

ASSEMBLY THIRD READING

AB 587 (Gabriel)

As Amended April 28, 2021

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 quarterly reports, as specified, starting July 1, 2022, to the Attorney General (AG).

Major Provisions

- 1) Requires a social media company to post its ToS in a manner reasonably designed to inform all users of the social media company's service of the existence and contents of the ToS and requires the ToS to include all of the following:
 - a) Contact information for the purpose of allowing users to ask the social media company questions about its ToS.
 - b) A description of the process users must follow to flag content, groups or other users they believe violate ToS, and the social media company's commitments on response and resolution time.
 - c) A list of potential actions the social media company may take against any item of content, or a user, or group of users, including, but not limited to, removal, demonetization and de-prioritization or banning.
- 2) Requires a social media company to submit to the Department of Justice (DOJ), on a quarterly basis, a ToS report, covering content moderation policies and practices within the previous three months and requires the DOJ to post on its website all submitted reports with the first report submitted no later than July 1, 2022.
- 3) Provides that a social media company is in violation of this bill if it fails to comply with its requirements within 30 days of being notified of noncompliance by the AG.
- 4) States that a violation of this bill is actionable under the Unfair Competition Law (UCL).

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 quarterly 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.

AB 587 specifies that a violation of its provisions is actionable under the Unfair Competition Law (UCL; Business & Professions Code Section 17200) in addition to any other applicable state or federal law. The UCL creates a private right of action, but allows individual plaintiffs to seek only injunctive relief or restitutionary disgorgement, and only in the event the plaintiff can demonstrate injury-in-fact resulting from the violation. The UCL also permits the AG and district attorneys to bring causes of action in the name of the people of the State of California, and, in these cases, adds civil penalties up to \$2,500 per violation as an available remedy. Opponents of the bill express concerns that the liability exposure as a result of this enforcement mechanism may be counterproductive, and potentially unlawful under Section 230 of the federal Communications Decency Act of 1996, which provides that an online platform generally cannot be held liable for content posted by third parties. Section 230 explicitly preempts any conflicting state law.

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 quarterly 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.

That said, it is not clear whether the UCL is the appropriate mechanism for enforcing this bill. Because it would be extremely difficult, if not impossible, for an individual to demonstrate injury-in-fact and loss of money or property as a result of a social media company's failure to submit a report or publish ToS, this leaves only public actions for injunctive relief or civil penalties. The bill does not make clear whether failure to submit a report in compliance with all specified requirements constitutes a single violation, or whether each non-compliant component is a separate violation. Assuming the former, the civil penalties available under the UCL are likely insufficient to enforce the bill, as complete noncompliance would result in a maximum annual penalty of \$12,500. For a company with gross annual revenue of over \$100,000,000, the threat of that penalty is not likely to ensure compliance.

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 biannual and quarterly public filings with the Attorney General.

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, Internet Association, and the Civil Justice Association of California argue in opposition unless amended:

In seeking to increase transparency around content moderation practices, AB 587 requires companies to report to the Attorney General the guidelines, practices, and even training materials companies use to moderate their platforms. This detailed information about content moderation practices, capabilities, and data regarding content moderation would not only threaten the security of these practices but provides bad actors with roadmaps to get around our protections. We believe that while well intentioned, these requirements will ultimately allow scammers, spammers, and other bad actors to exploit our systems and moderators. [...]

AB 587 opens companies up to the threat of liability and government investigation for routine moderation practices. Companies should not be subject to civil penalties or injunctive relief for the filing of a report, especially as comprehensive as the ones contemplated by this bill. Such litigation will deter investment in content moderation and suppress ongoing efforts to protect users from harmful content online. This extension of liability could also be interpreted to allow for lawsuits to be filed against platforms for the sufficiency of their moderation practices, which may be preempted by Section 230 of the Communications Decency Act (Section 230).

FISCAL COMMENTS

According to the Assembly Appropriations Committee, "[c]osts (General Fund (GF)) possibly in the low to mid hundreds of thousands of dollars for the DOJ to review and post TOS reports on its website on a quarterly basis. Additional possibly significant cost pressures to the GF in the low millions of dollars in staff and resources, to the extent this bill results in the DOJ taking legal action against any social media company that does not comply. This bill states its intent to apply meaningful remedies sufficient to induce compliance, that a violation of the requirements of this bill are actionable under the UCL and that DOJ notify a social media company when it is not in compliance with the requirements of this bill. DOJ currently enforces the UCL and other privacy laws and may be required to file for injunctive relief if a social media platform refuses to post its TOS as required by this bill."

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

UPDATED

VERSION: April 28, 2021

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

FN: 0000489