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

SB 981 (Wahab)
Version: April 22, 2024
Hearing Date: April 30, 2024
Fiscal: No
Urgency: No
CK

SUBJECT

Sexually explicit digital images

DIGEST

This bill requires social media platforms to provide a mechanism for reporting "digital identity theft," essentially the posting of nonconsensual, sexual deepfakes. The bill requires platforms to timely respond and investigate and to block instances of this material, as provided.

EXECUTIVE SUMMARY

With the technological advances in digital editing capabilities, the use of such tools to alter audiovisual work to portray individuals in various states of undress and/or engaging in sexually explicit conduct in which they are not actually performing has become an increasingly widespread issue. The growth of generative artificial intelligence capabilities has intensified the incidence and impact of nonconsensual, sexual deepfakes.

This bill addresses the appearance of these deepfakes on social media platforms. The platforms are required to create a mechanism for persons depicted in this "covered material" to report it to the platform. The platform is required to communicate with the reporting person and make a determination of whether the reported material amounts to "digital identity theft," essentially the posting of these nonconsensual deepfakes on the platform. Social media platforms are required to temporarily block content as soon as it is reported through this mechanism. If the platform determines it is, in fact, "digital identity theft, they must permanently block that instance of the material.

This bill is author-sponsored. No timely support was received by the Committee. The bill is opposed by the Electronic Frontier Foundation.

PROPOSED CHANGES TO THE LAW

Existing federal law:

1) Provides that no provider or user of a website shall be treated as the publisher or speaker of any information provided by another information content provider, and that no provider of a website shall be held liable on account of any action voluntarily taken in good faith to restrict the availability of materials that the provider determines to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected. (47 U.S.C. § 230(c) (Section 230).)

2) 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 Section 230. (47 U.S.C. § 230(e).)

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

4) Recognizes certain judicially created exceptions to the rights of freedom of speech and expression. (*E.g., Virginia v. Black* (2003) 538 U.S. 343, 359.)

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 them to interact socially with each other within the service or application. (A service or application that provides email or direct messaging services does not meet this criterion based solely on that function.)
    b) The service or application allows users to do all of the following:
        i. Construct a public or semipublic profile for purposes of signing into and using the service or application.
        ii. Populate a list of other users with whom an individual shares a social connection within the system.
        iii. Create or post content viewable by other users, including 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(e).)

2) Defines "deepfake" to mean audio or visual content that has been generated or manipulated by artificial intelligence (AI) which would falsely appear to be

authentic or truthful and which features depictions of people appearing to say or do things they did not say or do without their consent. Requires the Secretary of Government Operations to evaluate the impact of the proliferation of deepfakes on the state. (Gov. Code § 11547.5.)

3) Authorizes a depicted individual to bring a cause of action against a person who does either of the following:
    a) creates and intentionally discloses sexually explicit material and the person knows or reasonably should have known the depicted individual in that material did not consent to its creation or disclosure; or
    b) intentionally discloses sexually explicit material that the person did not create and the person knows the depicted individual in that material did not consent to the creation of the sexually explicit material. (Civ. Code § 1708.86.)

4) Defines, for the preceding statute, "depicted individual" as an individual who appears, as a result of digitization, to be giving a performance they did not actually perform or to be performing in an altered depiction. (Civ. Code § 1708.86.)

5) Prohibits a person who intentionally distributes the image of the intimate body part or parts of another identifiable person, or an image of the person depicted engaged in an act of sexual intercourse, sodomy, oral copulation, sexual penetration, or an image of masturbation by the person depicted or in which the person depicted participates, under circumstances in which the persons agree or understand that the image shall remain private, the person distributing the image knows or should know that distribution of the image will cause serious emotional distress, and the person depicted suffers that distress. (Pen. Code § 647(j)(4)(A).)

