Date of Hearing: April 22, 2021

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Ed Chau, Chair
AB 587 (Gabriel) – As Amended March 25, 2021

AS PROPOSED TO BE AMENDED

SUBJECT: Social media companies: terms of service

SUMMARY: This bill would require social media companies, as defined, to post their terms of service 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). Specifically, **this bill would**:

1) Require a social media company to post their terms of service (ToS) in a manner reasonably designed to inform all users of the internet-based service owned or operated by the social media company of the existence and contents of the terms of service, and require the ToS to include all of the following:

    - Contact information for the purpose of allowing users to ask the social media company questions about the terms of service.

    - 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.

    - 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, deprioritization, or banning.

2) Require the ToS to be available in all languages in which the social media company offers product features, including but not limited to menus and prompts.

3) Provide that a social media company is in violation of the above provisions if it fails to comply with the provisions of this section within 30 days of being notified of noncompliance by the AG.

4) Beginning July 1, 2022, require a social media company to submit to the AG, on a quarterly basis, a terms of service report, covering activity within the previous three months. The AG shall post on its website all ToS reports submitted pursuant to this requirement. The report shall include:

    - The current ToS of the social media company.

    - A complete and detailed description of any changes to the ToS since the last report, as specified;

    - A statement of whether the current version of the ToS defines each of the following categories of content, including their definitions, if applicable: (1) hate speech or racism;

(2) extremism or radicalization; (3) disinformation or misinformation; (4) harassment; or (5) foreign political interference.

- A complete and detailed description of content moderation practices used by the social media company, including, but not limited to: (1) any existing policies intended to address the categories of content described immediately above; (2) any rules or guidelines regarding automated content moderation systems, as specified; (3) any training materials provided to content moderators, including educational materials; (4) responses to user reports of violations of the ToS; (5) any rules, guidelines, product changes and content moderator training materials that cover how the social media company would remove individual pieces of content, users, or groups that violate the ToS, or take broader action against individual users or against groups of users that violate the ToS; and (6) the languages in which the social media company offers product features, as specified.

- Information, deidentified and disaggregated, as specified, on content that was flagged by the social media company as content belonging to any of the categories described above, including all of the following: (1) the total number of flagged items of content; (2) the total number of actioned items of content, including the total number of actioned items of content that were removed, demonetized, or deprioritized; (3) the number of times actioned items of content were viewed by users, shared, and the number of users that viewed that content before it was actioned; (4) the number of times users appealed social media company actions and reversals of those actions on appeal, as specified.

5) Provide that any violation of the above provisions shall be actionable under the Unfair Competition Law in addition to any other applicable state or federal law.

6) Provide various definitions including:

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

   - "Content" to mean media, including, but not limited to, text, images, videos, and groups of users that are created, posted, shared, or otherwise interacted with by users on an internet-based service.

   - "Social media company" to mean a person or entity that owns or operates a public-facing internet-based service that generated at least $100,000,000 in gross revenue during the preceding year, and that allows users in the State to do all of the following: (1) construct a public or semipublic profile within a bounded system created by the service; (2) populate a list of other users with whom an individual shares a connection within the system; and (3) view and navigate a list of the individual's connections and the connections made by other individuals within the system.

7) State that it is the intent of the Legislature that a social media company that violates this chapter shall be subject to meaningful remedies sufficient to induce compliance with the provisions above.

EXISTING LAW:

1) Provides, under the U.S. Constitution, that "Congress shall make no law . . . 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., 1st Amend., as applied to the states through the 14th Amendment's Due Process Clause; *see Gitlow v. New York* (1925) 268 U.S. 652.)

2) Pursuant to the Communications Decency Act of 1996, provides, that "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider," and affords broad protection from civil liability for the good faith content moderation decisions of interactive computer services. (47 U.S.C. Sec. 230(c)(1) and (2).)

3) Provides under the California Constitution for the right of every person to freely speak, write and publish his or her 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, Sec. 2(a).)

4) Establishes the Unfair Competition Law, which, among other things, provides for specific or preventive relief to enforce a penalty, forfeiture, or penal law in the case of unfair competition; and defines unfair competition to mean any unlawful, unfair or fraudulent business act or practice and unfair, deceptive, and untrue or misleading advertising. (Bus. & Prof. Code Sec. 17200, et seq.)

5) Permits actions for relief pursuant to 4), above, to be prosecuted exclusively by the Attorney General, a district attorney, a county counsel as specified, a city attorney as specified, or a city prosecutor as specified, in the name of the people of the State of California, or by a person who has suffered injury in fact and has lost money or property as a result of the unfair competition. (Bus. & Prof. Code Sec. 17204.)

