Date of Hearing: April 30, 2024

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
AB 2481 (Lowenthal) – As Amended April 23, 2024

AS PROPOSED TO BE AMENDED

SUBJECT: Social media-related threats: reporting

SYNOPSIS

According to the Surgeon General, "[t]he current body of evidence indicates that while social media may have benefits for some children and adolescents, there are ample indicators that social media can also have a profound risk of harm to the mental health and well-being of children and adolescents. At this time, we do not yet have enough evidence to determine if social media is sufficiently safe for children and adolescents." Existing law, the Cyberbullying Protection Act requires social media platforms to disclose their cyberbullying reporting procedures and to implement a reporting mechanism for the reporting of cyberbullying between pupils as well as other conduct that violates the platform's terms of service. The Attorney General is authorized to bring suit against platforms for intentional failure to comply and seek a civil penalty of $7,500 per day, per violation.

This bill, modeled after the Cyberbullying Protection Act, would establish the Youth Social Media Protection Act, which would create a two-tiered reporting mechanism for "social media-related threats" —content posted on a social media platform that promotes, incites, facilitates, or perpetrates certain things, including cyberbullying, suicide, and drug trafficking. Under the bill, any person would be able to report such threats to social media platforms and would be entitled to receive a response as to whether the content violates the platform's terms and conditions. Additionally, with regard to the biggest social media platforms, certain "verified reporters"—school counselors, principals, and licensed mental health professionals—would be entitled to expedited review of any reports of social media-related threats. In cases where the reporter deems the risk severe, a human must review the content. Those platforms would be required to annually post on their websites information relating to reports received by verified reporters. The bill's provisions are enforceable by a private right of action.

The bill is sponsored by the Organization for Social Media Safety, which argues social media platforms are failing to properly respond to harmful content, at the expense of millions of children, and thus a more robust reporting system requiring an actual affirmative response from the platform is necessary.

The bill is opposed by a coalition of industry associations—Computer & Communications Industry Association, California Chamber of Commerce, and TechNet—as well as Electronic Frontier Foundation. Some of their concerns have been addressed by recent amendments that clarify the scope of a platform's obligation to respond; the prior version of the bill arguably required platforms to remove certain offending content. As amended, the bill requires nothing more than a response to the reporter as to whether the content violates the platform's terms and conditions. Still, concerns persist over the breadth of the bill's dual reporting mechanisms as well as the private right of action, the combination of which, in their view, could chill protected

speech. Committee amendments specify that standing to bring suit under the bill is limited to reporters and cap damages for individual violations at $10,000.

SUMMARY: Creates a two-tiered reporting mechanism for "social media-related threats" — content posted on a social media platform that promotes, incites, facilitates, or perpetrates certain enumerated harmful outcomes. Any person would be able to report such threats to social media platforms and receive a response under specified timeframes as to whether the content violates the platform's terms and conditions. Additionally, with regard to the biggest social media platforms, certain "verified reporters"—school counselors, principals, and licensed mental health professionals—would be entitled to expedited review of any reports of social media-related threats. Reports of severe risks from verified reporters must be undertaken by a human. Large platforms would be required to annually post on their websites information relating to reports received by verified reporters. Enforced via a private right of action. Specifically, **this bill**:

1) Defines "social media-related threat" as content that promotes, incites, facilitates, or perpetrates any of the following: suicide, disordered eating, drug trafficking, substance abuse, fraud, human trafficking, sexual abuse, cyberbullying, harassment, academic dishonesty, stalking, or identity theft.

2) Requires a social media platform to disclose all media-related threat reporting procedures in the social media platform's terms of service.

3) Requires a social media platform to establish a mechanism within its internet-based service that allows an individual, whether or not they have a profile on the platform, to report a "social media-related threat" or any content that violates the social media platform's terms of service that meets all of the following:

    a. The mechanism enables, but does not require an individual to upload a screenshot of the content that contains a social media-related threat or violates the terms of service.

    b. The mechanism offers an electronic point of contact specific to matters involving harms to a minor.

    c. The mechanism provides confirmation of the receipt of a submitted report and a means to track that report.

4) Requires a "large social media platform" —those with 100 million monthly global active users or $1 billion in gross revenue per year—to establish an internal process to receive and "substantively respond" —by informing the person whether the content violates the platform's terms and conditions—to a submitted report within 10 days of receipt of the report. Social media platforms not reaching those thresholds must respond within 21 days.