This bill:

1) Requires a social media platform to do all of the following:
    a) Provide a mechanism that is reasonably accessible to reporting persons to report digital identity theft to the social media platform.
    b) Collect information reasonably sufficient to enable the social media platform to locate the instance of digital identity theft and to contact a reporting person with both of the following:
        i. Confirmation that the social media platform received the reporting person's report within 48 hours of receipt of the report.
        ii. Within seven days of the date on which the confirmation is issued, a written update to the reporting person as to the status of the platform's handling of the reported digital identity theft.

c) Determine within 14 days of the date on which the confirmation is issued whether the reported digital identity theft is digital identity theft.
d) Temporarily block a reported instance of digital identity theft from being publicly viewable on the platform pending the above determination.
e) Permanently block an instance of digital identity theft from being publicly viewable on the social media platform.
f) Make reasonable efforts to remove and block unreported instances of digital identity theft from being publicly viewable on the platform.

2) Defines the relevant terms, including:
   a) "Covered material" means material that a reporting person reasonably believes meets all of the following criteria:
      i. The material is an image or video created or altered through digitization that would appear to a reasonable person to be an image or video of any of the following:
         1. An intimate body part of an identifiable person.
         2. An identifiable person engaged in an act of sexual intercourse, sodomy, oral copulation, or sexual penetration.
         3. An identifiable person engaged in masturbation.
      ii. The reporting person identifies the reporting person as the person depicted in the material and confirms that the reporting person did not consent to the use of the reporting person's likeness in the material.
      iii. The material is displayed, stored, or hosted on the social media platform.
   b) "Digital identity theft" means the posting of covered material on a social media platform.
   c) "Reporting person" means a natural person located within the geographic boundaries of the state who reports material to a social media platform using the mechanism provided by the social media platform.
   d) "Social media platform" has the same meaning as defined in Section 22675 of the Business and Professions Code. However, it does not include specified messaging services or internet-based services or applications owned or operated by nonprofit organizations, as provided.

COMMENTS

1. Stated intent of the bill

According to the author:

> Identity theft is a serious crime that continues to evolve. When an individual digitally alters images with the intent to distribute them on social media—and in order to create a false characterization of the

featured individual—it is identity theft. This form of identity theft harms the victim in a variety of ways by creating a tarnished reputation, financial ruin, professional ruin, and more. This harm is especially significant when the digitally altered images are non-consensual sexually explicit material. Social Media platforms do not make it easy for users to request the removal of non-consensual sexually explicit images. The distribution of these non-consensual images can contribute to mental health issues, decreased socialization, and increased online harassment.

It is imperative organizations take action to protect individuals from these forms of 21st century identity theft, but also have basic mechanisms for the reporting and removal of these non-consensual images, pages, and/or accounts.

Currently, there is no 24-hour hotline or number to call nor other alternative means of seeking recourse. Delays between a user reporting an image, and the image being removed create harm and perpetuate toxic online environments.

We deserve better, and we should expect these social media platforms to do better. Through collaboration between users, platforms, and lawmakers, this bill seeks to create a safer online environment by ending the spread of non-consensual sexually explicit material.

2. Combating deepfakes

This bill aims to address the growing concerns associated with what are called "deepfakes," a term drawn from "deep learning" plus "fake." There are various manifestations, but essentially all involve the digital manipulation of audiovisual material to falsely depict an individual engaging in certain conduct. One privacy expert describes the problem relevant here: "Although political deepfakes are relatively new, pornographic deepfakes have been a problem for some time. These often purport to show a famous actress or model involved in a sex act but actually show the subject's face superimposed onto another woman's body."[1]

With the rapid advancement in generative artificial intelligence (GenAI), deepfake images have become exponentially easier to produce, exacerbating the problem:

[1] Nicholas Schmidt, *Privacy law and resolving 'deepfakes' online* (Jan. 30, 2019) IAPP, https://iapp.org/news/a/privacy-law-and-resolving-deepfakes-online/. All further internet citations are available as of Apr. 22, 2024.

