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

AB 587 (Gabriel)
Version: June 23, 2022
Hearing Date: June 28, 2022
Fiscal: Yes
Urgency: No
CK

SUBJECT

Social media companies: terms of service

DIGEST

This bill requires social media companies, as defined, to post their terms of service and to submit quarterly reports to the Attorney General on their terms of service and content moderation policies and outcomes.

EXECUTIVE SUMMARY

In 2005, five percent of adults in the United States used social media. In just six years, that number jumped to half of all Americans. Today, over 70 percent of adults use at least one social media platform. Facebook alone is used by 69 percent of adults, and 70 percent of those adults say they use the platform on a daily basis.

Given the reach of social media platforms and the role they play in many people's lives, concerns have arisen over what content permeates these sites, entering the lives of the billions of users, and the effects that has on them and society as a whole. In particular, the sharpest calls for action focus on the rampant spread of misinformation, hate speech, and sexually explicit content. Social media companies' content moderation of a decade ago involved handfuls of individuals and user policies were minimal. These programs and policies have dramatically evolved over the years but the proliferation of objectionable content and "fake news" has led to calls for swifter and more aggressive action in response. However, there has also been backlash against perceived censorship in response to filtering of content and alleged "shadow banning."

This bill requires social media companies, as defined, to publicly post the terms of service, with certain required elements, for their social media platforms and to provide the Attorney General with a quarterly report on their content moderation procedures and outcomes.

This bill is sponsored by the Anti-Defamation League. It is supported by a variety of groups, including Common Sense and the Islamic Networks Group. It is opposed by various technology and business associations, including the California Chamber of Commerce and TechNet.

PROPOSED CHANGES TO THE LAW

Existing law:

1) Prohibits, through the United States Constitution, the enactment of any law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances. (U.S. Const. Amend. 1.)

2) Provides, through the California Constitution, for the right of every person to freely speak, write, and publish their sentiments on all subjects, being responsible for the abuse of this right. Existing law further provides that a law may not restrain or abridge liberty of speech or press. (Cal. Const., art. I, § 2(a).)

3) Provides, in federal law, that a provider or user of an interactive computer service shall not be treated as the publisher or speaker of any information provided by another information content provider. (47 U.S.C. § 230(c)(2).)

4) Provides that a provider or user of an interactive computer service shall not be held liable on account of:
    a) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
    b) any action taken to enable or make available to information content providers or others the technical means to restrict access to such material. (47 U.S.C. § 230(c)(2).)

5) Defines "interactive computer service" as any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions. (47 U.S.C. § 230(f)(2).)

6) Establishes the Unfair Competition Law (UCL) and defines "unfair competition" to mean and include any unlawful, unfair, or fraudulent business act or practice and unfair, deceptive, untrue, or misleading advertising and any act prohibited

by Chapter 1 (commencing with Section 17500) of Part 3 of Division 7 of the Business and Professions Code. (Bus. & Prof. Code § 17200 et seq.)

7) Provides that any person who engages, has engaged, or proposes to engage in unfair competition may be enjoined. Any person may pursue representative claims or relief on behalf of others only if the claimant meets specified standing requirements and complies with Section 382 of the Code of Civil Procedure, but these limitations do not apply to claims brought under this chapter by the Attorney General, or any district attorney, county counsel, city attorney, or city prosecutor in this state. (Bus. & Prof. Code § 17203.)

8) Requires actions for relief pursuant to the UCL be prosecuted exclusively in a court of competent jurisdiction and only by the following:
    a) the Attorney General;
    b) a district attorney;
    c) a county counsel authorized by agreement with the district attorney in actions involving violation of a county ordinance;
    d) a city attorney of a city having a population in excess of 750,000;
    e) a city attorney in a city and county;
    f) a city prosecutor in a city having a full-time city prosecutor in the name of the people of the State of California upon their own complaint or upon the complaint of a board, officer, person, corporation, or association with the consent of the district attorney; or
    g) a person who has suffered injury in fact and has lost money or property as a result of the unfair competition. (Bus. & Prof. Code § 17204.)

9) Holds any person who engages, has engaged, or proposes to engage in unfair competition liable for a civil penalty not to exceed $2,500 for each violation, which shall be assessed and recovered in a civil action brought by the Attorney General, or other public prosecutors. (Bus. & Prof. Code § 17206(a).)

10) Prohibits false or deceptive advertising to consumers about the nature of any property, product, or service, including false or misleading statements made in print, over the internet, or any other advertising method. (Bus. & Prof. Code § 17500.)

11) Defines libel as a false and unprivileged publication by writing, printing, or any other representation that exposes any person to hatred, contempt, ridicule, or obloquy, which causes that person to be shunned or avoided, or which has a tendency to injure that person in their occupation. (Civ. Code §§ 45, 47.)

