Date of Hearing: June 24, 2025
Fiscal: Yes

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
SB 771 (Stern) – As Amended June 19, 2025

SENATE VOTE: 29-4

SUBJECT: Personal rights: liability: social media platforms

SYNOPSIS

In 2022, after Elon Musk took over X, the platform formerly known as Twitter, he rolled back its content moderation policies. Rates of hate speech on the platform – including posts containing homophobic, transphobic, and racist slurs – spiked by about 50% shortly thereafter.[1] *In 2025, Meta CEO Mark Zuckerberg followed suit. Citing the reelection of Donald J. Trump, he announced Facebook, Instagram, and Threads would "remove restrictions on topics like immigration and gender that are out of touch with mainstream discourse."* [2] *Meta platform users soon thereafter reported a similar spike in harmful conduct.*

This bill is co-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, Rainbow Spaces, San Diego Democrats for Equality Executive Board, and Loma LGBTQA+ Alumni and Allies. They argue that changes to social media content moderation policies could not come at a worse time, as violence, threats, and intimidation specifically aimed at historically vulnerable populations appear to have reached record highs.

To help ensure social media companies are held accountable for the ways their policies may contribute to violence against such communities, this bill creates a civil cause of action against large social media platforms that violate certain existing civil rights laws or act as algorithmic accomplices to such violations. In addition to the remedies available for the violation of the underlying law, the bill would allow prevailing plaintiffs to seek a civil penalty of up to $1,000,000 for intentional, knowing, or willful violations, or $500,000 for reckless violations – amounts that may be doubled if the platform knew, or should have known, that the plaintiff was a minor.

The bill is opposed by the California Chamber of Commerce, the Computer and Communications Industry Association, and TechNet, who raise concerns relating to freedom of speech and federal preemption. If passed by this Committee, the bill will next be heard by the Judiciary Committee.

---

[1] UC Berkeley News, *Study finds persistent spike in hate speech on X* (Feb. 13, 2025), https://news.berkeley.edu/2025/02/13/study-finds-persistent-spike-in-hate-speech-on-x/.
[2] Barbara Ortutay, "Meta rolls back hate speech rules as Zuckerberg cites 'recent elections' as a catalyst" *Chicago Sun-Times* (Jan. 9, 2025) https://chicago.suntimes.com/nation-world/2025/01/09/meta-rolls-back-hate-speech-rules-zuckerberg-election-trump-x-twitter-musk-immigration-gender.

THIS BILL:

1) Provides that a social media platform that earns more than $100,000,000 in annual revenue and that violates specified civil rights laws (described below), including through recommendation algorithms, or that aid, abet, act in concert, or conspire in a violation of any of those laws, or is a joint tortfeasor in an action alleging a violation of any of those laws, is, in addition to any other remedy, subject to a civil penalty of up to $1,000,000 for intentional, knowing, or willful violations, or $500,000 for reckless violations. Such amounts maybe doubled if the platform knew or should have known the plaintiff was a minor.

2) Provides that deploying an algorithm that relays content to users may be considered an act of the platform independent from the message of the content relayed.

3) Deems a platform to have actual knowledge of the operations of its own algorithms, including how and under what circumstances its algorithms deliver content to some users but not to others.

4) Includes a severability clause and deems waivers of the bill's provisions void and unenforceable as contrary to public policy.

EXISTING 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) Establishes the Ralph Civil Rights Act of 1976, 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 protected characteristic, or position in a labor dispute, or because another person perceives them to have one or more of those characteristics. (Civ. Code § 51.7.)

3) 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.)

4) Provides that a person who denies a right as provided in 2) or 3), 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.)

5) Establishes the Tom Bane Civil Rights 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) Clarifies that 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) Provides that no order issued in any proceeding the Bane Act may 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.)

COMMENTS:

1) **Author's statement**. 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 targeted 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.