> Recent improvements in artificial intelligence software have made it surprisingly easy to graft the heads of stars, and ordinary women, to the bodies of X-rated actresses to create realistic videos.
>
> These explicit movies are just one strain of so-called "deepfakes," which are clips that have been doctored so well they look real. Their arrival poses a threat to democracy; mischief makers can, and already have, used them to spread fake news. But another great danger of deepfakes is their use as a tool to harass and humiliate women.[2]

Infamously, in January of this year, Taylor Swift was the victim of sexually explicit, nonconsensual deepfake images using GenAI that were widely spread across social media platforms.[3] Perhaps more disturbingly, a trend has emerged in schools of students creating such images: "At schools across the country, people have used deepfake technology combined with real images of female students to create fraudulent images of nude bodies. The deepfake images can be produced using a cellphone."[4]

3. Targeting deepfakes and "digital identity theft" on social media

This bill aims to address this problem by ensuring there are tools to combat the appearance of these nonconsensual, sexual deepfakes on social media platforms. It requires social media platforms to provide a mechanism for reporting "digital identity theft" to them. "Digital identity theft" is defined as the posting of covered material on a social media platform. "Covered material" is defined as material that the reporting person reasonably believes meets several criteria. First, that the content is a digitized image or video that appears to be a person's intimate body part or the person engaged in sexually explicit conduct. Second, the reporting person is the one depicted in the material without their consent. And finally, that the material is displayed, stored, or hosted on the social media platform.

Platforms are required to confirm they received a report through this mechanism within 48 hours and to collect sufficient information for them to find the reported instance of digital identity theft. Within a week, they are to provide the reporting person an update and then at two weeks must make a determination if the content is in fact digital identity theft.

---

[2] Jeff John Roberts, *Fake Porn Videos Are Terrorizing Women. Do We Need a Law to Stop Them?* (Jan. 15, 2019) Fortune, http://fortune.com/2019/01/15/deepfakes-law/.
[3] Brian Contreras, *Tougher AI Policies Could Protect Taylor Swift – And Everyone Else – From Deepfakes* (February 8, 2024) Scientific American, https://www.scientificamerican.com/article/tougher-ai-policies-could-protect-taylor-swift-and-everyone-else-from-deepfakes/.
[4] Hannah Fry, Laguna Beach High School investigates 'inappropriate' AI-generated images of students (April 2, 2024) Los Angeles Times, https://www.latimes.com/california/story/2024-04-02/laguna-beach-high-school-investigating-creation-of-ai-generated-images-of-students.

The bill also requires social media platforms to block this content from being publicly viewable on their platform both temporarily, pending the platform's determination, and then permanently, if it is an instance of digital identity theft. Platforms must also make reasonable efforts to remove and block unreported instances of digital identity theft on their platforms.

One concern is that the definition of "covered material" includes what the reporting person "reasonably believes" is included in the material, not what the material actually is. To ensure the bill is getting at actual nonconsensual deepfakes, the author has agreed to the following amendments that take that phrase, and other subjective elements, out of the definition:

> 22670. (a) "Covered material" means material ~~that a reporting person reasonably believes~~ that meets all of the following criteria:
> (1) The material is an image or video created, created or altered through, through digitization that would appear to a reasonable person to be an image or video of any of the following:
> (A) An intimate body part of an identifiable person.
> (B) An identifiable person engaged in an act of sexual intercourse, sodomy, oral copulation, or sexual penetration.
> (C) An identifiable person engaged in masturbation.
> (2) The reporting person is ~~identifies the reporting person as~~ the person depicted in the material and ~~confirms that~~ the
> reporting person did not consent to the use of the reporting person's likeness in the material.
> (3) The material is displayed, stored, or hosted on the social media platform.

4. Legal considerations

As with most of the legislation seeking to govern content on social media, legal questions arise around whether the specific approach of any proposed law runs afoul of the First Amendment or is preempted by Section 230.

a. First Amendment

The First Amendment, as applied to the states through the Fourteenth Amendment, prohibits Congress or the states from passing any law "abridging the freedom of speech."[5] "[A]s a general matter, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content."[6] However, while the amendment is written in absolute terms, the courts have

[5] U.S. Const., 1st & 14th amends.
[6] *Ashcroft v. American Civil Liberties Union* (2002) 535 U.S. 564, 573.

created a handful of narrow exceptions to the First Amendment's protections, including "true threats,"[7] "fighting words,"[8] incitement to imminent lawless action,[9] defamation,[10] and obscenity.[11]

