SENATE JUDICIARY COMMITTEE
Senator Thomas Umberg, Chair
2021-2022 Regular Session

SB 1056 (Umberg)
Version: April 7, 2022
Hearing Date: April 19, 2022
Fiscal: No
Urgency: No
AWM

SUBJECT

Violent posts

DIGEST

This bill requires a social media platform, as defined, with 1,000,000 or more monthly users 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.

EXECUTIVE SUMMARY

Social media platforms have existed for fewer than 25 years, yet are ubiquitous and affect nearly every facet of many people's lives. The size and scope of many social media platforms give them unprecedented power over what people read, what they believe, and how they vote. It is not an exaggeration to say that social media platforms are one of the driving forces behind the current state of the country—for better and for worse.

One of the negative aspects of social media is the extent to which it is used to threaten violence or plan violent incidents. The horrific January 6, 2021, attack on the United States Capitol was largely planned over social media. Individuals used social media to give advice about how to illegally bring guns into Washington D.C., and to call for violence against elected officials, including Vice President Michael Pence, Senator Mitt Romney, and Speaker of the House Nancy Pelosi. Many of the individuals who track extremism on social media have commented that the attack was organized in plain sight, due to either the indifference or ineffectiveness of social media platforms' content moderation efforts.

Not all online violence is as high profile as the January 6 attack. Many social media users—particularly users who are people of color, LGBTQ+, or members of religious minorities—are subjected to threats of violence from other users on a regular basis.

Living with a constant barrage of threats of violence leads to negative mental health outcomes and often causes the recipients to think twice before speaking—which sadly vindicates the use of online threats as a tool for silencing voices disfavored by the person making the threats.

This bill creates a narrow cause of action for persons who are the target of violent posts on social media, defined as posts that make true threats or incitements to imminent lawless action that are unprotected by the First Amendment. The bill's requirements and the cause of action created are limited to social media platforms with 1,000,000 or more discrete monthly users.

First, the bill requires that social media platforms post in a clear location whether they have such a reporting mechanism to ensure that individuals are aware whether they should first report a post to the platform.

The bill then authorizes a person who is, or believes they are, the target of a violent post on social media to file an action for a court order requiring the social media platform to remove the post. When the social media platform has a mechanism for reporting violent posts, the person must first report the post to the platform, and the court must wait 48 hours before ruling on the action to give the platform time to act on it; when the social media platform does not have a reporting mechanism, the person may file, and the court may rule on, the action at any time. If a court determines that the post is a violent post, the court may order the social media platform to remove it. The court may also order related violent posts removed, if doing so furthers the interests of justice, as long as the related posts also qualify as violent posts. When the plaintiff succeeds in their action to have a post removed, the court must award them reasonable attorney fees. If the plaintiff's request is denied, the court may award the platform reasonable attorney fees if it determines that the action was not prosecuted in good faith.

This bill is sponsored by the author. There is no known opposition.

PROPOSED CHANGES TO THE LAW

Existing constitutional law:

1) Provides a right to free speech and expression. (U.S. Const., 1st amend; Cal. Const., art 1, § 2.)

2) Recognizes certain judicially created exceptions to the rights of freedom of speech and expression, including for true threats and incitement to imminent violence. (*E.g.*, *Virginia v. Black* (2003) 538 U.S. 343, 359.)

Existing federal law:

1) 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).)

2) 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).)

3) Provides that no cause of action may be brought and no liability may be imposed under any state or local law that is inconsistent with items 1) and 2). (47 U.S.C. § 230(e)(3).)

Existing state law:

1) Authorizes a court to issue a temporary restraining order on an ex parte basis, or a restraining order after a noticed hearing, against a person who has harassed another.
    a) Conduct that may warrant a restraining order includes a credible threat of violence, defined as a knowing and willful statement or course of conduct that would place a reasonable person in fear for their safety or the safety of their immediate family, and that serves no legitimate purpose, and harassment, which includes a credible threat of violence that would cause a reasonable person to suffer substantial emotional distress. (Code Civ. Proc., § 527.6.)

2) Makes it a crime to willfully threaten to commit a crime that will result in the death or great bodily injury with the specific intent that the statement is to be taken as a threat, even if there is no intent to carry out the threat, when the threat on its face and under the circumstances is so unequivocal, unconditional, immediate, and specific that it conveys to the person threatened a gravity of purpose and an immediate prospect of execution of the threat and thereby causes that person reasonably to be in sustained fear for their own safety or for the safety of their family. (Pen. Code, § 422.)

This bill:

1) Establishes the Online Violence Prevention Act.

2) Defines the following terms:
    a) "Content" is media that are created, posted, shared, or otherwise interacted with by users on an internet-based service, but does not include media put online exclusively for file sharing.
    b) "Social media platform" means an internet-based service or application that has users in California and that meets all of the following criteria:
        i. The primary purpose of the service or application is to connect users and allow users to interact with each other within the service or application.
        ii. The service or application allows users to do all of the following:
            1. Construct a public or semipublic profile within a bounded system created by the service or application.
            2. Populate a list of other users with whom an individual shares a connection within the system.
            3. View and navigate a list of connections made by other individuals within the system.
            4. Create or post content viewable by other users.
    c) "User" is a person with an account on a social media platform.
    d) "Violent post" is content on a social media platform that contains a true threat against a specific person or an incitement to imminent lawless action that is not protected by the First Amendment to the United States Constitution.

3) Requires a social media platform to clearly and conspicuously state whether it has a mechanism for reporting violent posts that is available to users and nonusers of the platform. If the social media platform has such a reporting mechanism, the statement must contain a link to the mechanism.

4) Allows a person who is the target of a violent post, or a person who believes they are the target of a violent post, to seek a court order requiring the social media platform to remove the violent post and any related violent post the court determines shall be removed in the interests of justice. The person may bring the action at any time.