5) Requires a large social media platform to create a process by which principals (or positions of similar responsibility), school counselors (or positions of similar responsibility), and licensed mental health professionals who serve minors in the state may be designated as verified reporters who can report social media-related threats or violations of the platform's terms of service that, in the verified reporter's opinion, poses a risk or a severe risk to the health and safety of a minor.

6) Requires large social media platforms to receive and substantively respond within 72 hours to reported social media-related threats, or within 24 hours if the report is of a severe risk to a minor. Requires the platform to retain the content for at least six months. Requires that review of severe risks are by a natural person.

7) Requires large social media platforms that receive reports from verified reporters to annually report on the total number of reports received, the percentages of social media-related threats that formed the basis for the total number of reports from a verified reporter for the calendar year, and the percentage of reports from verified reporters for which the large social media platform took further action.

8) Provides for enforcement via a civil action with no specified statutory damages. Specifies that each day a social media platform is in violation of the bill's provisions constitutes a separate violation.

EXISTING LAW:

1) Establishes the Cyberbullying Protection Act, which requires a social media platform to disclose all cyberbullying reporting procedures in its terms of service. The Act only applies to platforms that generated more than $100,000,000 in gross revenue during the preceding calendar year. (Bus. & Prof. Code § 22589 *et seq*.)

2) Requires a social media platform to establish a mechanism within its internet-based service that allows any individual, whether or not that individual has a profile on the internet-based service, to report cyberbullying or any content that violates the existing terms of service. The reporting mechanism must allow, but not require, an individual to upload a screenshot of the content that contains cyberbullying or violates the terms of service. (Bus. & Prof. Code § 22589.1.)

3) Defines "cyberbullying" as any severe or pervasive conduct made by an electronic act or acts, as defined, committed by a pupil or group of pupils directed toward one or more pupils that has or can reasonably be predicted to have the effect of one or more of the following:

    a) Placing a reasonable pupil or pupils in fear of harm of their person or property,

    b) Causing a reasonable pupil to experience a substantially detrimental effect on the pupil's physical or mental health,

    c) Causing a reasonable pupil to experience substantial interference with the pupil's academic performance, or

    d) Causing a reasonable pupil to experience substantial interference with the pupil's ability to participate in or benefit from the services, activities, or privileges provided by a school. (Bus. & Prof. Code § 22589.)

4) Provides that the Attorney General may bring an action against a social media platform that intentionally violates the provisions of the Act and to recover a civil penalty of up to $7,500 for each intentional violation per day that the violation was incurred. The Attorney General may also seek injunctive relief. (Bus. & Prof. Code § 22589.3.)

5) Defines "social media platform" as a public or semipublic internet-based service or application that has users in California and that meets both of the following criteria:

   a) A substantial function of the service or application is to connect users in order to allow them to interact socially with each other within the service or application. (A service or application that provides email or direct messaging services does not meet this criterion based solely on that function.)

   b) The service or application allows users to do all of the following:

      i) Construct a public or semipublic profile for purposes of signing into and using the service or application.

      ii) Populate a list of other users with whom an individual shares a social connection within the system.

      iii) Create or post content viewable by other users, including, but not limited to, on message boards, in chat rooms, or through a landing page or main feed that presents the user with content generated by other users. (Bus. & Prof. Code § 22945(a)(3).)

FISCAL EFFECT: As currently in print, the bill is keyed non-fiscal.

COMMENTS:

1) **Author's statement.** According to the author:

> Severe, pervasive harm is being inflicted on California's youth through social media, yet social media platforms continue to fail in providing a minimally sufficient response to the threats. That failure includes a generally inadequate system of responding to dangerous content. Seriously dangerous content, including severe cyberbullying, fraud, and stalking, remain on platforms even after reports from at-risk, targeted youth. Either the social media platform provides no response or the response comes far too late after significant harm has done. AB 2481 will provide new protection for social media users requiring social media platforms to provide timely responses to reported content. It also creates a verified reporter mechanism through which our trusted school leaders and licensed mental health professionals can report content that in their professional judgment represents a material risk to a child user's health and safety. Such reports will trigger an expedited human review by the social media platform.