12) Requires certain businesses to disclose the existence and details of specified policies, including:

a) Operators of commercial websites or online services that collect personally identifiable information about individual consumers residing in California who use or visit the website must conspicuously post its privacy policy. (Bus. & Prof. Code § 22575.)
b) Retailers and manufacturers doing business in this state and having annual worldwide gross receipts over $100,000,000 must disclose online whether the business has a policy to combat human trafficking and, if so, certain details about that policy. (Civ. Code § 1714.43.)
c) End-users of automated license plate recognition technology must post its usage and privacy policy on its website. (Civ. Code § 1798.90.53.)
d) Campus bookstores at public postsecondary educational institutions must post in-store or online a disclosure of its retail pricing policy on new and used textbooks. (Educ. Code § 66406.7(f).)

This bill:

1) Requires a social media company to post terms of service for each social media platform owned or operated by the company in a manner reasonably designed to inform all users of the social media platform of the existence and contents of the terms of service. The terms of service shall include all of the following:
    a) contact information for the purpose of allowing users to ask the social media company questions about the terms of service;
    b) a description of the process that users must follow to flag content, groups, or other users that they believe violate the terms of service, and the social media company's commitments on response and resolution time; and
    c) a list of potential actions the social media company may take against an item of content or a user, including, but not limited to, removal, demonetization, deprioritization, or banning.

2) Requires the terms of service to be available in all Medi-Cal threshold languages, as defined, in which the social media platform offers product features, including, but not limited to, menus and prompts.

3) Requires social media companies to submit a terms of service report, quarterly, with the first report due July 1, 2022, to the Attorney General, who must post it on their website. The terms of service report must include, for each social media platform owned or operated by the company, all of the following:
    a) the current version of the terms of service of the social media platform;
    b) if a social media company has filed its first quarterly report, a complete and detailed description of any changes to the terms of service since the last quarterly report;
    c) a statement of whether the current version of the terms of service defines specified categories of content, and, if so, the definitions of those

categories, including any subcategories. This includes hate speech, racism, extremism, harassment, disinformation, and foreign political interference;

d) a detailed description of content moderation practices used by the social media company for that platform, including, but not limited to, all of the following:
    i. any policies intended to address the above categories of content;
    ii. how automated content moderation systems enforce terms of service and when these systems involve human review;
    iii. how the social media company responds to user reports of violations of the terms of service;
    iv. how the social media company would remove individual pieces of content, users, or groups that violate the terms of service, or take broader action against individual users or against groups of users that violate the terms of service; and
    v. the languages in which the social media platform does not make terms of service available, but does offer product features;
e) information on content that was flagged by the social media company as content belonging to any of the above categories, including the total number of all of the following:
    i. flagged items of content;
    ii. actioned items of content;
    iii. actioned items of content that resulted in action taken by the social media company against the user or users responsible;
    iv. actioned items of content that were removed, demonetized, or deprioritized by the social media company;
    v. times actioned items of content were viewed by users;
    vi. times actioned items of content were shared, and the number of users that viewed the content before it was actioned; and
    vii. times users appealed social media company actions taken on that platform and the number of reversals on appeal disaggregated by each action; and
f) all information required by (e) shall also be disaggregated into the category of content, the type of content, the type of media, and how the content was flagged and actioned.

4) Defines "social media company" as a person or entity that owns or operates one or more social media platforms, as defined. The bill exempts certain social media companies, including those making less than $100,000,000 in gross revenue during the preceding year.

5) Defines "actioned" to mean a social media company, due to a suspected or confirmed violation of the terms of service, has taken some form of action, including, but not limited to, removal, demonetization, deprioritization, or banning, against the relevant user or relevant item of content.

6) Defines "terms of service" as a policy or set of policies adopted by a social media company that specifies, at least, the user behavior and activities that are permitted on the internet-based service owned or operated by the social media company, and the user behavior and activities that may subject the user or an item of content to being actioned. This may include, but is not limited to, a terms of service document or agreement, rules or content moderation guidelines, community guidelines, acceptable uses, and other policies and established practices that outline these policies.

7) Subject companies in violation to penalties of up to $15,000 per violation per day to be sought by specified public prosecutors. The court is to assess the amount with consideration of whether the company made a reasonable, good faith attempt to comply. A social media company is in violation for each day it does any of the following:
   a) fails to post terms of service;
   b) fails to timely submit to the Attorney General the report required above; or
   c) materially omits or misrepresents required information in a submitted report.

8) Provides that the duties and obligations imposed are cumulative to any others imposed under local, state, or federal law. The remedies or penalties provided are also cumulative to each other and to any others available.