5) Provides that, if the social media platform has a reporting mechanism as described in 3), a person must notify the social media platform of the violent post and request that it be removed before bringing an action described in 4).

6) Provides that, if a social media platform does not have a reporting mechanism as described in 3), the court may rule on the request at any time.

7) Provides that, if a social media platform has a reporting mechanism as described in 3), the court may not rule on the request until 48 hours has passed since the person notified the social media platform of the violent post and requested its removal. If the social media platform removes the post after the action is filed but before the end of the 48-hour window, the court may dismiss the action.

8) Provides that a court shall award court costs and reasonable attorney fees to a prevailing plaintiff in an action described in 4), and that a court may award reasonable attorney fees to a prevailing defendant if the court finds that the plaintiff's prosecution of the action was not in good faith.

9) Provides that the bill does not apply to a social media platform with fewer than 1,000,000 discrete monthly users.

COMMENTS

1. Author's comment

According to the author:

> SB 1056 would establish the Online Violence Prevention Act to protect victims of threats on social media platforms. Social media platforms can reunite friends, foster meaningful relationships, and provide perspectives that traditional media might ignore. Unfortunately, social media platforms have also been used for violence. The January 6, 2021, attack on the United States Capitol was planned and coordinated largely through various social media channels, with some participants openly threatening to commit acts of violence against elected officials. Many individuals on social media platforms—especially people of color, women, and LGPTQ+ individuals—are also frequently barraged with threats of violence, simply for speaking out or sharing their perspectives.
>
> While the guarantees of free speech and free expression protect vehement disagreement, even to the point of unpleasantness, there is no constitutional protection for true threats or incitements to imminent acts of violence. Such statements stifle speech, by inducing fear in the speakers, and add nothing to the online discourse. Yet there is no clear avenue for the target of such violence to have the violent post removed from a social media platform. Therefore, SB 1056 provides a precisely tailored solution for individuals who are the targets of online threats or incitements to imminent violence.
>
> SB 1056 requires social media platforms with users in California, and with at least one million discrete monthly viewers, to post in an easily accessible location whether they have a mechanism for reporting threats or incitements to violence. The bill does not require a social media platform to have a reporting

mechanism—only to make clear if it does or does not. Lastly, SB 1056 allows the target, or someone who believes they are the target, of a threat or incitement to violence on a social media platform to go to court and seek an injunction requiring the social media platform to remove the violent post. If the social media platform has a reporting mechanism, the court must give the platform 48 hours before issuing the order. In this way, SB 1056 helps to ensure that one should has to live in fear of online threats, or worry that a threatening post will inspire others to the same acts of violence.

2. The January 6, 2021, attack on the United States Capitol, and how online threats become real-world violence

The attack on the United States Capitol was a dark day for America. A mob of supporters of President Donald Trump, many of them armed, broke into the United States Capitol, roamed the halls looking for members of Congress and Vice President Pence, assaulted law enforcement officers, and attempted to prevent Congress from certifying the legitimate election results.[1] The attack resulted in at least seven deaths, hundreds of injuries, and untold trauma for the law enforcement officers who defended the Capitol against the mob.[2] The attack has been called " 'the most serious attack on the operations of the Federal Government since the Civil War.' "[3]

For many who watched footage of the attack, the images were stunning—an unprecedented display of political violence on American soil. But for those who had paid attention to the attackers' pre-January 6 social media activity, the attack was not a surprise. From the November 3, 2020, election until January 6, plans for the attack were made openly on numerous social media sites.[4] In the words of the Washington Post, "[t]he red flags were everywhere."[5]

One study by the Digital Forensics Lab found that the "stop the steal" movement on social media—which falsely argued that President Donald Trump had won the

---

[1] *E.g.*, Kanno-Youngs, Tavernise, & Cochrane, *After House Was Breached, a Fear 'We'd Have to Fight' to Get Out*, New York Times (Jan. 17, 2021), https://www.nytimes.com/2021/01/06/us/politics/capitol-breach-trump-protests.html (last visited Apr. 5, 2022).
[2] Cameron, *These Are the People Who Died in Connection with the Capitol Riot*, New York Times (Jan. 5, 2022), https://www.nytimes.com/2022/01/05/us/politics/jan-6-capitol-deaths.html (last visited Apr. 5, 2022).
[3] *Trump v. Thompson* (D.C. Cir. 2021) 20 F.4th 10, 20-21.
[4] *E.g.*, Silverman, et al., *Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading Up to Jan. 6 Attack, Records Show*, ProPublica (Jan. 4, 2022), https://www.propublica.org/article/facebook-hosted-surge-of-misinformation-and-insurrection-threats-in-months-leading-up-to-jan-6-attack-records-show (last visited Apr. 1, 2022). Facebook had established a task force to specifically target and eliminate election-related misinformation leading up to the November 3, 2020, election, but dissolved the task force in early December, allowing the "stop the steal" and Capitol attack groups to flourish. (*Id.*)
[5] Davis, *Red Flags*, Washington Post (Oct. 31, 2021), *available at* https://www.washingtonpost.com/politics/interactive/2021/warnings-jan-6-insurrection/ (last visited Apr. 1, 2021).

election—grew between November and January and was allowed by social media platforms to morph into a movement calling for civil war, violence, and the executions of "treasonous" government officials.[6] One of President Trump's attorneys tweeted that if Vice President Pence were arrested, Secretary of State Mike Pompeo would save the election and Pence "will face execution by firing squad."[7] A group on Parler asked users to share the locations of Black Lives Matter and antifa buses, so that the addresses could be sent to the Proud Boys who would "get them before they go out to the streets," while a group on Facebook called "Red State Succession" asked for the home and office addresses and travel routes of "political enemies."[8] A member of a Florida-based militia movement uploaded a video on Facebook giving advice on how to illegally bring guns into the District of Columbia.[9] Posters inserted versions of the word "peaceful" into their posts to avoid triggering content moderation, such as " 'Mitt Romney peacefully gets it first.' "[10] Parler sent "dozens" of posts containing violent content to the FBI, including specific threats against specific politicians.[11] In the aftermath, Twitter suspended President Trump's Twitter account due to the risk of further violence.[12]