2) **Social media threats.** From 2010 to 2019, "rates of [youth] depression and anxiety—fairly stable during the 2000s—rose by more than 50 percent in many studies" and "[t]he suicide rate rose 48 percent for adolescents ages 10 to 19." This trend tracks "the years when adolescents in rich countries traded their flip phones for smartphones and moved much more of their social lives online—particularly onto social-media platforms designed for virality and addiction."[1]

---

[1] Haidt, "End the Phone-Based Childhood Now" (March 13, 2024) The Atlantic, https://www.theatlantic.com/technology/archive/2024/03/teen-childhood-smartphone-use-mental-health-effects/677722/.

According to the recent U.S. Surgeon General's advisory on the impact of social media on children's mental health, social media use by youth is nearly universal. Up to 95% of youth ages 13-17 report using a social media platform, with more than a third saying they use social media "almost constantly." Although age 13 is commonly the required minimum age used by social media platforms in the U.S., nearly 40% of children ages 8–12 use social media. As of 2021, the Surgeon General notes that 8th and 10th graders spent an average of 3.5 hours per day on social media.[2]

Adolescents, in a critical formative period of brain development, are especially vulnerable to the mental health impacts of social media. Among these impacts are increased neuroticism and anxiety, higher rates of depression, lower self-esteem, decreased attention spans, impulsivity, and brain patterns that resemble attention-deficit hyperactivity disorder.[3] The studies reviewed by the Surgeon General's Office point to a higher risk of harm in adolescent girls and those already experiencing poor mental health. The Surgeon General concludes:

> [T]he current body of evidence indicates that while social media may have benefits for some children and adolescents, there are ample indicators that social media can also have a profound risk of harm to the mental health and well-being of children and adolescents. At this time, we do not yet have enough evidence to determine if social media is sufficiently safe for children and adolescents.[4]

Whereas the European Union requires platforms to take down certain illegal content, Section 230 of the Communications Decency Act of 1996 provides civil immunity for online platforms based on third-party content and for the removal of content in certain circumstances.[5] As the United States Department of Justice has stated, "[t]he combination of significant technological changes since 1996 and the expansive interpretation that courts have given Section 230. . . has left online platforms both immune for a wide array of illicit activity on their services and free to moderate content with little transparency or accountability."[6] Social media platforms in the United States thus have virtually no duty to remove deplorable, tortious, or even criminal content such as hate speech, harassment, misinformation, criminal incitement, sexually predatory content, and drug trafficking.[7] Inadequate content moderation exposes users, particularly adolescents, to enormous risks.

The author and sponsors point to a number of relatively recent studies that show the prevalence of the various types of harmful content on social media:

---

[2] "Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory" (May 23, 2023) p. 7, https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf.
[3] Center for Humane Technology, "Extractive Technology is Damaging our Attention and Mental Health," https://www.humanetech.com/attention-mental-health.
[4] "Social Media and Youth Mental Health," *supra*, p. 4.
[5] 47 U.S.C. § 230.
[6] "Section 230—Nurturing Innovation or Fostering Unaccountability" (June, 2020), https://www.justice.gov/ag/file/1072971/dl?inline=.
[7] See Rustad and Koenig, "The Case for a CDA Section 230 Notice-and-Takedown Duty" (Spring, 2023) 23 Nev.L.J. 533; Hoffman, "Fentanyl Tainted Pills Bought on Social Media Cause Youth Drug Deaths to Soar" (May 19, 2022) N.Y. Times, https://www.nytimes.com/2022/05/19/health/pills-fentanyl-social-media.html.

- *Cyberbullying:* A 2022 Pew Research Center survey found "Nearly half of U.S. teens ages 13 to 17 (46%) report ever experiencing at least one of six cyberbullying behaviors."[8]

- *Self-harm:* A 2018 review of several studies found that, compared to non-victims, victims of cyberbullying were more than twice as likely to engage in self harm or exhibit suicidal behaviors.[9]

- *Disordered eating:* In 2022, the Center for Countering Digital Hate found that TikTok recommended eating disorder and self-harm content to new teen accounts within minutes.[10]

- *"Sexting":* A 2019 study found that approximately 13% of students reported sending, and 18.5% reported receiving, a sexually explicit or sexually suggestive image or video.[11]

- *Hate speech:* According to a 2018 report by Common Sense Media, "nearly two-thirds (64 percent) of teen social media users in 2018 say they 'often' or 'sometimes' come across racist, sexist, homophobic, or religious-based hate content in social media; one in five (21 percent) say they 'often' do so."[12]

- *Illicit substances:* A 2019 study found that "[o]ne in four young people (24%) reported that they see illicit drugs advertised for sale on social media."[13]