COMMENTS

1. Social media content

In recent years, the clamor for more robust content moderation on social media has reached a fever pitch. This includes calls to control disinformation or "fake news," hate speech, political interference, and other online harassment.

The 2016 election was a major breaking point for many. Investigations uncovered attempted interference in the United States Presidential election through a social media "information warfare campaign designed to spread disinformation and societal division in the United States."[1] The United States Senate Select Committee on Intelligence issued a report detailing how Russian operatives carried out their plan:

> Masquerading as Americans, these operatives used targeted advertisements, intentionally falsified news articles, self-generated

[1] Select Committee on Intelligence, Russian Active Measures, Campaigns, and Interference in the 2016 U.S. Election, United States Senate, https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf. All internet citations are current as of June 24, 2022.

> content, and social media platform tools to interact with and attempt to deceive tens of millions of social media users in the United States. This campaign sought to polarize Americans on the basis of societal, ideological, and racial differences, provoked real world events, and was part of a foreign government's covert support of Russia's favored candidate in the U.S. presidential election.

This again became a threat in the 2020 election, with social media rife with misinformation such as the incorrect election date,[2] and then social media became a hotbed of misinformation about the results of the election.[3] The author points to investigations that have found the violent insurrectionists that stormed the Capitol on January 6, 2021, were abetted and encouraged by posts on social media sites.[4] In response to indications that social media provided a venue for those who overran and assaulted police officers, Facebook deflected blame, asserting that "these events were largely organized on platforms that don't have our abilities to stop hate, don't have our standards, and don't have our transparency."[5] However, later indictments of those perpetrating the attack "made it clear just how large a part Facebook had played, both in spreading misinformation about election fraud to fuel anger among the Jan. 6 protesters, and in aiding the extremist militia's communication ahead of the riots."[6]

One area the author specifically focuses in on as motivation for the bill is the rise of hate speech online and the real world consequences. The author points to a recent study of over 500 million Twitter posts from 100 cities in the United States that found that "more targeted, discriminatory tweets posted in a city related to a higher number of hate crimes."[7]

---

[2] Pam Fessler, *Robocalls, Rumors And Emails: Last-Minute Election Disinformation Floods Voters*, NPR (October 24, 2020), https://www.npr.org/2020/10/24/927300432/robocalls-rumors-and-emails-last-minute-election-disinformation-floods-voters.

[3] Sheera Frenkel, *How Misinformation 'Superspreaders' Seed False Election Theories*, New York Times (November 23, 2020), https://www.nytimes.com/2020/11/23/technology/election-misinformation-facebook-twitter.html; Philip Bump, *The chain between Trump's misinformation and violent anger remains unbroken*, Washington Post (May 12, 2021), https://www.washingtonpost.com/politics/2021/05/12/chain-between-trumps-misinformation-violent-anger-remains-unbroken/.

[4] Ken Dilanian & Ben Collins, *There are hundreds of posts about plans to attack the Capitol. Why hasn't this evidence been used in court?* (April 20, 2021) NBC News, https://www.nbcnews.com/politics/justice-department/we-found-hundreds-posts-about-plans-attack-capitol-why-aren-n1264291.

[5] Sheera Frenkel & Cecilia Kang, *Mark Zuckerberg and Sheryl Sandberg's Partnership Did Not Survive Trump* (July 8, 2021) The New York Times, https://www.nytimes.com/2021/07/08/business/mark-zuckerberg-sheryl-sandberg-facebook.html.

[6] *Ibid.*

[7] Press Release, *Hate speech on Twitter predicts frequency of real-life hate crimes* (June 24, 2019) NYU Tanden School of Engineering, https://engineering.nyu.edu/news/hate-speech-twitter-predicts-frequency-real-life-hate-crimes.

Misinformation also poses a danger to public health: One study found that the more people rely on social media as their main news source, the more likely they are to believe misinformation about the COVID-19 pandemic.[8] Another found that a mere 12 people are responsible for 65 percent of the false and misleading claims about COVID-19 vaccines on Facebook, Instagram, and Twitter.[9] Misinformation hinders emergency responses to natural responses when social media posts contain incorrect or out-of-date information.[10]

The author frames the problem:

> Over the past several years, there has been growing concern around the role of social media in promoting hate speech, disinformation, conspiracy theories, violent extremism, and severe political polarization. If properly managed, the ability for social media to amplify ideas and messages that would otherwise lack widespread exposure can give voice to otherwise marginalized populations and improve the public discourse, but the same capacity can feed the propagation of misinformation and dangerous rhetoric.

