

**SENATE JUDICIARY COMMITTEE**  
**Senator Thomas Umberg, Chair**  
**2025-2026 Regular Session**

SB 771 (Stern)  
Version: March 24, 2025  
Hearing Date: April 29, 2025  
Fiscal: Yes  
Urgency: No  
AWM

**SUBJECT**

Pupils: use of social media

**DIGEST**

This bill creates a civil action against a social media platform, as defined, with over \$100 million in gross annual revenues that aids and abets the commission of, conspires with a person to violate, or is a joint tortfeasor for a violation of, specified civil rights and hate crime laws.

**EXECUTIVE SUMMARY**

In 2005, five percent of adults in the United States used social media. In just six years, that number jumped to half of all Americans. Today, over 70 percent of adults use at least one social media platform. Facebook alone is used by 69 percent of adults, and 70 percent of those adults say they use the platform on a daily basis. Minors are also using social media at startling rates: a recent advisory from former U.S. Surgeon General Vivek Murthy found that up to 95 percent of minors aged 13 to 17 use social media platforms, with one-third of them reporting that they use social media “almost constantly.”

The Legislature has grappled with the risks and harms stemming from social media for years now; this Committee has heard numerous bills aimed at protecting children and users from the harmful content and effects of social media. One constant barrier to Legislative action is the federal law known as “Section 230,” which has been interpreted by the courts to provide social media platforms with broad immunity from any injuries arising from user-generated content.

Two recent cases have called into question the scope of Section 230 when applied to a social media platform’s presentation of content. The United States Supreme Court, in *Moody v. NetChoice LLC*, held that social media platforms can create expressive products when they create a personalized feed for a user; these expressive products can amount

to the platforms' own speech. And the United States Court of Appeal for the Third Circuit recently allowed a wrongful death suit against TikTok on the basis that TikTok's presentation of content to the user – separate and apart from the content itself – could have risen to the level of tortious conduct.

This bill does not alter the law on when, or how, a social media platform – or any other party – could be held criminally or civilly liable for a crime or tort committed by a user. Instead, this bill creates a civil cause of action against a social media platform that has, under existing law, aided and abetted, conspired in, or is otherwise jointly civilly or criminally liable for specified civil rights laws. In addition to the remedies available for the violation of the underlying law, this bill would allow the court to assess a civil penalty of one to three times the platform's gross monthly revenue, depending on the mental state of the platform; those amounts can be doubled if the platform knew, or should have known, that the plaintiff was a minor. This cause of action is intended to impose meaningful consequences on social media platforms that continue to push hate speech and other harassing conduct, but which are too profitable to be affected by a normal-sized civil damages award.

This bill is sponsored by the Children's Advocacy Institute at the University of San Diego School of Law, the Consumer Federation of California, Jewish Family and Children's Services of San Francisco, Loma LGBTQA+ Alumni and Allies, Rainbow Spaces, the Executive Board of San Diego Democrats for Equality, and is supported by the Association of University Women of California and the Center for Countering Digital Hate. This bill is opposed by the California Chamber of Commerce, the Computer and Communications Industry Association, and TechNet.

### **PROPOSED CHANGES TO THE LAW**

Existing constitutional law:

- 1) Provides that the U.S. Constitution, and the Laws of the United States, are the supreme law of the land. (U.S. Const., art. VI, cl. 2.)
- 2) Provides that Congress shall make no law abridging the freedom of speech. (U.S. Const., 1st amend. (the First Amendment) & 14th amends.; *see Gitlow v. People of State of New York* (1925) 268 U.S. 652, 666 (First Amendment guarantees apply to the states through the due process clause of the Fourteenth Amendment).)
- 3) Provides that every person may freely speak, write, and publish their sentiments on all subjects, and that a law may not restrain or abridge liberty of speech. (Cal. Const., art. I, § 2.)

Existing federal law:

- 1) Establishes the Children's Online Privacy Protection Act (COPPA) to provide protections and regulations regarding the collection of personal information from children under the age of 13. (15 U.S.C. §§ 6501 et seq.)
- 2) Provides that a provider or user of an interactive computer service shall not 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 a provider or user of an interactive computer service shall not be held liable on account of:
  - a) 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; or
  - b) any action taken to enable or make available to information content providers or others the technical means to restrict access to such material. (47 U.S.C. § 230(c)(2).)

