Date of Hearing: June 21, 2022

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Jesse Gabriel, Chair
SB 1018 (Pan) – As Amended May 19, 2022

AS PROPOSED TO BE AMENDED

SENATE VOTE: 30-9

SUBJECT: Platform Accountability and Transparency Act

SUMMARY: This bill would require a social media platform, as defined, to disclose to the public on or before July 1, 2023, and annually thereafter, statistics regarding the extent to which, in the preceding 12-month period, items of content that the platform determined violated its policies were recommended or otherwise amplified by platform algorithms before and after those items were identified as in violation of the platform's policies, disaggregated by category of policy violated. Specifically, **this bill would**:

1) Require a social media platform to disclose to the public, on or before July 1, 2023, and annually thereafter, statistics regarding the extent to which, in the preceding 12-month period, items of content that the platform determined violated its policies were recommended or otherwise amplified by platform algorithms before and after those items were identified as in violation of the platform's policies, disaggregated by category of policy violated.

2) Provide that a violation of 1), above, shall subject the violator to a civil penalty of up to $100,000 for each violation that may be recovered only in a civil action brought by the Attorney General.

3) Specify that the required disclosure pursuant to 1), above, does not require the dissemination of confidential business information or trade secrets.

4) Exempt from the requirements of the bill a social media platform with fewer than 1,000,000 discrete monthly users.

5) Define "social media platform", for the purposes of the bill's provisions, to mean an internet-based service or application that has users in California and that meets all of the following criteria:

   - A substantial function of the service or application is to connect users and allow users to interact socially with each other within the service or application. A service or application for which a substantial function is the conveyance of email or direct messages shall not be considered to meet this criterion on the basis of that function alone.

   - The service or application allows users to do all of the following:

     - Construct a public or semipublic profile for purposes of signing into and using the service or application.

     - Populate a list of other users with whom an individual shares a connection within the system.

- 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.

6) Define "content", for the purposes of the bill's provisions, to mean statements or comments made by users and media that are created, posted, shared, or otherwise interacted with by users on an internet-based service; and specify that content does not include media put online exclusively for the purpose of cloud storage, transmitting documents, or file collaboration.

7) Define "user", for the purposes of the bill's provisions, to mean a person with an account on a social media platform.

8) Specify that the bill shall be known as the "Platform Accountability and Transparency Act."

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) 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).)

3) 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).)

4) Specifies that no cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with 3), above. (47 U.S.C. Sec. 230(e)(3).)

5) Prohibits an employer from requiring or requesting an employee or applicant for employment, and prohibits a public or private postsecondary educational institution from requiring or requesting a student, prospective student, or student group, to do any of the following: disclose a username or password for the purpose of accessing personal social media; access personal social media in the presence of the employer or institution's representative; or divulge any personal social media, except as specified. (Lab. Code Sec. 980; Educ. Code Sec. 99121.)

6) Defines "social media" for the above purposes 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); Educ. Code Sec. 99120.)

FISCAL EFFECT: According to the Senate Appropriations Committee, with respect to a previous version of the bill, "[t]he Department of Justice (DOJ), reports ongoing costs of $221,000 in FY 2022-23, and $388,000 in FY 2023-24 and ongoing, to enforce the provisions of SB 1018 (General Fund). Unknown cost pressures resulting from increased workload on the courts to adjudicate filings generated by the provisions of this bill (Trial court Trust Fund, General Fund)."

COMMENTS:

1) **Purpose of this bill**: This bill seeks to improve transparency with respect to the role algorithmic amplification plays in the propagation of harmful content on social media platforms in order to better inform future policymaking and discourage undesirable practices. This bill is sponsored by ProtectUS.

2) **Author's statement**: According to the author:

> In October of 2021, Frances Haugen, a former employee of Facebook, was the latest person to expose the callous operating policies of the major social media platforms. In her testimony to Congress, she explained how Facebook knowingly used their algorithms to prioritize profits over their civic responsibilities. The detrimental effects of these policies result in the proliferation of disinformation as well as a myriad of mental health issues that are affecting our most vulnerable. […] The pandemic has only exacerbated these issues. During the last two years, an unprecedented level of medical misinformation has proliferated and undermined the messaging from public health officials. […] Addressing the many public policy concerns regarding social media begins with more transparency.