Writing in support, the Anti-Defamation League, the sponsor of this bill, further explains the context of the bill:

> In recent years, there has been growing concern around the role of social media in promoting hate speech, disinformation, conspiracy theories, violent extremism, harassment, and severe political polarization. According to ADL's 2021 Online Hate and Harassment Survey, 41% of individuals experience online harassment and one in three of those individuals attribute at least some harassment to their identity. Identity-based harassment remains worrisome, affecting the ability of already marginalized communities to be safe in digital spaces.
>
> Importantly, this hate and harassment isn't only taking place in the dark corners of the internet. 75% of ADL's 2021 Online Hate and Harassment Survey respondents who were harassed said at least some harassment

---

[8] Yan Su, *It doesn't take a village to fall for misinformation: Social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation belief* (May 2021) Telematics and Information, Vol. 58, https://www.sciencedirect.com/science/article/abs/pii/S0736585320302069?via%3Dihub.

[9] Shannon Bond, *Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows* (May 14, 2021) NPR, https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes.

[10] United States Department of Homeland Security, *Countering False Information on Social Media in Disasters and Emergencies* (March 2018), https://www.dhs.gov/sites/default/files/publications/SMWG_Countering-False-Info-Social-Media-Disasters-Emergencies_Mar2018-508.pdf.

> happened on Facebook – and many also attributed harassment to other mainstream social media platforms. And online extremism is also front and center: Facebook's own researchers found that 64% of people who joined an extremist group on Facebook only did so because the company's algorithm recommended it to them.

A recent Congressional Research Services Report discussed the issue of content moderation and specifically the spread of misinformation and the role that social media companies play in worsening the issue:

> Two features of social media platforms—the user networks and the algorithmic filtering used to manage content—can contribute to the spread of misinformation. Users can build their own social networks, which affect the content that they see, including the types of misinformation they may be exposed to. Most social media operators use algorithms to sort and prioritize the content placed on their sites. These algorithms are generally built to increase user engagement, such as clicking links or commenting on posts. In particular, social media operators that rely on advertising placed next to user-generated content as their primary source of revenue have incentives to increase user engagement. These operators may be able to increase their revenue by serving more ads to users and potentially charging higher fees to advertisers. Thus, algorithms may amplify certain content, which can include misinformation, if it captures users' attention.[11]

The role that content moderation, or the lack of it, has in alleviating or exacerbating these issues has been a source of much debate. A policy paper published by the Shorenstein Center on Media, Politics, and Public Policy at the Harvard Kennedy School, *Countering Negative Externalities in Digital Platforms*, focuses on the costs associated with various internet platforms that are not absorbed by the companies themselves:

> Today, in addition to the carcinogenic effects of chemical runoffs and first and second hand tobacco smoke, we have to contend with a new problem: the poisoning of our democratic system through foreign influence campaigns, intentional dissemination of misinformation, and incitements to violence inadvertently enabled by Facebook, YouTube and our other major digital platform companies.[12]

---

[11] Jason A. Gallo & Clare Y. Cho, *Social Media: Misinformation and Content Moderation Issues for Congress* (January 27, 2021) Congressional Research Service, https://crsreports.congress.gov/product/pdf/R/R46662.

[12] *Countering Negative Externalities in Digital Platforms* (October 7, 2019) Shorenstein Center on Media, Politics and Public Policy, https://shorensteincenter.org/countering-negative-externalities-in-digital-platforms/.

The paper asserts that these major platform companies "enable exceptionally malign activities" and "experience shows that the companies have not made sufficient investments to eliminate or reduce these negative externalities."

As pointed out by recent Wall Street Journal reporting, the companies' employees are aware of the dangers:

> A Facebook Inc. team had a blunt message for senior executives. The company's algorithms weren't bringing people together. They were driving people apart.
>
> "Our algorithms exploit the human brain's attraction to divisiveness," read a slide from a 2018 presentation. "If left unchecked," it warned, Facebook would feed users "more and more divisive content in an effort to gain user attention & increase time on the platform."
>
> That presentation went to the heart of a question dogging Facebook almost since its founding: Does its platform aggravate polarization and tribal behavior?
>
> The answer it found, in some cases, was yes.[13]

A recent New York Times article on leadership at Facebook elaborates:

> To achieve its record-setting growth, the [Facebook] had continued building on its core technology, making business decisions based on how many hours of the day people spent on Facebook and how many times a day they returned. Facebook's algorithms didn't measure if the magnetic force pulling them back to Facebook was the habit of wishing a friend happy birthday, or a rabbit hole of conspiracies and misinformation.
>
> Facebook's problems were features, not bugs.[14]

Another paper recently released provides "Recommendations to the Biden Administration," and is relevant to the considerations here:

> The Administration should work with Congress to develop a system of financial incentives to encourage greater industry attention to the social

[13] Jeff Horowitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive* (May 26, 2020) Wall Street Journal, https://www.wsj.com/articles/facebook-knows-it-encourages-division-topexecutives-nixed-solutions-11590507499.
[14] Sheera Frenkel & Cecilia Kang, *Mark Zuckerberg and Sheryl Sandberg's Partnership Did Not Survive Trump* (July 8, 2021) The New York Times, https://www.nytimes.com/2021/07/08/business/mark-zuckerberg-sheryl-sandberg-facebook.html.