Existing state law:

- 1) 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 users to interact socially with each other within the service or application; a service or application that provides email or direct messaging services, without more, does not meet this criterion.
  - b) The service or application allows users to do all of the following: (1) construct a public or semipublic profile for purposes of signing into and using the service or application; (2) populate a list of other users with whom an individual shares a connection within the system; (3) 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, § 22675(f).)
- 2) Defines the following additional relevant terms:
  - a) "Content" means statements or comments made by users and media that are created, posted, shared, or otherwise interacted with by users on an internet-based service or application; "content" does not include media put on a service or application exclusively for the purpose of cloud storage, transmitting files, or file collaboration.
  - b) "Public or semipublic internet-based service or application" excludes a service or application used to facilitate communication within a business or

- enterprise among employees or affiliates of the business or enterprise, provided that access to the service or application is restricted to employees or affiliates of the business or enterprise using the service or application. (Bus. & Prof. Code, § 22675(c), (d).)
- 3) Establishes the Ralph Civil Rights Act of 1976 (Ralph Act), which provides that all persons within the jurisdiction of this state have the right to be free from any violence, or intimidation by threat of violence, committed against their persons or property because of political affiliation, or on account of any characteristic listed in the Unruh Civil Rights Act (set forth below), or position in a labor dispute, or because another person perceives them to have one or more of those characteristics. (Civ. Code, § 51.7.)
- a) The perceived characteristics imported from the Unruh Civil Rights Act are: sex, race, color, religion, ancestry, national origin, disability, medical condition, genetic information, marital status, citizenship, primary language, and immigration status, as defined. (Civ. Code, § 51(a), (e).)
  - b) The bases of discrimination set forth in the Ralph Civil Rights Act are illustrative rather than restrictive. (Civ. Code, § 51.7(b)(1).)
- 4) Provides that a person is liable in a cause of action for sexual harassment when the plaintiff proves all of the following elements:
- a) There is a business, service, or professional relationship between the plaintiff and defendant or the defendant holds themselves out as being able to help the plaintiff establish a business, service, or professional relationship with the defendant or a third party; such relationships may exist between a plaintiff and persons including a physician, psychotherapist, dentist, attorney, real estate agent, banker, building contractor, executor, trustee, landlord or property manager, teacher, elected official, lobbyist, or director.
  - b) The defendant has made sexual advances, solicitations, sexual requests, demands for sexual compliance by the plaintiff, or engaged in other verbal, visual, or physical contact of a sexual nature or of a hostile nature based on gender, that were unwelcome and pervasive or severe.
  - c) The plaintiff has suffered, or will suffer, economic loss or disadvantage or personal injury, including, but not limited to, emotional distress or the violation of a statutory or constitutional right, as a result of the defendant's conduct. (Civ. Code, § 51.9.)
- 5) Provides that a person who denies a right as provided in 3) or 4), or aids, incites, or conspires in that denial, is liable for each and every offense for the actual damages suffered by any person denied that right and, in addition, the following:
- a) An amount to be determined by a jury, or a court sitting without a jury, for exemplary damages.

- b) A civil penalty of \$25,000 to be awarded to any person denied the right provided under 3) in any action brought by the person denied the right, or by the Attorney General, a district attorney, or a city attorney.
  - c) Attorney's fees as may be determined by the court. (Civ. Code, § 52.)
- 6) Establishes the Tom Bane Civil Rights Act (Tom Bane Act), which provides that if a person or persons, whether or not acting under color of law, interferes by threat, intimidation, or coercion, or attempts to interfere by threat, intimidation, or coercion, with the exercise or enjoyment by any individual or individuals of rights secured by the Constitution or laws of the United States, or of the rights secured by the Constitution or laws of this state, the Attorney General, or any district attorney or city attorney, or the person whose exercise or enjoyment of rights was interfered with, or attempted to be interfered with, may institute a civil action for damages, including a \$25,000 civil penalty, injunctive relief, and other appropriate equitable relief to protect the peaceable exercise enjoyment of the rights secured, and the court may award the petitioner or plaintiff reasonable attorney's fees.
  - a) Speech alone is not sufficient to support an action, except upon a showing that the speech itself threatens violence against a specific person or group of persons; and that the person or group of persons against whom the threat is directed reasonably fears that, because of the speech, violence will be committed against them or their property and that the person threatening violence had the apparent ability to carry out the threat.
  - b) No order issued in any proceeding under 6) shall restrict the content of any person's speech; an order restricting the time, place, or manner of any person's speech shall do so only to the extent reasonably necessary to protect the peaceable exercise or enjoyment of constitutional or statutory rights consistent with the constitutional rights of the person sought to be enjoined. (Civ. Code, § 52.1.)
- 7) Defines a "hate crime" as a criminal act committed, in whole or in part, because of one or more of the following actual or perceived characteristics of the victim: disability; gender; nationality; race or ethnicity; religion; sexual orientation; or association with a person or group with one or more of these actual or perceived characteristics. A hate crime includes, but is not limited to, a violation of 8). (Pen. Code, § 422.55.)
- 8) Provides that a person, whether or not acting under color of law, shall not:
  - a) By force or threat of force, willfully injure, intimidate, interfere with, oppress, or threaten any other person in the free exercise or enjoyment of a right or privilege secured by the Constitution or laws of this state or by the Constitution or laws of the United States in whole or in part because of one or more of the actual or perceived characteristics of the victim listed in 8).
  - b) Knowingly deface, damage, or destroy the real or personal property of any other person for the purpose of intimidating or interfering with the free