Despite the clear evidence that the January 6 attack was planned on social media, many social media platforms argued that their content moderation efforts leading up to the attack were adequate.[13] The CEOs of Facebook (now Meta, and the parent company of Instagram and WhatsApp) and Alphabet (which owns YouTube) have denied that their platforms bore any responsibility for the misinformation that led to the attack, though Twitter's CEO conceded Twitter shared some of the blame.[14] Efforts to determine the full extent of social media's role in the attack are ongoing: the House committee investigating the January 6 attack has issued subpoenas to Alphabet, Meta, Reddit, and Twitter to ascertain their social media platforms' precise roles in, and culpability for, the

---

[6] Atlantic Council's DFR Lab, #*StoptheSteal: Timeline of Social Media and Extremist Activities Leading to 1/6 Insurrection* (Feb. 10, 2021), https://www.justsecurity.org/74622/stopthesteal-timeline-of-social-media-and-extremist-activities-leading-to-1-6-insurrection/ (last visited Apr. 1, 2022).
[7] *Ibid.*
[8] *Ibid.*
[9] *Ibid.*
[10] Davis, *supra*, fn. 5.
[11] Benner, *Parler says it sent the F.B.> posts about threats to the Capitol ahead of Jan. 6*, New York Times (Mar. 25, 2021), *available at* https://www.nytimes.com/2021/03/25/us/parler-fbi-capitol-attack.html (last visited Apr. 1, 2022).
[12] Tiku, Romm, & Timberg, *Twitter bans Trump's account, citing risk of further violence*, Washington Post (Jan. 8, 2021), *available at* https://www.washingtonpost.com/technology/2021/01/08/twitter-trump-dorsey/ (last visited Apr. 5, 2022).
[13] *E.g.*, Timberg, Dwoskin, & Albergotti, *Inside Facebook, Jan. 6 violence fueled anger, regret over missed warning signs*, Washington Post (Oct. 22, 2021), *available at* https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/ (last visited Apr. 1, 2022).
[14] McCabe & Kang, *Lawmakers Grill Tech C.E.O.s on Capitol Riot, Getting Few Direct Answers*, New York Times (Mar. 25, 2021, updated Oct. 23, 2021), *available at* https://www.nytimes.com/2021/03/25/technology/facebook-twitter-google-capitol-riots-hearing.html (last visited Apr. 1, 2022).

attack.[15] The committee issued the subpoenas after the companies failed to cooperate with requests to voluntarily turn over documents and information.[16]

The January 6 attack was unique in its size and scope, but it was not an outlier in terms of the use of social media to post and plan for acts of violence. As this Committee heard at its November 9, 2021, informational hearing, *The State of Social Media Regulation: Misinformation, Exploitation, Harassment, and Radicalization*, online violence is all too common. A study conducted by the Anti-Defamation League found that:

- 27 percent of respondents reported experiencing "severe" online harassment in the last year, which includes sexual harassment, stalking, physical threats, swatting, doxing, and sustained harassment.[17]
- 18 percent of respondents reported being physically threatened on social media.[18]
- Of the respondents who received physical threats, 41 percent reported that the social media platform did not take any action on the threatening content, and only 14 percent reported that the platform deleted the threatening content.[19]
- LGBTQ+ individuals and Muslims made up the largest percentage of individuals harassed online, and Jewish and Asian-American individuals report disproportionately high rates of harassment.[20]

As FBI Director Christopher Wray told the federal Senate Judiciary Committee, " 'The amount of angry, hateful, combative, violent even, rhetoric on social media exceeds what anybody in their worst imagination [thinks] is out there.' "[21]

Threatening to harm another person is a crime.[22] Yet it appears that many have accepted threats and plans for violence as the cost of doing business, even when it disproportionately burdens members of protected classes. This bill is intended to push back against the assumption that California can do nothing to protect its residents from the worst types of online violence.

3. <u>This bill authorizes a person who is the target of a violent threat on social media to seek a court order requiring the platform to remove the threat</u>

---

[15] Broadwater & Isaac, *Jan. 6 Committee Subpoenas Four Big Tech Firms* New York Times (Jan. 13, 2022), *available at* https://www.nytimes.com/2022/01/13/us/politics/jan-6-tech-subpoenas.html (last visited Apr. 1, 2022).
[16] *Ibid.*
[17] ADL Center for Technology, Report, *Online Hate and Harassment: The American Experience 2021* (2021), *available at* https://www.adl.org/online-hate-2021#results (last visited Apr. 5, 2022), at p. 8.
[18] *Id.* at p. 8.
[19] *Id.* at p. 15.
[20] *Id.* at p. 14.
[21] Dilanian, *Why did the FBI miss the threats about Jan. 6 on social media?*, NBC News (Mar. 8, 2021), https://www.nbcnews.com/politics/justice-department/fbi-official-told-congress-bureau-can-t-monitor-americans-social-n1259769 (last visited Apr. 1, 2022).
[22] *See, e.g.*, Pen. Code, § 422.

This bill provides a remedy for individuals who are the target of threats of violence and incitements to imminent violence on social media platforms. As discussed further below, the bill's remedy is limited to the targets of "violent posts," defined as true threats or incitements to imminent lawless action, to ensure that only speech unprotected by the First Amendment is affected. The bill defines "social media platform" as an online platform whose primary purpose is to connect users for social interactions, which should make clear that this bill applies to true social media sites—Facebook, Twitter, YouTube, etc.—and not online platforms that incidentally involve user profiles and connections, such as e-commerce sites, news sites that allow comments, or media sites that do not allow users to upload their own content (e.g., HBO Max, Netflix, Spotify). The bill limits its application to social media platforms with 1,000,000 discrete monthly users or more, to ensure that nascent sites are not overly burdened.

