, 15 PREVENTIVE MED. REPS. 1, 1, 3 (2019), http://www.sciencedirect.com/science/article/pii/S2211335519300993 [http://perma.cc/3W2B-BHPX]. Female adolescents with social media profiles have significantly higher levels of depressed mood and lower self-esteem compared to young females that do not have a social media profile. *See id.* at 1.

10 *See id.*

11 There is an increasing number of adolescents experiencing adverse effects due to PSMU. *Id.* Evidence suggests that an increase in adolescent girls' PSMU is related to an increase in depressive symptoms. *Id.* at 3.

12 U.S. PUB. HEALTH SERV., PROTECTING YOUTH MENTAL HEALTH: THE U.S. SURGEON GENERAL'S ADVISORY 3 (2021). The U.S. Surgeon General also called attention to research supporting the linkage between social media usage and mental health challenges. *Id.* at 8 (citing 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*, CLINICAL PSYCH. SCIENCE 1, 3–17 (2018), http://journals.sagepub.com/doi/10.1177/2167702617723376).

13 *See* Complaint at 2, Roberts v. Meta Platforms, Inc., No. 3:22-cv-04210 (N.D. Cal. filed July 20, 2022); Complaint at 2, Rodriguez v. Meta Platforms, Inc., No. 3:22-cv-00401 (N.D. Cal. filed Jan. 20, 2022). These lawsuits were filed in federal court on a theory of defective design. *See* Complaint at 2, Roberts v. Meta Platforms, Inc., No. 3:22-cv-04210 (N.D. Cal. filed July 20, 2022); Complaint at 2, Rodriguez v. Meta Platforms, Inc., No. 3:22-cv-00401 (N.D. Cal. filed Jan. 20, 2022). This Note does not explore this theory because it focuses on providing relief to those currently suffering from mental distress unaccompanied by physical injury, rather than providing relief for a wrongful death.

14 *See* Keach Hagey et al., *Facebook's Pushback: Stem the Leaks, Spin the Politics, Don't Say Sorry*, THE WALL ST. J. (Dec. 29, 2021, 10:14 AM), http://www.wsj.com/articles/facebook-whistleblower-pushback-political-spin-zuckerberg-11640786831?mod=article_inline. In 2012, Facebook acquired Instagram, a social media application that allows users to share phots and add distinctive filters and visual flair to them. *See* Laurie Segall, *Facebook Acquires Instagram for $1 Billion*, CNNMONEY (Apr. 9, 2012), http://money.cnn.com/2012/04/09/technology/facebook_acquires_instagram/index.htm [http://perma.cc/E5VH-D683].

Facebook product manager, released a host of internal reports demonstrating that Facebook's amplification algorithm,[15] including its engagement-based ranking on Instagram, negatively affects teen mental health and well-being.[16] The algorithm's engagement-based ranking enables Instagram to present specific content to the viewer based on personal user data collected by the platform and then amplify the user's preferences.[17] Haugen's testimony illustrates that the algorithm is harmful; for example, it leads children from innocuous topics, like healthy recipes, to anorexia-promoting content.[18]

So, is a typically reasonable business objective—to engage users—still reasonable when it is set in the context of social media and achieved by deploying artificial intelligence that lacks any sense of moral consequence?[19] Is it reasonable when the underlying strategy causes harm to teen users in the form of depression, suicide, anxiety, and other emotional disorders, and the platforms are aware of these harms?

The existing legal landscape is ill-equipped to provide relief to teens suffering from mental harm caused by the algorithms and to hold Facebook and other social media companies accountable for such mental harm. The circuit courts' current interpretation of section 230 of the Communications Decency Act broadly immunizes these providers, even if they deploy algorithms.[20] This interpretation rejects any possibility that certain algorithmic functions may take providers out of the purview of immunity.[21] Additionally, existing tort jurisprudence does not address the issue of mental harm caused by algorithmic capabilities, so courts would have to extend tort law to provide relief to teens suffering from mental harm.[22] As for government regulation, the House of Representatives and Senate proposed

---

15 *See* Hagey et al*., supra* note 14. While some allege that Ms. Haugen had a political motive to release internal company documents, she denied any partisan motivations. *See id.* Additionally, the information reported is not contested. *See id.* On the contrary, the greater controversy was that Facebook's research into Instagram's effects on teen girls was hidden from the public and even some company advisory board members. *See id.*

16 *See Focusing on Testimony from a Facebook Whistleblower: Hearings to Examine Protecting Kids Online Before the Subcomm. on Consumer Prot., Prod. Safety, & Data Sec. of the S. Comm. on Com., Sci. & Transp.*, 117th Cong. 8 (2021) [hereinafter *Focusing on Testimony from a Facebook Whistleblower*] (statement of Frances Haugen, former Product Manager of Facebook Inc.).

17 *See id.* at 13, 35.

18 *Id.* at 8.

19 *See* Hidden Forces, *supra* note 3, at 36:43, 39:12, 39:50.

20 *See* 47 U.S.C. § 230; *see also* discussion *infra* Part II.A.

21 *See* 47 U.S.C. § 230; *see also* discussion *infra* Part II.A.

22 *See* discussion *infra* Part II.B.

bills to address the mental health crisis from social media and provide relief to teens, but progress is slow.[23]

Neither government nor society anticipated the serious, harmful effects that excessive use of social media would have on teen mental health today.[24] Teens cannot protect themselves from depression, anxiety, addiction, and other negative side effects of Instagram's engagement-based algorithm because they cannot control the content that they view; rather, the algorithm does.[25] On the one hand, there is a need to protect teenage users against the negative consequences of Instagram, to deter social media giants from knowingly developing harmful algorithms, and to prevent further harm to teens. On the other hand, there is a competing interest to ensure that social media businesses are not unduly regulated or disadvantaged by overly broad mandates.

This Note proposes a roadmap for two non-mutually exclusive solutions to the problem of a deficient legal landscape for mental harm caused by certain social media algorithms. Part I leads the discussion with a focus on Facebook and Instagram, by exploring Facebook's business model and the various externalities of Instagram's algorithm. Part II describes the problem, arising out of courts' broad interpretation of section 230(c)(1), existing tort law, and Congress' proposed bill. Part III synthesizes a new reading of section 230(c)(1) and suggests extending tort law to provide relief in conjunction with the proposed interpretation of section 230(c)(1). Part III also proposes a legislative solution to hold Facebook and other companies like it accountable for writing algorithms that cause mental harm, noting the advantages and disadvantages of a legislative approach.

---

23 *See* S. 2917, 117th Cong. (2021) (no action has been taken since the bill was introduced to the in the Senate in 2021); H.R. 5449, 117th Cong. (2021) (no action has been taken since the bill was referred to the Subcommittee on the Constitution, Civil Rights, and Civil Liberties on November 1, 2022).

24 *See Hearing on "Algorithms and Amplification: How Social Media Platforms' Design Choices Shape our Discourse and our Minds" Before the Subcomm. on Priv., Tech. & the L., of the S. Comm. on the Judiciary*, 117th Cong. (2021) [hereinafter *Social Media Design Discourse Hearing*], http://www.judiciary.senate.gov/imo/media/doc/Harris%20Testimony.pdf [http://perma.cc/5WRE-6CV3] (statement of Tristan Harris, President and Co-Founder of Center for Humane Technology) ("We are raising entire generations of young people who will have come up under these exaggerated . . . mental health problems. . . . If this continues, we will see . . . more children with ADHD, more suicides and depression—deficits that are cultivated and exploited by [social media] platforms.") (alteration in original).

25 *See Focusing on Testimony from a Facebook Whistleblower*, *supra* note 16, at 28. To increase the control that people have over their News Feed, Facebook enables users to "reject the personalized ranking algorithm altogether and instead view their feed chronologically, meaning that their [feed] simply shows them the most recent posts from their eligible sources of content in reverse chronological order." *See Social Media Design Discourse Hearing*, *supra* note 24. *But see* Levin, *supra* note 7, at 17 ("[T]he social media platform is in the best (perhaps the only) position to control what users see, so solutions premised on the free-market ideal of individuals choosing what content they view are unrealistic . . . .").

I. META: A CASE STUDY[26]

A. The Algorithm as a Business Strategy

The risk of mental harm to social media users is exacerbated by the way algorithms are evolving and being utilized in the platform. Facebook did not use an algorithm at its inception in 2004;[27] the platform was merely a collection of disconnected profiles.[28] Facebook played a passive role in the user experience, allowing users to independently search for friends or strangers without any active input from Facebook. Thus, a user was largely in control of their experience. In 2009, Facebook introduced an algorithm that "determined the order of stories for each user" to display the most "juicy" posts near the top of the page.[29] This straightforward ranking system helped users stay engaged on the platform without taking control from the user.[30] By 2016, Facebook was joined by other social media platforms like Snapchat (owned by Snap, Inc.) and was forced to compete for the attention of young users.[31] To keep from losing young users' attention, Facebook used the algorithm to implement a user retention strategy to help users form meaningful social interactions.[32] The algorithm executed this strategy by showing users the posts with greater comments and replies.[33] These posts tended to be more extreme in nature, leading to adverse effects that perhaps were not anticipated.[34] Today, Instagram deploys amplification algorithms, including engagement-based ranking.[35] These algorithms bombard users with content the user wants to see based on the personal data collected.[36] The danger is the development of feedback cycles, where teens are using Instagram to self-soothe, but then are exposed to more content that preys on their fears and insecurities.[37] The

---

[26] Facebook's CEO changed the company's name to Meta Platforms, Inc. *See* Hagey et al., *supra* note 15. For clarity and consistency, I will refer to the company as Meta and to the platforms as Facebook and Instagram respectively throughout this Note.

[27] *See* Megan Rose Dickey, *It's Hard to Believe How Drastic the Changes to Facebook Have Been over the Years*, BUS. INSIDER (Mar. 11, 2013, 6:16 PM), http://www.businessinsider.com/facebook-evolution-2013-3 [http://perma.cc/M4BV-PCKR].

[28] *See id.*

[29] *See* Will Oremus et al., *How Facebook Shapes your Feed*, THE WASH. POST (Oct. 26, 2021, 7:00 AM), http://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/ [http://perma.cc/KXS5-QURB].

[30] *See id.*

[31] *See id.*

[32] *See id.*

[33] *See id.*

[34] *See id.*; *see also Focusing on Testimony from a Facebook Whistleblower*, *supra* note 16, at 8 ("[T]o be able to share fun photos of your kids with old friends, you must also be inundated with anger-driven virality.").