> costs, or "externalities," imposed by social media platforms. A system of meaningful fines for violating industry standards of conduct regarding harmful content on the internet is one example. In addition, the Administration should promote greater transparency of the placement of digital advertising, the dominant source of social media revenue. This would create an incentive for social media companies to modify their algorithms and practices related to harmful content, which their advertisers generally seek to avoid.[15]

2. Content moderation, transparency, and the low-grade war on our cognitive security

There are a number of considerations when addressing how to approach the proliferation of these undesirable social media posts and the companies' practices that fuel the flames. A number of methods of content moderation are being deployed and have evolved from simply blocking content or banning accounts to quarantining topics, removing posts from search results, barring recommendations, and down ranking posts in priority. However, there is a lack of transparency and understanding of exactly what companies are doing and why it does not seem to be enough. An article in the MIT Technology Review articulates the issues with content moderation behind the curtain:

> As social media companies suspended accounts and labeled and deleted posts, many researchers, civil society organizations, and journalists scrambled to understand their decisions. The lack of transparency about those decisions and processes means that—for many—the election results end up with an asterisk this year, just as they did in 2016.
>
> What actions did these companies take? How do their moderation teams work? What is the process for making decisions? Over the last few years, platform companies put together large task forces dedicated to removing election misinformation and labeling early declarations of victory. Sarah Roberts, a professor at UCLA, has written about the invisible labor of platform content moderators as a shadow industry, a labyrinth of contractors and complex rules which the public knows little about. Why don't we know more?
>
> In the post-election fog, social media has become the terrain for a low-grade war on our cognitive security, with misinformation campaigns and conspiracy theories proliferating. When the broadcast news business

[15] Caroline Atkinson, et al., *Recommendations to the Biden Administration On Regulating Disinformation and Other Harmful Content on Social Media* (March 2021) Harvard Kennedy School & New York University Stern School of Business, https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/6058a456ca24454a73370dc8/1616421974691/TechnologyRecommendations_2021final.pdf.

> served the role of information gatekeeper, it was saddled with public interest obligations such as sharing timely, local, and relevant information. Social media companies have inherited a similar position in society, but they have not taken on those same responsibilities. This situation has loaded the cannons for claims of bias and censorship in how they moderated election-related content.

This bill seeks to increase transparency around what terms of service social media companies are setting out and how it ensures those terms are abided by. The goal is to learn more about the methods of content moderation and how successful they are. According to the author:

> The line between providing an open forum for productive discourse and permitting the proliferation of hate speech and misinformation is a fine one, and depends largely on the structure and practices of the platform. However, these platforms rarely provide detailed insight into such practices, and into the relative effectiveness of different approaches. This, along with constraints imposed by existing federal law, has historically made policy-making in this space remarkably difficult. This bill seeks to provide critical transparency to both inform the public as to the policies and practices governing the content they post and engage with on social media, and to allow for comparative assessment of content moderation approaches to better equip both social media companies and policymakers to address these growing concerns.

A coalition of groups, including ADL, Equality California, NAACP, and Esperanza Immigrant Rights Project, emphasizes the need for the bill:

> Despite the widespread nature of these concerns, efforts by social media companies to self-police such content have been widely criticized as opaque, arbitrary, biased, and inadequate. While some platforms share limited information about their efforts, the current lack of transparency has exacerbated concerns about the intent, enforcement, and impact of corporate policies, and deprived policymakers and the general public of critical data and metrics regarding the scope and scale of online hate and disinformation. Additional transparency is needed to allow consumers to make informed choices about the impact of these products (including the impact on their children) and so that researchers, civil society leaders, and policymakers can determine the best means to address this growing threat to our democracy.
>
> AB 587 would address this troubling lack of transparency by requiring social media platforms to publicly disclose their policies and report key data and metrics around the enforcement of their policies. This disclosure

would be accomplished through quarterly public filings with the Attorney General.

This bill starts with a baseline requirement to have social media platforms post their terms of service. These policies must include information about how users can ask questions, how they can flag content or users in violation, and a list of potential actions that the company might take in response. To ensure meaningful access, the terms of service must be posted in a manner reasonably designed to inform all users of their existence and contents and available in all Medi-Cal threshold languages in which the social media platform offers product features.