The bill first requires a social media platform to post in a clear and accessible location whether it has a mechanism for reporting violent posts. If the platform does have such a mechanism, the disclosure must include a link to the mechanism. The bill does not require social media platforms to have any such mechanism, or impose any requirements for content moderation.

The bill then authorizes any person—whether or not they are a user of the social media platform—who is or believes themselves to be the target of a violent post to file an action in court for a court order requiring the social media platform to remove the post. If the social media platform has a reporting mechanism, the person must use the mechanism before filing the suit, and the court must wait 48 hours before ruling on the request for an order to remove the post. If the social media platform does not have a reporting mechanism, the person may file the action, and the court may rule on it, at any time. If the court determines that the post is a violent post not protected by the First Amendment, the court may order the social media platform to remove it, along with any other related violent posts the court determines should be removed in the interests of justice. Any related posts ordered to be removed must also fall within the definition of "violent post," to ensure that protected speech is not removed.

Finally, the bill provides that the court must award reasonable attorney fees to a prevailing plaintiff in this action. If the court rules against the plaintiff, the court may award reasonable attorney fees to the defendant if the court determines that the plaintiff's prosecution of the action was not in good faith.

4. <u>First Amendment implications</u>

    a. *The bill's requirement that social media platforms post a notice of whether they have a reporting mechanism for threats is likely permissible compelled commercial speech*

The bill does not impose any content moderation obligations on a social media platform or dictate what content must be moderated. It does, however, require a platform to

clearly and conspicuously state whether it has a mechanism for reporting violent posts available to users and nonusers, and to include a link to the mechanism if it has one. The bill thus imposes a limited form of government-mandated speech on social media platforms.

Commercial speech is protected under the state and federal guarantees of free speech, but to a lesser degree than noncommercial speech.[23] Generally speaking, requiring a commercial actor to provide factual, uncontroversial product information is permissible "as long as the disclosure requirements are reasonably related to the State's interest in preventing deception of consumers."[24] California currently imposes similar disclosure requirements on companies doing business online, such as the provision of the California Consumer Privacy Act of 2018 that requires a company to inform consumers about the categories of personal information it collects and the purposes for which the information is collected.[25] Accordingly, it appears that the bill's requirement that a social media company post whether it has a reporting mechanism is consistent with the First Amendment's framework for mandated commercial disclosures.

b. The bill's authorization for a court to order the removal of posts constituting true threats or imminent incitements to violence is likely consistent with the boundaries of speech unprotected by the First Amendment

This bill targets "violent posts" on a social media platform, by allowing a court to order their removal from a social media platform if the platform fails to do so on its own. Posts on social media platforms constitute speech. As explained below, however, the bill is likely adequately narrowly tailored so as not to run afoul of the First Amendment.

The First Amendment of the United States Constitution and the corresponding provision of the California Constitution generally guarantee the freedoms of speech and expression.[26] "The hallmark of the protection of free speech is to allow 'free trade in ideas'—even ideas that the overwhelming majority of people might find distasteful or discomforting."[27]

But "the right of free speech is not absolute at all times and under all circumstances."[28] Restrictions on the content of speech are permissible "in a few limited areas, which are 'of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.' "[29]

---

[23] *See Gerawan Farming, Inc. v. Lyons* (2004) 33 Cal.4th 1, 22.
[24] *Zauderer v. Office of Disciplinary Counsel of Supreme Court of Ohio* (1985) 471 U.S. 626, 651.
[25] *See* Civ. Code, § 1798.100.
[26] U.S. Const., 1st amend; Cal. Const., art 1, § 2; *e.g.*, *Texas v. Johnson* (1989) 491 U.S. 397, 404-405.
[27] *Virginia v. Black* (2003) 538 U.S. 343, 358 (lead opn. of O'Connor, J.).
[28] *Chaplinsky v. State of New Hampshire* (1942) 315 U.S. 568, 571.
[29] *R.A.V. v. City of Saint Paul, Minnesota* (1992) 505 U.S. 377, 382-383.)

This bill specifically targets two such categories of unprotected speech: true threats and incitements to imminent lawless action. True threats are "statements where the speaker means to communicate a serious expression of an intent to commit an act of unlawful violence to a particular individual or group of individuals."[30] The "prohibition on true threats 'protect[s] individuals from the fear of violence' and 'from the disruption that fear engenders,' in addition to protecting people 'from the possibility that the threatened violence will occur.' "[31] Incitements to imminent lawless action, or "fighting words," is speech " 'that itself inflicts injury or tends to incite immediate violence.' "[32]

Because the bill is expressly targeted at these two categories of unprotected speech, the bill does not facially proscribe any protected speech. The bill also appears to be consistent with judicially created tests for when a restriction on speech is permissible. The bill is viewpoint-neutral in that it does not prohibit speech "based on hostility—or favoritism—towards the underlying message expressed."[33] The bill is explicit in its application only to unprotected threats and incitements to imminent violence that fall outside the First Amendment, so it does not appear that the bill is at risk of overbreadth or vagueness.[34] The bill also does not impose an impermissible prior restraint because a post can be removed only after the fact.[35] And the bill appears unlikely to chill legitimate speech because the speaker is not directly targeted or penalized for posting a violent post that is subsequently ordered removed.[36]

One potential issue that may arise is whether a court can determine whether a threat constitutes an unprotected "true threat" without knowing the intent of the poster, i.e., the subjective intent of the person who posted the threat. There is currently an open context as to whether the First Amendment exempts threats only when the speaker had a subjective intent to threaten. The United States Supreme Court held in *Elonis v. United States* that the federal criminal threats statute[37] requires a subjective intent to threaten—whether or not the speaker intended to carry out the threat—in addition to a finding that the speech would have been understood as a threat to a reasonable person.[38]