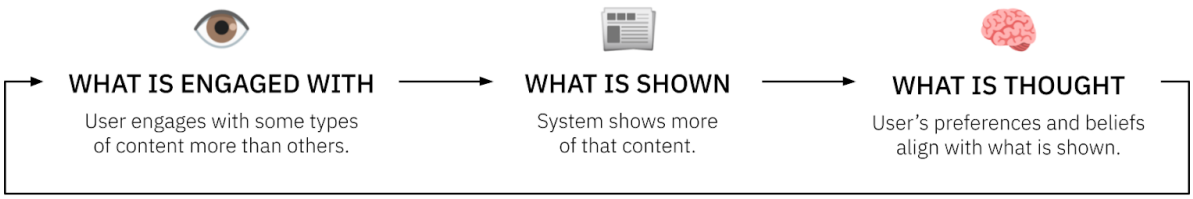
2) **Background**. Recommendation algorithms are a form of artificial intelligence that platforms like Facebook, Instagram, TikTok, and Twitter use to filter, rank, and recommend content. Using data about a user's past behavior – such as likes, comments, shares, watch time, and interaction frequency – these algorithms can predict the type of content that the user is most likely to find engaging.

While recommendation algorithms can create a more enjoyable, tailored experience for users, they can also lead to "social media bubbles". These bubbles exist because the algorithms that drive content feeds optimize for user engagement, as described by Dr. Luke Thorburn:

> When faced with a slate of recommendations, a user will engage with some types of content more than others. This is called selective exposure. For example, they may engage with ingroup content more than outgroup content (due to homophily or confirmation bias), with mainstream content more than indie content (due to cultural pressures), or with radical content more than moderate content (due to sensationalism or outrage). The recommender, because it is at least partially optimizing for engagement, starts to show more of the types of content that *are* engaged with. This changes what the users see, which causes their beliefs and preferences to change, becoming more aligned with what they see. Finally, this preference change causes the user to be even more selective, perhaps by explicitly following more homogeneous sources, or more consistently ignoring other types of content. Then the loop begins again.[3]

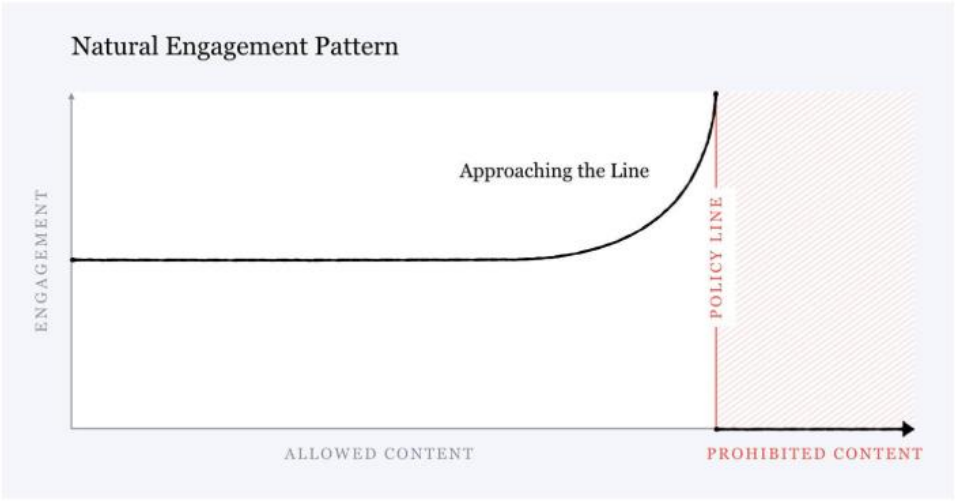
---

[3] Jonathan Stray, Luke Thorburn and Priyanjana Bengani, "From 'Filter Bubbles', 'Echo Chambers', and 'Rabbit Holes' to 'Feedback Loops,'" *TechPolicy.press* (April 17, 2023), www.techpolicy.press/from-filter-bubbles-echo-chambers-and-rabbit-holes-to-feedback-loops/.



Source: https://www.techpolicy.press/from-filter-bubbles-echo-chambers-and-rabbit-holes-to-feedback-loops/

Additionally, recommendation algorithms that seek to maximize engagement tend to prioritize sensationalized content that captures the user's attention for longer; as a result, extreme and polarizing content tends to be disproportionately promoted and amplified. As Meta CEO Mark Zuckerberg has observed, content that is closer to violating content moderation policies is more likely to increase engagement – a phenomenon he termed "the natural engagement pattern."[4]



Source: https://perma.cc/ZK5C-ZTSX.