- *Sextortion:* A 2018 study found that "[a]pproximately 5% of students reported that they had been the victim of sextortion."[14]

The author and sponsor also point to an annual survey by the Anti-Defamation League (ADL). In 2023, ADL reported that "online hate and harassment rose sharply for adults and teens ages 13-17." According to ADL, reports of online harassment for teens had increased 36% to 51% for teens. Overall, reports of each type of hate and harassment increased by nearly every measure and within almost every demographic group."[15] ADL recommends that social media platforms take three key steps to mitigate hate and harassment:

- Push hate out of the mainstream by enacting strong policies against hate and harassment and enforce them transparently, equitably, and at scale.

---

[8] Pew Research Center, Teens and Cyberbullying 2022, https://www.pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/.
[9] https://www.jmir.org/2018/4/e129/.
[10] "Deadly by Design," CCDH-Deadly-by-Design_120922.pdf (counterhate.com).
[11] Patchin and Hinduja, "The Nature and Extent of Sexting Among a National Sample of Middle and High School Students in the U.S," https://pubmed.ncbi.nlm.nih.gov/31309428/.
[12] "Social Media, Social Life: Teens Reveal Their Experiences," p. 15 2018-social-media-social-life-executive-summary-web.pdf (commonsensemedia.org).
[13] "DM for Details: Selling Drugs in the Age of Social Media," DM for Details: Selling Drugs in the Age of Social Media - Volteface
[14] "Sextortion Among Adolescents: Results from a National Survey of U.S. Youth," Sextortion Among Adolescents: Results From a National Survey of U.S. Youth - Justin W. Patchin, Sameer Hinduja, 2020 (sagepub.com)
[15] "Online Hate and Harassment: The American Experience 2023," Online Hate and Harassment: The American Experience 2023 | Center on Extremism (adl.org)

- Support targets of harassment through use of trust and safety teams, by ensuring reporting features are effective, and by implementing anti-hate by design and principles.

- Foster trust and accountability by being transparent with users, lawmakers, and civil society.[16]

3) **Failure to respond.** A major impetus for this bill arises from widespread reports about the lack of responsiveness of social media platforms to complaints from parents and caretakers of children about harmful content on the platforms. The bill's sponsor, the Organization for Social Media Safety, cites statistics from a 2021 ADL survey:

- 41% of respondents who experienced a physical threat stated that the platform took no action on a threatening post, an increase from the 38% who had reported a similar lack of action the year before.

- 38% said they did not flag the threatening post to the platform, up from 33% the prior year.

- Only 14% of those who experienced a physical threat said the platform deleted the threatening content, a significant drop from 22% the prior year.

- Just 17% of those who experienced a physical threat stated that the platform blocked the perpetrator who posted the content, a sharp decrease from the prior year's 28%.[17]

The sponsor concludes: "Seriously dangerous content, including severe cyberbullying, fraud, and stalking, either indefinitely remain on platforms even after reports from at-risk, targeted youth, or the response from reports come far too late after significant harm has done."

In opposition, Computer & Communications Industry Association, California Chamber of Commerce, and TechNet jointly write to detail current content moderation practices:

> Responsible digital service providers have already taken aggressive steps to moderate dangerous and illegal content, consistent with their terms of service. The companies deliver on the commitments made to their user communities with a mix of automated tools and human review. Doing so is an evolving industry practice: since its launch, the Digital Trust & Safety Partnership (DTSP) has quickly developed and executed initial assessments of how participating companies implement the DTSP Best Practices Framework, which provides a roadmap to increase trust and safety online meaningfully. (Footnote omitted.)

The framework referenced by opposition includes the following commitments:

- Anticipate the risks for misuse as part of product design, and develop ways to prevent misuse or abuse.

- Adopt rules for user conduct and content that are clear and consistent.

---

[16] *Ibid.*
[17] "Online Hate and Harassment: The American Experience 2021," Online Hate and Harassment: The American Experience 2021 | ADL

- Enforce the rules. The framework includes examples of best practices for enforcement operations, including investing in wellness and resilience of teams dealing with sensitive materials.

- Keep up with changing risks and review whether policies are effective in limiting harmful content or conduct.