The bill next requires an extremely detailed report to be compiled by these companies and submitted to the Attorney General on a quarterly basis. This report must include information on the terms of service, any changes made, and whether they define certain categories of content, including hate speech or racism; extremism or radicalization; disinformation or misinformation; harassment; and foreign political interference.

The bill also requires the report to contain a "detailed description of content moderation practices" used by the platform. There must also be outcome-focused information included. Platforms must report on the number of flagged items of content and the number of times the company took action in response. To understand the impact of the reported content, the report must detail the number of times this content was viewed and shared by users. The data must also include these details broken down by content category, the type of media, and other factors.

The bill leaves enforcement to public prosecutors who may seek injunctive relief and civil penalties of up to $15,000 per violation per day, with the amount based on a consideration of whether the platform made a reasonable, good faith attempt to comply. A social media platform is subject to liability if it does the following:

- fails to post terms of service;
- fails to timely submit the quarterly report to the Attorney General; or
- materially omits or misrepresents required information in that report.

3. Legal Obstacles

As the author references above, all of this occurs within tight quarters due to federal statutory and constitutional law.

a. 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).[16] 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 statements might be false, defamatory, or otherwise actionable.[17] 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.[18]

The crux of Section 230 is 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."[19] 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."[20]

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. Section 230 specifies 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."[21] Courts have applied Section 230 in a vast range of cases to immunize internet platforms from "virtually all suits arising from third-party content."[22]

---

[16] *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 bookstores 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.)

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

[18] *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.)

[19] *Id.*, § 230(c)(1).

[20] *Id.*, § 230(c)(1) & (2).

[21] *Id.*, § 230(e)(1) & (3).

[22] 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.

The author argues that the bill walks this line carefully:

> AB 587 does not impose liability based on the nature of content moderation decisions taken by social media platforms. Rather, the requirements of AB 587 are focused exclusively on disclosure of information relating to those practices, with liability imposed based on failure to disclose the specified information. By taking this transparency approach, AB 587 is thus unlikely to run afoul of the liability protections provided by Section 230, and would be far less susceptible to a preemption challenge than most attempts to regulate in this space.

b. First Amendment

In addition, any specific mandates to remove some subset of this broad swath of content could run afoul of the First Amendment. The United States Supreme Court has held that posting on social networking and/or social media sites constitutes communicative activity protected by the First Amendment.[23] As a general rule, the government "may not suppress lawful speech as the means to suppress unlawful speech."[24] In addition, the First Amendment places restrictions on compelled speech. However, the case law generally affords a wide berth to laws that regulate commercial speech and that involve disclosure requirements that involve conveying factual information that has sound public policy justification, such as food labeling.[25]

Because this bill simply seeks transparency into what content moderation practices are being deployed and their outcomes, it likely does not run afoul of these laws. No specific content moderation is required or penalized. The information required to be disclosed can play a key role in informing future legislative action and public debate of these issues.

4. Defining social media platforms

With a multitude of bills currently moving through the legislative process that seek to regulate social media platforms, efforts have been made to harmonize the various definitions that exist. The author has agreed to amend in the following definition, which will be going into the various other bills. However, the Committee and authors will continue to engage with stakeholders to further refine the definition as necessary.

---

[23] *E.g.*, *Packingham v. North Carolina* (2017) 137 S.Ct. 1730, 1735-1736.
[24] *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]).
[25] *See Central Hudson Gas & Elec. Corp. v. Public Serv. Comm'n* (1980) 447 U.S. 557; *Zauderer v. Office of Disciplinary Counsel of Supreme Court* (1985) 471 U.S. 626.

Amendment

Replace definition of "content":

(b) (1) "Content" means statements or comments made by users and media that are created, posted, shared, or otherwise interacted with by users on an internet-based service or application.
(2) "Content" does not include media put online exclusively for the purpose of cloud storage, transmitting documents, or file collaboration.

Replace definition of "social media platform:

(d) "Social media platform" means a public or semipublic internet-based service or application that has users in California and that meets all of the following criteria:
(1) (A) A substantial function of the service or application is to connect users in order to allow users to interact socially with each other within the service or application.
(B) A service or application that provides email or direct messaging services shall not be considered to meet this criterion on the basis of that function alone.
(2) The service or application allows users to do all of the following:
(A) Construct a public or semipublic profile for purposes of signing into and using the service.
(B) Populate a list of other users with whom an individual shares a social connection within the system.
(C) Create or post content viewable by other users, including, but not limited to, on message boards, in chat rooms, or through a landing page or main feed that presents the user with content generated by other users.
(e) "Public or semipublic internet-based service or application" excludes a service or application used to facilitate communication within a business or enterprise among employees or affiliates of the business or enterprise, provided that access to the service or application is restricted to employees or affiliates of the business or enterprise using the service or application.