According to Zuckerberg: "Our research suggests that no matter where we draw the lines for what is allowed, as a piece of content gets close to that line, people will engage with it more on average – even when they tell us afterwards they don't like the content."[5]

Following the reelection of President Donald J. Trump, Zuckerberg announced that these lines would be redrawn on Facebook, Instagram, and Threads to allow for a more lax approach to content moderation, including "on topics like immigration, gender identity and gender that are the subject of frequent political discourse and debate."[6] A leaked internal document set forth examples of the type of hateful content that would be newly allowed, including:

---

[4] Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement" (May 5, 2021), *https://perma.cc/ZK5C-ZTSX*.
[5] *Id.*
[6] *See* Meta, *More Speech and Fewer Mistakes* (Jan. 7, 2025), https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/.

- "Immigrants are grubby, filthy pieces of shit"
- "Mexican immigrants are trash!"
- "Jews are flat out greedier than Christians"
- "Gay people are sinners"
- "Trans people are mentally ill"[7]

Predictably, reports of harmful content on Meta platforms have since spiked. According to a recent survey of around 7,000 users, 92% of respondents "felt less protected from being exposed to or targeted by" harmful content, including hateful or violent material, and 77% felt "less safe" expressing themselves on Meta's platforms. One in six respondents reported being the victim of gender-based or sexual violence on Meta platforms, while 2/3 said they had witnessed harmful content.[8]

Failure to moderate hate speech on social media has had lethal consequences, as demonstrated in several developing countries where social media platforms failed to adequately invest in content moderation. This was most starkly exemplified by the ethnic cleansing against Rohingya Muslims in Myanmar beginning in 2017. According to Amnesty International:

> In the months and years leading up to and during the 2017 atrocities, Facebook in Myanmar became an echo chamber of virulent anti-Rohingya content. Actors linked to the Myanmar military and radical Buddhist nationalist groups systematically flooded the Facebook platform with incitement targeting the Rohingya, sowing disinformation regarding an impending Muslim takeover of the country and seeking to portray the Rohingya as sub-human invaders. The mass dissemination of messages that advocated hatred inciting violence and discrimination against the Rohingya, as well as other dehumanizing and discriminatory anti-Rohingya content, poured fuel on the fire of long-standing discrimination and substantially increased the risk of an outbreak of mass violence.
>
> [. . .]
>
> Amnesty International's analysis shows how Meta's content-shaping algorithms and reckless business practices facilitated and enabled discrimination and violence against the Rohingya. Meta's algorithms directly contributed to harm by amplifying harmful anti-Rohingya content, including advocacy of hatred against the Rohingya. They also indirectly contributed to real-world violence against the Rohingya, including violations of the right to life, the right to be free from torture, and the right to adequate housing, by enabling, facilitating, and incentivizing the actions of the Myanmar military. Furthermore, Meta utterly failed to engage in appropriate human rights due diligence in respect of its operations in Myanmar ahead of the 2017 atrocities. This analysis leaves little room for doubt: Meta substantially contributed

---

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

[8] Agence France-Presse, "Harmful content' in Meta apps rises – poll" *MSN.com* (Jun. 18, 2021), https://www.msn.com/en-ph/news/world/harmful-content-in-meta-apps-rises-poll/ar-AA1GTqGu?ocid=BingNewsSerp.

> to adverse human rights impacts suffered by the Rohingya and has a responsibility to provide survivors with an effective remedy.[9]

Similar events have occurred in several countries. As Princeton researchers Arvind Narayanan and Sayash Kapoor write:

> Hate speech on Facebook fueled a civil war in Tigray, Ethiopia, in which over a half a million people have been killed. In India, inflammatory content and calls to violence on WhatsApp form the backdrop of constant communal violence. In Sri Lanka, hate speech on Facebook played a role in anti-Muslim violence. In Afghanistan, less than 1 percent of hate speech was taken down, according to Facebook's own metrics. Content moderation failures in various postconflict countries, including Bosnia and Herzegovina, Indonesia, and Kenya, have also been documented.[10]

Meanwhile, hate crimes against some groups appear to have been increasing domestically. According to the Los Angeles County Commission on Human Relations report on hate crimes, following three years of double-digit increases, reported hate crimes in Los Angeles County drastically increased 45% from 930 to 1,350 in 2023 – the highest total of hate crimes in the history of the report.[11] For 2023, while Attorney General Rob Bonta's office reports an overall decrease from 2022, hate crimes against LGBTQ+, Jewish, and Muslim communities increased.[12]

3) **What this bill would do.** This bill provides that a social media platform that earns more than $100,000,000 in annual revenue and that violates specified civil rights laws – including the Ralph Civil Rights Act of 1976 (Ralph Act)[13] and the Tom Bane Civil Rights Act (Bane Act)[14] – is, in addition to any other remedy, subject to a civil penalty of up to $1,000,000 for intentional, knowing, or willful violations, or $500,000 for reckless violations. Such amounts maybe doubled if the platform knew or should have known the plaintiff was a minor.

