

CONCURRENCE IN SENATE AMENDMENTS  
AB 587 (Gabriel)  
As Amended August 24, 2022  
Majority vote

**SUMMARY**

This bill requires social media companies, as defined, to post their terms of service (ToS) in a manner reasonably designed to inform all users of specified policies and would require a social media company to submit semiannual reports, as specified, starting January 1, 2024, to the Attorney General (AG).

**Senate Amendments**

- 1) Reduce the frequency of required reports from quarterly to semiannually.
- 2) Limit the languages in which terms of service must be posted to only the Medi-Cal threshold languages.
- 3) Remove requirements to disclose specific rules, guidelines, product changes, or training materials, and remove the requirement that the detailed description of content moderation practices within a report be "complete."
- 4) Remove the 30-day right to cure violations.
- 5) Replace enforcement through the Unfair Competition Law with a specified enforcement mechanism that permits specified public attorneys to bring actions seeking injunctive relief in addition to a civil penalty not to exceed \$15,000 per violation per day.
- 6) Specify that a social media company shall be considered in violation of the provisions of the bill for each day the social media company fails to post terms of service, as specified, fails to timely submit to the AG a report as required, or materially omits or misrepresents required information in a report.
- 7) Limit enforcement of the bill's provisions to actions brought by the AG or by city attorneys of cities with populations over 750,000.
- 8) Offset reporting such that a submitted report need not be released until one full quarter after the period of time covered by the report, and specify the timing of these reports.
- 9) Clarify that the bill shall not be construed to apply to an internet-based service or application for which interactions between users are limited to direct messages, commercial transactions, consumer reviews of products, sellers, services, events, or places, or any combination thereof.
- 10) Consistent with the definition adopted for all pending bills pertaining to social media this legislative session, define "social media platform" to mean a public or semipublic internet-based service or application that has users in California and that meets both of the following criteria: a) a substantial function of the service or application is to connect users in order to allow users to interact socially with each other within the service or application, except as specified; and b) the service or application allows users to construct a public or semipublic

profile for purpose of signing into and using the service, populate a list of other users with whom an individual shares a social connection within the system, and create or post content viewable by others.

- 11) Specify exclusions from the provided definitions for "content" and "public or semipublic internet-based service or application", respectively, to exclude media put on a service or application exclusively for the purpose of cloud storage, transmitting files, or file collaboration, and services or applications used to facilitate communication within a business or enterprise exclusively among employees or affiliates of the business or enterprise.

## COMMENTS

As online social media become increasingly central to the public discourse, the companies responsible for managing social media platforms are faced with a complex dilemma regarding content moderation, i.e., how the platforms determine what content warrants disciplinary action such as removal of the item or banning of the user. In broad terms, there is a general public consensus that certain types of content, such as child pornography, depictions of graphic violence, emotional abuse, and threats of physical harm, are undesirable, and should be mitigated on these platforms to the extent possible. Many other categories of information, however, such as hate speech, racism, extremism, misinformation, political interference, and harassment, are far more difficult to reliably define, and assignment of their boundaries is often fraught with political bias. In such cases, both action and inaction by these companies seems to be equally maligned: too much moderation and accusations of censorship and suppressed speech arise; too little, and the platform risks fostering a toxic, sometimes dangerous community.

AB 587 seeks to confront issues around social media content moderation practices by requiring the publication of ToS with specified information, and by requiring social media companies to submit semiannual reports containing information related to content moderation policies and data related to the application of those policies in practice. Though content moderation on social media is a notoriously difficult problem to tackle, AB 587 seemingly adopts a unique, data driven approach to progressing public policy in that space. Rather than placing specific content moderation requirements on companies, which in many cases raises constitutional issues, the bill instead provides for transparency and public accountability with respect to these practices, and establishes a timely, comprehensive dataset of untoward content on social media. This dataset can support research into the ever-changing social media ecosystem to help inform policies designed to root out its most problematic components while preserving its benefits for expression and connection.

Though granularity in this information can be useful for understanding the landscape and establishing transparency, however, opponents of the bill point out that too much granularity could put the platforms at risk. Indeed, in the past few years, the social media ecosystem has seen the emergence of sophisticated, sometimes state-sponsored actors seeking to exploit the design of their platforms toward nefarious ends. In this respect, it does not seem outlandish to presume that a large, detailed, public repository of information related to how content is moderated may increase sophistication of attempts to subvert content moderation systems. That said, in much the same way as policies for assessment and disclosure of security vulnerabilities is considered a best practice for cybersecurity, this same repository could enhance public scrutiny in a manner that would expose shortcomings in content moderation practices before they become catastrophic. Additionally, such information in aggregate from several platforms may facilitate

comparison and meta-analysis that can help establish best practices that, even if transparent, are nonetheless secure. Accordingly, on balance, it is difficult to determine whether extensive, detailed publication of moderation practices would increase or decrease the vulnerability of these platforms to exploitation by bad actors. Nonetheless, amendments taken in the Senate make minor changes to language enumerating required disclosures in order to clarify that disclosure of such highly granular information is not expected.

Senate amendments divorce the bill's provisions from enforcement through the Unfair Competition Law, instead specifying an enforcement mechanism that permits actions for relief to be prosecuted by a wide range of specified public attorneys. Specifically, if a social media company fails to post terms of service in accordance with the bill's requirements, fails to timely submit to the AG a required report, or materially omits or misrepresents required information in a report, that company shall be liable for a civil penalty not to exceed \$15,000 per violation per day, and may be enjoined in a court of competent jurisdiction.