[35] *See Focusing on Testimony from a Facebook Whistleblower*, *supra* note 16, at 28.

[36] *See id.*

[37] *See id.*

engagement-based ranking system is different than the straightforward ranking system because it takes control away from how the user experiences the platform.[38] Rather than allowing the user to experience content on the platform under their own volition, the amplification algorithm dictates how the user experiences the content, creating a greater risk of harm by preying on the user's vulnerabilities without the user even realizing.

While an engagement-based algorithm poses more risk to users, it helps brands advertise to a highly active audience that is more likely to be interested in the advertisement.[39] Instagram overwhelmingly helps small businesses by affording them the opportunity to reach millions of viewers at a low cost, an opportunity that would not exist without the algorithm's capabilities.[40] In 2020, the platform supported about 2 million monthly advertisers and over 25 million business accounts.[41] Since Instagram's service is funded by advertisers, Instagram is encouraged to deploy the engagement-based algorithm because it attracts more advertisers, and thus aggressively generates more revenue.[42] The result is that users engage with more businesses on the platform.[43] But the opportunity for harm forms when targeted messaging comes not from these advertisers, but from Instagram itself by promoting a specific message to the user that is perhaps unhealthy or dangerous to keep the user engaged.

B.   The Impact of the Algorithm

Two aspects of social media platforms like Facebook and Instagram give rise to the risk of user harm: (1) a business model based on advertising revenue, and (2) the need to compete for engagement with competitors, such as Snapchat, Twitter, and

---

[38] *See id.* at 8 (noting that users are self-identifying that they do not have control over their usage and that their usage is materially harming their health); *see also Social Media Design Discourse Hearing*, *supra* note 24, at 3.(statement of Monika Bickert, Vice President for Content Policy, Facebook) (trying to give more control back to users through various solutions).

[39] *See   generally   Advertising   on   Instagram,*   INSTAGRAM, http://business.instagram.com/advertising/ [http://perma.cc/4RCU-JK4U] (last visited Sept. 11, 2022) (describing Instagram's advertising service and promoting its precise targeting).

[40] *See* Nick Wyatt, *A Small Business Guide to Advertising on Instagram*, MEDIUM (Apr. 19, 2020), http://nwyatt227.medium.com/a-small-business-guide-to-advertising-on-instagram-74c7023d7ac2 [http://perma.cc/SH2P-LGUG].

[41] *See id.*

[42] *See Social Media Design Discourse Hearing*, *supra* note 24, at 1(statement of Joan Donovan, Ph.D. Research Director at Harvard Kennedy School's Shorenstein Center on Media, Politics, and Public Policy), http://www.judiciary.senate.gov/meetings/algorithms-and-amplification-how-social-media-platforms-design-choices-shape-our-discourse-and-our-minds [http://perma.cc/V5R5-EBLE] ("Over the last decade, social networking (connecting people to people) morphed into social media (connecting people to people and to content), which resulted in exponential profits and growth.").

[43] *See* Wyatt, *supra* note 40.

TikTok. First, a revenue model based on third-party providers will inherently motivate a business to consider those providers' interests.[44] Thus, even though Instagram's stated mission is "[t]o bring you closer to the people and things you love,"[45] the means employed by Instagram are actually motivated to help third-party advertisers—which may involve bringing users closer to content with implicit harmful messaging from Instagram.[46] Second, competition in the social media space makes it more difficult to keep users engaged.[47] A solution that addicts users to the platform—such as deployment of an amplification algorithm—is good for advertisers because it promises more traction over their content, and keeps Meta in the game as a competitor. However, it is the algorithm's addictive effect that contributes to users' mental harm.[48]

In 2019 and 2020, Facebook's in-house analysts became aware of the intense social pressure, addiction, body image issues, eating disorders, anxiety, depression, and suicidal thoughts resulting from teen girls' Facebook addiction.[49] For eighteen months in 2019-2020, Facebook conducted a "teen mental-health deep dive" which included focus groups, online surveys, and diary studies.[50] The research concluded that problems of mental health were specific to Instagram, coining an issue of "social comparison," defined as a person's assessment of their own value in relation to the attractiveness, wealth, and success of others.[51] The large cause of social comparison is the algorithm's curation of photos and videos on the Explore Page.[52] A presentation posted to Facebook's internal message board indicated that 32% of teen girls feel worse about their bodies after using Instagram, and "40% of teen boys

---

[44] *See Social Media Design Discourse Hearing*, *supra* note 24, at 2 (statement of Joan Donovan, Ph.D. Research Director at Harvard Kennedy School's Shorenstein Center on Media, Politics, and Public Policy), http://www.judiciary.senate.gov/meetings/algorithms-and-amplification-how-social-media-platforms-design-choices-shape-our-discourse-and-our-minds [http://perma.cc/V5R5-EBLE] (noting Facebook's advertising revenue at $84 billion and growing).

[45] *Terms of Use*, INSTAGRAM, http://help.instagram.com/581066165581870 [http://perma.cc/4H8N-X7LM] (last updated July 26, 2022).

[46] *See supra* notes 39–42 and accompanying text.

[47] *See supra* note 31 and accompanying text.

[48] *See* Force v. Facebook, Inc., 934 F.3d 53, 86–87 (2d Cir. 2019) (noting that the algorithms deployed by social platforms such as Facebook and Twitter are designed to keep users using, and such manipulation of news feeds influences users' moods).

[49] *See* 167 CONG. REC. S6759 (daily ed. Sept. 29, 2021) (statement of Sen. Marsha Blackburn). *But see* Hagey et al., *supra* note 14 (noting that Facebook invests billions of dollars to protect the safety of its users).

[50] *See* 167 CONG. REC. S6759 (daily ed. Sept. 29, 2021) (statement of Sen. Marsha Blackburn); Wells et al., *supra* note 8.

[51] *See* Wells et al., *supra* note 8.

[52] *See id.*

experience[d] negative social comparison."[53] However, Instagram researchers found it challenging to convince other colleagues of the gravity of the findings, who instead pointed to studies from the Oxford Internet Institute showing little correlation between social media use and depression.[54]

In 2021, a teenager shared her story with the *Wall Street Journal*, explaining her belief that Instagram caused her eating disorder.[55] She started using the platform when she was thirteen-years-old and was repeatedly bombarded by images of "perfect abs and women doing 100 burpees in 10 minutes."[56] The harm that people experience from social media use can rise to clinical-level depression that requires treatment and can even extend to self-harm.[57] In fact, a director for the eating-disorders program at Johns Hopkins Hospital expressed that she commonly hears from patients that their condition was caused by social media tips.[58] For those vulnerable to negative emotional distress, Instagram escalates it.[59]

C. Problems with Leaving Regulation to the Platform or to Teen Users

Hoping that either the social media platform will self-regulate or that teen users will regulate themselves is ineffective to protect teen mental health.[60] For example, Facebook and Instagram cannot be trusted to prioritize mental health over user engagement goals because they've chosen to deploy an addictive algorithm despite awareness of the harmful effects.[61] Facebook has acknowledged that the platform is a "sensory experience of communication that helps us connect to others, without having to

---

[53] *Id.* Facebook's own researchers were aware that "[t]eens blame Instagram for increases in anxiety and depression." *Id.*

[54] *See id.*; *see also* Zoe Kleinman, *Teens, Tech, and Mental Health: Oxford Study Finds No Link*, BBC NEWS (May 4, 2021), http://www.bbc.com/news/technology-56970368# [http://perma.cc/459Y-KLT7]. *But see* Wells et al., *supra* note 8 (noting that Facebook donated to a researcher at the Oxford Internet Institute).

[55] *See* Wells et al., *supra* note 8.

[56] *See id.*

[57] *See id.*

[58] *See id.*

[59] *See id.*

[60] *See* Levin, *supra* note 7, at 14. Ms. Levin justifies government regulation because solutions premised on the free-market ideal of individuals choosing what content they view is unrealistic, and the option of "opting out" of personal data collection is unrealistic given the ability of modern algorithms to identify users' identities. *See id.*

[61] *See* Wells et al., *supra* note 8 (describing Instagram as an "addictive product"); *see also* 167 CONG. REC. S6759 (daily ed. Sept. 29, 2021) (statement of Sen. Marsha Blackburn) (noting that Instagram "manifests itself in the minds of teenagers in the form of intense social pressure [and] addiction") (alteration in original). Facebook publicly undermines the app's negative effects on teens, and Instagram tells reporters that the research suggests the app's effects on teen well-being are "quite small." *See* Wells et al., *supra* note 8. However, Meta's internal research represents a clear discrepancy between Meta's "understanding of itself and its public position." *See id.*

look away."[62] It may be difficult to understand the algorithm,[63] but rather than take real steps to mitigate harms caused by the algorithm, platforms like Instagram merely warn users that services are provided "as is," with no guarantee that they will work perfectly all the time.[64] The algorithm's unpredictability and lack of any moral sense, coupled with Facebook's lack of motivation to protect teen health, does not lead towards improved mental health absent legal deterrence.

Additionally, despite Facebook's attempt to help users improve their experience by allowing them to alter their account settings,[65] teens are not making these changes because they are already addicted to the algorithm experience. Facebook's own research showed that those struggling with the platform's harmful psychological effects weren't logging off, even if they wanted to, because they lacked the self-control.[66] Some teens have shared that they often feel addicted and know that their mental health is deteriorating but are unable to stop themselves from using the application.[67] Between 2009 and 2019, the number of high school students who experienced "persistent feelings of sadness or hopelessness" increased by more than ten percent.[68] One could argue it's unreasonable to require Facebook to protect users from the negative effects that result from the mere act of scrolling over content, even if that scrolling is excessive, and hold Facebook liable when it falls short. However, the addictive effect of the amplification algorithm may be as harmful to teen mental health as the addictive effect of nicotine is to teen physical health, and the public's knowledge of tobacco's harm necessitated federal legislation to reduce harm to teens.[69] Moreover, studies show that

---

[62] Cole F. Watson, *Protecting Children in the Frontier of Surveillance Capitalism*, 27 RICH. J.L. & TECH. 1, 23 (2021) (noting that the platform intends for users to "enter a mental state called the 'machine zone': a connection between user and device that invokes a 'loss of self-awareness, automatic behavior, and a total rhythmic absorption carried along a wave of compulsion'") (citing SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 449–50 (2019)).

[63] *See Focusing on Testimony from a Facebook Whistleblower*, *supra* note 16, at 22.

[64] *Terms of Use*, *supra* note 45.

[65] *See supra* note 25 and accompanying text.