The Ralph Act provides that all persons within the jurisdiction of California 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 protected characteristic, or position in a labor dispute, or because another person perceives them to have one or more of those characteristics.[15] The rights protected by section 51.7 may be enforced by a private action for damages.[16]

---

[9] *Myanmar: The social atrocity: Meta and the right to remedy for the Rohingya* (2022), pp. 6, 9, https://www.amnesty.org/en/documents/ASA16/5933/2022/en/.
[10] Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, Can't Do, and How to Tell the Difference* (1st ed. 2024), pp. 189-190.
[11] *LA County Releases Annual Report on Hate Crimes: Highest Total of Hate Crimes Ever Reported* (2024), https://lacounty.gov/2024/12/11/highest-total-of-hate-crimes-ever-reported/.
[12] California Department of Justice, *Attorney General Bonta Releases 2023 Hate Crime Report, Highlights Continued Efforts to Combat Hate* (2024), https://oag.ca.gov/news/press-releases/attorney-general-bonta-releases-2023-hate-crime-report-highlights-continued.
[13] The Ralph Civil Rights Act of 1976 (Civ. Code, § 51.7; Ralph Act) Tom Bane Civil Rights Act (Civ. Code, § 52.1; Bane Act)
[14] Civ. Code, § 52.1
[15] Civ. Code, § 51.7.
[16] See § 52, subd. (b).

The Bane Act prohibits "threat[s], intimidation or coercion" designed to prevent a person's exercise or enjoyment of their constitutional or statutory rights. "A plaintiff must show (1) intentional interference or attempted interference with a state or federal constitutional or legal right, and (2) the interference or attempted interference was by threats, intimidation or coercion."[17] Speech alone is not sufficient to show a violation, "except upon a showing that the speech itself threatens violence against a specific person or group of persons and 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."[18] The section provides for civil damages and equitable relief, as well as criminal sanctions for violations.

The bill provides that a predicate violation can occur directly, by conspiracy, or by aiding and abetting. Conspiracy requires an actual agreement to commit a crime. Aiding and abetting requires participation in the crime. "Aid" requires some conduct by which one becomes concerned in the commission of a crime, whether it be to aid (i.e., assist or supplement), promote, encourage, or instigate. "Abet," on the other hand, requires that this conduct be accompanied by the requisite criminal state of mind – that is, knowledge of the perpetrator's unlawful purpose and the intent that it be facilitated.[19]

To assist in proving such claims, the bill specifies that deploying an algorithm that relays content to users may be considered an act of the platform independent from the message of the content relayed. The bill also deems a platform to have actual knowledge of the operations of its own algorithms, including how and under what circumstances its algorithms deliver content to some users but not to others.

4) **Opposition concerns.** Opponents of the bill raise concerns relating to freedom of speech and federal preemption.

First Amendment. The United States and California Constitutions prohibit abridging, among other fundamental rights, freedom of speech.[20] "The Free Speech Clause of the First Amendment . . . can serve as a defense in state tort suits."[21] "[T]he basic principles of freedom of speech and the press, like the First Amendment's command, do not vary when a new and different medium for communication appears."[22] Additionally, "the creation and dissemination of information are speech . . . ."[23] Dissemination of speech is different from "expressive conduct," which is conduct that has its own expressive purpose and may be entitled to First Amendment protection.[24] The United States Supreme Court in *Moody v. NetChoice, LLC*[25] recently held that algorithmic curation of content can constitute expressive activity protected under the First Amendment.