Expression on the internet is given the same measure of protection granted to in-person speech or statements published in a physical medium.[12] Accordingly, a social media user may generally post content and comments free from government regulation, but may incur civil or criminal liability if their comment falls within one of the First Amendment exceptions. At the same time, social media platforms themselves—as private businesses—are not subject to the constraints of the First Amendment and may limit or prohibit users' speech on their sites as they see fit.[13]

The United States Supreme Court has held that posting on social networks and/or social media sites constitutes communicative activity protected by the First Amendment.[14] As a general rule, the government "may not suppress lawful speech as the means to suppress unlawful speech."[15]

A constitutional challenge to a restriction on speech is generally analyzed under one of two frameworks, depending on whether the courts deem it to be "content neutral" or "content based," i.e., targeting a particular type of speech. A law is content neutral when it "serves purposes unrelated to the content of the expression."[16] On the other hand, a law is content based when the proscribed speech is "defined solely on the basis of the content of the suppressed speech."[17]

If a law is determined to be content neutral it will be subject to intermediate scrutiny, which requires that the law "be 'narrowly tailored to serve a significant government interest.' "[18] 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

---

[7] *Snyder v. Phelps* (2011) 562 U.S. 443, 452.
[8] *Cohen v. California* (1971) 403 U.S. 15, 20.
[9] *Virginia v. Black* (2003) 538 U.S. 343, 359.
[10] *R.A.V. v. St. Paul* (1992) 505 U.S. 377, 383.
[11] *Ibid.*
[12] *Reno v. ACLU* (1997) 521 U.S. 844, 870.
[13] *E.g.*, *Hudgens v. NLRB* (1976) 424 U.S. 507, 513. Some have argued that certain social media platforms are so essential to the freedom of expression that they should be treated as common carriers subject to the First Amendment.
[14] *E.g.*, *Packingham v. North Carolina* (2017) 137 S.Ct. 1730, 1735-1736.
[15] *Ashcroft v. Free Speech Coalition* (2002) 535 U.S. 234, 255; see also *United States v. Alvarez* (2012) 567 U.S. 709, 717 (Supreme Court "has rejected as 'startling and dangerous' a 'free-floating test for First Amendment coverage…[based on] an ad hoc balancing of relative social costs and benefits' " [alterations in original]).
[16] *Ward v. Rock Against Racism* (1989) 491 U.S. 781, 791.
[17] *FCC v. League of Women Voters* (1984) 468 U.S. 364, 383.
[18] *Packingham*, *supra*, 137 S.Ct. at p. 1736.

such a manner that a substantial portion of the burden on speech does not serve to advance its goals.'"[19]

If a restriction on speech is determined to be content based, it will be subject to strict scrutiny.[20] A restriction is content based "if it require[s] 'enforcement authorities' to 'examine the content of the message that is conveyed to determine whether' a violation has occurred."[21] Content-based restrictions subject to strict scrutiny are "presumptively unconstitutional."[22] A restriction can survive strict scrutiny only if it uses the least-restrictive means available to achieve a compelling government purpose.[23]

Writing in support of past legislation seeking to outlaw nonconsensual sexually explicit deepfakes, SB 564 (Leyva, 2019), noted First Amendment scholar Erwin Chemerinsky argues that this type of "speech" is not expression that has historically received First Amendment protection:

> While the content regulated by this statute may be "speech" insofar as the First Amendment is concerned, it is a longstanding precept of First Amendment doctrine that "not all speech is of equal First Amendment importance." *See, e.g., Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc.*, 472 U.S. 749, 758-759 (1985). And it is likewise true that "speech on matters of purely private concern is of less First Amendment concern." *Id.* at 759. The U.S. Supreme Court has correctly recognized that "sexual behavior" is "the most private human conduct." *Lawrence v. Texas*, 539 U.S. 558, 567 (2003). By regulating the nonconsensual use of an individual's persona in sexually explicit audiovisual works, Senate Bill 564 does not target any forms of expression that have historically received First Amendment protection.

He goes on to assert that protecting victims from the "serious reputational and economic harms" that can result from this conduct more than justifies regulating this false and misleading material that arguably lacks social value.