---

[30] *Virginia*, *supra*, 538 U.S. at p. 359 (lead opn. of O'Connor, J.).
[31] *Id.* at p. 360; *see also Watts v. U.S.* (1969) 394 U.S. 705, 708.
[32] *R.A.V.*, *supra*, 505 U.S. at p. 380.
[33] *Id.* at p. 386.
[34] *See Bd. of Airport Commissioners of City of Los Angeles v. Jews for Jesus* (1987) 482 U.S. 569, 574 (statute may be invalidated under the overbreadth doctrine when it is overbroad to the point of roping in protected speech as well as unprotected speech, or is at risk of dissuading protected speech).
[35] *See Bantam Books, Inc. v. Sullivan* (1963) 372 U.S. 58, 70 ("Any system of prior restraints of expression comes to this Court bearing a heavy presumption against its constitutional validity").
[36] *See Reno v. ACLU* (1997) 521 U.S. 844, 872 (statute that raised significant risk that speakers would "remain silent rather than communicate ever arguably unlawful words" posed particular First Amendment concerns).
[37] 18 U.S.C. § 875.
[38] *Elonis v. United States* (2015) 575 U.S. 723, 740.

The Court did not, however, rule on the issue of whether the First Amendment always requires a finding of a subjective intent to threaten, specifically declining to do so until another case presented the issue directly.[39] The Court also has not ruled on the question of whether subjective intent is necessary when speech is penalized outside the criminal context. The Ninth Circuit appears to have approved of an objective-only test in the civil context, noting that a state university could lawfully penalize conduct that a reasonable person would have interpreted as threatening without examining what the speaker intended.[40]

Additionally, it appears that the courts have yet to consider the nature of threats in the specific realm of the internet—particularly on social media platforms and when the threats are coming from total strangers. The unique nature of threats over social media militates in favor of a test that looks only at the plain language of the message conveyed.

Consider the experience of female journalists. A UNESCO study of 901 women journalists worldwide found that 73 percent had experienced online violence; 25 percent experienced threats of physical violence, including death threats; 18 percent experienced threats of sexual violence; and 13 percent had received threats of violence to persons close to them, including children and infants.[41] The harassment is worse for women journalists who are not white, straight, or members of the religious majority.[42] The barrage of online violence caused a range of negative professional and personal outcomes, ranging from needing to increase physical security, missing work to recover from threats, making themselves less visible, and psychological effects including PTSD.[43] The two platforms used most commonly to threaten the women surveyed were Facebook and Twitter.[44] Many female journalists report that the platforms have been insufficiently responsive when they report online threats and abuse.[45]

Threats of violence conveyed over the internet are likely to lack the context clues that are present in in-person interactions and that can inform whether violent words should be interpreted as a threat. In *Watts v. U.S.*, the Court reviewed a conviction for criminal threats where the defendant stated " 'They always holler at us to get an education. And now I have already received my draft classification as 1-A and I have got to report for my physical this Monday coming. I am not going. If they ever make me carry a rifle the first man I want to get in my sights is L.B.J.' "[46] The Court looked at the context—such

---

[39] *Id.* at pp. 741-742.
[40] *O'Brien v. Welty* (9th Cir. 2016) 818 F.3d 920, 932.
[41] Posetti, et al., *The Chilling: Global trends in online violence against women journalists*, UNESCO (Apr. 2021), *available at* https://en.unesco.org/sites/default/files/the-chilling.pdf (last visited Mar. 28, 2022), at p. 12 (*The Chilling*).
[42] *Id.* at p. 16.
[43] *Id.* at p. 13.
[44] *Id.* at p. 14.
[45] *Id.* at p. 36.
[46] *Watts*, *supra*, 394 U.S. at p. 706.

as the fact that the speaker was at an anti-war rally, and speaking to other rallygoers—the frequently hyperbolic nature of political speech, and "the expressly conditional nature of the statement and the reaction of the listeners" to conclude the statement could be interpreted only as a statement of political opposition, not a threat.[47]

A threat online, out of the blue and from an unknown or anonymous source, lacks those context clues. So, for example, when a journalist receives a message on a social media account that the sender is going to rape and murder her,[48] she likely has no way to know what the sender's intent is—whether the message is a misguided attempt at legitimate discourse, or intended to terrorize her, or is an actual threat of violence to come. The ambiguity is part of the point: threats against women journalists (and other women, people of color, trans people, etc.) are frequently intended to frighten their targets into silence.[49] In other words, these threats are speech intended to prevent the speech of others—usually because of their gender, race, sexuality, gender identity, or other characteristic the poster believes gives them less of a right to speak out.

As noted above, speech falls outside the protection of the First Amendment when it is " 'of such slight social vale as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.' "[50] In the context of threats, the Court has determined that any value derived from threats is outweighed by the interest in " 'protect[ing] individuals from the fear of violence' and 'from the disruption that fear engenders,' in addition to protecting people 'from the possibility that the threatened violence will occur.' "[51] The kind of speech this bill is intended to target—bald threats transmitted over social media, often to a stranger—appears to have little redeeming social or political value, whether the speaker intended to threaten or not. And the effect of such threats implicates several interests the state is dedicated to protecting: not only is the state interested in preventing individuals from suffering fear and distress from receiving threats, but the state has a strong interest in protecting free speech for *all*, including on the internet. There is, therefore, no apparent

---

[47] *Id.* at p. 708.
[48] *The Chilling*, *supra*, fn. 41, at p. 31. Many of the threats received by women journalists (and women in general) are significantly more graphic and disturbing. (*E.g.*, *id.* at p. 25.)
[49] *Id.*, at p. 6; *see* Trollbusters & International Women's Media Foundation, *Attacks and Harassment: The Impact on Female Journalists and Their Reporting* (2018) *available at* https://www.iwmf.org/wp-content/uploads/2018/09/Attacks-and-Harassment.pdf (last visited Mar. 28, 2022), at p. 44; Amnesty International, *Toxic Twitter – A Toxic Place for Women* (2018), https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/ (last visited Mar. 28, 2022); Vilk, et al., *No Excuse for Abuse: What Social Media Companies Can Do Now to Combat Online Harassment and Empower Users*, PEN America (2020) https://pen.org/report/no-excuse-for-abuse/ (last visited Mar. 28, 2022).
[50] *R.A.V.*, *supra*, 505 U.S. at pp. 382-383.
[51] *Virginia*, *supra*, 538 U.S. at p. 360 (lead opn. of O'Connor, J.).