6) Permits any person specified in 5), above, to seek injunctive relief and actual damages, and permits any person specified in 5) except for a person who has suffered injury in fact to pursue civil penalties, as specified, for violations of the provisions of the Unfair Competition Law. (Bus. & Prof. Code Secs. 17204 and 17206.)

7) Defines "social media" for the above proposes to mean an electronic service or account, or electronic content, including, but not limited to, videos, still photographs, blogs, video blogs, podcasts, instant and text messages, email, online services or accounts, or internet website profiles or locations. (Lab. Code Sec. 980(a).)

FISCAL EFFECT: Unknown

COMMENTS:

1) **Purpose of this bill**: This bill seeks to increase transparency and accountability with respect to content moderation policies on social media platforms by requiring social media companies to maintain ToS containing specified information, and mandating the submission of quarterly reports to the AG detailing content moderation policies and data related to content moderation practices and objectionable content. This bill is author sponsored.

2) **Author's statement**: 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 U.S. 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 polices 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.

3) **Social media and content moderation**: 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.

This dilemma has been at the forefront of the public conscience since, in the wake of the attack on the nation's capital on January 6, 2021, the sitting President of the United States was banned from some social media platforms for incitement of violence and propagation of misinformation. But the largest social media platforms are faced with thousands, if not millions of similarly difficult decisions related to content moderation on a daily basis. Despite the problem being more visible than ever, the machinations of content moderation in many ways remain a mystery. As a coalition of civil, minority, and immigrant rights organizations in support of the bill argues:

> Despite the widespread nature of these concerns, efforts by social media companies to self-police such 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 polices and report key data and metrics around the enforcement of their policies.

Efforts to address online content moderation at the state level have often been frustrated by issues of federal preemption. Specifically, 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, explicitly preempts any conflicting state law. The law was designed to permit online platforms to freely moderate content in good faith without the risk of liability for content moderation decisions. But in effect, the liability shield provided by Section 230, coupled with its preemption of state law, makes it remarkably difficult to legislate at the state level with respect to content moderation. As a result, attempts to impose specific guidelines, restrictions, or requirements on social media platforms have thus far been unsuccessful.

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.

4) **AB 587 would require social media companies to submit detailed reports on content moderation policies and practices to the AG**: AB 587 seeks to increase transparency with respect to content moderation policies and practices by large social media companies. Specifically, the bill consists of three main components: (1) ToS requirements; (2) reporting on content moderation policies and procedures; and (3) reporting on content moderation practices, including the data relating to the types of objectionable content being moderated.

The bill would require a social media company to post their ToS in a manner reasonably designed to inform users of their existence and contents, including in all languages in which the company offers product features, and would require those ToS to include all of the following information: (1) contact information for user inquiries relating to the ToS; (2) a description of the process users must follow to flag content, groups, or other users, that they believe violate the social media company's ToS, as well as commitments by the company with respect to response and resolution times for flagged items; and (3) a list of potential actions the social media company may take against an item of content, user, or group, including but not limited to removal, demonetization, deprioritization, or banning. The bill specifies that failure to comply with these provisions within 30 days of notification of noncompliance by the AG would constitute a violation.

Next, the bill would require a quarterly report to be submitted to the AG by the social media company consisting of specified information related to content moderation policies, including all of the following: (1) the current version of their ToS; (2) a complete and detailed description of any changes to the ToS since the previous report; (3) a statement of whether the ToS defines "hate speech or racism," "extremism or radicalization," "disinformation or misinformation," harassment," or "foreign political interference," and if so, the definitions of those categories including any subcategories; (4) a complete and

detailed description of content moderation practices, including any policies intended to address categories described in (3), rules or guidelines regarding how automated content moderation systems enforce ToS and when and how those systems involve human review, training materials provided to content moderators, and a description of how the company responds to user reports of violations of the ToS; and (5) the languages in which the company offers product features and the languages for which the company has ToS.

Finally, the bill would require, in the same quarterly report, that the company provide information on content that was flagged by the company as content belonging to any of the categories described in (3), above, including all of the following: (1) the total number of flagged items of content; (2) the total number of actioned items of content, as defined; (3) the total number of actioned items of content that resulted in action taken by the company against the user or group of users responsible for the content; (4) the total number of actioned items of content that were removed, demonetized, or deprioritized by the social media company; (5) the number of times actioned items of content were viewed by users; (6) the number of times actioned items of content were shared, and the number of users that viewed the content before it was actioned; and (7) the number of times users appealed the company's actions and the number of reversals of the company's actions on appeal, disaggregated by each type of action. The bill would require that all such information be deidentified and disaggregated into the category of content, the type of content (e.g. posts, comments, messages, groups), the type of media of the content (e.g., text, images, videos), how the content was flagged (e.g., by human moderators, by AI software, by users), and how the content was actioned.