The author wishes to continue to limit the application of the bill to only platforms controlled by a business entity that generated at least $100,000,000 in gross revenue during the preceding calendar year, so that exemption continues to apply. Most of the other exemptions are incorporated into the definition of social media platform, excluding email and direct messaging services from the social interaction element, for instance.

5. Concerns with the bill

A coalition, including TechNet, the Civil Justice Association, and the California Chamber of Commerce, objects to the granularity and depth of the data that is required to be provided and argues that disclosing their practices will allow bad actors to exploit these platforms:

> AB 587 requires companies to publicly disclose more than just content moderation policies, which are already available to the public. The bill requires companies to report to the Attorney General sensitive information about how we implement policies, detect activity, train employees, and use technology to detect content in need of moderation. The language makes it explicit that the bill is seeking "detailed" information about content moderation practices, capabilities, and data regarding content moderation.

The coalition also argues the reporting requirements are too onerous:

> AB 587 requires businesses to report detailed metrics on a quarterly basis regarding not only the numerical scale of content moderation practices, but also details about how content is flagged and acted against. It would be nearly impossible to report this information quarterly due to the need to review, analyze, and adjudicate actioned content. Further, the sheer volume of content our companies review makes it similarly difficult and costly to implement these disclosures, particularly the number of times actioned items of content were viewed or shared. Producing this information quarterly is unworkable and unreasonable.
>
> Furthermore, there is little justification to require a report to the Attorney General. Instead, the bill should simply require our companies to post these reports on their website or platform. If the goal is increased transparency, it is unlikely that a consumer would ever look to the Attorney General's website for information about a company's terms of service or community guidelines. [. . .]
>
> To address these concerns, we suggest amending the bill to require an annual report, to be posted in a manner reasonably designed to inform a platform's users and to strike the requirements for specific information about company content moderation practices and training materials.

6. Support for this transparency measure

The Simon Wiesenthal Center writes in support:

> AB 587 is a much-needed bill that addresses this public policy need. It will require social media platforms to publicly disclose their corporate polices and report key data and metrics around the enforcement of their policies. This disclosure would require corporate policies to be disclosed to the Attorney General on such issues as hate speech and racism, extremism and harassment. Corporations would be compelled to disclose how they enforce these policies and any changes they make to these policies or enforcement.
>
> AB 587 is long overdue and the Simon Wiesenthal Center fully supports it.

Writing in support, the California Labor Federation points to evidence more needs to be done:

> A GLAAD study in 2021 reported that 64 percent of LGBTQ social media users polled stated that they had experienced harassment and hate speech at a much higher rate than other identity groups on social media. Investigations have also shown that the violent riots at the U.S. Capitol in early January of 2021 were abetted and encouraged by posts on social media sites. Lack of accountability for social media platforms severely affects a wide range of marginalized communities—including, but not limited to, people of color, women, the LGBTQ+ community, Jewish and Muslim communities, and people seeking reproductive justice—who are disproportionately targeted by online hate and adversely affected by misinformation on social media.

The Los Angeles County Democratic Party explains its support:

> Social media companies have behaved irresponsibly, allowing disinformation, misinformation and hate to add fuel to some of America's greatest challenges, like combating a global pandemic or grappling with longstanding social wounds around racism and misogyny, and they have also enabled the micro targeting of vulnerable individuals.
>
> The requirements of AB 587, clear Terms of Service and their enforcement, regular reporting, and significant penalties for non-compliance, are necessary measures to bring some measure of corporate accountability to our increasingly virtual lives.

SUPPORT

Anti-Defamation League (sponsor)
Accountable Tech
Alameda County Democratic Party
American Academy of Pediatrics, California
American Association of University Women - California
American Association of University Women, Camarillo Branch
American Federation of State, County and Municipal Employees, AFL-CIO
American Jewish Committee - Los Angeles
American Jewish Committee - San Francisco
American Muslim & Multifaith Women's Empowerment Council
The Arc and United Cerebral Palsy California Collaboration
Armenian Assembly of America
Armenian National Committee of America - Western Region
Asian Americans in Action
Asian Law Alliance
Bend the Arc: Jewish Action
Buen Vecino
California Asian Pacific American Bar Association
California Federation of Teachers AFL-CIO
California Hawaii State Conference National Association for the Advancement of Colored People
California Labor Federation, AFL-CIO
California League of United Latin American Citizens
California Nurses Association
California State Council of Service Employees International Union (SEIU California)
California Women's Law Center
Center for LGBTQ Economic Advancement & Research (CLEAR)
Center for the Study of Hate & Extremism - California State University, San Bernardino
College Democrats at UC Irvine
Common Sense
Consumer Reports Advocacy
Courage California
Davis College Democrats
Decode Democracy
Democratic Party of the San Fernando Valley
Democrats for Israel-Orange County
East Bay Young Democrats
Equality California
Esperanza Immigrant Rights Project, Catholic Charities of Los Angeles
The Greenlining Institute
Harvey Milk LGBTQ Democratic Club
Hindu American Foundation, Inc.