[66] *See* Wells et al., *supra* note 8.

[67] *See id.*

[68] 167 CONG. REC. S6759 (daily ed. Sept. 29, 2021) (statement of Sen. Marsha Blackburn).

[69] *See* Jennifer McCullough, *Lighting up the Battle Against the Tobacco Industry: New Regulations Prohibiting Cigarette Sales to Minors*, 28 RUTGERS L.J. 709, 710 (1997). Professor Turley believed that tobacco was a "factional dispute involving fundamental questions of personal responsibility versus corporate conduct." Jonathan Turley, *A Crisis of Faith: Tobacco and the Madisonian Democracy*, 37 HARV. J. ON LEGIS. 433, 449 (2000). Social media involves these same questions, exacerbated by social media companies' profit-focused decisions to drive innovation, rather than to optimize for the public interest. *See Social Media Design Discourse Hearing*, *supra* note 24 at 1 (statement of Joan Donovan, Ph.D. Research Director at Harvard Kennedy School's Shorenstein Center on Media, Politics, and Public Policy).

teens implicitly trust social media companies.[70] Thus, Facebook and other companies like it should be held accountable for harm caused to its users' mental well-being because it is aware of the risk of serious harm and affirmatively contributes to it by deploying the amplification algorithm. We may never fully quantify the impact of social media on the communicative and behavioral development of teens. But leaving the choice to the user about whether to use the service is not much of a choice at all, since the service is free and more than a socially acceptable habit—it is a prerequisite of daily encounter.[71]

II. The Problem of the Existing Legal Landscape to Deter Social Media Giants

The existing challenge is two-fold. First, the circuit courts' current interpretation of section 230 of the Communications Decency Act ("CDA") shields interactive computer service providers, like Facebook and Instagram, from liability for harm caused by its algorithms.[72] Second, even if the courts reinterpret section 230 in a manner that does not put the function of algorithms within the scope of protection, the court must still extend the tort theory of negligent infliction of emotional distress to provide relief to teens that suffer mental distress, with or without any physical injury. While Congress has proposed a bill to address the issue of mental harm caused by social media companies, the language of the proposed bill imposes broad liability on these providers by providing relief for mental harms caused to teens by mere usage of the platform.[73]

A. Current Interpretation of Section 230(c)(1) of the Communications Decency Act Shields Social Media Businesses from Liability

Congress enacted the "CDA" "to protect children from sexually explicit Internet content."[74] But since the public policy of the United States is to prevent "content regulation by the Federal Government of what is on the Internet,"[75] section 230 was added as an amendment to the CDA "to maintain the robust nature of

---

[70] *See, e.g.*, Watson, *supra* note 62, at 24 (noting that teenagers presume that technological companies act in the user's best interest).

[71] *See Terms of Use*, *supra* note 45; *see also* Danielle Keats Citron, *How to Fix Section 230*, B.U. L. Rev. (forthcoming 2022) (manuscript at 9) (on file with author) (describing the Internet's "totalizing impact," inextricable from daily life).

[72] *See* discussion *infra* Part II.A.

[73] *See* S. 2917, 117th Cong. (2021); H.R. 5449, 117th Cong. (2021).

[74] *See, e.g.*, Fed. Trade Comm'n v. LeadClick Media, LLC, 838 F.3d 158, 173 (2d Cir. 2016) (citing 141 Cong. Rec. S1953 (daily ed. Feb. 1, 1995) (statement of Sen. J. James Exon)); *see also* Force v. Facebook, Inc., 934 F.3d 53, 63 (2d Cir. 2019).

[75] *Force*, 934 F.3d at 78–79.

Internet communication and, accordingly, to keep government interference in the medium to a minimum."[76] The hope was that interactive computer service providers would "self-regulate" and "provide tools for parents to regulate."[77] Section 230(c)(1) immunizes interactive computer services against liability arising from content created by third-parties: "No provider . . . of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[78] An "interactive computer service" means any "information service, system, or access software provider that provides or enables computer access by multiple users to a computer server . . . ."[79] A content provider is a "person or entity that "is responsible, in whole or in part, for the creation or development" of the content,[80] but a website provider "can be both a service provider and content provider."[81]

Shortly after its enactment, in interpreting section 230, the Fourth Circuit stated that Congress's objective was to immunize service providers from potential liability for messages republished by their services to prevent these service providers from severely restricting third-party messages.[82] Since then, circuit courts have construed section 230(c)(1) broadly in favor of immunity.[83]

The Second Circuit created a three-part test to determine whether section 230(c) shields the defendant from civil liability.[84] The defendant is immune from liability for state law claims if: (1) it is a "provider or user of an interactive computer service"; (2) the plaintiff's claims treat the defendant as the publisher or speaker of content; and (3) that content is provided by a content provider other than the defendant interactive computer service.[85] Social media companies like Facebook are considered interactive computer service providers ("providers").[86] The problem is that courts equate algorithmic functions as functions of a publisher of third-party content, satisfying the second and third elements to immunize the provider.[87]

---

[76] Ricci v. Teamsters Union Loc. 456, 781 F.3d 25, 28 (2d Cir. 2015) (quoting Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997)).

[77] *See Force*, 934 F.3d at 79.

[78] 47 U.S.C. § 230(c).

[79] *Id.* § 230(f)(2).

[80] *Id.* § 230(f)(3).

[81] Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1162 (9th Cir. 2008).

[82] *See* Zeran v. Am. Online, Inc., 129 F.3d 327, 331 (4th Cir. 1997).

[83] *See, e.g.*, Force v. Facebook, Inc., 934 F.3d 53, 64 (2d Cir. 2019).

[84] *See id.*

[85] *Id.*

[86] *See id.*

[87] *See, e.g.*, *id.* at 67–68.

The case *Force v. Facebook* was the first to address the effect of Facebook's algorithm on Facebook's status as a publisher.[88] In *Force*, the Second Circuit determined that Facebook acted as a "publisher" within the meaning of section 230(c) when Facebook provided third-parties with a forum to communicate messages to interested parties.[89] The court did not believe that the algorithm changed the nature of Facebook's role as a publisher because many of the algorithm's functions like the "matchmaking" equated to editorial decisions that providers "have made since the early days of the Internet."[90] The court implicitly classified Facebook's algorithm as a "neutral tool[]" because it matches third-party content to users based on their preferences.[91] To support this finding, the court cited to precedent which concluded that such neutral tools merely perform the job that is an inherent part of publishing: "organizing and displaying content exclusively provided by third parties."[92] The problem with such a conclusion is that, as Judge Katzmann pointed out in his dissent, the "majority . . . 'cuts off all possibility for relief based on algorithms like Facebook's, even if . . . future plaintiffs could prove a sufficient nexus between those algorithms and their injuries.'"[93] Certain algorithms, like Instagram's amplification algorithm, are unlike ordinary editorial decisions; they do not merely determine where third-party content should appear on the site, who should see it, and in what form, as the Second Circuit suggests is the traditional result of editorial decision-making.[94] The court even pointed out that the algorithm's capability goes beyond the capability of editorial decisions by presenting users with targeted content of more interest to them.[95]

At the time section 230(c) was enacted, and later when *Force* was decided, the full extent of an algorithm's capability was unknown. Control was an important underlying presumption motivating Congress's decision to give broad protection to

---

[88] In *Force v. Facebook, Inc.*, users claimed that Facebook was civilly liable for aiding and abetting acts of international terrorism. *Id.* at 61. The plaintiffs argued that Facebook's algorithm, exploiting user engagement to predict and show third-party content most likely to interest and engage the user, makes it so that Facebook is not a "publisher" within the meaning of section 230(c)(1) of the CDA. *Id.* at 65. The majority struck down their claim in finding that Facebook was immunized from liability under section 230. *See id.* at 68.

[89] *See id.* at 65.

[90] *See id.* at 66–67.

[91] *See id.* at 66 (citing Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1172 (9th Cir. 2008)).

[92] *Id.*

[93] *See id.* at 77.

[94] *See id.*

[95] *See id.* at 67; *see also* Swathi Meenakshi Sadagopan, *Feedback Loops and Echo Chambers: How Algorithms Amplify Viewpoints*, THE CONVERSATION (Feb. 4, 2019, 4:18 PM), http://theconversation.com/feedback-loops-and-echo-chambers-how-algorithms-amplify-viewpoints-107935 [http://perma.cc/GE3N-4MTM] (discussing how users view lighter versions of topics, then are recommended more hardcore content).

providers under section 230(c), but modern users do not have a great degree of control over their experience with certain providers that deploy amplification-type algorithms.[96] Congress also presumed that Internet services have "flourished, to the benefit of all Americans,"[97] but the rise in mental health problems among teens contradicts Congress's presumption that all Americans are benefitting. Thus, including harmful algorithms within the scope of section 230 immunity, as the courts have done, undermines the underlying presumptions of the defense.

Two years after *Force*, the Ninth Circuit took up the issue of the effect of algorithmic functions on Google's status as a publisher in *Gonzalez v. Google LLC*.[98] The court concluded that an algorithm that shows particular content to a user based on that user's inputs does not strip the provider of immunity as a publisher of third-party content.[99] The court determined that by providing a neutral platform, not prompting the submission of certain content, and not determining the "types of content its algorithm[] would promote," Google did nothing more than republish third-party content.[100]

By viewing these recommendation capabilities as editorial functions, negligence claims based on the provider's algorithm will continue to be dismissed under section 230.[101] But as we better understand algorithms' capabilities, a generalization that the algorithm does nothing more than help providers perform ordinary editorial decisions, as articulated by the majority in *Force*, does

---

[96] *See Force*, 934 F.3d at 68; *see also* 47 U.S.C. § 230(a)(2) (providing immunity based on the presumption that the services "offer users a greater degree of control over the information they receive, as well as the potential for even greater control in the future . . . .").

[97] 47 U.S.C. § 230(a)(4).

[98] In *Gonzalez v. Google LLC*, plaintiffs asserted that Google was not immune under the CDA for using computer algorithms to match and suggest content to users based on their viewing history. Specifically, they alleged that by recommending ISIS videos to users, Google assisted ISIS in spreading its message, going beyond its role as a publisher of third-party content. *See* Gonzalez v. Google LLC, 2 F.4th 871, 881 (9th Cir. 2021). The United States Supreme Court granted plaintiffs' writ of certiorari and heard oral arguments on February 21, 2023. *See* Transcript of Oral Argument, Gonzalez v. Google LLC, 143 S.Ct. 762 (2023) (No. 21-1333), http://www.supremecourt.gov/oral_arguments/argument_transcripts/2022/21-1333_f2ag.pdf. During questioning, Justice Sotomayor stated "[T]here is a line at which affirmative action by an Internet provider should not get them protection under 230(c)." *See id.* at 97. Later Justice Gorsuch added "Is an algorithm always neutral? Don't many [providers] seek to profit-maximize or promote their own products? Some might even prefer one point of view over another." *See id.* at 101. Finally, Chief Justice Roberts commented to respondents that the third-party content appears "pursuant to the algorithms that [providers] have. And those algorithms must be targeted to something. And their targeting . . . is fairly called a recommendation, and that is [the providers']. That's not the provider of the underlying information." *See id.* at 119.

