

SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2023-2024 Regular Session

AB 2481 (Lowenthal)
Version: May 7, 2024
Hearing Date: July 2, 2024
Fiscal: No
Urgency: No
CK

SUBJECT

Social media-related threats: reporting

DIGEST

This bill requires social media platforms to establish reporting mechanisms, including expedited reporting from “verified reporters,” for “social media-related threats” and to substantively respond to such reports, as provided. Platforms are also required to disclose these procedures in their terms of service and post annual reports detailing reports from verified reporters.

EXECUTIVE SUMMARY

Social media can connect people – for good or for ill. Survey data found that overall screen use among teens and tweens increased by 17 percent from 2019 to 2021, with the number of hours spent online spiking sharply during the pandemic. A recent report found that nearly all teens, 95 percent, use social media.

Given the reach of social media platforms and the increasing role they play in many children’s lives, concerns have arisen over the connection between social media usage and mental health, drug use, and other self-harming conduct. Social media has also opened up new avenues for bullies of all ages to harass their victims; research indicates that cyberbullying is now common among teens and tweens.

This bill looks to expand on other social media reporting mechanism laws by requiring social media platforms to develop mechanisms for individuals to report “social media-related threats” and to “substantively respond” to such reports within specified timelines. An expedited process shall be provided for reporters that are verified by the platforms, including school principals and licensed mental health providers.

This bill is supported by the Children’s Advocacy Institute and CFT. It is opposed by various industry associations, including the Tri-County Chamber Alliance.

PROPOSED CHANGES TO THE LAW

Existing federal law:

- 1) Provides a right to free speech and expression. (U.S. Const., 1st amend; Cal. Const., art 1, § 2.)
- 2) Provides that 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. (47 U.S.C. § 230(c)(1).)
- 3) Provides that no provider or user of an interactive computer service shall be held liable on account of:
 - a) Any action voluntarily taken in good faith to restrict access to or availability of material that users consider to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.
 - b) Any action taken to enable or make available to content providers or others the technical means to restrict access to material described above. (47 U.S.C. § 230(c)(2).)
- 4) Defines “interactive computer service,” for purposes of 2) and 3), as any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions. (47 U.S.C. § 230(f)(2).)

Existing state law:

- 1) Establishes the Student 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 an "electronic act," as the creation or transmission originated on or off the schoolsite, by means of an electronic device of a communication, including, a post on a social network internet website. (Ed. Code § 48900(r)(2).)
- 6) Provides that a pupil cannot be suspended or expelled for an act of bullying or cyberbullying unless the act is related to a school activity or school attendance occurring within a school under the jurisdiction of the superintendent of the school district or principal or occurring within any other school district, which includes being on school grounds, going to or coming from school, during the lunch period whether on or off campus, and during, or while going to or coming from, a school-sponsored activity. (Ed. Code § 48900(s).)
- 7) Requires a social media platform to provide, in a mechanism that is reasonably accessible to users, a means for a user who is a California resident to report material to the platform that the user reasonably believes meets all of the following criteria:
- a) The reported material is child sexual assault material (CSAM);
 - b) The reporting user is an identifiable minor depicted in the reported material;
 - c) The reported material is displayed, stored, or hosted on the social media platform. (Civ. Code § 3273.66.)
- 8) Requires platforms to respond and take specified actions in response to reports above and subject them to liability for violations in actions brought by reporters. (Civ. Code §§ 3273.66, 3273.67.)

This bill:

- 1) Requires a social media platform, as defined, to disclose all social media-related threat reporting procedures in their terms of service.

- 2) Requires platforms to establish a mechanism that allows an individual, whether or not that individual has a profile, to report a social media-related threat or any content that violates the social media platform's terms of service that meets specified criteria.
- 3) Requires a large social media platform to establish an internal process to receive and substantively respond to a submitted report within 10 days of the receipt of the report. A social media platform that is not a large social media platform shall establish an internal process to receive and substantively respond to a submitted report within 21 days of the receipt of the report.
- 4) Requires a large social media platform to create a process to verify all of the following individuals as verified reporters:
 - a. A principal, or a position of similar responsibility, of any school lawfully operating any programs from kindergarten and grades 1 to 12, inclusive, in the state.
 - b. A school counselor, or position of similar responsibility, of any school lawfully operating any programs from kindergarten and grades 1 to 12, inclusive, in the state.
 - c. A licensed mental health professional who provides mental health services to minors in the state.
- 5) Requires a large social media platform to create a process by which a verified reporter can make a report of a social media-related threat or a violation of the large social media 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 these platforms to do all of the following in connection with this process:
 - a. Establish an internal process to receive and substantively respond within 72 hours, or within 24 hours if the report is of a severe risk, to a report by a verified reporter of content the verified reporter deems to be a risk to a minor.
 - b. Preserve content reported by a verified reporter for a period of at least six months from the date of the report.
 - c. Ensure that a report submitted by a verified reporter and deemed by the verified reporter to be of a severe risk receives a review by a natural person.
- 7) Requires a large social media platform that receives a report from a verified reporter to report annually, on a publicly accessible internet website, all of the following:
 - a. The total number of reports from a verified reporter received for the calendar year.