---

[17] *Allen v. City of Sacramento* (2015) 234 Cal.App.4th 41, 67.
[18] Civ. Code § 52.1(k).
[19] *People v. Campbell* (Cal. App. 6th Dist. 1994), 25 Cal. App. 4th 402.
[20] U.S. Const., 1st and 14th Amends; Cal. Const. art. I, § 2.
[21] *Snyder v. Phelps* (2011) 562 U.S. 443, 451.
[22] *Joseph Burstyn v. Wilson* (1952) 343 U.S. 495, 503.
[23] *Sorrell v. IMS Health Inc.* (2011) 564 U.S. 552, 570.
[24] *Ibid.*
[25] (2024) 603 U.S. 707.

Laws that are not content specific are generally subject to "intermediate scrutiny," which requires that the law "be 'narrowly tailored to serve a significant government interest.'"[26] In other words, the law "'need not be the least restrictive or least intrusive means of' serving the government's interests," but "'may not *regulate expression* in such a manner that a substantial portion of the burden on speech does not serve to advance its goals.'"[27] This bill does not regulate expression; it augments liability for large platforms that violate an existing civil rights statutes.

Furthermore, as concluded by the Senate Judiciary Committee:

> . . . there should be no real concern that this bill will penalize a platform for protected speech. The statutes referred to in the bill . . . are longstanding civil rights laws. 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, hostile sexual advances, or coercion —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.

Federal preemption. Section 230(c)(1) of the federal Communications Decency Act of 1996 shields online platforms from liability for third-party content: "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."[28] This provision has been hailed as the law that created the modern internet, fostering free expression online and allowing an array of innovative services and spaces to flourish, from search engines to social media.[29] It has also come with a destructive side, absolving platforms of responsibility for virtually all third-party harms arising from the use of their services – "a protection not available to print material or television broadcasts."[30]

Section 230 was intended to promote investment in online companies and encourage "'Good Samaritan' blocking and screening of offensive material" [31] without fear of liability for defamation.[32] Courts soon adopted an expansive interpretation – a key early decision construed "publisher" immunity as encompassing "traditional editorial functions" such as deciding whether to publish, remove, or even alter content.[33] Consequently, the plaintiff, a victim of online defamation by an anonymous user, had no recourse against the platform despite its failure to remove the content in a timely manner, which would have resulted in liability in the offline world. Following this logic, courts have extended Section 230 well beyond the defamation context, routinely concluding that online intermediaries are not liable for harms related to third-party illicit content.[34] Such harms, Professor Danielle Keats Citron argues, are disproportionately

---

[26] *Packingham v. North Carolina* (2017) 582 U.S. 98, 98.
[27] *McCullen v. Coakley* (2014) 573 U.S. 464, 486, emphasis added.
[28] 42 U.S.C. § 230(c)(1). Section 230 also (1) provides a safe harbor for good faith content moderation, (2) preempts contrary state laws, and (3) enumerates exemptions for enforcement of federal criminal statutes, intellectual property laws, communications privacy laws, and sex trafficking.
[29] See e.g., Kosseff, *The Twenty-Six Words that Created the Internet* (2019).
[30] Quinta Jurecic, "The politics of Section 230 reform: Learning from FOSTA's mistakes" *Brookings* (Mar. 1, 2022), https://www.brookings.edu/articles/the-politics-of-section-230-reform-learning-from-fostas-mistakes.
[31] § 230(c).
[32] *Fair Hous. Council v. Roommates.com, LLC* (9th Cir. 2008) 521 F.3d 1157, 1163.
[33] *Zeran v. Am. Online, Inc* (4th Cir. 1997) 129 F.3d 327.
[34] Michael Rustad & Thomas Koenig, "The Case for a CDA Section 230 Notice-and-Takedown Duty" (2023) 23 Nev.L.J. 533, 561-574.

visited upon women, children, and historically marginalized groups, depriving them of the opportunity to participate fully in online life at the expense of their civil rights and liberties.[35] "The common thread weaving through these cases is that the courts have sapped §230's Good Samaritan concept of its meaning."[36]

This sweeping grant of immunity has been the subject of widespread criticism and calls for reform.[37] Senators Lindsey Graham and Dick Durbin are planning to introduce a bill that would sunset Section 230.[38] Justice Clarence Thomas has called for the Supreme Court to review the scope of Section 230.[39] Ninth Circuit Judge Ryan Nelson recently stated that courts have "stretch[ed] the statute's plain meaning beyond recognition," leading to "perverse effects."[40] The Ninth Circuit "should revisit our precedent," he urged, particularly in light of "artificial intelligence raising the specter of lawless and limitless protections."[41]