- exercise or enjoyment of a right or privilege secured by the Constitution or laws of this state or by the Constitution or laws of the United States, in whole or in part because of one or more of the actual or perceived characteristics of the victim listed in 8). (Pen. Code, § 422.6(a), (b).)
- 9) Provides that a person convicted of violating 8) shall be punished either by imprisonment in a county jail not to exceed one year, or by a fine not to exceed \$5,000, or both, or pursuant to Penal Code section 1170(h); and that, in addition, the court shall order the defendant to perform a minimum of community service, not to exceed 400 hours, to be performed over a period not to exceed 350 days. (Pen. Code, § 422.6(c).)
- 10) Provides that a person shall not be convicted of violating 8)(a) based upon speech alone, except upon a showing that the speech itself threatened violence against a specific person or group of persons and that the defendant had the apparent ability to carry out the threat. (Pen. Code, § 422.6(c).)

This bill:

- 1) States that the Legislature finds and declares the following:
- a) In California, corporations, like natural people, do not have a right to participate, through aiding, abetting, or conspiring, in hate crimes directed at vulnerable California populations, which is made unlawful by Section 422.6 of the Penal Code, to intimidate or terrorize Californians because of their sex, race, color, religion, ancestry, national origin, disability, medical condition, genetic information, marital status, sexual orientation, citizenship, primary language, or immigration status, which is made unlawful by Sections 51.7 and 51.9 of the Civil Code, or to interfere with the ability of those Californians to exercise their constitutional rights, which is made unlawful by Sections 52 and 52.1 of the Civil Code.
  - b) Social media platforms' failure to moderate content is a matter of life and death for minorities historically subject to discrimination.
  - c) According to a December 2024 report by the County of Los Angeles Commission on Human Relations, "[h]ate crimes in which anti-immigrant slurs were used climbed 31%...the largest number ever recorded since we began tracking this phenomenon in 2007."
  - d) According to American University, when it comes to being bullied on social media, "[a]mong the most victimized are people in the LGBTQ (lesbian, gay, bisexual, transgender, queer/questioning) community, especially teens and young adults." Anti-LGBTQ+ misinformation and inflammatory accusations of "grooming" have recently surged by over 400 percent across social media platforms, according to a recent report released by the Human Rights Campaign and the Center for Countering Digital Hate.

- e) Anti-Semitic incidents are spiking dramatically and are at all-time highs, according to the Anti-Defamation League (ADL). More than 10,000 anti-Semitic incidents occurred between October 7, 2023, and September 2024, which is up from 3,325 incidents during the prior year. That marks the most incidents recorded in a 12-month period by the organization since it began tracking threats in 1979.
  - f) Social media platforms have a documented record of accepting money from advertisements that plainly and violently threaten women and that clearly violate their own terms of service. In one 2023 study, a nonprofit placed advertisements that “referred to women as prostitutes, psychopaths, or vermin, and called for them to be beaten and killed.” Researchers recently said they detected a four-fold increase in the level of misogynistic content suggested by TikTok over a five-day period of monitoring as the algorithm served more extreme videos often focused on anger and blame directed at women.
  - g) As illustrated in cases such as *Sindell v. Abbott Laboratories* (1980) 26 Cal.3d 588, 604, California has adopted Prosser and Keaton on The Law of Torts (4th ed. 1971), Section 46, p. 292, which states that “those who, in pursuance of a common plan or design to commit a tortious act, actively take part in it, or further it by cooperation or request, or who lend aid or encouragement to the wrongdoer, or ratify and adopt his acts done for their benefit, are equally liable with him.” This principle of liability is reflected in Section 876 of the Restatement of Torts.
  - h) As illustrated in such cases as *Summers v. Tice* (1948) 33 Cal.2d 80, it has long been the law in California that “if a party cannot identify which of the two or more defendants caused an injury, the burden of proof may shift to the defendants to show that they were not responsible for the harm.” (*Sindell v. Abbott Laboratories* (1980) 26 Cal.3d 588, 598.)
  - i) For all of these reasons, the Legislature declares it necessary in order to preserve the peace, welfare, and lives of its residents to clarify the applicability of current laws against hate crime, terrorism, and intimidation to social media platforms.
- 2) Establishes Title 23 within Part 4 of Division 3 of the Civil Code (Title 23).
  - 3) Defines, for purposes of Title 23, a “social media platform” as a social media platform as defined in Section 22675 of the Business and Professions Code that generates more than \$100,000,000 per year in gross revenues.
  - 4) Provides that a social media platform that violates Section 31 or 422.6 of the Penal Code, the Ralph Act, the Tom Bane Act, or Section 51.9 or 52 of the Civil Code, including through its algorithms that relay content to users, or aids, abets, acts in concert, or conspires in a violation of any of those sections, or is a joint tortfeasor in an action alleging a violation of any of those sections, shall, in addition to any other

remedy, in an action brought pursuant to this provision, be liable for a civil penalty for each violation sufficient to deter future violations, but not to exceed the following:

- a) For an intentional violation, an amount equal to three months of the gross revenue of the social media platform preceding the judgment.
  - b) For a knowing or willful violation, an amount equal to two months of the gross revenue of the platform preceding the judgment.
  - c) For a reckless violation, an amount equal to one month of the gross revenue of the platform preceding the judgment.
  - d) If the evidence demonstrates that the platform knew, or should have known, that the plaintiff was a minor, the court may award up to twice the penalties described in 4)(a)-(c).
- 5) Provides that, for purposes of 4), deploying an artificial intelligence or algorithm that relays content to users may be considered an act of the platform independent from the message of the content relayed.
- 6) Provides that a platform shall be deemed to have actual knowledge of the operations of its own artificial intelligence and algorithms, including how and under what circumstances its artificial intelligence and algorithms deliver content to some users but not to others..
- 7) Includes a severability clause.
- 8) Provides that any waiver of 1)-7) shall be void and unenforceable as contrary to public policy.

### COMMENTS

#### 1. Author's comment

According to the author:

Violence, threats, and intimidation targeting certain historically vulnerable populations – Jews, LGBTQ+ community members, women, immigrants, and people of color especially – are at historic highs and rising at record-shattering rates. A recent Harvard study found a causal relationship between widespread violence against historically target groups and the practices of social media platforms.

Notwithstanding the escalating danger, social media platforms have announced dramatic retreats in screening and moderation practices to protect targeted populations. This change could not have come at a more dangerous time for groups that are historically targeted. L.A. County's most recent hate crime



report reflected double or triple digit increases in hate crimes resulting in “the largest number[s] ever recorded” against the LGBTQ+ community, Jews, Asians, Blacks, Latinos, and immigrants. This is a national trend that is accelerating.

California must respond to protect its most vulnerable residents. The least California can do is ensure that our existing laws against hate crimes, intimidation, and harassment, including conduct aimed at preventing our neighbors from exercising their constitutional rights, unambiguously apply to platform practices and offer penalties sufficient to prompt compliance with our laws without the necessity of a lawsuit.

## 2. Social media: the bad, the bad, and the ugly

According to the 2023 U.S. Surgeon General’s advisory on social media:

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...

More research is needed to fully understand the impact of social media; however, the 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.<sup>1</sup>

The “current body of evidence” includes:

- *Evidence of correlation:* a vast body of research, summarized in large literature reviews, establishes a correlation between social media use and poor mental health outcomes for youth.<sup>2</sup>
- *Known harm to teens’ body image and suicidal thoughts:* leaked research from Meta, Instagram’s parent company, revealed that Instagram made “body image issues worse for one in three teen girls.”<sup>3</sup> The same report found that, of users who

---

<sup>1</sup> United States Surgeon General, The U.S. Surgeon General’s Advisory: Social Media and Youth Mental Health (2023) p. 4, available at <https://www.hhs.gov/surgeongeneral/reports-and-publications/youth-mental-health/social-media/index.html>. All links in this analysis are current as of April 24, 2025.

<sup>2</sup> Costello et. al, *Algorithms, Addiction, and Adolescent Mental Health: An Interdisciplinary Study to Inform State-level Policy Action to Protect Youth from the Dangers of Social Media* (2023) 49 Am. J. L. and Med. 135, fn. 60 (listing studies).

<sup>3</sup> Wells et al. *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show* (Sept. 14, 2021) Wall Street Journal, available at [https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp\\_lead\\_pos7&mod=article\\_inline](https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp_lead_pos7&mod=article_inline).

reported having suicidal thoughts, 13% of UK and 6% of US teenagers suggested that they were the result of Instagram use.<sup>4</sup>

- *Bullying and harassment*: roughly two-thirds of adolescents are “often” or “sometimes” exposed to hate-based content, and nearly 75% of adolescents state that social media sites do a fair to poor job of addressing online harassment and bullying.<sup>5</sup>
- *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.<sup>6</sup>

The author argues that platforms have become less, not more, concerned about the effects of content on their users as time goes on.