- b. 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.
 - c. The percentage of reports from verified reporters for which the large social media platform took further action.
- 8) Prohibits a large social media platform from requiring a verified reporter to reverify the verified reporter's qualifications more often than once every two years.
- 9) Defines the relevant terms, including:
- a. "Large social media platform" means a social media platform that meets all of the following criteria:
 - i. The social media platform's terms of service do not prohibit the use of the social media platform by a child.
 - ii. The social media platform includes features that enable a child to share images, text, or video through the internet with other users of the social media platform whom the child has met, identified, or become aware of solely through the use of the social media platform.
 - iii. The social media platform has more than 100,000,000 monthly global active users or generates more than one billion dollars (\$1,000,000,000) in gross revenue per year, adjusted yearly for inflation.
 - b. "Severe risk" means a social media-related threat that more likely than not will cause serious bodily or mental harm to a child.
 - c. "Social media-related threat" means content that promotes, incites, facilitates, or perpetrates any of the following:
 - i. Suicide.
 - ii. Disordered eating.
 - iii. Drug trafficking.
 - iv. Substance abuse.
 - v. Fraud.
 - vi. Human trafficking punishable pursuant to Section 236.1 of the Penal Code.
 - vii. Sexual abuse.
 - viii. Cyberbullying.
 - ix. Harassment.
 - x. Distribution of harmful matter, as defined by Section 313 of the Penal Code.
 - xi. Academic dishonesty.
 - xii. An offense punishable pursuant to Section 653.2 of the Penal Code.
 - xiii. An offense punishable pursuant to Section 530.5 of the Penal Code.
 - xiv. An offense punishable pursuant to Section 529 of the Penal Code.

- d. "Substantively respond" means to inform a person who makes a report that the content being reported meets either of the following criteria:
 - i. The content does not violate the terms and conditions of the social media platform or the large social media platform.
 - ii. The content violates the terms and conditions of the social media platform or the large social media platform.
- 10) Authorizes civil actions seeking statutory penalties of up to \$10,000 per violation per day the violation persists, along with injunctive relief and attorneys' fees and costs. Such actions may be brought by either of the following:
 - a. A person who has, pursuant to this chapter, made a report with respect to which a violation of this chapter occurred.
 - b. A person who, as a result of a social media platform's actions or omissions, was unable to make a report pursuant to this chapter.
- 11) Clarifies that the duties, obligations, and remedies are cumulative and do not interfere with any others that exist.
- 12) Includes a severability clause.

COMMENTS

1. Social media and children

The effects of social media on our mental health and what should and can be done about it are pressing policy and societal questions that have become increasingly urgent. Evidence shows that engagement on social media has a clear effect on our emotions and with 95 percent of teens reporting some usage of social media, the issue has prompted the United States Surgeon General to issue a mental health advisory:

At this time, we do not yet have enough evidence to determine if social media is sufficiently safe for children and adolescents. We must acknowledge the growing body of research about potential harms, increase our collective understanding of the risks associated with social media use, and urgently take action to create safe and healthy digital environments that minimize harm and safeguard children's and adolescents' mental health and well-being during critical stages of development.¹

¹ U.S. Surgeon General's Advisory, *Social Media and Youth Mental Health* (May 23, 2023) <https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf>. All internet citations are current as of June 22, 2024.

A meta-analysis of research on social networking site (SNS) use concluded the studies supported an association between problematic SNS use and psychiatric disorder symptoms, particularly in adolescents.² The study found most associations were with depression and anxiety.