Arguing Section 230 preempts the bill, opponents cite to *Twitter, Inc. v. Taamneh*,[42] a case that did not address section 230. Plaintiffs in that case were families of an individual killed in a terrorist attack who contended that Google, Twitter, and Facebook were liable for aiding and abetting international terrorism by failing to take meaningful action to remove the terrorists' recruitment and fundraising posts. Although the Ninth Circuit had concluded that Section 230 largely preempted plaintiffs' claims – other than the claim in a consolidated case that Google shared advertising revenue with ISIS[43] – the United States Supreme Court concluded that the allegations failed to show that the platforms consciously, voluntarily, and culpably participated in or supported terrorism.[44] The case is thus relevant to this bill, albeit not for the reason opponents assert.

Proponents, on the other hand, point to the Third Circuit's decision in *Anderson v. TikTok Inc.*,[45] which held that TikTok's recommendation algorithm—which promoted a "Blackout Challenge" to a 10-year-old girl who then died from self-asphyxiation—was the platform's own expressive conduct that falls outside of section 230. The court drew on the Supreme Court holding in *Moody* that an algorithm that editorially curates third-party speech is protected by the First Amendment.[46] As to the interplay between section 230 and the First Amendment, the Third

---

[35] *See* Danielle Keats Citron, "How to Fix Section 230" (2023) 103 B.U.L. Rev. 713.
[36] *Id.* at p. 727.
[37] *See id. See also* e.g., John Lucas, "AG Moody Joins with Other Attorneys General to Urge Congress to Stop Protecting Illegal Activity on the Net," *Capitolist* (May 23, 2019), https://thecapitolist.com/ag-moody-joins-with-other-attorneys-general-to-urge-congress-to-stop-protecting-illegal-activity-on-the-net.
[38] Lauren Feiner, "Lawmakers are tyring to repeal section 230 again" *The Verge* (Mar. 21, 2025), https://www.msn.com/en-us/politics/government/lawmakers-are-trying-to-repeal-section-230-again/ar-AA1BptAI?ocid=BingNewsVerp.
[39] *Doe ex rel. Roe v. Snap, Inc.* (2024) 144 S. Ct. 2493 (Thomas, J., dissenting from denial of certiorari).
[40] *Calise v. Meta Platforms, Inc.* (9th Cir. 2024) 103 F.4th 732, 747 (Nelson, J. concurring) (*Calise*).
[41] *Ibid.*
[42] (2023) 598 U.S. 471
[43] *Gonzalez v. Google LLC* (9th Cir. 2021) 2 F.4th 871, 898 (stating "We conclude that § 230 does not immunize Google from the claims premised on revenue-sharing"). Like *Taamneh*, this case was eventually resolved on the grounds that the plaintiffs did not allege facts sufficient to show a violation of anti-terrorism laws. (*Gonzalez v. Google* LLC (2023) 598 U.S. 617, 622.)
[44] *Id.* at p. 505.
[45] (3d Cir. 2024) 116 F.4th 180.
[46] *Id.*, discussing *Moody v. NetChoice, LLC* (2024) 603 U.S. 707. When social media companies shut down President Trump's accounts following the January 6, 2021, insurrection, Texas and Florida passed laws purporting to restrict social media companies from "censoring" user content by overriding their content moderation choices.

Circuit quoted Justice Clarence Thomas's observation that "[i]n the platforms' world, they are fully responsible for their websites when it results in constitutional protections, but the moment that responsibility could lead to liability, they can disclaim any obligations and enjoy greater protections from suit than nearly any other industry."[47] Proponents argue the Ninth Circuit may adopt the reasoning of this case. While early returns are not promising – the Northern District of California rejected such an argument[48] – the issue has not yet been addressed at the appellate level.

Proponents acknowledge the formidable obstacles this bill faces, contending that, in a time of increasing hate, any possibility of protecting vulnerable Californians is worth a shot.