3) **Social media and content amplification**: As online social media becomes 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 foster 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 early as September 2021, The Wall Street Journal began publishing articles detailing otherwise opaque machinations of Facebook, referring to a trove of internal documents received by the media outlet, along with a consortium of other news organizations. These articles mainly detailed fatal flaws in content moderation and algorithmic prioritization by the company that underlie known toxic effects on individual users and on the greater public discourse at large. In October, these articles were revealed to have resulted from redacted documents provided by Frances Haugen, a former lead product manager for Facebook's division on civic integrity, who had disclosed the documents to the U.S. Securities and Exchange Commission and applied for whistleblower protection. Since then, Haugen has made a number of media appearances and testified before Congress and the U.K. Parliament to answer questions pertaining to the documents and to elaborate on their content.

Reporting on the documents provided by Haugen and Haugen's own testimony publicized some of the first explicit examples of how prioritization and amplification algorithms are often calibrated in order to maximize virality, emotional salience of content, and user engagement, often to the detriment of the public discourse. Haugen's testimony detailed the use of so-called "downstream meaningful social interaction" (MSI) as the primary metric governing exposure to content, meaning the more likely a piece of content posted by a user is to elicit engagement from other users, the higher its priority. This means that more inflammatory content is generally prioritized, as it is more likely to elicit responses from other users. Exacerbating this preference for inflammatory content, the documents revealed that beginning in 2017, Facebook's algorithm gave emoji reactions such as "angry" five times the weight of "likes" in prioritizing content in users' feeds. Facebook's data scientists confirmed that "angry," "wow," and "haha" emoji reactions occurred more frequently on toxic content and misinformation, but the company nonetheless failed to rectify this obvious flaw until 2020, in part due to pressures leading up to the 2020 election.

The documents submitted by Haugen include hundreds of pages of internal research demonstrating that downstream MSI as a prioritization mechanism expands hate speech, misinformation, incitement of violence, and graphic content on the platform. But absent the revelations by Haugen, the relationship between Facebook's particular parameters for algorithmic amplification and the prominence of undesirable content would have remained private.

As Oakland Privacy and Media Alliance argue in support of the bill:

> Despite mechanisms designed to assist users to manage their feeds to highlight content that they want to receive, platform priorities to serve advertisers and maximize engagement have led to algorithmic formulas that deliver content. Engagement based metrics value user response over all, leading to content that is inflammatory, sensational, controversial, and emotional being widely distributed. […] Worse, such content is not exempt from being misleading, propagandistic or flat out false. Whether such content is, at best, clickbait or at worst, a significant violation of content policies, will generally not prevent it from being seen by large numbers of people – even if moderation eventually kicks in to later remove it, slow its spread, or mark it as disinformation.
>
> Disinformation content has several concrete negative impacts on social media users, and on society at large. False content can drown out well-sourced journalism, undermine

> public health initiatives and democratic processes, and feed hate-based movements targeted at religious or ethnic minorities.
>
> By bringing sunshine to […] the spread and viewership of problematic content, SB 1018 will help policy-makers, researchers and the general public better understand what is happening on social media platforms and what is needed for more effective management of the disinformation epidemic.

Efforts to address online content moderation and amplification 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 and amplification. As a result, attempts to impose specific guidelines, restrictions, or requirements on social media platforms have thus far been unsuccessful.

This bill seeks to improve public accountability and transparency regarding the relationship between a social media platform's parameters for algorithmic amplification and the propagation of problematic content by requiring social media platforms to annually report statistics regarding the extent to which content that violates the platform's policies was recommended or otherwise amplified by the platform's algorithms before and after it was identified as violating content.

4) **Bill would require reporting on the algorithmic amplification of violating content before and after identifying the violation**: As it is proposed to be amended, this bill would require a social media platform with more than 1,000,000 discrete monthly users to disclose annually to the public statistics regarding the extent to which, in the preceding 12-month period, items of content that the platform determined violated its policies were recommended or otherwise amplified by platform algorithms before and after those items were identified as in violation of the platform's policies. The bill would require these statistics to be disaggregated by the category of policy violated (e.g. hate speech, misinformation, harassment, etc.), and would require the first disclosure on or before July 1, 2023. The bill specifies that its provisions do not require the dissemination of confidential business information or trade secrets, and provides that a violation of the bill is subject to a civil penalty of up to $100,000 that may be recovered only in a civil action brought by the Attorney General.