It should be noted that the bill requires material to be "temporarily blocked" pending a determination by the social media platform of whether the content amounts to "digital identity theft." Therefore, the bill requires content to be taken down immediately based solely on a report; this will likely result in protected speech on social media being blocked by this statute, at least temporarily. The bill also requires examination of whether or not the underlying material is digital identity theft, therefore it is likely to be considered content-based, notwithstanding that nonconsensual, sexual deepfakes are entitled to very little First Amendment protection, if any.

---

[19] *McCullen v. Coakley* (2014) 573 U.S. 464, 486 (*McCullen*).
[20] *Id.* at p. 478.
[21] *Id.* at p. 479.
[22] *Reed v. Town of Gilbert* (2015) 135 S.Ct. 2218, 2226 (*Reed*).
[23] *United States v. Playboy Entertainment Group* (2000) 529 U.S. 803, 813.

Writing in an oppose unless amended position, Oakland Privacy raises concerns with the "permanently block language":

> A change that needs to be made to Senate Bill [981] is to remove the term "permanently block" from the bill language. We completely understand the intent of the author to avoid content popping up in other places after being removed. In fact, this exact same conversation ensued in 2023 with a bill from Assemblymember Wicks which focused on child sexual abuse material (AB 1394).
>
> That bill's "permanently block" language was amended to specify that the language refers only to the reported instance of the content and only that reasonable efforts be made to remove other instances of the content. This clarification is necessary because it is not necessarily possible to "permanently block" content that can be reposted and attempts to try to do so can be problematic.

To ensure that the requirement for permanently blocking content is more narrowly tailored to its goal, the author has agreed to amend the bill to require the platform to permanently block only when it determines there is a reasonable basis to believe reported content is digital identity theft.

> Amendment
>
> Amend Section 22671(c) and (d) as follows:
>
> (c) Determine within 14 days of the date on which the confirmation required by paragraph (1) of subdivision (b) is issued whether there is a reasonable basis to believe the reported digital identity theft is digital identity theft.
>
> (d) (1) Temporarily block a reported instance of digital identity theft from being publicly viewable on the social media platform pending a determination pursuant to subdivision (c).
> (2) Permanently block a reported instance of digital identity theft from being publicly viewable on the social media platform if the social media platform determines there is a reasonable basis to believe the reported digital identity theft is digital identity theft.

With the amendments discussed here and below that the author has agreed to accept, Oakland Privacy has agreed to go neutral.

b. Conflict with Section 230 of the Communications Decency Act, 47 U.S.C. § 230

In addition to the First Amendment, the other primary source governing content on social media is Section 230. Section 230 does not apply to the *users* of social media (or the internet generally), but rather applies to the *platforms themselves*. In the early 1990s, prior to the enactment of Section 230, two trial court orders—one in the United States District Court for the Southern District of New York, and New York state court—suggested that internet platforms could be held liable for allegedly defamatory statements made by the platforms' users if the platforms engaged in any sort of content moderation (e.g., filtering out offensive material).[24] In response, two federal legislators and members of the burgeoning internet industry crafted a law that would give internet platforms immunity from liability for users' statements, even if they might have reason to know that the statements might be false, defamatory, or otherwise actionable.[25] The result—Section 230—was relatively uncontroversial at the time, in part because of the relative novelty of the internet and in part because Section 230 was incorporated into a much more controversial internet regulation scheme that was the subject of greater debate.[26]

Section 230 begins with findings and a statement of policy that extol the value of the internet and the intention to let the internet develop without significant government regulation.[27] The crux of Section 230 is then laid out in two parts. The first provides that "[n]o 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] The second provides a safe harbor for content moderation, by 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 protected."[29] Together, these two provisions give platforms immunity from any civil or criminal liability that could be incurred by user statements, while explicitly authorizing platforms to engage in their own content moderation without risking that immunity.

---

[24] *See Cubby, Inc. v. Compuserve, Inc.* (S.D.N.Y. 1991) 776 F.Supp. 135, 141; *Stratton Oakmont v. Prodigy Servs. Co.* (N.Y. Sup. Ct., May 26, 1995) 1995 N.Y. Misc. LEXIS 229, *10-14. These opinions relied on case law developed in the context of other media, such as whether book stores and libraries could be held liable for distributing defamatory material when they had no reason to know the material was defamatory. (*See Cubby, Inc.*, 776 F. Supp. at p. 139; *Smith v. California* (1959) 361 U.S. 147, 152-153.)