[99] *See Gonzalez*, 2 F.4th at 895.

[100] *See id.*

[101] *See, e.g.*, Herrick v. Grindr, LLC, 306 F. Supp. 3d 579, 593 (S.D.N.Y. 2018).

not comport with reality.[102] Providers act beyond the functions of publishers and play active roles in the user experience—they make and send curated messages to achieve effective targeted messaging for third-party advertisers.[103] Courts should adopt an interpretation of section 230 that does not categorically treat all algorithmic functions as publishing functions. If Congress adopts a carve-out for harmful algorithms, plaintiffs can survive a section 230 immunity defense and seek recovery for mental harm caused by certain algorithms.

B.   Challenges Applying Existing Tort Law to Social Media Algorithms

To provide a remedy for mental harm caused by certain social media algorithms, state courts must extend existing tort law, specifically under the theory of negligent infliction of emotional distress ("NIED"). Under existing law, the weight of a plaintiff's burden varies from state to state depending on 'the characterization of the elements that must be established to bring an NIED claim.[104] In California, the plaintiff must establish the traditional tort elements of duty, breach of duty, causation, and damages.[105] A duty's existence depends on reasonably foreseeable risks of emotional injury and a weighing of policy considerations for and against liability.[106] Additionally, the right to recover as a "direct victim" for emotional distress arises from the breach of a duty that is assumed by the defendant or imposed on the defendant as a matter of law, or that arises out of the defendant's preexisting relationship with the plaintiff.[107]

The issue as to whether a duty of care for algorithms exists or should exist remains open for courts to address. Today, section 230 theorizes a duty of care in the general social media context,

---

[102] *See, e.g.*, *Focusing on Testimony from a Facebook Whistleblower*, *supra* note 16, at 23–24 (calling attention to algorithmic biases and computer-driven content under amplification algorithms); *see also* Alina Glaubitz, How Should Liability be Attributed for Harms Caused by Biases in Artificial Intelligence? 13 (Apr. 29, 2021) (Senior Thesis, Yale Dep't of Pol. Sci.) (noting that some algorithms can appear to be "facially neutral" when in reality they are discriminatory in application).

[103] *See* discussion *infra* Part III.A.

[104] *See, e.g.*, Alicea v. Commonwealth, 993 N.E.2d 725, 730 n.9 (Mass. 2013) (requiring a plaintiff to establish negligence, emotional distress, causation, physical harm, and that a reasonable person would have suffered emotional distress under the circumstances to prevail on an NIED claim). *But see* Stancuna v. Schaffer, 998 A.2d 1221, 1226 (Conn. App. Ct. 2010) (requiring a plaintiff to establish that: (1) defendant's conduct created an unreasonable risk of causing emotional distress; (2) plaintiff's distress was foreseeable; (3) the emotional distress was severe enough that it could result in illness or bodily harm; and (4) defendant's conduct caused the plaintiff's distress).

[105] *See* Huggins v. Longs Drug Stores Cal., Inc., 862 P.2d 148, 151 (Cal. 1993).

[106] *See* Molien v. Kaiser Found. Hosps., 616 P.2d 813, 816 (Cal. 1980) (en banc); *see also* Burgess v. Superior Ct., 831 P.2d 1197, 1200 (Cal. 1992).

[107] *See Molien*, 616 P.2d at 816.

but limits the duty to moderation of illegal content.[108] Additionally, courts have raised concerns about imposing a duty of care.[109] The Ninth Circuit stated that "[n]o website could function if a duty of care was created when a website facilitates communication, in a content-neutral fashion, of its users' content."[110] The plaintiff's challenge, it seems, is to persuade the court to impose a duty of care on the interactive service provider to refrain from deploying algorithms that cause mental anguish.[111] Since the original goal behind section 230 immunity was to protect minors from harmful material by incentivizing providers to block and screen such content, imposing a duty on providers to police their *own actions*, rather than the actions of third-parties, would continue to protect minors without chilling third-party speech.[112] As it becomes more apparent that providers are, in fact, aware of the negative effects of their service's algorithm on teens, an argument for the imposition of a duty of care for algorithms can create an avenue for redress while not imposing unreasonable burdens on providers. This Note addresses in Part III that the courts should impose a duty on social media companies to the extent they deploy amplification-type algorithms, given the foreseeable risk of mental harm caused to teens.

Another obstacle to bringing a successful NIED claim is establishing causation—that the algorithm caused the plaintiff's mental harm. There is a great risk that social media litigation might mirror tobacco litigation. Tobacco litigation, under common law causes of action, was unsuccessful for over thirty years because the scientific evidence was insufficient to establish a causal link between tobacco and cancer.[113] Although the scientific community recognizes the link between social media and mental harm, the evidence is still developing and social media businesses are downplaying the linkage.[114]

---

[108] *See* Glaubitz, *supra* note 102, at 29 (noting that social media platforms only have a duty to remove content that is prohibited by law).

[109] *See, e.g.*, Hayes v. SpectorSoft Corp., No. 1:08-cv-187, 2009 WL 3713284, at *1, 11–12 (E.D. Tenn. Nov. 3, 2009) (declining to find that a software provider owes a duty to avoid emotional injury to third-parties harmed by misuse of the software absent prior legal authority).

[110] Dyroff v. Ultimate Software Grp., Inc., 934 F.3d 1093, 1101 (9th Cir. 2019).

[111] *See generally* Artiglio v. Corning Inc., 957 P.2d 1313, 1318 (Cal. 1998) (noting that the existence of a duty to use due care toward an interest that enjoys legal protection against unintentional invasion is a threshold element of a cause of action for negligence).

[112] *See* In re Apple Inc. App Store Simulated Casino-Style Games Litig., No. 5:21-CV-02777-EJD, 2022 WL 4009918, at *4–5 (N.D. Cal. Sept. 2, 2022).

[113] *See* Turley, *supra* note 69, at 446.

[114] *See* Wells et al., *supra* note 8. Adam Mosseri, Instagram head, reported that the app's effects on teen health are likely "quite small" despite evidence showing that Instagram is damaging for many. *Id.*

The third challenge is the element of damages. Leaders in modern health recognize an array of mental health hardships that persist among teens,[115] yet the common law contemplates only those of a sufficient severity that are chronic, or that are more traditionally understood as mental health disorders.[116] Additionally, some states do not permit recovery for emotional distress alone without any accompanying physical injury.[117] A handful of states, however, have turned the page, recognizing NIED as a means to recover for mental anguish without physical injury.[118] In *Rodrigues v. State*, the Hawaii Supreme Court supported extension of the law by noting an important legal interest in protecting individual freedom from "the debilitating effect[s] mental distress may have on an individual's capacity to carry on the functions of life."[119] In jurisdictions that recognize recovery under NIED for emotional distress alone, the court need only apply existing law in determining the damages element to a claim alleging mental harm caused by social media algorithms. Alternatively, if the state court has not modified the traditional rule requiring physical injury, the plaintiff must persuade the court to extend the law to impose a duty of care and allow the plaintiff to recover for mental harm unaccompanied by a physical injury.

C. Congress' Proposed Bill

To address mental harm caused by social media, Congress proposed a bill in September 2021 to create a federal tort against social media companies.[120] The purpose of the tort is limited to the

---

[115] *See Adolescent Mental Health*, WORLD HEALTH ORGANIZATION (Nov. 17, 2021), http://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [http://perma.cc/MZ5Z-PDMG] (noting that depression, anxiety, and behavioral disorders are the leading causes of illness and disability among adolescents, and failure to address adolescent mental health conditions leads to impairment of physical and mental health in adulthood).

[116] *See, e.g.*, Jarrett v. Jones, 258 S.W.3d 442, 448 (Mo. 2008) (en banc) (requiring proof that emotional distress is medically diagnosable and of sufficient severity to be medically significant). *But see* McAllister v. Ha, 496 S.E.2d 577, 583 (N.C. 1998) (noting that emotional distress "means any emotional or mental disorder, such as . . . neurosis, psychosis, chronic depression . . . or any other type of severe and disabling emotional or mental condition which may be generally recognized and diagnosed by professionals . . . .").

[117] *See, e.g.*, Kallstrom v. U.S., 43 P.3d 162, 165 (Alaska 2002) (requiring proof of physical injury to award damages for NIED since plaintiff's case did not fall under Alaska's two narrow exceptions); *see also* Anderson v. Scheffler, 752 P.2d 667, 669 (Kan. 1988) (emphasizing that a plaintiff cannot recover for emotional distress unless that distress results in actual physical injury, and headaches and insomnia are insufficient proof of physical injury).

[118] *See, e.g.*, Rodrigues v. State, 472 P.2d 509, 519–20 (Haw. 1970) (finding that traditional policy concerns limiting NIED to the establishment of physical injury are unpersuasive); *see also* Shuamber v. Henderson, 579 N.E.2d 452, 456 (Ind. 1991) (expanding NIED law to entitle a plaintiff to recover for emotional harm irrespective of whether the emotional harm arises out of or accompanies any physical injury); *see also* Molien v. Kaiser Found. Hosps., 616 P.2d 813, 820 (Cal. 1980) (en banc) (holding that the unqualified requirement of physical injury for NIED is no longer justifiable).

[119] *Rodrigues*, 472 P.2d at 520.

[120] S. 2917, 117th Cong. (2021); H.R. 5449, 117th Cong. (2021).

deterrence of physical and mental harm caused to children less than sixteen years of age by social media companies.[121] While imposing liability for harm caused to teenagers is beneficial to prevent harm to a vulnerable and targeted user group, the companies may actually be incentivized by the language of this regulation to bury their heads in the sand, avoiding liability by asserting lack of knowledge of the harmed user's age.[122] Moreover, the tort is not narrowly tailored to meet the root of the problem: the deployment of amplification-type algorithms.[123] Instead, the tort imposes liability for harm caused merely by *use*.[124] Since social media companies like Instagram and Facebook have the resources and knowhow to alter their platforms to provide more beneficial services to users, liability should be narrowly imposed for harm caused by detrimental capabilities of the algorithm, rather than broadly imposed for harm caused by mere usage.