- Report periodically on actions taken on complaints, enforcement and other activities related to trust and safety.[18]

4) **Comparison to the Cyberbullying Protection Act.** AB 2879 (Low; Chap. 700, Stats. 2022) added the Cyberbullying Protection Act (CPA), which requires social media platforms to establish a reporting mechanism for the reporting of cyberbullying and other conduct that violates the platform's terms of service. The reporting mechanism must include two features: it must be useable by individuals who do not have an account on the platform, and it must permit, but not require, the report to include a screenshot of the problematic post. These measures are designed to make the mechanism as useful as possible for, e.g., a parent who might not have an account on a particular platform but who wishes to protect their child. Platforms are also required to disclose in their terms of service the procedures for using the reporting mechanism. The Attorney General is authorized to bring suit against platforms for intentional failure to comply and seek a $7,500 per day, per violation civil penalty.

This bill encompasses a much wider range of harmful content. Whereas the CPA applies only to "severe or pervasive" conduct, this bill applies to conduct that "promotes, incites, facilitates, or perpetrates" one of several enumerated harms: suicide, disordered eating, drug trafficking, substance abuse, fraud, human trafficking, sexual abuse, cyberbullying, harassment, academic dishonesty, stalking or identity theft. This bill also imposes an affirmative requirement that social media platforms "substantively respond" to reporters as follows:

- For reporters who are not "verified reporters," responses must be made within 10 days by large social media platforms (those with over 100 million users or $1 billion in global revenues) or 21 days by smaller social media platforms.

- For "verified reporters"—school counselors, principals, and licensed mental health professionals—large social media platforms must respond within 72 hours. However, if the reporter deems the risk severe, the large social media platform must respond within 24 hours and review must be undertaken by a human being. Smaller social media platforms are not subject to specific requirements relating to verified reporters.

Large social media platforms must annually report on the total number of reports from verified reporters, the percentage of threat types that formed the basis for the total number of reports, and the percentage of threats for which the platform took further action. Finally, instead of enforcement by a public prosecutor for $7,500 per day per violation, the bill grants a private right of action for unspecified damages.

---

[18] McGill, "Tech giants list principles for handling harmful content" (Feb. 18, 2021) Axios https://www.axios.com/2021/02/18/tech-giants-list-principles-for-handling-harmful-content.

Meanwhile, related legislation is progressing in the Senate. SB 1504 (Stern) expands the Cyberbullying Protection Act by, in part, extending its reach to cover cyberbullying of any minor by any person and establishing a timeline and process for the platform's response to those reports. The bill increases the civil penalties to $75,000 per violation per day and authorizes anyone reporting cyberbullying to bring civil actions to recover those penalties. The bill also includes a non-exclusive list of content that meets the definition of "severe or pervasive conduct," an element of cyberbullying. The bill is pending in the Senate Appropriations Committee.

5) **Concerns over breadth.** Under the bill, a "social media-related threat" is defined as anything that could lead to suicide, disordered eating, drug trafficking, substance abuse, fraud, human trafficking, sexual abuse, cyberbullying, harassment, academic dishonesty, stalking or identity theft. Opponents argue these terms are vague and subjective. In opposition, Electronic Frontier Foundation writes:

> Defining "social media related threat" as content that "promotes, incites, facilitates, or perpetrates," a broad category of content that includes suicide, disordered eating, drug trafficking, and substance abuse is both vague and overbroad. Californians do not agree, and never have agreed about what types of content "promote, incite, facilitate, or perpetrate," for example, may be considered disordered eating. People of all ages may wish to see content about losing weight that is not problematic or does not encourage eating disorders. Does a news article about policing policy around an open-air drug market in a San Francisco neighborhood facilitate drug trafficking? A "social media related threat" is incredibly subjective.
>
> A.B. 2481's broad authorization for private reporting exacerbates this problem. When any individual can report, an incredible degree of variance is created on what is reported. Children's health, safety, and sexuality, for example, are the subject of widespread debate largely because there is no general consensus in California or otherwise about what is harmful to children. Elected officials in both California and other states have said that access to LGBTQ+ content harms children. More than 20 states have passed recent laws that ban the provision of gender-affirming care to minors. A.B. 2481 could empower anti-LGBTQ+ advocates to report content on gender-affirming care as sexual abuse, and then force platforms to substantively respond. (Footnotes omitted.)

While it is possible that abusive reports could be made over content that is beneficial to vulnerable communities, the bill requires nothing more than a "substantive response" from a social media platform—that is, the platform, within a given timeframe, must inform the reporter as to whether the content violates its terms of service. It is unlikely that LGBTQ+ content, such as information about gender-affirming care for minors, would actually violate the terms of service of any mainstream social media platform.