In addition, cyberbullying – bullying tactics made through online means – is remarkably prevalent. Studies suggest that around 15 percent of teens and tweens have experienced cyberbullying.³ Bullying of any kind is associated with negative health effects, but cyberbullying presents unique risks to its victims in light of the nature of social media and the internet in general. Social media platforms can be used to create a false profile for a person, disseminate embarrassing photos or videos, or engage in bullying anonymously in ways that are not available in the real world. Without a reliable mechanism to report cyberbullying, parents and other concerned adults might not have alternative means to protect their children from ongoing online harassment.

Many believe that social media platforms are not doing enough to combat this and reporting reveals that it is not a lack of awareness of the problem, as a series of startling revelations unfolded after a Facebook whistle-blower, Frances Haugen, began sharing internal documents. The Wall Street Journal published many of the findings:

About a year ago, teenager Anastasia Vlasova started seeing a therapist. She had developed an eating disorder, and had a clear idea of what led to it: her time on Instagram.

She joined the platform at 13, and eventually was spending three hours a day entranced by the seemingly perfect lives and bodies of the fitness influencers who posted on the app.

“When I went on Instagram, all I saw were images of chiseled bodies, perfect abs and women doing 100 burpees in 10 minutes,” said Ms. Vlasova, now 18, who lives in Reston, Va.

Around that time, researchers inside Instagram, which is owned by Facebook Inc., were studying this kind of experience and asking whether it was part of a broader phenomenon. Their findings confirmed some serious problems.

“Thirty-two percent of teen girls said that when they felt bad about their bodies, Instagram made them feel worse,” the researchers said in a March

² *Ibid.*

³ See Basile, et al., *Interpersonal Violence Victimization Among High School Students – Youth Risk Behavior Survey, United States, 2019*, CDC National Center for Injury Prevention and Control, Division of Violence Prevention (Aug. 21, 2020), at p. 1; Patchin & Hinduja, *Tween Cyberbullying in 2020*, Cyberbullying Research Center (2020) at p. 4.

2020 slide presentation posted to Facebook's internal message board, reviewed by The Wall Street Journal. "Comparisons on Instagram can change how young women view and describe themselves."

For the past three years, Facebook has been conducting studies into how its photo-sharing app affects its millions of young users. Repeatedly, the company's researchers found that Instagram is harmful for a sizable percentage of them, most notably teenage girls.

"We make body image issues worse for one in three teen girls," said one slide from 2019, summarizing research about teen girls who experience the issues.

"Teens blame Instagram for increases in the rate of anxiety and depression," said another slide. "This reaction was unprompted and consistent across all groups."

Among teens who reported suicidal thoughts, 13% of British users and 6% of American users traced the desire to kill themselves to Instagram, one presentation showed.

Expanding its base of young users is vital to the company's more than \$100 billion in annual revenue, and it doesn't want to jeopardize their engagement with the platform.

More than 40% of Instagram's users are 22 years old and younger, and about 22 million teens log onto Instagram in the U.S. each day⁴

The released documents from Instagram make clear that "Facebook is acutely aware that the products and systems central to its business success routinely fail":

The features that Instagram identifies as most harmful to teens appear to be at the platform's core.

The tendency to share only the best moments, a pressure to look perfect and an addictive product can send teens spiraling toward eating disorders, an unhealthy sense of their own bodies and depression, March 2020 internal research states. It warns that the Explore page, which serves users photos and videos curated by an algorithm, can send users deep into content that can be harmful.

⁴ Georgia Wells et al., *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show* (September 14, 2021) The Wall Street Journal, https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=article_inline.

“Aspects of Instagram exacerbate each other to create a perfect storm,” the research states.⁵

The referenced documents revealed that Facebook’s own internal research found “1 in 8 of its users reported compulsive social media use that interfered with their sleep, work, and relationships – what the social media platform calls ‘problematic use’ but is more commonly known as ‘internet addiction.’”⁶

2. Responding to the crisis

In the U.S. Surgeon General’s advisory, he provides a number of suggestions for responding to the crisis, including the following recommendation for what technology companies can do:

Create effective and timely systems and processes to adjudicate requests and complaints from young people, families, educators, and others to address online abuse, harmful content and interactions, and other threats to children’s health and safety. Social media platforms should take these complaints seriously, thoroughly investigate and consider them, and respond in a timely and transparent manner.