constitutional imperative to allow a threat of violence to remain posted on the internet once a court has determined that it is, objectively, such a threat.[52]

To be clear, this analysis would be different in the context of a measure to penalize the poster of a threat without regard for the poster's intent. This analysis does not take a position on whether sending an anonymous threat like "I am going to rape you" could be punished criminally without regard to the speaker's subjective intent, or whether the statement itself could be understood as prima facie evidence of intent to threaten.[53] For purposes of this bill, however, the question is only whether a court can order a threat transmitted over social media to be removed without knowing whether the speaker's subjective intent. For all the reasons discussed above, it appears that the answer is yes.

5. Section 230 and federal preemption issues

a. The history of Section 230

The federal law governing social media platforms was enacted years before social platforms existed. Section 230 of the Communications Decency Act,[54] also known as Section 230, was enacted in 1996.[55] Section 230 was passed in response to two trial court orders that could have hampered the development of the then-fledgling internet.

The first of these orders granted an online subscription-based service's motion for summary judgment in a libel suit filed against it by a user of the service.[56] The allegedly defamatory statements were not made by the service itself, but instead were made by writers on one of the forums on the service and who were not directly employed by the service.[57] In granting the service's motion for summary judgment, the court used the republisher/distributor framework developed for print media: a republisher is liable for any defamatory statements in the republished work, but a distributor is liable only if it was unaware and had no reason to know of the defamation in the work it

---

[52] It does not follow, however, that any unprotected speech online can be regulated First Amendment (*R.A.V., supra*, 505 U.S. at p. 383), a determination of whether a statement is defamatory requires a determination of whether a statement is true or false. If social media platforms were put in the position of facing liability for failing to remove allegedly defamatory content, the platform would have little incentive to keep up any post flagged as defamatory, even if it is not. Such a measure would likely result in a significant chilling of political speech, which "central to the meaning and purpose of the First Amendment." (*Citizens United v. Federal Election Com'n* (2010) 558 U.S. 310, 329.)

[53] To the extent courts have required a subjective intent for online threats, many cases have looked at threats made "into the ether"—threats that were not sent directly to the supposed target, rather than directly to the target. (*E.g., U.S. v. Bagdasarian* (9th Cir. 2011) 652 F.3d 1113.) To the extent a person transmitted an explicit threat like the ones discussed in this analysis directly to the target, it is difficult to see how the subjective intent could be anything *but* to threaten, regardless of whether the speaker intended to carry out the threat.

[54] 47 U.S.C. § 230.

[55] Pub. L. 104-104, 110 Stat. 56 (1996).

[56] *See Cubby, Inc. v. CompuServe, Inc.* (S.D.N.Y. 1991) 776 F.Supp. 135.

[57] *Id.* at pp. 137-138.

distributed.[58] Because it was undisputed that the online service did not have editorial control over the forum or engage in any content moderation whatsoever, the court ruled that the service was only a distributor and therefore could not be liable absent evidence it knew or had reason to know of the defamation.[59]

The second order followed the logic of the first: another user had sued an online network for the allegedly defamatory statements of a third party on the online network's service, and the court applied the republisher/distributor framework to determine if the online network could be liable for the third party's statements.[60] Here, evidence established the online network engaged in a modest degree of content moderation within the network.[61] The court found that, even though the online network did not exercise total control over the content on its network, the network had "uniquely arrogated to itself the role of determining what is proper for its members to post and read on its bulletin boards."[62] On that basis, the court ruled that the online network should be treated as a republisher, not a distributor, of the allegedly defamatory statements, thereby allowing it to be liable for the third party's statements on its network even though it had no reason to know of them.[63]

The reaction to this second order—though it was only a single trial court order in New York state court—was swift. If internet providers could be held liable as publishers if they engaged in *any* level of content moderation, online platforms would have the untenable choice of abandoning content moderation all together (likely allowing their pages to be overrun with slurs, obscenities, and other unwanted content) or policing their pages so tightly as to restrict any meaningful discourse. The solution, from two federal legislators and members of the burgeoning internet industry, was Section 230.[64]

The crux of Section 230 is simple:

> 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.[65]

Section 230 also provides a safe harbor for platforms engaging in content moderation, stating that no provider or user shall be held liable because of good-faith efforts to restrict access to material that is "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally

---

[58] *Id.* at pp. 139-140.
[59] *Id.* at pp. 140-141.
[60] *See Stratton Oakmont, Inc. v. Prodigy Services Co.* (N.Y. Sup. Ct., May 24, 1995) 1995 WL 323710, at *3.
[61] *Id.* at p. *4.
[62] *Ibid.*
[63] *Id.*, at p. *5.
[64] Kosseff, The Twenty-Six Words That Created The Internet (2019) pp. 68-73.
[65] 47 U.S.C. § 230(c)(1).

protected."[66] Section 230 is intended to occupy the field of content moderation on the internet, by stating that "[n]o cause of action may be brought and no liability may be imposed under any State law that is inconsistent with this section."[67]

There is virtually no legislative history for Section 230. It was passed as a last-minute addition to a much more controversial bill, and the relative novelty of the internet meant that few were paying attention to Section 230's potential impact.[68] Yet despite a clear statement of legislative intent, and Section 230's use of terms associated with defamation cases, Section 230 has been interpreted by courts to "reach[] far beyond the initial state court decision that sparked its enactment."[69] Courts almost immediately began to interpret Section 230 in a range of cases to immunize internet platforms from "virtually all suits arising from third-party content."[70] Courts even extended Section 230 immunity to situations where the platform's moderator affirmatively solicited the information, selected the user's statement for publication, and/or edited the content.[71]

Of course, Section 230 does not prevent a user from seeking redress from the actual *speaker* of a threatening, or otherwise actionable, statement. As a practical matter, however, this is rarely a realistic option. Even if a plaintiff obtains the speaker's IP address from the social media platform via subpoena, the IP address often cannot be connected to a single user, leaving the plaintiff with no defendant from whom to seek recourse.[72] Thus, counterintuitively, Section 230 appears to provide *greater* protection for speech on social media *greater* protections than speech elsewhere.