For example, Meta CEO Mark Zuckerberg announced this year that Facebook, Instagram, and Threads would reduce their content-moderation efforts, including “a number of restrictions on topics like immigration, gender identity and gender that are the subject of frequent political discourse and debate.”<sup>7</sup> Meta’s stated “Hateful Conduct” policy online still purports to “remove dehumanizing speech, allegations of serious immorality or criminality, and slurs,” along with “harmful stereotypes, which [they] define as dehumanizing comparisons that have historically been used to attack, intimidate, or exclude specific groups” and “serious insults, expressions of contempt or disgust, cursing, and calls for exclusion or segregation when targeting people based on protected characteristics.”<sup>8</sup> A leaked Meta document, however, told a different story: the memo to Meta workers gave specific examples of the type of content that would be newly allowed, including:

- “Immigrants are grubby, filthy pieces of shit”
- “Mexican immigrants are trash!”
- “Jews are flat out [*sic*] greedier than Christians”<sup>9</sup>

Additionally, although Meta’s new policy broadly prohibits attacks on groups of people based on “mental characteristics,” including “allegations of stupidity, intellectual

---

<sup>4</sup> *Ibid.*

<sup>5</sup> Alhajji et al., *Cyberbullying, Mental Health, and Violence in Adolescents and Associations With Sex and Race: Data From the 2015 Youth Risk Behavior Survey* (2019) *Global pediatric health*, <https://journals.sagepub.com/doi/10.1177/2333794X19868887>; Vogels, *Teens and Cyberbullying* (2022) Pew Research Center: Internet, Science & Tech, <https://www.pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/>.

<sup>6</sup> CCDH, *Report: Deadly by Design*, (2022) p. 7, available at [CCDH-Deadly-by-Design\\_120922.pdf](https://counterhate.com/CCDH-Deadly-by-Design_120922.pdf) ([counterhate.com](https://counterhate.com/)).

<sup>7</sup> See Meta, *More Speech and Fewer Mistakes* (Jan. 7, 2025), <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.

<sup>8</sup> Meta, *Hateful Conduct* (Jan. 7, 2025), <https://transparency.meta.com/policies/community-standards/hateful-conduct/>.

<sup>9</sup> Biddle, *Leaked Meta Rules: Users Are Free to Post “Mexican Immigrants Are Trash!” or “Trans People Are Immoral* (Jan. 9, 2025) *The Intercept*, <https://theintercept.com/2025/01/09/facebook-instagram-meta-hate-speech-content-moderation/>.

capacity, and mental illness,” Meta specifically carved out “allegations of mental illness or abnormality when based on gender or sexual orientation.”<sup>10</sup> So while Meta’s policy would flag “men are crazy,” posts such as “gay people are sinners,” “trans people are immoral” and “trans people are mentally ill” are all specifically allowed.<sup>11</sup> Meta’s stated rationale is that the “political and religious discourse about transgenderism and homosexuality” justifies allowing hate speech against LGBTQ individuals;<sup>12</sup> but given that there is “religious and political discourse” about other kinds of speech that Meta prohibits, it is difficult to take this justification seriously.

On April 23, 2025, the Oversight Board – an independent board created by Meta in 2020 to review challenges to content-moderation decisions after a user has exhausted the appeals process on Facebook, Instagram, or Threads<sup>13</sup> – announced 11 decisions in user appeals relating to content moderation decisions.<sup>14</sup> In its summary of its rulings, the Oversight Board noted “concerns that Meta’s January 7, 2025, policy and enforcement changes were announced hastily, in a departure from regular procedure, with no public information shared as to what, if any, prior human rights due diligence the company performed.”<sup>15</sup> The Oversight Board further called on Meta to “live up to its public commitment to uphold the UN Guiding Principles on Business and Human Rights,” and recommended that Meta assess the human impact of its January 7 policy updates, improve how it enforces its bullying and harassment policies, and clarify the references to hateful ideologies not permitted on its platforms.<sup>16</sup>

### 3. The ongoing Section 230 debate: when does a platform’s presentation of user content become the platform’s own expressive conduct?

For users of social media, the First Amendment to the United States Constitution is their first line of defense: the state cannot proscribe protected speech simply because it is on the internet.<sup>17</sup> Platforms, however, are generally free to limit or prohibit speech by their users as they see fit: as private companies, they are not bound by the First Amendment.<sup>18</sup> Platforms also have broad immunity for content posted by their users, under the federal law known as 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

---

<sup>10</sup> Meta, *Hateful Conduct*, *supra*.

<sup>11</sup> Biddle, *supra*.

<sup>12</sup> Meta, *Hateful Conduct*, *supra*.

<sup>13</sup> Oversight Board, *How We Do Our Work*, <https://www.oversightboard.com/our-work/>.

<sup>14</sup> Oversight Board, *Wide-Ranging Decisions Protect Speech and Address Harms* (Apr. 23, 2025), <https://www.oversightboard.com/news/wide-ranging-decisions-protect-speech-and-address-harms/>.

<sup>15</sup> *Ibid*.

<sup>16</sup> *Ibid*.

<sup>17</sup> *Packingham v. North Carolina* (2017) 582 U.S. 98, 107-108.