   The language of the bill is not entirely clear with respect to specifically what statistics are being requested, nor how those statistics must be presented to the public. Indeed, in opposition to a previous version of the bill, a coalition of industry trade groups consisting of TechNet, NetChoice, and the California Chamber of Commerce argues that the requirement "to disclose 'the extent of dissemination of or engagement with the content' [is] needlessly vague. Requiring "statistics regarding the extent to which" violating content is amplified or recommended, without further clarification, could result in under informative or misrepresentative disclosures, potentially undermining the efficacy of the disclosure in informing the public.

Additionally, the bill does not clearly define how or where the public disclosure must be made. While a reasonable presumption would be that the disclosure is made on the website of the social media platform, a social media platform could theoretically comply with the bill by making a verbal public statement regarding these statistics once a year, reducing the availability of the disclosed information. To ensure the bill effectively accomplishes the author's intent, should this bill pass out of this Committee, as the bill moves through the legislative process, the author may wish to consider clarifying these requirements of the bill.

Still, given the lack of public transparency with respect to the roles algorithmic amplification and recommendation play in the propagation of problematic content on social media, this bill seems likely to facilitate both systematic research and public assessment of algorithmic content prioritization practices. Since all large social media platforms would be required to provide these statistics, users could more effectively compare how content amplification priorities of social media platforms manifest in practice, and assess the likelihood of exposure to problematic content accordingly. As a result, consumers may be able to make more informed choices with respect to the social media with they engage, and independent researchers would be able to more effectively identify best algorithmic practices to inform businesses and policymakers alike.

5) **Definition of "social media platform"**: As issues pertaining to social media have come into focus in the policymaking arena, this Legislature has generally struggled to consistently define what constitutes a "social media platform" for regulatory purposes. While certain services clearly constitute social media platforms, some services maintain social components but may not be appropriately subject to the same regulations. For instance, while canonical social media platforms such as Facebook and Twitter invariably fall within scope, those that permit sharing fitness information with friends or transferring money along with descriptive messages may or may not. Depending on the nature of the legislation in question, the appropriate contours of services captured may indeed vary, but a consistent starting point to define the universe of services being discussed would arguably facilitate thoughtful policymaking.

   Toward this end, this Committee, in collaboration with the Senate Judiciary Committee, endeavored to develop a uniform definition of "social media platform" to use consistently across bills pertaining to social media that are currently pending. The definition fundamentally relies on essential aspects of social media platforms, including a substantial function of interacting socially, the ability to establish connections with others, and the creation or sharing of content. Importantly, this definition is not intended to preclude this bill or future legislation from further defining the contours of the policy's scope, including through additional exemptions where appropriate (e.g. based on platform size or revenue).

   The definition, which is predicated on an accompanying definition of "content," has been incorporated into this bill as it is proposed to be amended and reads as follows:

   > (a)(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.

(2) "Content" does not include media put online exclusively for the purpose of cloud storage, transmitting documents, or file collaboration.

(b) "Social media platform" means an 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 and allow users to interact socially with each other within the service or application.

(B) A service or application for which a substantial function is the conveyance of email or direct messages 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 or application.

(B) Populate a list of other users with whom an individual shares a 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.

6) **Bill does not appear to raise First Amendment concerns or federal preemption concerns under Section 230**: Section 230 of the federal Communications Decency Act of 1996 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," 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).) Though Section 230 was originally passed in response to judicial inconsistency with respect to the liability of internet service providers under statutes pertaining to "publishers" of content created by others, it has since been interpreted to confer operators of social media platforms and other online services with broad immunity from liability for content posted on their platforms by others.