III. A ROADMAP TO PREVENT FUTURE HARM

Two different routes may prevent social media companies from deploying harmful algorithms that cause mental harm to teens: a common law approach and a legislative approach. Under a common law approach, a plaintiff's success on an NIED claim depends on two important variables: (1) whether the court is willing to adopt an interpretation of section 230(c)(1) that does not treat all algorithmic functions as the function of a publisher; and (2) whether the court is willing to extend tort law as needed to provide relief, including finding that social media companies owe a duty of care in algorithmic development.[125] The alternative route to protect teen' mental health is a legislative approach: Congress allowing the states to regulate under section 230.[126] States could enact laws broad enough to target the harmful conduct—deployment of dangerous algorithms like amplification algorithms—yet impose a burden that is narrowly tailored to solve the problem, consistent

---

[121] H.R. 5449, 117th Cong. (2021).

[122] *See id.* (providing social media companies with an affirmative defense to the federal tort by assertion that the company took reasonable steps to ascertain the age of each user, or that the company did not know or had no reason to know of the user's age).

[123] As discussed in Part I, the issue of amplification algorithms stems from social media platforms' third-party advertising revenue model. The business model is at the heart of the problem. If the platforms were less concerned with engaging users to third-party advertising, a shift away from amplification algorithms would be easier to make. Some scholars have proposed structural reforms as a means to reduce harms caused by the platforms. *See Social Media Design Discourse Hearing*, *supra* note 24 (statement of Tristan Harris, President and Co-Founder of Center for Humane Tech., proposing structural reforms for tech platforms' incentives that would strengthen our capacity to solve problems like addiction and mental health problems).

[124] *See id.*

[125] *See* discussion *supra* Part II.A–B.

[126] *See* 47 U.S.C. § 230(e)(3).

with Congress's policy under section 230.[127] States could even draft such laws in ways that would not implicate section 230 by not premising liability on whether the provider was acting as a publisher of third-party content. Since the providers would not be able to raise section 230 in response to the state law claim, a new interpretation of section 230 would not be necessary to ensure the success of a plaintiff's claim under state law.

These two approaches are not mutually exclusive. However, the legislative approach is preferable because legislators can contemplate business interests along with societal interests to achieve the ultimate goal: preventing harm to teen mental health caused by social media platforms. Additionally, one state's law can be adopted by various states over time to create uniformity. This will ultimately put pressure on social media companies to return to the drawing board to deploy safer algorithms that do not endanger teen mental health.

A.   Incorporate New Understanding of Algorithms into Interpretation of Section 230(c)(1)

This Part III.A proposes an interpretation of section 230, as it applies to algorithms, inspired by the minority opinions in *Force v. Facebook* and *Gonzalez v. Google LLC*. Courts should adopt the following interpretation because a social media company becomes a form of provider-created content and is not exempt from liability under section 230 when it deploys an algorithm that enables it to use third-party content amplifying its *own* message to users to further its own goals.

In the dissent of *Force*, Chief Judge Katzmann suggested that the section 230 does not protect Facebook from claims based on its suggestion algorithms because these claims do not inherently treat Facebook as the publisher of third-party content.[128] To determine whether the claim inherently treats Facebook as the publisher of third-party content, the appropriate question is whether a plaintiff's claim arises from a third-party's information and whether that inquiry requires the court to view the provider as *the* publisher of that third-party information.[129] Even though a provider may publish third-party content, that provider's liability is limited to the harmful function it performs; liability is not based on the provider's identity.[130] Chief Judge Katzmann seemed to recognize that the

---

[127] *See supra* notes 74–77 and accompanying text.

[128] Force v. Facebook, Inc., 934 F.3d 53, 82 (2d Cir. 2019) (Katzmann, C.J., concurring in part and dissenting in part).

[129] *See id.* at 81.

[130] *See id.* (citing Fed. Trade Comm'n v. LeadClick Media, LLC, 838 F.3d 158, 174 (2d

actions of the interactive computer service provider fall on a continuum, where the provider may be the publisher of third-party content on one end, and the third-party may be the publisher of their own content on the other end (when the provider transforms into the speaker of its own message by way of certain algorithms).[131] If the provider transforms into the speaker of its own message, the provider is not the *publisher* of that information but rather a *promoter* of its *own message*. This is because, in this case, the provider is only using the third-party content to promote its message through the process of amplification. While Chief Judge Katzmann focused on harms that Facebook's algorithm causes by connecting users, the same idea—that an algorithm enables a provider to play an "affirmative role" in causing harm—is pointedly applicable to mental health harms that the algorithm causes.[132] It is the basis for arguing why social media platforms perform non-editorial functions when they deploy these algorithms and are thus not within the scope of section 230.

Similarly, a concurring opinion by Judge Berzon in *Gonzalez v. Google LLC* suggests that some algorithms enable providers to perform functions that are not within the scope of traditional publication.[133] Judge Berzon of the Ninth Circuit explained why targeted recommendations and affirmative promotion of interactions among independent users are outside the scope of the traditional publication, and thus are not protected by section 230.[134] Under her view, there is a difference between distributing content to anyone who engages with it and connecting users to specific content, treating the latter as more analogous to a direct marketer than to a publisher.[135] Going a step further, Judge Gould, in his dissent, correctly points out that providers like Google and Facebook can act affirmatively through algorithms to repeatedly direct content to susceptible users, and when plaintiffs' alleged harm is caused by such action, those allegations do not treat the provider as a publisher of the third-party content.[136]

Cir. 2016) (noting that the CDA only bars lawsuits seeking to hold providers liable for exercising traditional editorial functions, such as deciding whether to publish, withdraw, or alter content)).

[131] *See id.* at 76–77 (explaining, through a hypothetical, that it "strains the English language" to say that when the provider targets and recommends information to users, it is acting as the publisher of that information).

[132] *See id.* at 77.

[133] *See* Gonzalez v. Google LLC, 2 F.4th 871, 914, 920 (9th Cir. 2021) (Berzon, J., concurring).

[134] *See id.* at 914.

[135] *Id.* ("Traditional publication has never included selecting the news, opinion pieces, or classified ads to send each individual reader based on guesses as to their preferences and interests . . . .") (alteration in original).

[136] *See id.* at 921 (Gould, J., concurring in part and dissenting in part).

Synthesizing the foregoing opinions, an interactive computer service provider becomes a form of provider-created content and is thus not immune under section 230 when (1) the algorithm enables the provider to select third-party content to affirmatively promote its own message, to (2) targeted or susceptible users, and (3) the provider's suggestions immerse the user in a universe of ideas that gives rise to the probability of harm.

Under the first factor, the question is whether the algorithm merely facilitates communication and content of others or enables the provider to actively communicate with users. Purely neutral search functions exemplify the former, and amplification algorithms, such as recommendation and social connectivity algorithms, exemplify the latter.[137] Even though Facebook's algorithm relies on and displays third-party user content,[138] the anxiety and depression that may result from ordinary use of the platform is caused by the specific algorithm—the engagement-based ranking system—that synthesizes the user data to send a targeted message to the user.[139] A claim containing this allegation does not inherently fault Facebook's activity as the publisher of specific third-party content, but rather as the promoter of Facebook's own message.[140] The recent cases brought against providers involved third-party content that was itself harmful or offensive.[141] Yet, for users suffering from the engagement-based ranking system, it may be the case where each piece of content, on its own and viewed independently, is not itself harmful or offensive.[142] It is in these cases where it is more apparent that the provider plays an active role as a promoter of its own message, rather than as a passive arranger of content. For example, one photo of "how to lose weight" is reasonably not harmful, but impounding a user with similar media several times per day for endless days intensifies and magnifies a message, one that cannot be ignored or assuaged by the user, impacting the user's overall

---

[137] *See id.* at 917 (Berzon, J., concurring).

[138] *See id.* at 914, 917.

[139] *See Focusing on Testimony from a Facebook Whistleblower*, *supra* note 16, at 28.

[140] *See, e.g.*, Force v. Facebook, Inc., 934 F.3d 53, 83 (2d Cir. 2019) (Katzmann, J., concurring in part and dissenting in part).

[141] *See, e.g.*, Cohen v. Facebook, Inc., 252 F. Supp. 3d 140, 146 (E.D.N.Y. 2017) (involving harmful content from third-party terrorist organization); *Force*, 934 F.3d at 59 (involving harmful content from third-party terrorist organization); *Gonzalez*, 2 F.4th at 881 (involving harmful ISIS messaging and videos); Herrick v. Grindr, LLC, 306 F. Supp. 3d 579, 586–87, 589 (S.D.N.Y. 2018) (involving harmful third-party content: impersonating profiles).

[142] *See* 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, 1144 (2020) (recognizing the idea that the content itself is not harmful but rather the way it is shown to the user).

mental health.[143] This supports Judge Katzmann's conclusion that Facebook plays two roles as a service provider: the publisher of third-party content and the promoter of its own message to target the user based on statistical analysis of user information.[144] The latter is not protected under section 230.

Under the second factor, the question is whether the algorithm acts on user-generated data. For example, Google (through YouTube), Facebook, and Twitter promote content to users who are susceptible to the harmful consequences of repeatedly viewing a subject of media.[145] Suggesting content to users without any determination of user interest does not pose the same harm because the user is not as vulnerable to the provider's message. To not protect interactive computer service providers merely because they suggest content would be detrimental to the service models that rely on advertising revenue. However, where the algorithm displays curated content to a user it has determined is engaged with the content, this aspect contributes to the dominating effect of the provider over the user and thus sets the stage for harm to occur.

Under the third factor, the question is whether the cumulative effect of suggestive content dominates the user experience.[146] Where the algorithm enables the provider to interject its own message through its suggestive content, the provider may envelop the user, "immersing her in an entire universe filled with people, ideas, and events she may never have discovered on her own."[147] Facebook's purpose is to build tools to help people connect.[148] However, the current algorithm metrics do not put Facebook in the category of a passive service provider, providing the user with neutral features to build and maintain relationships with other users. On the contrary, Facebook is more like a promoter, interjecting a targeted viewpoint through the display of content that immerses the viewer with ideas that are not of the user's own volition. This function, executed by the algorithm, is beyond the traditional editorial functions that section 230 immunizes.[149] The interjection may be as simple as "you may be interested in viewing this content or connecting with these people," but it is a message that the user would not have received on a platform deploying a "neutral" algorithm. Similarly, YouTube's algorithm recalibrates

---

[143] *See Gonzalez*, 2 F.4th at 921 (Gould, J., concurring in part and dissenting in part) (describing how a seemingly neutral algorithm amplifies messages).