This bill attempts to do just that by requiring social media platforms to establish mechanisms for reporting “social media-related threats,” defined as content that promotes, incites, facilitates, or perpetrates any of the following:

- Suicide.
- Disordered eating.
- Drug trafficking.
- Substance abuse.
- Fraud.
- Human trafficking punishable pursuant to Section 236.1 of the Penal Code.
- Sexual abuse.
- Cyberbullying.
- Harassment.
- Distribution of harmful matter, as defined by Section 313 of the Penal Code.
- Academic dishonesty.
- An offense punishable pursuant to Section 653.2 of the Penal Code.
- An offense punishable pursuant to Section 530.5 of the Penal Code.
- An offense punishable pursuant to Section 529 of the Penal Code.

⁵ *Ibid.*

⁶ Kim Lyons, *Facebook reportedly is aware of the level of ‘problematic use’ among its users* (November 6, 2021) The Verge, www.theverge.com/2021/11/6/22766935/facebook-meta-aware-problematic-use-addiction-wellbeing.

Platforms must establish a mechanism for any individual, whether or not they are a user of the platform, to report social media-related threats or any content that violates the platform's terms of service. This mechanism is required to meet specified criteria and platforms are required to "substantively respond" within a specified time frame depending on their size, as provided.

"Substantively respond" means to inform a person who makes a report that the content being reported meets either of the following criteria:

- The content does not violate the terms and conditions of the social media platform or the large social media platform.
- The content violates the terms and conditions of the social media platform or the large social media platform.

The bill also provides for a separate reporting process for "verified reporters." *Large* social media platforms, defined as social media platforms that have more than 100 million monthly active users or generate more than \$1 billion in gross annual revenue and that meet other criteria, are required to verify the following individuals for this status:

- A principal, or a position of similar responsibility, of any school lawfully operating any programs from kindergarten and grades 1 to 12, inclusive, in the state.
- A school counselor, or position of similar responsibility, of any school lawfully operating any programs from kindergarten and grades 1 to 12, inclusive, in the state.
- A licensed mental health professional who provides mental health services to minors in the state.

The verified reporter process must allow reporters to make a report of a social media-related threat or a violation of the large social media platform's terms of service that in the verified reporter's opinion poses a risk to the health and safety of a minor. Large platforms must establish an internal process to receive and substantively respond within 72 hours, or within 24 hours if the report is of a *severe* risk, to a report by a verified reporter of content the reporter deems to be a risk to a minor. "Severe risk" is a social media-related threat that more likely than not will cause serious bodily or mental harm to a child. These reports of severe risks are required to receive review by a natural person. Platforms are required to annually report on their websites statistics regarding reports from verified reporters.

Those unable to make a report due to a violation of these provisions and those making a report with respect to which a violation occurs are authorized to bring a civil action against platforms for statutory damages not to exceed \$10,000 per violation per day the violation persists, in addition to injunctive relief and fees and costs.

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.

Writing in support, the Children's Advocacy Institute asserts:

AB 2481's approach to threats to children online is a sensible one. It creates a two-tiered reporting mechanism for "social media-related threats," meaning content posted on a social media platform that promotes, incites, facilitates, or perpetrates certain listed harms. As current law permits "any individual" to report cyberbullying, this bill 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.

Regarding the biggest social media platforms, "verified reporters" consisting of 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.

3. Amendments to focus the approach of the bill

Concerns have been raised about the approach of the bill, its breadth, and its enforcement mechanism. Writing in opposition, the Tri-County Chamber Alliance states:

While we support measures to protect minors online, AB 2481's approach places undue burdens on social media platforms, potentially disrupting the tech industry's growth and economic contributions to the state. We recommend reconsidering the bill and exploring alternative solutions that achieve the same protective goals without imposing excessive regulatory demands.

In response to opposition concerns, the author has agreed to a series of amendments that will focus the bill on the most problematic threats being flagged by trusted reporters. These amendments involve the following:

- Removing the requirement to establish a mechanism for any individual to report social media-related threats.
- Removing the private right of action.
- Narrowing the scope of what is considered a social media-related threat.
- Removing school counselors and "positions of similar responsibility" from the list of verified reporters.
- Removing the requirement to preserve content reported by verified reporters.
- Delaying implementation by one year.
- Other technical and clarifying amendments, including honing definitions within the bill.