Staff notes that the bill does not appear to require any particular actions on the part of the company other than: 1) posting terms of service in accordance with specified criteria; and 2) submitting semiannual reports containing specified information. As such, it would appear that violations of the bill would only occur if the company failed to perform one or both of these requirements, and that so long as the reports and ToS conform to the specifications, the actual content moderation itself is not subject to enforcement. It therefore does not appear likely that liability imposed by this bill would allow for lawsuits to be filed against platforms for the sufficiency of their moderation practices, arguably making the risk of preemption under Section 230 on these grounds minimal.

### **According to the Author**

In recent 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. Twitter, along with other social media platforms, has been implicated as a venue for hate groups to safely grow. A recent study of Twitter posts from 100 [United States] cities found that the greater proportion of tweets related to race- and ethnicity-based discrimination in a given city, the more hate crimes were occurring in that city. Robert Bowers, accused of murdering 11 elderly worshipers at a Pennsylvania synagogue in 2018, had been active on Gab, a Twitter-like site used by white supremacists. Most recently, investigations have shown that the violent riots at the Capitol in early January of this year were abetted and encouraged by posts on social media sites.

AB 587 would require social media platforms to publicly disclose their corporate policies and report key data and metrics around the enforcement of their policies. This disclosure would be accomplished through semiannual public filings with the AG.

### **Arguments in Support**

A coalition of civil, minority, and immigrant rights organizations including the Anti-Defamation League, Common Sense, and the California League of United Latin American Citizens argues:

[E]fforts by social media companies to self-police [problematic] content have been 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 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 corporate policies and report key data and metrics around the enforcement of their policies.

### **Arguments in Opposition**

A coalition of groups representing business interests including CalChamber, TechNet, and the Civil Justice Association of California argue in opposition unless amended:

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. [...] This requirement would not only threaten the security of these practices but provides extremists, terrorist organizations, child predators, scammers, and serial abusers of our policies with roadmaps to get around our protections. [...] AB 587 will undermine the extensive work our companies have already undertaken to combat harmful content.

### **FISCAL COMMENTS**

According to the Senate Appropriations Committee:

- 1) *DOJ*: The Department of Justice (DOJ) reports costs of \$414,000 in 2022-23 and \$711,000 annually thereafter to enforce the provisions of AB 587 and for IT resources to allow for submission of terms of service (General Fund)
- 2) *Judicial Branch*: Unknown cost pressures due to increased court workload (Special Fund - Trial Court Trust Fund, General Fund)
- 3)

### **VOTES:**

#### **ASM PRIVACY AND CONSUMER PROTECTION: 9-0-2**

**YES:** Chau, Bauer-Kahan, Bennett, Carrillo, Cunningham, Gabriel, Irwin, Lee, Wicks

**ABS, ABST OR NV:** Kiley, Gallagher

#### **ASM JUDICIARY: 10-0-1**

**YES:** Stone, Gallagher, Chau, Chiu, Davies, Lorena Gonzalez, Holden, Kalra, Maienschein, Reyes

**ABS, ABST OR NV:** Kiley

#### **ASM APPROPRIATIONS: 13-0-3**

**YES:** Lorena Gonzalez, Calderon, Carrillo, Chau, Davies, Gabriel, Eduardo Garcia, Levine, Quirk, Robert Rivas, Akilah Weber, Holden, Luz Rivas

**ABS, ABST OR NV:** Bigelow, Megan Dahle, Fong

**ASSEMBLY FLOOR: 64-1-14**

**YES:** Aguiar-Curry, Arambula, Bauer-Kahan, Bennett, Berman, Bloom, Boerner Horvath, Bryan, Burke, Calderon, Carrillo, Cervantes, Chau, Chiu, Cooley, Cooper, Cunningham, Daly, Davies, Frazier, Friedman, Gabriel, Gallagher, Cristina Garcia, Eduardo Garcia, Gipson, Lorena Gonzalez, Grayson, Holden, Irwin, Jones-Sawyer, Kalra, Lackey, Lee, Levine, Low, Maienschein, McCarty, Medina, Mullin, Muratsuchi, Nazarian, O'Donnell, Petrie-Norris, Quirk, Quirk-Silva, Ramos, Reyes, Luz Rivas, Robert Rivas, Rodriguez, Blanca Rubio, Salas, Santiago, Stone, Ting, Valladares, Villapudua, Waldron, Ward, Akilah Weber, Wicks, Wood, Rendon

**NO:** Gray

**ABS, ABST OR NV:** Bigelow, Chen, Choi, Megan Dahle, Flora, Fong, Kiley, Mathis, Mayes, Nguyen, Patterson, Seyarto, Smith, Voepel

**SENATE FLOOR: 33-3-4**

**YES:** Allen, Archuleta, Atkins, Becker, Bradford, Caballero, Cortese, Dahle, Dodd, Durazo, Eggman, Glazer, Gonzalez, Hertzberg, Hueso, Hurtado, Kamlager, Laird, Leyva, Limón, McGuire, Min, Newman, Ochoa Bogh, Pan, Portantino, Roth, Rubio, Skinner, Stern, Umberg, Wieckowski, Wiener

**NO:** Melendez, Nielsen, Wilk

**ABS, ABST OR NV:** Bates, Borgeas, Grove, Jones

**UPDATED**

VERSION: August 24, 2022

CONSULTANT: Landon Klein / P. & C.P. / (916) 319-2200

FN: 0004499