<sup>18</sup> *E.g., Manhattan Community Access Corp. v. Hallek* (2019) 587 U.S. 802, 808.

content provider.”<sup>19</sup> That section also provides a safe harbor for “any action voluntarily taken in good faith to restrict access to, or the 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.”<sup>20</sup> 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.”<sup>21</sup>

Section 230 was passed in 1996, when the internet was in its relative infancy.<sup>22</sup> Congress enacted Section 230 in response to a lone unpublished New York trial court order, which ruled that, because an internet provider engaged in some forms of content moderation, the provider could be held liable for all the content on its bulletin boards.<sup>23</sup> There is little legislative history elucidating Congress’s intention for Section 230, but the courts applied Section 230 to give platforms broad immunity for harms caused by any user-posted content,<sup>24</sup> even when the platform was aware that their user was using their platform to harass or stalk a third party.<sup>25</sup>

The internet today bears little resemblance to the internet of 1996. Goecities pages and message boards read on a desktop computer over dial-up have given way to social media platforms and apps that stream video content to users on their phones. And instead of passively hosting third-party content, many social media platforms now select, algorithmically, what content to present to their users. This presentational element – the selection of a series of videos or posts, out of billions of options, to display to users – was unheard-of when Section 230 was enacted.

The United States Supreme Court, in *Moody v. NetChoice LLC*, held that social media platforms appear to “create expressive products” when they create a personalized feed for a user.<sup>26</sup> The way that platforms “make choices about what third-party speech to display and how to display it,” and determine what to “include and exclude, organize and prioritize...produce their own distinctive compilations of expression.”<sup>27</sup> The Court held that these platforms’ compilation and presentation of content can be protected by the First Amendment.<sup>28</sup> In doing so, the Court reignited a debate over whether Section 230 protects a platform from liability from harm caused by the overall effect of its presentation of content to users.

---

<sup>19</sup> 47 U.S.C. § 230(c)(1).

<sup>20</sup> *Id.*, § 230(c)(2)(A)

<sup>21</sup> *Id.*, § 230(e)(3).

<sup>22</sup> See, e.g., *Reno v. American Civil Liberties Union* (1997) 521 U.S. 844, 849-850 (explaining that the internet “is an international network of interconnected computers” and that “about 40 million people used the internet at the time of the trial [in 1996], a number that is expected to mushroom to 200 million by 1999”).

<sup>23</sup> (N.Y.S.C., May 24, 1995) 1995 WL 323710, \*3.

<sup>24</sup> E.g., *Zeran v. America Online, Inc.* (9th Cir. 1997) 129 F.3d 327, 331;

<sup>25</sup> E.g., *Jones v. DirtyWorld Entertainment Recordings LLC* (6th Cir. 2014) 755 F.3d 398, 415-416; *Carafano v. Metrosplash.com, Inc.* (9th Cir. 2003) 339 F.3d 1119, 1125.

<sup>26</sup> *Moody v. NetChoice LLC* (2024) 603 U.S. 707, 716.

<sup>27</sup> *Ibid.*

<sup>28</sup> *Id.* at pp. 726-727.

The Third Circuit said no. In *Anderson v. TikTok, Inc.*, the Third Circuit held that a mother could proceed with her suit alleging that TikTok, by feeding her daughter a “Blackout Challenge” video “that encourage[d] viewers to record themselves engaging in acts of self-asphyxiation,” was liable for her daughter’s death; ten-year-old Nylah died attempting to replicate the challenge.<sup>29</sup> The court held that TikTok’s “For You Page” (FYP), through which TikTok creates a “tailored compilation of videos for a user...based on a variety of factors, including the user’s age and other demographics, online interactions and other metadata” was “TikTok’s own expressive activity” and therefore not covered by Section 230.<sup>30</sup>

The Fourth Circuit, on the other hand, affirmed the dismissal of a complaint that argued that Dylan Roof, “who shot and killed nine people at Mother Emanuel AME Church in South Carolina” after being fed white supremacist content by Facebook and Google.<sup>31</sup> The majority opinion does not mention *Moody*; Judge Rushing, concurring in part and dissenting in part, would have permitted claims arising from Facebook’s recommendations to move forward.<sup>32</sup> In Judge Rushing’s analysis, “[t]he statement ‘ “You Should Join” this hate group’ is Facebook’s own speech” and is not entitled to Section 230 immunity.<sup>33</sup>

The Ninth Circuit has yet to interpret *Moody* in this context.<sup>34</sup>

4. This bill creates a cause of action against very large social media companies that aid and abet, or are otherwise jointly responsible for, a violation of specified hate crime and civil rights laws

Consistent with the Third Circuit’s interpretation of Section 230 discussed above, it appears that there is some room for a social media platform, through its own conduct, to aid and abet, or otherwise be jointly guilty or liable for, crimes committed by its users. Take the example of a TikTok FYP—if, unprompted by the user, their FYP began serving up explicitly and exclusively violent white supremacist content encouraging the viewers to commit hate crimes, there could theoretically be a point at which TikTok’s creation of the FYP could amount to aiding and abetting a hate crime committed by that user, or at least result in civil liability under one of the state’s civil rights laws.

