

AB 587 – SOCIAL MEDIA TRANSPARENCY AND ACCOUNTABILITY

AB 587 HELPS ADL COMBAT ONLINE HATE

We know that social media platforms play a large role in spreading hate speech, racism, and misinformation. In Facebook's case, blockbuster disclosures by whistleblower Frances Haugen, a former product manager on the company's now-disbanded civic integrity team, revealed that the company buried internal research showing it was aware of the ways the platform creates societal harm. **AB 587 helps fight online hate by making social media platforms more accountable by requiring them to publicly disclose their policies and enforcement regarding online hate/racism, disinformation, extremism, harassment and foreign political interference.**

WHY NOW?

In recent years, there has been growing concern around the role of social media in promoting hate speech, conspiracy theories, violent extremism, and severe political polarization. Further, investigations have shown that recent violent hate crimes have their origins on social media platforms. **For example, the deadly insurrection at the Capitol on January 6, 2021 was planned out in the open on mainstream social media platforms.**

Despite these concerns, social media platforms have taken inadequate steps to manage harmful content or provide consistent transparency about their policies and enforcement. Currently, platforms' transparency reports have failed to provide meaningful information about the prevalence and impact of hateful content. These reports are confusing, uninformative, and evasive. Platform transparency reports end up serving as a deflection away from the fact that we have very little insight into how platforms manage the ecosystem of hate online—and they have little to no legal or financial incentives to give us that access.

BILL BASICS

AB 587 would require social media platforms to publicly disclose their corporate policies regarding online hate/racism, disinformation, extremism, harassment, and foreign political interference. It would also require disclosure of key metrics and data around the enforcement of those policies. The disclosure would be accomplished through quarterly public filings with the Attorney General as outlined below:

Quarterly Reports

1. Disclose corporate policies, if any, regarding:
 - a. Hate speech and racism
 - b. Disinformation or misinformation, including health disinformation, election disinformation, and conspiracy theories
 - c. Extremism, including threats of violence against government entities
 - d. Harassment
 - e. Foreign political interference
2. Disclose efforts to enforce those policies.
3. Disclose any changes to policies or enforcement.
4. Disclose key metrics and data regarding the categories of content above, including the number of pieces of content, groups, and users flagged for violation; the method of flagging (e.g., human reviewers, artificial intelligence, etc.); the number of actions; and the type of enforcement action.

NOTABLE BILL SUPPORTERS: ORGANIZATIONS

- Anti-Defamation League (Bill Sponsor)
- Common Sense Media
- Decode Democracy
- National Hispanic Media Coalition

AB 587 TIMELINE

- **February 11, 2021:** First reading in the Assembly
- **March 25, 2021:** Referred to Committee on Privacy and Consumer Protection and Judiciary
- **June 2, 2021:** Passed after third read in the Assembly
- **June 3, 2021:** First read in the Senate
- **June 16, 2021:** Referred to Judiciary Committee
- **Fall 2021:** Senate Judiciary Committee Informational Hearing
- **January 2022:** Senate Judiciary Committee vote
- **2022:** Appropriations
- **2022:** Senate vote
- **2022:** Governor's desk for signature