SUPPORT

CFT

Children's Advocacy Institute

OPPOSITION

California Chamber of Commerce

Computer & Communications Industry Association

Electronic Frontier Foundation

Technet

Tri-County Chamber Alliance

RELATED LEGISLATION

Pending Legislation:

SB 918 (Umberg, 2024) requires a social media platform to, at all times, make available by telephone to a law enforcement agency a law enforcement liaison for the purpose of receiving, and responding to, requests for information. It requires a social media platform to comply with a search warrant within 72 hours if the search warrant is provided to the social media platform by a law enforcement agency and the subject of

the search warrant is information associated with an account on the social media platform and that information is controlled by a user of the social media platform. SB 918 is currently in the Assembly Privacy and Consumer Protection Committee.

SB 976 (Skinner, 2024) prohibits operators of “internet-based services or applications” from providing “addictive feeds,” as those terms are defined, to minors without parental consent and from sending notifications to minors at night and during school hours without parental consent, as provided. This bill requires operators to make available to parents a series of protective measures for controlling access to and features of the platform for their children. This bill also requires reporting on data regarding children on their platforms, as specified. SB 976 is currently in the Assembly Privacy and Consumer Protection Committee.

SB 981 (Wahab, 2024) requires social media platforms to provide a mechanism for reporting “digital identity theft,” essentially the posting of nonconsensual, sexual deepfakes; and requires platforms to timely respond and investigate and to block instances of this material, as provided. SB 981 is currently in the Assembly Privacy and Consumer Protection Committee.

SB 1504 (Stern, 2024) amends the Cyberbullying Protection Act by requiring social media platforms to respond to reports of cyberbullying and disclose final determinations. It authorizes the parents or legal guardians of minors who report violations to bring civil actions for those violations and increases the applicable penalty tenfold. SB 1504 is currently in the Assembly Judiciary Committee.

Prior Legislation:

SB 287 (Skinner, 2023) would have subjected social media platforms to civil liability for damages caused by their designs, algorithms, or features, as provided. It would have provided a safe harbor where certain auditing practices are carried out. SB 287 was held in the Senate Appropriations Committee.

AB 1394 (Wicks, Ch. 579, Stats. 2023) required social media platforms to provide a reporting mechanism for suspected child sexual abuse material and requires them to permanently block the material, as provided. It also prohibits platforms from knowingly facilitating, aiding, or abetting minor’s commercial sexual exploitation.

SB 1056 (Umberg, Ch. 881, Stats. 2022) required a social media platform, as defined, to clearly and conspicuously state whether it has a mechanism for reporting violent posts, as defined; and allows a person who is the target, or who believes they are the target, of a violent post to seek an injunction to have the violent post removed.

AB 587 (Gabriel, Ch. 269, Stats. 2022) required social media companies, as defined, to post their terms of service and report certain information to the Attorney General on a quarterly basis.

AB 1628 (Ramos, Ch. 432, Stats. 2022) required a social media platform, as defined, that operates in this state to create and publicly post a policy statement including specified information pertaining to the use of the platform to illegally distribute controlled substances, until January 1, 2028.

AB 2273 (Wicks, Ch. 320, Stats. 2022) established the California Age-Appropriate Design Code Act, placing a series of obligations and restriction on businesses that provide online services, products, or features likely to be accessed by a child.

AB 2408 (Cunningham, 2022) would have prohibited a social media platform from using a design, feature, or affordance that the platform knew, or which by the exercise of reasonable care it should have known, causes child users to become addicted to the platform. AB 2408 died in the Senate Appropriations Committee.

AB 1114 (Gallagher, 2021) would have required a social media company located in California to develop a policy or mechanism to address content or communications that constitute unprotected speech, including obscenity, incitement of imminent lawless action, and true threats, or that purport to state factual information that is demonstrably false. AB 1114 died in the Assembly Arts, Entertainment, Sports, Tourism, and Internet Media Committee.

SB 388 (Stern, 2021) would have required a social media platform company, as defined, that, in combination with each subsidiary and affiliate of the service, has 25,000,000 or more unique monthly visitors or users for a majority of the preceding 12 months, to report to the Department of Justice by April 1, 2022, and annually thereafter, certain information relating to its efforts to prevent, mitigate the effects of, and remove potentially harmful content. This bill died in the Senate Judiciary Committee.

PRIOR VOTES:

Assembly Floor (Ayes 46, Noes 0)

Assembly Privacy and Consumer Protection Committee (Ayes 6, Noes 0)