Nevertheless, the sheer breadth of the types of content covered under this bill could dilute the bill's effectiveness and cause significant implementation challenges, even for ultra-profitable platforms. While each of these categories is a compelling concern, going forward the author may wish to consider narrowing the bill's scope to focus on content that poses the highest risk.

6) **Section 230.** Section 230 states, "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[19] That section also provides a safe harbor for "any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected."[20] Finally, it provides that "[n]o cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section."[21]

Through this statute, "Congress intended to create a blanket immunity from tort liability for online republication of third party content."[22] "The courts have consistently construed CDA Section 230 to eliminate all tort liability against websites, search engines, and other online intermediaries arising out of third-party postings on their services. The result is that large gatekeepers such as Facebook, Google, Twitter, and YouTube have no duty to respond to takedown notices, even if the deplorable content is a continuing tort or crime."[23]

Because this bill merely requires a platform to provide a mechanism for reporting social media-related threats and to disclose whether such threats violate the platform's terms of service, the bill does not treat a platform "as the publisher or speaker," nor hold it liable, for such content. Section 230 is thus not plainly implicated by this bill.

7) **First Amendment.** The United States and California Constitutions prohibit abridging, among other fundamental rights, freedom of speech.[24] This bill does not expressly regulate speech; it simply provides a mechanism for reporting harmful content and requires the platform to disclose to the reporter whether the content violates its terms and conditions. As noted above, industry opponents to the bill have indicated they are already taking "aggressive steps to moderate dangerous and illegal content, consistent with their terms of service." Thus, if the content violates the terms of service, the platform would, apparently, already be disposed to remove such content; as such, it seems that the only difference the bill would make with regard to content moderation is to accelerate a process that platforms are already supposedly undertaking.

Industry opponents also suggest that the provision in the bill that requires expedited human response for verified reporters unduly intrudes on the First Amendment. They reference *Murthy v. Missouri*,[25] a case pending before the United States Supreme Court involving the issue of whether the Biden administration's "jawboning" efforts to pressure social media platforms to restrict misinformation about the COVID-19 vaccine amounted to a violation of users' free speech rights. Opponents write:

> In that case, the Biden Administration is arguing that their communications with social media companies regarding user generated content were not coercive and did not violate the First Amendment because their reports were treated the same as any other report and did not receive priority treatment. In contrast, AB 2481 would create a streamlined process and

---

[19] 47 U.S.C. § 230(c)(1).
[20] *Id.* at § 230(c)(2)(A)
[21] *Id.* at (e)(3).
[22] *Barrett v. Rosenthal* (2006) 40 Cal.4th 33, 57.
[23] *The Case for a CDA Section 230 Notice-and-Takedown Duty*, *supra*, 23 Nev.L.J. at p. 536.
[24] U.S. Const., 1st and 14th Amends; Cal. Const. art. I, § 2.
[25] Docket No. 23-411.

elevate reports from K-12 principals and school counselors. The upcoming decision in *Murthy* will provide more guidance in this area but this bill seems to be at odds with the arguments made by the Biden Administration.

But this argument assumes that the power of local K-12 school officials and children's therapists to cow social media platforms into censoring speech is somehow comparable to that of officials from a Presidential Administration.

Finally, the bill implicates free speech principles by requiring social media platforms to report on the total number of reports received, the percentages of social media-related threats that formed the basis for the total number of reports from a verified reporter for the calendar year, and the percentage of reports from verified reporters for which the large social media platform took further action. Because the right to speak encompasses the right *not* to speak, this provision implicates the First Amendment.[26] Compelled speech in the commercial context, however, is subjected to much less exacting scrutiny than in other arenas; a law concerning commercial speech is generally upheld if the law advances a substantial government interest and directly advances that interest.[27]

Here, the state's interest in protecting children from social media-related threats is clearly substantial, and the requirement that large social media platforms detail the volume of, and response to, such content appears to directly advance this interest by ensuring transparency and accountability in the implementation of the bill's provisions.

8) **Amendments.** The author has agreed to clarify the bill's enforcement provisions by specifying who has standing and capping damages awarded for violations at $10,000. The amendments are as follows:

22588.4. (a) (1) Actions for relief pursuant to this chapter may be brought ~~by a minor~~ in a civil action~~.~~ ***by either of the following:***

(A) Any person who has, pursuant to this chapter, made a report for which a violation of this chapter occurred.