The bill would require the first of these reports to be submitted to the AG no later than July 1, 2022, and would require the AG to post on its official website all reports submitted pursuant to the bill.

Though content moderation on social media is a notoriously difficult problem to tackle, AB 587 seeming 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.

5) **Opposition raises security and workability concerns regarding granularity of data required in reports**: AB 587 requires regular reporting by social media companies with respect to a wide range of information related to content moderation. Though granularity in this information can be useful for understanding the landscape and establishing transparency, opponents of the bill point out that too much granularity could put the platforms at risk. As a coalition of groups representing business interests argue in opposition to the bill:

> 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.

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 are 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.

Opponents of the bill further contend that granularity in the reporting of practical moderation data would be unworkable due to the magnitude of information that must be evaluated. The opposing coalition argues:

> 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 to report data on individual pieces of content. Our companies address hundreds of millions of pieces of content across their platforms every few months. A requirement to collect, retain, and report information on individual pieces of content is unreasonable and unworkable. Furthermore, this volume of information is actually counterproductive to increasing transparency.

Notably, the provisions of this bill are limited to social media companies with over $100,000,000 in gross revenue from the preceding year, a limitation intended to ensure that companies subject to the bill have ample resources to absorb the reporting requirement. Though the amount of content many of these platforms receive is indeed enormous, most platforms of this size and maturity internally perform detailed evaluation of content for product optimization purposes, necessitating the expertise and technical capacity to manage large datasets. In that sense, requiring management of such data to comply with reporting requirements would arguably be unlikely to exceed the capabilities of these companies. That said, considering the amount of individual items of content a single report would require if each item needed to be discussed individually, the reports, which are released for public consumption, would be virtually unreadable, and provide little benefit to the general public. While the previous version of the bill implied each item was to be dealt with independently, the author prudently amended the bill to deal with these items of actioned content collectively, only disaggregated across specified criteria.

6) **Liability and enforcement**: AB 587 specifies that a violation of its provisions is actionable under the Unfair Competition Law (UCL; Bus. & Prof. Code Sec. 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. The coalition of business groups in opposition contends:

> 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).

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 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. If the bill passes out of this Committee, the author may wish to consider amending the bill to provide for enforcement via specified civil penalties sufficient to ensure compliance.

7) **Related legislation**: AB 13 (Chau) would enact the Automated Decision Systems Accountability Act of 2021 and state the intent of the Legislature that state agencies use an acquisition method that minimizes the risk of adverse and discriminatory impacts resulting from the design and application of automated decision systems.

AB 35 (Chau) would social media platforms, as defined, to disclose whether or not that social media platform has a policy or mechanism in place to address the spread of misinformation, as specified. The bill would require the disclosure to be made easily accessible on the social media platform's website and mobile application.

AB 1379 (E. Garcia) would prohibit a social media platform from amplifying, in a manner that violates its terms of service or written public promises, content that is in violation of the platform's terms of service.

8) **Prior legislation**: AB 1316 (Gallagher, 2019) would have prohibited social media internet website operators located in California, as defined, from removing or manipulating content from that site on the basis of the political affiliation or political viewpoint of that content, except as specified. This bill was held in the Assembly Rules Committee.

   AB 3169 (Gallagher, 2018) would have prohibited any person who operates a social media internet website or search engine located in California, as specified, from removing or manipulating content on the basis of the political affiliation or political viewpoint of that content. This bill failed passage in the Privacy & Consumer Protection Committee.

9) **Double referral**: This bill has been double-referred to the Assembly Judiciary Committee where it will be analyzed if passed by this Committee.

REGISTERED SUPPORT / OPPOSITION:

Support

AAUW Camarillo Branch
American Jewish Committee - Los Angeles
Anti-Defamation League
Armenian Assembly of America
Armenian National Committee of America - Western Region
Buen Vecino
California Asian Pacific American Bar Association
California League of United Latin American Citizens
Center for the Study of Hate & Extremism - California State University, San Bernardino
Common Sense
Hindu American Foundation, INC.
Islamic Networks INC.
Israeli-American Civic Action Network
Japanese American Citizens League, Berkeley Chapter
Jewish Center for Justice
Jewish Public Affairs Committee
Korean American Bar Association of Northern California
Maplight
National Hispanic Media Coalition
Progressive Zionists of California
Sikh American Legal Defense and Education Fund (SALDEF)
Simon Wiesenthal Center, INC.
Stonewall Democratic Club

Opposition

California Chamber of Commerce
Civil Justice Association of California
Internet Association
MPA - the Association of Magazine Media
TechNet

Analysis Prepared by: Landon Klein / P. & C.P. / (916) 319-2200