---

<sup>29</sup> *Anderson v. TikTok, Inc.* (3d Cir. 2024) 116 F.4th 180, 181.

<sup>30</sup> *Id.* at pp. 181, 184.

<sup>31</sup> *M.P. by and through Pinckney v. Meta Platforms, Inc.* (4th Cir. 2025) 127 F.4th 516, 521-522.

<sup>32</sup> *Id.* at p. 529 (conc. and dis. opn. of Rushing, J.).

<sup>33</sup> *Id.* at p. 531.

<sup>34</sup> The Ninth Circuit relied on *Moody*’s explanation of the legal bar to succeed on a facial First Amendment challenge in the context of a California law compelling speech from social media platforms (*see X Corp. v. Bonta* (9th Cir. 2024) 116 F.4th 888, 898-899), and the district court for the Northern District of California has relied on *Moody* to enjoin portions of California laws that require platforms to restrict or regulate their users’ engagement with the platforms (*see NetChoice, LLC v. Bonta* (N.D. Cal., Mar. 13, 2025) — F.Supp.3d —, 2025 WL 807961, \*1-2 (enjoining portions of the California Age Appropriate Design Act); *NetChoice v. Bonta* (9th Cir., Dec. 31, 2024) — F.Supp.3d —, 2024 WL 5264045, \*1, 8 (enjoining portions of the Protecting Our Kids from Social Media Addiction Act).

This bill does not alter the law on when, or how, a social media platform – or any other party – could be held criminally or civilly liable for a crime or tort committed by a user. To the extent *Moody* created a space in Section 230 for a social media platform to be liable for its own arrangement and recommendation of speech, it is a narrow one, and this bill does not change that balance.

Instead, this bill creates a civil cause of action against a social media platform that has, under existing law, aided and abetted, conspired in, or is otherwise jointly civilly or criminally liable for specific civil rights laws: the Ralph Civil Rights Act;<sup>35</sup> the Tom Bane Act;<sup>36</sup> Civil Code section 51.9, which prohibits sexual harassment in business, service, and professional relationships; and California’s hate crimes law.<sup>37</sup> Under this cause of action, a platform that is jointly civilly or criminally liable for one of the listed offenses would additionally be liable for a civil penalty in the following amounts:

- For an intentional violation, an amount equal to three months of the gross revenue of the social media platform preceding the judgment.
- For a knowing or willful violation, an amount equal to two months of the gross revenue of the platform preceding the judgment.
- For a reckless violation, an amount equal to one month of the gross revenue of the platform preceding the judgment.

Additionally, if the platform knew, or should have known, that the victim was a minor, the court may increase the penalties to up to double the amount permitted for the relevant mental state. The bill also states that deploying an artificial intelligence or algorithm that relays content to users may be considered an act of the platform independent from the message of the content conveyed; the meaning of that “may” will be subject to the limits of Section 230.

This cause of action is intended to provide a meaningful incentive for social media platforms to pay more attention to hate speech and other harassing content on their platforms, and to be more diligent about not serving such content up to users. The sponsors argue that the largest social media platforms are simply too profitable to be deterred by a regular civil damages award or criminal penalty, and that the revenue-based penalties set forth in this bill are the best way to impose meaningful consequences.

## 5. Constitutional questions

There are two main issues here: whether the bill infringes on a social media platform’s First Amendment rights, and whether the bill is prohibited by Section 230.

---

<sup>35</sup> See Civ. Code, § 51.7.

<sup>36</sup> *Id.*, § 52.1.

<sup>37</sup> See Pen. Code, § 422.66.

With respect to the first issue, there should be no real concern that this bill will penalize a platform for protected speech. The statutes referred to in the bill – the Ralph Civil Rights Act, the Tom Bane Act, and California’s hate speech law – are longstanding civil rights laws.<sup>38</sup> Each is narrowly tailored so as not to penalize purely protected speech, requiring an additional element – such as an actual threat of violence or force,<sup>39</sup> hostile sexual advances,<sup>40</sup> or coercion<sup>41</sup> – before liability or a penalty can attach. If a platform truly aids and abets the violation of one of these statutes, the First Amendment should not be a defense.

This, of course, begs the question: *can* a platform truly aid and abet a violation of California’s civil rights laws by presenting third-party content to users? Opponents argue that this bill is a clear violation of Section 230, because it makes the platform liable for harms caused by its users’ speech, not speech of its own. The sponsors of the bill, however, argue that platforms that deliver highly personalized content and recommendations are engaging in their own expressive content which can aid and abet actionable hate speech. As discussed above, *Moody* suggests there is a point at which mere “content moderation” morphs into more – an act of expression by the platform itself, and at least one federal appellate court has relied on *Moody* to permit a suit against TikTok for displaying a third-party video to move forward. This bill appears aligned with the approach suggested in *Moody*, though it is an open question how the Ninth Circuit or Supreme Court would ultimately decide the issue.