ARGUMENTS IN SUPPORT: The bill's cosponsors write: "One of the world's largest corporations controlled by the world's second wealthiest person ($232 billion) has, with perfect self-awareness, imperiled the lives and rights of the most vulnerable of all Californians. How California, to borrow the Governor's phrase, 'meets' this "moment" will properly determine how we are judged by history. SB 771 is a good step toward ensuring the judgment is a positive one."

The Center for Countering Digital Hate writes:

> 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…
>
> 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.

The American Association of University Women -- California writes:

NetChoice and Computer and Communications Industry Association challenged these laws, asserting, among other things, that they violated the First Amendment. In July 2024, the Supreme Court remanded these cases with instructions to lower courts to provide more analysis of this issue. In doing so, the majority, in an opinion written by Justice Elena Kagan, generally affirmed that the First Amendment protects the editorial decisions of social media platforms carried out by algorithmic recommendation systems. (*Id*. at p. 740.)

[47] *Doe ex rel. Roe v. Snap, Inc*. (2024) 144 S. Ct. 2493 (Thomas, J., dissenting from denial of certiorari).

[48] *Doe (K.B.) v. Backpage.com, LLC* (N.D.Cal. 2025) 768 F. Supp. 3d 1057.

> Regardless of a social media platform company's motivation in reducing or removing practices that cull out targeted, threatening content, the fact remains that this decision, and those of other companies that choose to facilitate illegal, threatening content delivery to intended targets by simply turning a blind eye, makes them culpable in aiding and abetting illegal behavior.
>
> These companies have not "failed" to provide safeguards, ***they have intentionally turned away from providing safeguards they previously provided*** for targeted vulnerable communities and should be held accountable for that decision, particularly in light of the real-world impacts of the proliferation of online hate and their complicity in unlawful intimidation. (Emphasis in original.)

ARGUMENTS IN OPPOSITION: In opposition to the bill, California Chamber of Commerce, the Computer and Communications Industry Association, and TechNet argue:

> Although SB 771 does not explicitly mandate content removal, it effectively incentivizes broad suppression of speech through the threat of legal action. In practice, the elevated liability risk could compel platforms to take down content based solely on unsubstantiated allegations of violence. This dynamic sets the stage for a heckler's veto, in which bad actors or politically motivated users can flag content they disagree with, knowing the platform may err on the side of removal to avoid potential lawsuits.
>
> This bill's implicit concern is harmful content. It is impossible for companies to identify and remove every potentially harmful piece of content because there's no clear consensus on what exactly constitutes harmful content, apart from clearly illicit content. Determining what is harmful is highly subjective and varies from person to person, making it impossible to make such judgments on behalf of millions of users. Faced with this impossible task and the liability imposed by this bill, some platforms may decide to aggressively over restrict content that could be considered harmful.
>
> Furthermore, platforms would need to evaluate whether to eliminate their fundamental features and functions, which are the reasons users go to their platforms, due to the legal risk involved. For instance, direct messaging features could potentially be misused for contacting and bullying other teens; such features would likely be removed.

REGISTERED SUPPORT / OPPOSITION:

Support

Childrens Advocacy Institute (Co-Sponsor)
Consumer Federation of California (Co-Sponsor)
Jewish Family and Children's Services of San Francisco, the Peninsula, Marin and Sonoma Counties (Co-Sponsor)
LOMA LGBTQIA+ Alumni and Allies Coalition (Co-Sponsor)
Rainbow Spaces (Co-Sponsor)
San Diego Democrats for Equality (Co-Sponsor)
American Association of University Women - California
American Association of University Women, San Jose Branch
Center for Countering Digital Hate

Easterseals Northern California
Hadassah
Jcc/federation of San Luis Obispo
Jcrc Bay Area
Jewish Community Federation and Endowment Fund
Jewish Community Relations Council, Santa Barbara
Jewish Democratic Club of Marin
Jewish Family & Community Services East Bay
Jewish Family and Children's Service of Long Beach and Orange County
Jewish Family Service LA
Jewish Family Service of San Diego
Jewish Family Services of Silicon Valley
Jewish Federation of Orange County
Jewish Federation of San Diego
Jewish Federation of the Greater San Gabriel and Pomona Valleys
Jewish Free Loan Association
Jewish Long Beach
Jewish Public Affairs Committee
Simon Wiesenthal Center and Museums of Tolerance

Oppose

California Chamber of Commerce
Computer & Communications Industry Association
Technet

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