[25] Kosseff, The Twenty-Six Words That Created The Internet (2019) pp. 57-65.

[26] *Id.* at pp. 68-73. Section 230 was added to the Communications Decency Act of 1996 (title 5 of the Telecommunications Act of 1996, Pub. L. 104-104, 110 Stat. 56), which would have imposed criminal liability on internet platforms if they did not take steps to prevent minors from obtaining "obscene or indecent" material online. The Supreme Court invalidated the CDA, except for Section 230, on the basis that it violated the First Amendment. (*See Reno*, *supra*, 521 U.S. at p. 874.)

[27] 47 U.S.C. § 230(a) & (b).

[28] *Id.*, § 230(c)(1).

[29] *Id.*, § 230(c)(1) & (2).

Section 230 specifies that it does not preempt federal criminal laws, but 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."[30]

Section 230 uses terminology generally applicable in defamation cases (e.g., "publisher," "speaker"), but courts interpreting Section 230 did not limit its application to the defamation context. Instead, courts have applied Section 230 in a vast range of cases to immunize internet platforms from "virtually all suits arising from third-party content."[31] Courts have 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.[32]

A brief look at recent, relevant case law is necessary to assess these Section 230 concerns.

First, the Ninth Circuit Court of Appeals in *Barnes v. Yahoo!, Inc.* (9th Cir. 2009) 570 F.3d 1096, 1100-01 established the prevailing three-part test for certain claims pursuant to Section 230: "[I]t appears that subsection (c)(1) only 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."

The United States Supreme Court in May 2023 issued opinions in *Gonzalez v. Google LLC* (2023) 143 S. Ct. 1191, and *Twitter, Inc. v. Taamneh* (2023) 143 S. Ct. 1206, a pair of highly anticipated decisions, which presented the highest court with the task of determining the scope of Section 230's protective shield and the valid bases for holding platforms liable for content and conduct carried out on their platforms. The cases below were brought by the families of several victims of ISIS attacks in various parts of the world. The defendants were several social media platforms. Relevant here, the Court in *Taamneh* found:

> In this case, the failure to allege that the platforms here do more than transmit information by billions of people—most of whom use the platforms for interactions that once took place via mail, on the phone, or in public areas—is insufficient to state a claim that defendants knowingly gave substantial assistance and thereby aided and abetted ISIS' acts. A contrary conclusion would effectively hold any sort of communications provider liable for any sort of wrongdoing merely for knowing that the

[30] *Id.*, § 230(e)(1) & (3).
[31] Kosseff, *supra*, fn. 13, 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.
[32] *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.

> wrongdoers were using its services and failing to stop them. That would run roughshod over the typical limits on tort liability and unmoor aiding and abetting from culpability.[33]

Although the case law in this area is ever-changing, the bill will likely face challenge as it exposes social media platforms to liability for failing to take down certain content posted by users. It also potentially exposes them to liability for failing to take reasonable steps to remove and block unreported instances posted by users. Writing in opposition, the Electronic Frontier Foundation argues the bill is preempted by Section 230:

> While S.B. 981 does not itself impose civil liability, any platform perceived to have failed to comply will find themselves subject to a suit. This imposes a cost on platforms, making S.B. 981 directly in conflict with Section 230's immunity: "no cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this Section." 47 U.S.C. Sec. 230 (e)(3).

SUPPORT

None received

OPPOSITION

Electronic Frontier Foundation
Oakland Privacy

RELATED LEGISLATION

Pending Legislation:

AB 3172 (Lowenthal, 2024) makes social media platforms liable for specified damages in addition to any other remedy provided by law, if the platform fails to exercise ordinary care or skill toward a child. AB 3172 is currently in the Assembly Judiciary Committee.

SB 646 (Cortese, 2023) creates liability for the distribution of certain "actionable material," which includes illicit pictures of minors and images or depictions of minors that serve as the basis for criminal and civil liability at the federal level. SB 646 is currently in the Assembly Appropriations Committee.

[33] *Taamneh*, 143 S. Ct. at 1213.

Prior Legislation:

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

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

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

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

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

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

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

****