

SENATE THIRD READING
SB 1018 (Pan)
As Amended August 24, 2022
Majority vote

SUMMARY

This bill requires a social media platform, as defined, to disclose to the public on or before October 1, 2024, and annually thereafter, statistics regarding the extent to which, during the third and fourth quarters of the preceding calendar year and the first and second quarters of the current calendar year, 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.

Major Provisions

- 1) Requires a social media platform to disclose to the public, on or before October 1, 2024, and annually thereafter, statistics regarding the extent to which, during the third and fourth quarters of the preceding calendar year and the first and second quarters of the current calendar year, 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) Provides 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 in the name of the people of the State of California by the Attorney General, by any district attorney, by any city attorney of a city having a population in excess of 750,000, by a county counsel of any county within which a city has a population in excess of 750,000, by any city attorney of any city and county, or, with the consent of the district attorney, by a city prosecutor in any city having a full-time city prosecutor, in any court of competent jurisdiction.
- 3) Specifies that the required disclosure pursuant to 1), above, does not require the dissemination of confidential business information or trade secrets.
- 4) Exempts from the requirements of this bill a social media platform with fewer than 1 million discrete monthly users.

COMMENTS

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. 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 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. 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 parameters for algorithmic amplification and the prominence of undesirable content would have remained private.

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.

Specifically, this bill would require a social media platform with more than 1,000,000 discrete monthly users to disclose to the public on or before October 1 annually statistics regarding the extent to which, during the third and fourth quarters of the preceding calendar year and the first and second quarters of the current calendar year, 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 October 1, 2024. 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, a district attorney, or another public attorney.

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, 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.

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.

Arguments in Support

Oakland Privacy and Media Alliance argue:

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.

Arguments in Opposition

TechNet, CalChamber, and NetChoice, who all oppose this bill unless amended, argue:

SB 1018 requires social media platforms to report statistics regarding items of content that were determined to have violated our policies both before *and* after they were identified as violative. While reporting whether items of content were recommended or amplified *after* they were identified as violative of one of our policies would be burdensome but possible for most social media platforms, reporting that information *before* will be extremely burdensome for even the largest platforms and nearly impossible for smaller platforms. Hundreds of millions of new items of content are added to social media platforms every day. This bill would require companies to engineer new data tracking and reporting processes for every piece of content that hits their site, rather than just for items that violate their policies, which will likely cost platforms tens of millions and significant employee hours.

FISCAL COMMENTS

According to the Assembly Appropriations Committee, with respect to a previous version of this bill:

- 1) Costs (General Fund (GF)) of \$221,000 in fiscal year (FY) 2022-23, \$388,000 in FY 2023-24, and annually thereafter in additional legal staff to file injunctive relief against any social media platform that violates the requirements of this bill.
- 2) Cost pressures (Trial Court Trust Fund (TCTF)) to the trial courts in the low-to-mid-hundreds of thousands of dollars to hear and adjudicate civil actions against social media platforms that do not provide annual statistics about content that violates platform policies before and after that content was amplified to users. It is unclear how many new claims will be filed statewide, but if 10 cases are filed in state civil court annually requiring three to five days, or 24 to 40 hours, of court time, at an average cost per hour of \$1,000 in workload costs, the cost to the trial courts would be between \$240,000 and \$400,000 annually. Although courts are not funded on the basis of workload, increased pressure on the TCTF and staff workload may create a need for increased funding for courts from the GF to perform existing duties.

VOTES

SENATE FLOOR: 30-9-1

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

NO: Bates, Borgeas, Dahle, Grove, Jones, Melendez, Nielsen, Ochoa Bogh, Wilk

ABS, ABST OR NV: Hertzberg

ASM PRIVACY AND CONSUMER PROTECTION: 9-2-0

YES: Gabriel, Bauer-Kahan, Bennett, Berman, Cunningham, Mike Fong, Irwin, Wicks, Wilson

NO: Kiley, Valladares

ASM JUDICIARY: 9-2-0

YES: Stone, Cunningham, Bloom, Haney, Kalra, Maienschein, Reyes, Mia Bonta, Wicks

NO: Davies, Kiley

ASM APPROPRIATIONS: 12-4-0

YES: Holden, Bryan, Calderon, Arambula, Mike Fong, Gabriel, Eduardo Garcia, Levine, Quirk, Robert Rivas, Akilah Weber, McCarty

NO: Bigelow, Megan Dahle, Davies, Fong

UPDATED

VERSION: August 24, 2022

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

FN: 0004072