Given that Section 230 was enacted in the very early stages of the internet, its authors could not anticipate the myriad ways social media would facilitate online harassment or the extent to which online harassment and threats would be used to silence speakers. In 1996, there were approximately 20 million Americans with internet access, and they spent an average of fewer than 30 minutes online each month.[73] In 2022, there are approximately 307.2 million internet users in America[74] and 4.95 billion internet users worldwide[75] who spend an average of approximately two-and-a-half hours on social

---

[66] *Id.*, § 230(c)(1) (2).
[67] *Id.*, § 230(e)(1), (3).
[68] Kossef, *supra*, fn. 64, at pp. 73-76.
[69] *HomeAway.com, Inc. v. City of Santa Monica* (9th Cir. 2019) 918 F.3d 676, 684.
[70] *Id.* at pp. 94-95; *see, e.g., Doe v. MySpace Inc.* (5th Cir. 2008) 528 F.3d 413, 421-422; *Carfano v. Metrospalsh.com, Inc.* (9th Cir. 2003); 339 F.3d 1119, 1125; *Zeran v. America Online, Inc.* (4th Cir. 1997) 129 F.3d 327, 333-334.
[71] *See, e.g., Jones v. Dirty World Entertainment Recordings LLC* (6th Cir. 2014) 755 F.3d 398, 415; *Batzel v. Smith* (9th Cir. 2003) 333 F.3d 1018, 1030-1031; *cf. Blumenthal v. Drudge* (D.D.C. 1998) 992 F.Supp. 44, 51-52.
[72] Kosseff, *supra*, fn. 64, at pp. 221-222.
[73] Manjoo, *Jurassic Web*, Slate (Feb. 9, 2009), https://slate.com/technology/2009/02/the-unrecognizable-internet-of-1996.html (last visited Apr. 5, 2022).
[74] DataReportal, *Digital 2022: The United States of America* (Feb. 2022), https://datareportal.com/reports/digital-2022-united-states-of-america (last visited Apr. 5, 2022).
[75] DataReportal, *Digital 2022: Global Overview Report* (Feb. 2022), https://datareportal.com/reports/digital-2022-global-overview-report (last visited Apr. 5, 022).

media each day.[76] The sheer volume of activity on social media makes it impossible for social media platforms to review and consider every post, so many platforms rely on artificial intelligence for most of their content moderation—with mixed results.[77]

The irony of the prevalence of online threats and violence is that the early internet was lauded as a "world where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity."[78] It appears that somewhere along the line, the breadth and scope of social media sites made truly effective content moderation impossible, and an absolutist interpretation of Section 230 prevents states from serving as a backstop for the most vile online statements.

Elected and governmental officials from both sides of the aisle, as well as many scholars and whistleblowers, have argued that Section 230 needs to be updated to properly regulate the internet as it is now, not the internet as it was over 25 years ago.[79] Legislation to reform Section 230—e.g., by allowing platform liability when the platform recommends content that result in user harm,[80] or by treating a platform as a content provider when it promotes content via certain algorithms[81]—and to repeal it entirely[82] is pending before Congress, though the prospects of any Section 230-related legislation are unclear. The delicate balance between protecting social media platform users and protecting vital free speech makes this issue difficult to resolve.[83]

---

[76] *Ibid.*
[77] *E.g.*, Seetharaman, Horwitz, & Scheck, *Facebook Says AI Will Clean Up the Platform. Its Own Engineers Have Doubts.*, Wall Street Journal (Oct. 17, 2021), https://www.wsj.com/articles/facebook-ai-enforce-rules-engineers-doubtful-artificial-intelligence-11634338184 (last visited Apr. 5, 2022).
[78] Barlow, *A Declaration of the Independence of Cyberspace* (Feb. 8, 1996), https://www.eff.org/cyberspace-independence (last visited Apr. 5, 2022).
[79] *See, e.g.*, Jurecic, Report, *The politics of Section 230 reform: Learning from FOSTA's mistakes*, The Brookings Institute (Mar. 1, 2022), https://www.brookings.edu/research/the-politics-of-section-230-reform-learning-from-fostas-mistakes/ (last visited Apr. 6, 2022) (Section 230 "is one of the few issues on which President Biden and his predecessor agree: President Trump tweeted dozens of times for § 230 to be repealed, and Biden commented during the 2020 Democratic primary should be done away with"); Citron & Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 Fordham L. Rev. 401, 403 (Nov. 2017).
[80] H.R. No. 5596—117th Congress (2021-2022).
[81] H.R. No. 2154—117th Congress (2021-2022).
[82] S.B. No. 2972—117th Congress (2021-2022).
[83] In 2018, Congress passed legislation amending Section 230 known as FOSTA or SESTA/FOSTA (P.L. 116-164 (2018)), which was intended to intended to make it easier to prosecute child sex trafficking. (Kosseff, *supra*, fn. 64, at pp. 269-270.) In practice, however, FOSTA did the opposite: the Government Accountability Office found that FOSTA made it more difficult for law enforcement to gather information about actual sex trafficking, and the main effect of FOSTA was to cause online platforms to shut down pages featuring legitimate activities for fear of liability. (United States Government Accountability Office, Report to Congressional Committees, *Sex Trafficking: Online Platforms and Federal Prosecutions*, No. 21-385 (June 2021), pp. 20-25.)