[144] *See Force*, 934 F.3d at 83 (Katzmann, J., concurring in part and dissenting in part).

[145] *See, e.g.*, *Gonzalez*, 2 F.4th at 921 (Gould, J., concurring in part and dissenting in part).

[146] *See id.* at 917 (Berzon, J., concurring).

[147] *Force*, 934 F.3d at 83 (Katzmann, J., concurring in part and dissenting in part).

[148] *Id.*

[149] *See id.*

the user's existing interests to steer them toward new interests, often by displaying more divisive and extreme content.[150] In both situations, the provider dominates the user by purposefully intercepting third-party content to convey a targeted message by the provider for the purpose of achieving any number of the provider's goals, like keeping users engaged on the platform for longer periods of time.

Adopting an interpretation that carves out certain algorithms from section 230's protection does not stunt the beneficial growth of the internet. Rather, such an adoption would help prevent the harmful effects of Internet use that were not understood at the time of its enactment. As Chief Judge Katzmann pointed out in his dissent in *Force*, where claims rest not on the content of the information but on the rules of the algorithm, the congressional intent of section 230 does not compel the judiciary to provide immunity.[151] Moreover, the suggested carve-out is itself narrow, and thus would still advance section 230's aim at giving providers breathing space to grow.[152] By broadly immunizing providers, they are not incentivized to make their algorithms safer, despite knowledge of the harmful impact on users.[153] Taking providers out of the purview of section 230 for deploying algorithms that fall within the narrow confines of the proposed factors would reasonably deter service providers from utilizing such algorithms and incentivize modifications to promote beneficial growth of the Internet, rather than plague users with emotional distress. Lastly, the narrow door would allow legitimate state law claims to be reviewed.[154]

B.   Impose a Duty of Care in Light of a New Understanding of Algorithms

By adopting this Note's proposal that some algorithmic capabilities treat social media companies as promoters of their own messages rather than as publishers of third-party content, remedial courses of action—such as NIED—should survive an immunity defense under section 230 if two issues are resolved in favor of the plaintiff. At this point, the first issue is whether providers owe a duty of care to users for deploying harmful algorithms.[155] If answered affirmatively, the second issue is

---

[150] *See id.* at 87.

[151] *See id.* at 77.

[152] *See* Gonzalez v. Google LLC, 2 F.4th 871, 921 (9th Cir. 2021).

[153] *See id.* at 920 (noting that a genuine factual issue exists as to whether social media companies are aware of the risks to the public stemming from content-generating algorithms).

[154] *See, e.g.*, Herrick v. Grindr, LLC, 306 F. Supp. 3d 579, 593 (S.D.N.Y. 2018) (barring NIED under the current reading of the CDA).

[155] *See* Artiglio v. Corning Inc., 957 P.2d 1313, 1318 (Cal. 1998).

whether the tortious conduct is framed such that the alleged duty does not treat the interactive computer service provider as a publisher or speaker of third-party content.[156] To evade the purview of section 230, this Part III.B will discuss how to frame the tortious conduct for an NIED claim by analogizing to two recent cases involving social media companies defending against negligent design claims.

The court should impose a duty of care on the defendant (interactive computer service provider) when (1) a person suffers severe mental harm from use of a social media platform, (2) the harm is caused by the platform's algorithm, and (3) the platform knew or should have known of the foreseeable risk of harm. To determine whether a duty of care exists, state courts consider various factors. For example, the California Supreme Court considers the following:

> The major [considerations] are the foreseeability of harm to the plaintiff, the degree of certainty that the plaintiff suffered injury, the closeness of the connection between the defendant's conduct and the injury suffered, the moral blame attached to the defendant's conduct, the policy of preventing future harm, the extent of the burden to the defendant and consequences to the community of imposing a duty to exercise care with resulting liability for breach, and the availability, cost, and prevalence of insurance for the risk involved.[157]

Applying these factors, a duty should be imposed on providers, like Facebook and Instagram, that deploy engagement-based ranking algorithms because there is a high risk of emotional distress and harm caused by such algorithms.[158] The extent of the burden on the defendant is appropriately limited if the duty is triggered only when the interactive computer service provider knows or has reason to know of the risk of harm from use of its platform. For example, with the revelation of Haugen's insights, it is evident that Facebook has knowledge of the harm posed by its conduct, yet it has not proposed a solution to prevent the harm. As for the consequences to the community for the imposition of a duty of care on social media companies, they likely weigh more in favor of imposition. If liability causes social media companies to rework algorithms to improve the user experience, we can help improve mental health for a generation of people currently suffering.[159] Also, liability would likely incentivize healthy technological innovation in the context of social media

---

156 *See* Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1102 (9th Cir. 2009) (noting that what matters is whether the claim "inherently requires the court to treat the defendant as the 'publisher or speaker' of content provided by another").

157 *See* Randi W. v. Muroc Joint Unified Sch. Dist., 929 P.2d 582, 588 (Cal. 1997).

158 *See* discussion *supra* Part I.B.

159 *See supra* note 24 and accompanying text.

rather than hinder it, or worse, promote innovation that does not consider mental wellness at all.[160] One of the difficulties establishing the duty is the closeness of the connection between the algorithm and the injury.[161] The degree of closeness is exemplified by answering whether modification of the algorithm would prevent the emotional distress, since this is the obligation that would be imposed on Facebook.[162] It may be difficult for a plaintiff to establish that the risk of harm could be prevented by modifying the algorithm when the claim is against a social media company whose internal research is not publicized, or where the company's knowledge of the risk is not publicly apparent. But according to Haugen, Facebook's internal reports show that modifying the amplification algorithm would alleviate the harms caused to users, and outside studies tend to show that the risk could be prevented.[163]

A criticism to imposing a duty on social media companies is that these social media companies may be encouraged to be less vigilant or proactive in conducting internal studies. This is problematic because social media companies possess the data, resources, and workforce to conduct accurate research efficiently,[164] so they are in the best position to assess the quality of their service and its impact on users.[165] The state legislature is thus likely the more appropriate forum to simultaneously (1) encourage social media businesses to study the use of their platforms and develop their algorithms in pursuit of healthier

---

[160] In the context of AI development for autonomous vehicles, the prospect of tort liability could hinder innovation because the market is still developing. *See, e.g.*, Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1326 (2020). This economic concern is not as great for the social media industry because it is not as new of a market. *See, e.g.*, Brian Dean, *Instagram Demographic Statistics: How Many People Use Instagram in 2022?*, BACKLINKO, http://backlinko.com/instagram-users [http://perma.cc/F8HB-PJWD] (last updated Jan. 5, 2022) (noting that about 500 million users around the world access Instagram daily).

[161] On the one hand, a "tight causal nexus" between conduct and its consequences is fundamental to a fair assignment of liability; however, on the other hand, an economic theorist may argue "that the goals of tort law lie in optimal deterrence or efficient risk allocation." *See, e.g.*, Selbst, *supra* note 160, at 1321.

[162] *See* Nathalie Dalzell, *Telecommunications Law - Facebook Immunized from Civil Liability under Communications Decency Act Despite Using Algorithms to Recommend Content - Force v. Facebook*, 934 F.3d 53 (2d Cir. 2019), cert. denied, 140 S. Ct. 2761 (2020) (mem.), 54 Suffolk U. L. Rev. 599, 610 n.54 (2021) ("Most common torts regarding Facebook's algorithms arise from . . . publishing material that inflicts emotional distress.").

[163] *See Focusing on Testimony from a Facebook Whistleblower*, *supra* note. 16, at 6.

[164] Instagram uses the information it gathers to study its service and "collaborate with others on research to make [it] better and contribute to the well-being of [the] community." *Terms of Use*, *supra* note 45.

[165] Notably, Facebook does not make its research public, even for academics and lawmakers who have asked for it. Wells et al., *supra* note 8.

user experiences, and (2) hold these businesses accountable for mental harm caused to users.[166]

Under Colorado state law, the court considers a different set of relevant factors and reserves consideration of any other relevant factors based on competing individual and societal interests implicated by the facts of the case.[167] In *English v. Griffith*, parents asserted an NIED claim against a woman for engaging in an argument with their son, allegedly causing their son such severe emotional distress to the point of causing him to take his life.[168] The Colorado Court of Appeals was asked to impose a duty on an individual not to cause another, who was known to be susceptible to emotional distress, to take his life.[169] The court did not find that the defendant owed a duty because the defendant could not "reasonably be expected to anticipate the mental health consequences that may flow from otherwise ordinary conduct such as the argument that allegedly occurred" in the case.[170] Under this line of reasoning, one might similarly argue that providers like Facebook and Instagram cannot reasonably be expected to foresee the mental health consequences that may flow from otherwise ordinary conduct— the use of social media—and therefore, a duty should not be imposed. However, unlike in *Griffith*, where the likelihood of injury resulting from the ordinary conduct was "extremely low,"[171] the likelihood of mental harm among teens caused by usage of social media tied to the amplification algorithm is high. Moreover, *Griffith* involved a defendant who was an individual, not a business entity.[172] Society may be more hesitant to burden individuals with legal duties to guard against mental harm. Conversely, society may have a greater interest in imposing a legal duty on a multibillion-dollar entity[173] that holds tremendous power over users, wields user trust, and knowingly

---

166 *See, e.g.*, Levin, *supra* note 7, at 16–17 (noting that government interference is justified where platforms can use the algorithm to set the agenda in harmful ways without government parameters); *see also* discussion *infra* Part IV.C (describing additional advantages to a state legislative approach).

167 *See* English v. Griffith, 99 P.3d 90, 94 (Colo. App. 2004) (considering, for purposes of imposition of a legal duty: "(1) the risk involved; (2) the foreseeability of harm to others and likelihood of injury as weighed against the social utility of the actor's conduct; (3) the magnitude of the burden of guarding against the injury or harm; and (4) the consequences of placing the burden on the actor").

168 *See id.* at 92.

169 *See id.* at 94.

170 *Id.*

171 *Id.*

172 *Id.*

173 *See Facebook | Meta | FB - Market Capitalization*, TRADING ECON. (Apr. 2023), http://tradingeconomics.com/fb:us:market-capitalization [http://perma.cc/E4RS-T673].

deploys an algorithm that exploits users' personal vulnerabilities to control their experience of the platform.