(B) Any person who, as a result of a social media platform's actions or omissions, was denied the opportunity to make a report permitted under this chapter.

(2) In a successful action brought pursuant to this subdivision, the court may order injunctive relief to obtain compliance with this chapter and may award ***up to $10,000 per violation as well as*** reasonable attorney's fees and costs to the prevailing plaintiff.

(b) Each day a social media platform is in violation of this chapter constitutes a separate violation.

(c) (1) The remedies provided by this section are in addition to any other civil, criminal, and administrative remedies, penalties, or sanctions provided by law and do not supplant, but are cumulative to, other remedies, penalties, or sanctions.

---

[26] *U.S. v. United Foods, Inc.* (2001) 533 U.S. 405, 410.
[27] *Central Hudson Gas & Elec. Corp. v. Public Service Commission of New York* (1980) 477 U.S. 556, 566.

(2) The duties and obligations imposed by this section are cumulative with any other duties or obligations imposed under other law and shall not be construed to relieve any party from any duties or obligations imposed under other law.

9) **Related legislation.** SB 1504 (Stern) expands the Cyberbullying Protection Act by, in part, extending its reach to cover cyberbullying of any minor by any person and establishing a timeline and process for the platform's response to those reports. The bill increases the civil penalties to $75,000 per violation per day and authorizes anyone reporting cyberbullying to bring civil actions to recover those penalties. The bill also includes a non-exclusive list of content that meets the definition of "severe or pervasive conduct," an element of cyberbullying. The bill is pending in the Senate Appropriations Committee.

ARGUMENTS IN SUPPORT:

The bill's sponsor, the Organization for Social Media Safety, writes:

> AB 2481 requires social media platforms to provide timely responses to reported content. It also creates a verified reporter mechanism through which school leaders and licensed mental health professionals can report content that in their professional judgment represents a material risk to a child user. Such reports will trigger an expedited human review by the social media platform.
>
> California's educational leaders and licensed mental health professionals regularly bear direct witness to severe harms being inflicted upon youth due to content on social media; while having no means to address such content and protect the children in their care. By creating a verified reporter process, California can enable its trusted professionals with their expert, experienced judgment to alert social media platforms to impending, severe risk facing children. Under the bill, social media platforms will have the reasonable, necessary obligation of reviewing and responding to these experts' professional assessments of imminent risk in a timely manner.
>
> The extent of harm that social media is inflicting upon children is severe. And even with rising awareness of these harms over recent years and various legislative efforts, the risk to children's health and welfare from social media use continues. The status quo is simply unacceptable; we must enable a workable system to report risks to children.

ARGUMENTS IN OPPOSITION:

Computer & Communications Industry Association, California Chamber of Commerce, and TechNet jointly write:

> Currently, AB 2481 requires "large social media platforms" to substantively respond to a submitted report within 10 days but requires platforms that are not "large social media platforms" to substantively respond to a submitted report within 21 days. We urge legislators to reconsider a statutory definition gerrymandered around particular businesses with user thresholds and an assortment of carve-outs, and instead, craft compliance obligations that are manageable by all entities operating in the relevant sector, regardless of the number of users they have.

> The application of such definitions regularly leads to ambiguities about scope. The current definition, for example, leaves questions regarding how to treat user thresholds in international markets that operate in California. Given the disproportionate penalties contemplated for compliance failures, these definitions demand greater clarity.

They further assert that:

> "[V]erified reporters" are defined as a K-12 school principal, counselor, or a licensed mental health professional who provides services to minors. However, the bill lacks a clear definition of what constitutes a "readily accessible and easy-to-use verification process" to confirm the employment status claimed by these individuals. In the absence of a defined "verification process," covered entities face a moving compliance target, risking inadvertent verification of individuals who do not qualify as "verified reporters." Uncertainties also arise concerning potential scenarios, such as when a "verified reporter" no longer maintains their employment status. There is also no defined duration for which someone can retain their verified status. This is concerning considering there are potentially tens of thousands of people who could qualify as verified reporters under this bill.

REGISTERED SUPPORT / OPPOSITION:

Support

Organization for Social Media Safety (sponsor)

Opposition

California Chamber of Commerce
Chamber of Progress
Computer & Communications Industry Association
Computer and Communications Industry Association
Electronic Frontier Foundation
Technet

Analysis Prepared by: Josh Tosney / P. & C.P. / (916) 319-2200