Section 230 also indicates that "[n]othing in this section shall be construed to prevent any State from enforcing any State law that is consistent with this section," but further provides that "[n]o cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section." (47 U.S.C. Sec. 230(e)(3).) The latter provision has generally been interpreted to expressly preempt any state law that has the effect of treating a social media or other online platform as the publisher of information posted by other users, including prescriptive requirements relating to content moderation. This is consistent with the law's original intent, which was to ensure that internet platforms facilitating the sharing of content can do so without considerable risk of liability in the event that content is not meticulously policed.

The First Amendment of the U.S. Constitution provides that "Congress shall make no law […] abridging the freedom of speech […]" (U.S. Const., 1st Amend.), and courts have

consistently held that this prohibition on legislation abridging speech applies to state and local governments. (*See,* e.g., *Gitlow v. New York* (1925) 268 U.S. 652.) Courts have further established the contours of First Amendment protection of speech to include prohibitions against government compellation of speech and against laws that serve the purpose of chilling speech on the basis of content, even if the law itself does not explicitly ban certain speech.

The sole requirement of this bill, that social media platforms disclose statistics regarding the extent to which content violating the platform's policies has been amplified or recommended by the platform's algorithms before and after being identified as violating policies, does not appear to run afoul of either of these laws. The bill imposes no particular requirements with respect to the nature of content moderation or amplification practices a platform must adopt, and does not prohibit the platform from hosting any speech based on its content. Rather, the bill simply requires transparency with respect to these practices. While the bill does compel the disclosure of certain business information, leniency with respect to compelled commercial speech, along with the state interest served by publishing the required information, seems likely to survive the appropriate level of scrutiny to comply with the First Amendment.

7) **Double referral**: This bill has been double-referred to the Assembly Judiciary Committee, where the bill will be analyzed should it pass out of this Committee. The Assembly Judiciary Committee has historically been responsible for analyzing issues of federal preemption and First Amendment constitutionality across a broad range of contexts. While the constitutionality of the disclosure required by this bill is a critical policy consideration, in this case, it is arguably more appropriately addressed by the committee of second referral based on jurisdictional precedent.

8) **Related legislation**: AB 587 (Gabriel) 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 concerning specified content moderation practices to the Attorney General.

   AB 1628 (Ramos) would require an online platform, as defined, that operates in this state to create and publicly post a policy statement including specified information pertaining to the use of the platform to illegally distribute controlled substances.

   AB 2826 (Muratsuchi) would require the Department of Technology to establish a program to identify qualified research projects and require online platforms to turn over research material for those projects; and would require the Department of Technology to submit annual reports to the Legislature concerning research projects approved and conducted.

   SB 1056 (Umberg) would require a social media platform, as defined, to state whether it has a mechanism for reporting violent posts, as defined, and, if it does, to include a link to the reporting mechanism in that statement; and would permit a person who is the target of a violent post to seek a court order requiring the social media platform to remove the violent post.

9) **Prior legislation**: AB 13 (Chau, 2021) would have enacted the Automated Decision Systems Accountability Act of 2021 and stated 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. This bill was held under submission in the Senate Appropriations Committee.

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

AB 1379 (E. Garcia, 2021) would have prohibited 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. This bill died in the Assembly Elections Committee.

SB 388 (Stern, 2021) would have required a social media platform company, as defined, with 25,000,000 or more unique monthly users, as specified, to report to the Department of Justice specified information pertaining to its efforts to prevent, mitigate the effects of, and remove potentially harmful content. This bill died in the Senate Judiciary Committee.

AB 2442 (Chau, 2020) would have required social media companies to disclose whether or not they have a policy concerning misinformation. This bill died in the Senate Judiciary Committee.

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.

SB 1424 (Pan, 2018) would have established a privately-funded advisory group to study the spread of false information on social media platforms, and would have tasked the advisory group with drafting a model strategic plan for social media platforms to use to mitigate the problem. This bill was vetoed by Governor Brown, whose veto message indicated that a statutory advisory group was not necessary because there is already extensive research and investigation concerning the spread of false information on social media.

REGISTERED SUPPORT / OPPOSITION:

Support

ProtectUS (sponsor)
Oakland Privacy

Opposition

California Chamber of Commerce (previous version)
NetChoice (previous version)
Stand Up Sacramento County (previous version)
TechNet (previous version)
The People's Movement – California (previous version)

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