## 6. Arguments in support

According to the Center for Countering Digital Hate:

Violence, threats, and intimidation targeting certain historically vulnerable populations – Jews, LGBTQ+ community members, women, immigrants, and people of color, especially – are at new highs and rising at record-shattering rates. For example, in L.A. County’s most recent hate crime report, the County documented both double or triple-digit increases in hate crimes resulting in “the largest number[s] ever recorded” against the LGBTQ+ community, Jews, Asians, Blacks, Latinos, and immigrants.

Notwithstanding the escalating danger, the market-dominant social media platform, Meta, has announced a dramatic retreat in screening and moderation practices to protect targeted populations. CCDH’s in-depth analysis of Meta’s policy changes shows that the company could stop as much as 97% of its content enforcement in key policy areas, including hate speech, bullying and harassment, and violence or incitement of violence...

---

<sup>38</sup> See Civ. Code, §§ 51.7, 51.9, 52, 52.1; Pen. Code, § 422.6.

<sup>39</sup> Civ. Code, § 51.7; Pen. Code, § 422.6.

<sup>40</sup> Civ. Code, § 51.9.

<sup>41</sup> *Id.*, § 52.1.

California law already prohibits every person and every corporation from engaging in hate crimes, harassment, and intimidation aimed at frightening people out of exercising their legal rights. It is urgent to update and clarify the application of these pre-Internet laws to ensure they meet the challenges of the modern era.

SB 771 will do just that while offering financial consequences minimally proportional to the vast wealth of the corporations and the need to ensure they are motivated to comply

## 7. Arguments in opposition?

According to the California Chamber of Commerce, the Computer and Communications Industry Association, and TechNet:

It is well established that the companies covered by this legislation have constitutional rights related to content moderation, including the right to curate, prioritize, and remove content in accordance with their terms of service. By exposing these companies to civil liability for content they do not remove, SB 771 creates a chilling effect on their editorial discretion. The significant, prescribed civil penalties - potentially amounting into the billions for each violation - would lead platforms to over-remove lawful content to mitigate legal exposure. Therefore, if this law passes, it will almost certainly be struck down in court (see *NetChoice v Paxton*) because it imposes liability on social media platforms for whether certain types of third-party content are shown to users, as well as the expressive choices social media platforms make in designing the user experience. This violates the First Amendment rights of users and social media platforms.

Moreover, the proposed liability framework likely conflicts with Section 230 of the Communications Decency Act, which provides strong federal protections for platforms against civil liability for third-party content and for good-faith content moderation. Courts (see *Twitter,inc v. Taamneh*, 598 U.S.\_\_ (2023)) have consistently upheld Section 230 as preempting state-level attempts to impose liability for content hosting or moderation decisions.

## SUPPORT

Children's Advocacy Institute, University of San Diego School of Law (co-sponsor)  
Consumer Federation of California (co-sponsor)  
Jewish Family and Children's Services of San Francisco (co-sponsor)  
Loma LGBTQA+ Alumni and Allies (co-sponsor)  
Rainbow Spaces (co-sponsor)  
San Diego Democrats for Equality, Executive Board (co-sponsor)  
Association of University Women of California



Center for Countering Digital Hate

### **OPPOSITION**

California Chamber of Commerce  
Computer and Communications Industry Association  
TechNet

### **RELATED LEGISLATION**

#### **Pending legislation:**

SB 659 (Reyes, 2025) increases the penalties available under the Parent’s Accountability and Child Protection Act in an action brought by a public prosecutor or parent or legal guardian or a minor. SB 659 is pending before the Senate Appropriations Committee.

AB 56 (Bauer-Kahan, 2024) requires social media platforms to display a warning label to its users at specified intervals. AB 56 is pending before the Assembly Judiciary Committee.

AB 2 (Lowenthal, 2024) makes a social media platform liable for specified damages if the platform fails to exercise ordinary care or skill by causing injury to a child. AB 2 is pending on the Assembly Floor.

#### **Prior legislation:**

SB 976 (Skinner, Ch. 321, Stats. 2024) enacted the Protecting Our Kids from Social Media Addiction Act, which restricted when and how social media platforms could provide access and content to minors, as specified.

SB 1504 (Stern, Ch. 900, Stats. 2024) amended the Cyberbullying Protection Act by requiring social media platforms to respond to reports of cyberbullying and disclose final determinations and increased the available penalties for a violation.

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 died in the Senate Appropriations Committee.

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 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.

\*\*\*\*\*