To defeat a section 230 defense to an NIED claim, the plaintiff must ensure that its allegations do not treat the provider as a publisher of third-party content but rather as a promoter of its *own* message.[174] The following two recent cases exemplify the differences between the former and the latter. In *Doe v. Twitter*, two thirteen-year-olds were manipulated into providing pornographic videos to a third-party sex trafficker, and the videos were posted on Twitter a few years later.[175] They asserted a state law claim based on negligent design, seeking to hold Twitter liable for enabling users to disseminate information quickly to large numbers of people, as well as for failing to deploy measures that prevent suspended users from opening new accounts and disseminating harmful content.[176] The district court held that these allegations treated Twitter as a publisher protected by the CDA because "Twitter would have to alter the content posted by its users" to meet the obligation plaintiffs sought to impose.[177] In reaching this conclusion, the court distinguished the allegations from those made in *Lemmon v. Snap*, where a negligent design claim was not barred by section 230.[178]

In *Lemmon v. Snap*, the plaintiffs "were parents of two boys who were killed in a high-speed car accident."[179] They brought the action against Snap, Inc., the owner of Snapchat.[180] The parents alleged that Snapchat's "speed filter incentivized young drivers to drive at high speeds" and that Snapchat "was aware of the danger" of the filter from news articles and other accidents linked to Snapchat users' high-speed snaps.[181] In this case, the negligent design was not barred by section 230(c)(1) because the claim sought to hold Snapchat liable for its conduct as a manufacturer rather than as a publisher of third-party content.[182] The primary reason for this conclusion was the fact that Snapchat could have "take[n] reasonable measures to design

---

[174] *Cf.* Lemmon v. Snap, Inc., 995 F.3d 1085, 1092 (9th Cir. 2021) (finding that a negligent design lawsuit treats the social media company as a products manufacturer, and the duty underlying such claims differs from the duties of publishers as defined in the CDA); *see also* Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1102 (9th Cir. 2009) (finding that a determination of whether a provider is a publisher protected by the CDA is based on "whether the duty that the plaintiff alleges the defendant violated derives from the defendant's status or conduct as a 'publisher or speaker'" (quoting 47 U.S.C. § 230(c)(1)).

[175] Doe v. Twitter, Inc., 555 F. Supp. 3d 889, 893–94 (N.D. Cal. 2021).

[176] *Id.* at 930.

[177] *Id.*

[178] *Id.*

[179] *Id.* at 929 (describing the facts of *Lemmon v. Snap*).

[180] *Id.*

[181] *Id.*

[182] *See id.* at 929–30.

a product more useful than it was foreseeably dangerous . . . without altering the content that Snapchat's users generate."[183] Since the speed filter was affirmatively created by Snapchat, the flaw was dependent on Snapchat's actions, rather than any posting of third-party content.[184]

An NIED claim against social media platforms in which plaintiffs allege that the algorithm, like engagement-based ranking, causes mental harm is more like the claim made in *Lemmon* and should withstand a section 230 defense where courts adopt this Note's proposal for an algorithm carve-out.[185] For example, an NIED claim against Instagram would seek to hold Instagram liable for its promotional action: targeting third-party content at users to send a message from Instagram meant to keep the user engaged on the platform which, as a result, harms the user.[186] This framing of Instagram's conduct does not treat Instagram's duty as that of a publisher of third-party content within the scope of section 230 immunity because the alleged duty does not rest on any affirmative obligation to remove, alter, monitor, or edit third-party content.[187] Rather, it is a duty to use reasonable care to refrain from writing algorithms that enable Instagram to send messages to targeted users that foreseeably cause mental distress.[188] Like in *Lemmon*, where the allegations treated Snap as liable for its conduct as a manufacturer, an NIED claim alleging that Instagram acted unreasonably by failing to deploy a safer algorithm, given foreseeable risks of harm, holds Instagram liable for its conduct as a business deploying a harmful algorithm, rather than for its conduct as a publisher.[189]

---

[183] *Id.* at 929.

[184] *Id.*

[185] *See* Lemmon v. Snap, Inc., 995 F.3d 1085, 1094 (9th Cir. 2021) ("CDA immunity is available only to the extent a plaintiff's claim implicates third-party content.").

[186] *Cf. id.* (finding that "even if [the social media company were] acting as a publisher in releasing . . . its various features to the public, the . . . claim still rests on nothing more than [the company's] 'own acts'" (quoting Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1165 (9th Cir. 2008)).

[187] *See Doe*, 555 F. Supp. 3d at 925–26; *cf. Lemmon*, 995 F.3d at 1092 (noting that the negligent design lawsuit did not treat the social media company liable as a publisher because the alleged duty had nothing to do with editing, monitoring, or removing third-party content).

[188] Scholars have argued that social media companies' economic motivation, combined with the lack of an internal ethical code, is one theory for justifying regulatory intervention. *See, e.g.*, Levin, *supra* note 7, at 32–33. The same reasoning supports the argument for the imposition of a duty. *See id.* If these companies are focused on generating revenue from third-party advertisers, the people behind the business should be held to a reasonable standard of care in the development of the platform to prevent harm to users for whom the platform exists.

[189] *See Lemmon*, 995 F.3d at 1092 (noting that the duty on the social media company arose from its capacity as a product designer, as evidenced by the fact that the company failed to take reasonable measures to design a product more useful than was foreseeably dangerous).

C.   State Legislature or Courts? Set the Parameters for Social Media Businesses

For social media platforms that follow an advertising-based revenue model, maximizing revenue will naturally put third-party advertiser interests at the forefront of algorithm development.[190] Absent an economic motivation to otherwise prioritize users' mental health, teenagers are at the mercy of the platforms. Thus, passing legislation that incentivizes social media companies to turn their attention back to the users may be the most effective approach to protect teenage mental health and well-being, especially in a world where, for many, the thought of dissolution of social media is unimaginable.

One advantage to a state legislative approach, as opposed to a judicial approach, is that deterrence of harmful social media practices is wrapped up in complex policy questions that are best left to each state.[191] Although social media and tobacco are uniquely similar in their addictive qualities targeting teens, regulating social media is more convoluted than tobacco regulation because social media *can* be positive,[192] and it *is* largely good for small businesses and other stakeholders—including the workforce, supply chain of businesses, and other advertisers.[193] The citizens of every state may feel differently about the extent of the burden that should be imposed on social media companies. For example, some states may wish to impose liability only for harms caused to vulnerable user groups, like teenagers, which is a limitation that cannot be imposed under an NIED cause of action. Rather than asking courts to extend tort law and create a zone of liability without considering the public voice, states can enact more optimal solutions that reflect competing interests. Although social media businesses would face fifty different remedies from state legislation, the first state law will serve as the blueprint for other states. Moreover, any patchwork of laws and judgments that may result would not likely contort the national market any more than state common law courses of action.[194]

---

190 *See supra* note 44 and accompanying text.

191 *See, e.g.*, Levin, *supra* note 7, at 34 (stating that government intervention "could be used to create ethical rules and norms that apply to all social media platforms, combined with the means to enforce them").

192 *See, e.g.*, Jacqueline Tabas, *How Nonprofits Can Use Social Media to Increase Donations and Boost Visibility*, FORBES (Mar. 6, 2021, 09:00 AM), http://www.forbes.com/sites/allbusiness/2021/03/06/how-nonprofits-can-use-social-media-to-increase-donations-and-boost-visibility/?sh=5800326a2bb7 [http://perma.cc/35HP-82G5] (noting how social media helped nonprofits achieve fundraising goals). Notably, the problem with the tobacco crisis was that states could enact laws that "eliminated the core defense needed by the tobacco industry to defend itself." Turley, *supra* note 69, at 472.

193 *See supra* notes 40–41 and accompanying text.

194 *See* Turley, *supra* note 69, at 468.

Another advantage to a state legislative approach is that neither re-interpretation nor reform of section 230 is necessary to allow relief to teens for mental harm; thus, a broad interpretation of section 230 in conjunction with a narrow state law may cohesively work to achieve, deter, and prevent future mental harm. Congress gave states implicit permission in section 230(e)(3) to enact law pertaining to interactive computer service providers, stating that "[n]othing in this section shall be construed to prevent any State from enforcing any State law that is consistent with this section."[195] As long as the enforcement action doesn't conflict with or undermine section 230,[196] states may address challenges of interactive computer service providers under their general police power.[197] A broad reading of section 230 affords social media companies protection while states are granted latitude to protect users. This is important because, as Professor Jonathan Turley has noted, states have an "interest in private litigation" and, if capable, can "construct procedures that can act like legal speedtraps to capture wealth."[198] Since the states cannot enact law that is inconsistent with section 230, social media companies would still be protected from allegations of liability for conduct that is outside their control, like the posting of harmful content by a third-party. States will then be afforded the opportunity to enact law that holds these businesses accountable for conduct that is within their control, like algorithm development. A critique of this argument is that even legislation will struggle to effectively regulate platforms given the fast-paced development of technology and business operations. However, the nuances of technology and the harms it causes are more appropriately handled by the legislature—as opposed to courts—since the legislature can rewrite, repeal, and amend, and is not bound by precedent.

On the other hand, why not a federal legislative solution? Federal regulation would establish uniform liability, eliminating the burden on social media companies of sifting through state laws to ensure compliance. However, state legislatures are the appropriate forum to craft a creative solution for a national

---

[195] 47 U.S.C. § 230(e)(3).

[196] *See* Google LLC v. Equustek Sols. Inc., No. 5:17-CV-04207-EJD, 2017 WL 5000834, at *2, *4 (N.D. Cal. Nov. 2, 2017) (allowing the interactive service provider to assert section 230 offensively to enjoin enforcement of a court order that undermined the policy goals of section 230).

[197] *See, e.g.*, Barbier v. Connolly, 113 U.S. 27, 31 (1884) (noting states' power, termed as the "police power," to "prescribe regulations to promote the health, peace, morals, education, and good order of the people, and to legislate so as to increase the industries of the state, develop its resources, and add to its wealth").

[198] Turley, *supra* note 69, at 471.

problem, balancing various stakeholder interests.[199] Additionally, if the states legislate to solve for mental harm by imposing liability for the deployment of harmful algorithms—even if the damages differ, the standard of causation differs, or the pool of plaintiffs eligible to take advantage of the law differs—the laws would nonetheless have the same effect on the businesses by pushing them to deploy less harmful algorithms. Lastly, providers like Facebook are already accustomed to navigating unique state laws, like data privacy laws, and they make changes to their business to comply with these laws because it is in their best interest. For example, California enacted a privacy law that gives Californians special privacy rights.[200] The law applies to Internet providers that "operate in the state, collect personal data for commercial purposes[,] and meet other criteria" like generating revenue that exceeds a threshold.[201] In response to the new legislation, many providers, like Microsoft, decided to "apply their changes to all users in the United States rather than give Californians special treatment."[202] Similarly, if providers were faced with a state law that imposed liability for deploying algorithms that harm teens who reside in the state, the providers could act in a manner that benefits all teen users.