Islamic Networks Group
Islamic Networks Inc.
Israeli-American Civic Action Network
Japanese American Citizens League, Berkeley Chapter
Jewish Center for Justice
Jewish Family and Children's Services of San Francisco, the Peninsula, Marin and Sonoma Counties
Jewish Federation of Greater Los Angeles
Jewish Federation of The Sacramento Region and The Sacramento Jewish Community Relations Council
Jewish Public Affairs Committee
Korean American Bar Association of Northern California
Korean American Coalition - Los Angeles
League of United Latin American Citizens
Los Angeles County Democratic Party
Maplight
Miracle Mile Democratic Club
National Association for the Advancement of Colored People, SV/SJ
Nailing It for America
National Center for Lesbian Rights
National Council of Jewish Women, California
National Hispanic Media Coalition
Oakland Privacy
Orange County Racial Justice Collaborative
Pakistani-American Democratic Club of Orange County
Pilipino American Los Angeles Democrats (PALAD)
Progressive Zionists of California
Protect US
Rabbis and Cantor of Congregation or Ami
Sacramento County Young Democrats
Sacramento LGBT Community Center
San Fernando Valley Young Democrats
San Francisco Democratic Party
Santa Barbara Women's Political Committee
Sikh American Legal Defense and Education Fund (SALDEF)
Simon Wiesenthal Center, Inc.
The Source LGBT+ Center
Stonewall Democratic Club
United Food and Commercial Workers, Western States Council
United Nurses Associations of California/union of Health Care Professionals
Voices for Progress

OPPOSITION

California Chamber of Commerce
Chamber of Progress
Civil Justice Association of California
Computer and Communications Industry Association
Consumer Technology Association
Internet Coalition
MPA - the Association of Magazine Media
Netchoice
TechNet

RELATED LEGISLATION

Pending Legislation:

SB 1056 (Umberg, 2022) requires a social media platform, as defined, to clearly and conspicuously state whether it has a mechanism for reporting violent posts, as defined; and allows a person who is the target, or who believes they are the target, of a violent post to seek an injunction to have the violent post removed. This bill is currently in the Assembly Judiciary Committee.

AB 1628 (Ramos, 2022) requires online platforms to create and post a policy that includes policies regarding distribution of controlled substances and its prevention, reporting mechanisms, and resources. AB 1628 is currently pending before this Committee and will be heard on the same day as this bill.

AB 2273 (Wicks, 2022) establishes 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. The bill tasks the California Privacy Protection Agency with establishing a taskforce to evaluate best practice and to adopt regulations. AB 2273 is currently pending before this Committee and will be heard on the same day as this bill.

AB 2408 (Cunningham, 2022) establishes a negligence cause of action for a platform's use of any design, feature, or affordance that causes a child user to become addicted to the platform. It also provides for heightened civil penalties in actions brought by public prosecutors. AB 2408 is currently pending before this Committee and will be heard on the same day as this bill.

AB 2879 (Low, 2022) requires social media platforms to implement a mechanism by which school administrators can report instances of cyberbullying, and to disclose specified data related to reported instances of cyberbullying and the platform's

response. AB 2879 is currently pending before this Committee and will be heard on the same day as this bill.

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. This bill died in the Senate Judiciary Committee.

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 this bill 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 due to the COVID-19 pandemic.

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 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 Consumer Protection.

SB 1361 (Corbett, 2010) would have prohibited social networking websites from displaying, to the public or other registered users, the home address or telephone number of a registered user of that site who is under 18 years of age, and imposed a civil penalty of up to $10,000 for each willful and knowing violation of this prohibition. SB 1361 died in the Assembly Committee on Entertainment, Sports, Tourism, and Internet Media.

PRIOR VOTES:

Assembly Floor (Ayes 64, Noes 1)
Assembly Appropriations Committee (Ayes 13, Noes 0)
Assembly Judiciary Committee (Ayes 10, Noes 0)
Assembly Privacy and Consumer Protection Committee (Ayes 9, Noes 0)

****