b. Preemption concerns regarding this bill

As noted above, Section 230 contains a preemption clause that prohibits any state law from holding an online platform liable as the publisher or speaker of a statement.[84] The United States Court of Appeals for the Ninth Circuit has held that Section 230 does not grant online platforms a general immunity for third-party content, but instead "protects from liability (1) a provider or user of an interactive computer service (2) whom a plaintiff seeks to treat, under a state law cause of action, as a publisher or speaker (3) of information provided by another information content provider."[85] The California Supreme Court has taken a different tack, holding that Section 230 provides blanket immunity to online platforms for third-party content; a majority of justices, however, has declined to hold that an injunction requiring the removal of third-party content automatically runs afoul of Section 230.[86]

This bill does not require that social media platforms moderate their platforms. Nor does it treat a social media platform as the publisher or speaker in the traditional sense, because the bill does not allow the target to seek redress for any pain and suffering, or reputational harm. This bill simply provides a safety valve for the worst of the worst speech on social media, by giving the targets of violent threats and incitements to violence—posts that serve no purpose other than to terrify and harm their recipients—the right to have the violence removed from the platform. This action could be brought in conjunction with a motion for a restraining order against an abuser, or as one step in an eventual prosecution for criminal threats, or as a standalone measure.

To the extent that the current case law interpreting Section 230 could be interpreted to preempt a straightforward state measure to protect its residents from the harm of online violence that lies outside the protection of the First Amendment, perhaps SB 1056's narrow reach would warrant a reexamination of the language of Section 230. Section 230 should not render a state powerless to protect its residents who are subjected to online threats of violence.

SUPPORT

None known

OPPOSITION

None known

---

[84] 47 U.S.C. § 230(e)(3).
[85] *Barnes*, *supra*, 570 F.3d at pp. 1100-1101.
[86] *See Hassell v. Bird* (2018) 5 Cal.5th 522, 527 (plur. opn. of Cantil-Sakauye, C.J.) (order to remove defamatory Yelp posts violated Section 230), 548 (conc. opn. of Krueger, J.) (concurring in the judgment but declining to rule on whether injunctions to remove content always violate Section 230), 560-561 (dis. opn. of Liu, J.), 566-567 (dis. opn. of Cuéllar, J.).

RELATED LEGISLATION

Pending Legislation:

SB 1390 (Pan, 2022) prohibits a social media platform, as defined, from recommending content to persons who did not specifically seek it out if the content is "harmful," defined as libel, slander, threats of imminent violence against government entities, disinformation, and misinformation; and requires a social media platform to establish a process for users to report improperly amplified harmful content and a database that tracks complaints about harmful content that must be submitted to the Attorney General. SB 1390 is pending before the Senate Judiciary Committee.

AB 587 (Gabriel, 2021) requires social media companies, as defined, to post their terms of service and report certain information to the Attorney General on a quarterly basis. AB 587 is pending before the Senate Judiciary Committee.

Prior Legislation:

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. SB 388 died in the Senate Judiciary 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.

AB 613 (Cristina Garcia, 2021) would have required a social media platform, as defined, or users or advertisers posting on a social media platform, to place text or marking within or adjacent to retouched images that have been posted on the platform for promotional or commercial purposes, and specify how that retouched image was altered. AB 613 died in the Assembly Privacy and Consumer Protection Committee.

AB 35 (Chau, 2021) would have required social media platforms, as defined, to disclose in an easy-to-find location whether they have a policy to combat misinformation, and imposed civil penalties for the failure to comply. AB 35 was pending before the Senate Judiciary Committee when the author left the Assembly.

SB 890 (Pan, 2020) would have required social media companies to remove images and videos depicting crimes, as specified, and imposed civil penalties for failing to do so. SB 890 died in the Senate Judiciary Committee.

AB 2391 (Gallagher, 2020) would have prohibited social media sites from removing user-posted content on the basis of the political affiliation or viewpoint of that content, except where the social media site is, by its terms and conditions, limited to the promotion of only certain viewpoints and values and the removed content conflicts with those viewpoints or values. AB 2931 died in the Assembly Committee on Arts, Entertainment, Sports, Tourism, and Media.

AB 2442 (Chau, 2020) was substantially similar to AB 35 (Chau, 2020) and would have required social media companies to disclose the existence, or lack thereof, of a misinformation policy, and imposed civil penalties for failing to do so. AB 2442 died in the Senate Judiciary Committee.

AB 1316 (Gallagher, 2019) would have prohibited social media sites from removing user-posted content on the basis of the political affiliation or viewpoint of that content, except where the social media site is, by its terms and conditions, limited to the promotion of only certain viewpoints and values and the removed content conflicts with those viewpoints or values. AB 1316 was held on the floor of the Assembly and was re-introduced as AB 2931 (2020).

AB 288 (Cunningham, 2019) would have required a social networking service, at the request of a user, to permanently remove personally identifiable information and not sell the information to third parties, within a commercially reasonable time of the request. AB 288 died in the Assembly Committee on Privacy and Consumer Protection.

SB 1424 (Pan, 2018) would have established a privately funded advisory group to study the problem of the spread of false information through Internet-based social media platforms and draft a model strategic plan for Internet-based social media platforms to use to mitigate this problem. SB 1424 was vetoed by Governor Edmund J. Brown, whose veto message stated that, as evidenced by the numerous studies by academic and policy groups on the spread of false information, the creation of a statutory advisory group to examine this issue is not necessary.

AB 3169 (Gallagher, 2018) would have prohibited social media sites from removing content on the basis of the political affiliation or viewpoint of the content, and prohibited internet search engines from removing or manipulating content from search results on the basis of the political affiliation or viewpoint of the content. AB 3169 died in the Assembly Committee on Privacy and Protection.

****