D.   Enact Law that Encourages Businesses to Play an Active Role in a Healthier World

If we accept the premise that some government intervention is necessary and desirable to ensure that all persons do, in fact, benefit from the use of the Internet, as Congress believed was already the case,[203] then the question is *how* to intervene. Social media platforms can be designed to foster community safety,[204] even with the help of algorithms.[205] A law that is broad enough to meet today's problem of mental harms arising from social media use and prevent the problem of advanced targeted messaging

---

199 *See* New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting) ("It is one of the happy incidents of the federal system that a single courageous state may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country.").

200 *See* Natasha Singer, *What Does California's New Data Privacy Law Mean? Nobody Agrees*, N.Y. TIMES (Dec. 29, 2019), http://www.nytimes.com/2019/12/29/technology/california-privacy-law.html?msclkid=908c0927d0bf11ec9e11cd47c28c9bf4.

201 *Id.*

202 *Id.*

203 *See* 47 U.S.C. § 230(a)(4).

204 *See Social Media Design Discourse Hearing*, *supra* note 24 (statement of Joan Donovan, Ph.D. Research Director at Harvard Kennedy School's Shorenstein Center on Media, Politics and Public Policy).

205 *See, e.g.*, Sadagopan, *supra* note 95 (explaining how algorithms can draw inspiration from human intelligence to break harmful feedback loops).

tomorrow may be the best solution.[206] This solution requires that we meet the root of the problem: algorithms.[207]

A state law should be directed at the creators of algorithms to encompass interactive service providers, as well as businesses that do not satisfy Congress's definition of an interactive computer service.[208] The law should be articulated as follows: A creator of an algorithm shall be liable to any consumer who suffers bodily injury or harm to mental health when the consumer was less than twenty-years-old that is attributable, in whole or in part, to the individual's use of technology that deploys a covered algorithm, where the creator of the algorithm knew or should have known of the risk of harm to the user. The term "creator of an algorithm" means an interactive computer service or other business that uses a covered algorithm to enhance a service or product provided to consumers.[209] The term "consumer" means purchasers, users, patrons, and clients.[210] The term "interactive computer service" has the meaning given to the term in section 230 of the Communications Decency Act (47 U.S.C. § 230).[211] The term "covered algorithm" means reinforcement algorithms,[212] amplification algorithms,[213] and any other

---

[206] *See* Watson, *supra* note 62, at 19 (noting how the Internet service provider, Google, has already introduced a new artificial intelligence that significantly improves clickthrough rate predictions); *see also* Glaubitz, *supra* note 102, at 6–7 (explaining the four generations of artificial intelligence and pointing out that engineers have only begun to develop the second generation).

[207] *See generally* Adam Beam, *Social Media Addiction Bill Fails in California Legislature*, AP NEWS (Aug. 11, 2022), http://apnews.com/article/social-media-california-legislature-f5fd4c8ac90546c506bc3a685ab58b2b [http://perma.cc/AU8U-EC2G] (discussing the failure of a bill that would hold social media companies accountable for knowingly using features that cause addiction). Since software development is at the core of Instagram's business, and the company has decided that the benefits of its algorithm outweigh the costs of harm, it is appropriate to hold it accountable for its intentional development and deployment of the algorithm. Additionally, limiting liability to the deployment of an algorithm not only narrows the scope, but it also accounts for future algorithm-caused harms known to businesses beyond the social media space.

[208] 47 U.S.C. § 230(f)(2) (defining "interactive computer service" as "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server").

[209] *See, e.g.*, *Terms of Use*, *supra* note 45 (discussing how Instagram uses automated technologies to ensure the functionality and integrity of the service).

[210] *See Consumer*, OXFORD LEARNER'S DICTIONARIES, http://www.oxfordlearners dictionaries.com/us/definition/american_english/consumer [http://perma.cc/2ZYQ-KC69] (last visited Sept. 16, 2022).

[211] 47 U.S.C. § 230(f)(2).

[212] *See Social Media Design Discourse Hearing*, *supra* note 24 (statement of Joan Donovan, Ph.D. Research Director at Harvard Kennedy School's Shorenstein Center on Media, Politics and Public Policy) (noting that reinforcement algorithms pattern the distribution of content based on user signals to reinforce user interests).

[213] *See Focusing on Testimony from a Facebook Whistleblower*, *supra* note 16, at 28; *see also Social Media Design Discourse Hearing*, *supra* note 24 (statement of Monika Bickert, Vice President for Content Policy, Facebook) (noting that amplification algorithms use a personalized ranking process driven by user choices and actions to sort content).

algorithm that enables the creator to select third-party content to communicate its own message to targeted consumers. An individual who suffers bodily injury or harm to mental health that is attributable, in whole or in part, to the individual's use of technology that deploys a covered algorithm where the creator knew or should have known of the risk of harm to the user may bring a civil action against the creator in an appropriate State court of competent jurisdiction for compensatory damages or actual damages, punitive damages, and attorney's fees and costs. If the user shows that the user's mental harm is attributable to the algorithm, and the creator knew or should have known of the risk of harm, the burden shifts to the creator to show that it acted reasonably in the deployment of the algorithm.[214]

The language of such a law is advantageous for several reasons. First, it is broad enough to remedy mental health harms caused to teenagers using social media platforms, yet also include other harms.[215] Second, by targeting algorithms, the law does not interfere with content moderation practices or regulation thereof—which currently stand amidst the crossfire of differing policy viewpoints.[216] This is because incentivizing safe algorithmic development doesn't impact the flow of third-party content itself on the platforms. Third, the language solves for unknown future harms caused by harmful algorithms by deterrence and through ease of amendment. As we discover more about the types of algorithms that cause harm to users, the legislature could amend the definition of "covered algorithms" to remain relevant and effective. Fourth, the law is reasonably tailored in two ways: (1) it imposes liability only for harm caused to teenagers, a more vulnerable and targeted group; and (2) it also limits liability to knowledge or scienter of the provider, which is in accordance with the literal language of section 230.[217] Fifth, the burden on businesses is also reasonable because it does not impose liability for mere usage of the technology, like

---

214 *See generally* Citron, *supra* note 71, at 16–20 (introducing a "reasonable steps approach" as one way to reform section 230(c)(1) to solve for harm caused to users by third-party content).

215 *See, e.g.*, Glaubitz, *supra* note 102, at 29 (describing disparate impact caused by algorithms); *see also* Gonzalez v. Google LLC, 2 F.4th 871, 921 (9th Cir. 2021) (Gould, J., concurring) (describing political violence caused by algorithms).

216 *See, e.g.*, Nina I. Brown & Jonathan Peters, *Say This, Not That: Government Regulation and Control of Social Media*, 68 SYRACUSE L. REV. 521, 541–42 (2018) (arguing that regulation of content moderation risks First Amendment violations). *But see* Citron, *supra* note 71, at 22 (arguing that leaving the Internet under current regulation actually "*chills* valuable speech").

217 *See Gonzalez*, 2 F.4th at 920–21 (Gould, J., concurring) (arguing that the text of section 230 does not suggest immunizing providers from liability for serious harms knowingly caused by their conduct).

Congress's bill proposes,[218] but rather limits liability to the deployment of particular algorithms described in the law.

As for the affirmative defense of reasonableness, the creator must be able to point either to robust internal research that does not show an association between the creator's algorithm and depression, anxiety, suicidal thoughts, or other emotional distress, or to steps it took to prevent such mental harm. Regarding the latter, the law should articulate examples of reasonable steps. For example, if the creator deploys an algorithm that manipulates a person toward a targeted message, the creator can show that it took reasonable steps by alerting the person—while the person was using the service—that he or she received a targeted message by the computer service from the choice of content displayed. The creator could also set up a system on the platform where users answer survey questions aimed at understanding user mental health, then regularly post findings to public bulletins on the platform. Liability is ultimately imposed if the algorithm's creator fails to show that it acted reasonably. As Professor Danielle Citron points out in her argument for section 230 reform, a reasonableness approach is "valuable precisely because it is flexible."[219] This kind of burden-shifting law may be the best way to balance society's interest in protecting teenage mental health and the market's interest in connecting small business advertisers with an engaged audience—all while incentivizing businesses to innovate algorithms in a healthier direction.

CONCLUSION

Innovation and creativity drive the world of marketing and business. New strategies will be developed to help businesses reach more people faster and at the least expense. Artificial intelligence is one such proven strategy, yet its value to some businesses is at a great cost to teen mental health. As we continue to discover the potential of artificial intelligence for targeted marketing, questions of law and ethics must be at the forefront. Today, we face a teen mental health crisis, partly impacted by social media algorithms. Social media platforms are best suited to change the nature of their algorithms to reduce harm, but change is not on the horizon where business models are based on third-party advertising.

---

[218] *See* discussion *supra* Part III.C.

[219] *See* Citron, *supra* note 71, at 19–20 (noting that requiring businesses to show reasonableness pressures platforms to keep up with best practices and defend those practices in litigation, ultimately establishing industry standards "that have the force of law to back them up").

The legal landscape is currently ill-equipped to help teens seek legal redress for mental harm caused by social media algorithms. But we must find a way to hold social media companies accountable for the harmful externalities of tech development and protect teens from ongoing mental harm. Courts should adopt an interpretation of section 230 of the Communications Decency Act that does not bar claims seeking relief for mental distress caused by harmful algorithms. The amplification algorithm—and others like it—executes user engagement strategies that treat the social media platform as a promoter, not as a publisher of third-party content. A duty of care should be imposed on social media companies for algorithm deployment because these providers are in the best position to deploy alternative, less harmful algorithms. Furthermore, severe harm to teen mental health outweighs any associated cost to advertisers. Beyond the court system, state legislatures can directly target the root of the problem—algorithms—with laws that balance competing stakeholder interests. A state legislative approach is probably favorable to a common law approach, since the legislature can craft unique laws that consider both society's stance on the extent of regulation and the future of algorithm development in the context of targeted messaging. If robust protective measures guard the stairs of technological innovation, we can take big steps toward ensuring teen safety and improving the lives of many.