

More State Content Moderation Laws Coming to Social Media Platforms

California and New York recently passed laws that seek to change how "social media platforms" and "social media networks" (both of which are broadly defined terms) design and report their content moderation practices. In doing so, they become the latest states—following Texas and Florida—to try their hands at platform regulation.

The new California laws require social media platforms to provide terms of service and extensive detail on their content moderation practices in transparency reports (AB 587), create content policies on the illegal distribution of controlled substances (AB 1628), and state whether they have reporting mechanisms for violent threats (SB 1056). New York's law requires social media networks to provide hateful conduct policies and reporting mechanisms for such conduct (S.4511-A/A.7865 or the NY law).

Constitutional challenges are expected in both states. In the meantime, service providers that may be subject to these new state laws can refer to an overview of their requirements and some practical steps to consider below.

Providers in Scope: California

The California laws define a social media platform as an internet-based service or application that meets both of the following criteria:

- **The service or application offers social interaction features.** This is defined as online services or applications that allow users to (1) make a public profile, (2) create a list of shared connections that they have with other users, and (3) create or post content that others can see (such as in chat rooms, on message boards, or on home or landing pages that show content created by other users).
- **A "substantial function" of the service or application is to connect users and allow them to interact socially.** The laws do not define substantial function, although the definition excludes companies that solely allow employees and affiliates of an organization to communicate internally, as well as any platforms that solely allow email or direct messaging services.

While the largest and highest-profile platforms may generate the most scrutiny, these broad definitions may subject other online service providers with substantial social components to investigation or enforcement. However, small platforms that do not exceed certain thresholds are excluded from the laws. AB 587 and AB 1628 exclude platforms that generated less than \$100 million in gross revenue in the preceding calendar year, and SB 1056 excludes platforms with fewer than one million discrete monthly users.

Under all three California laws, "content" is defined as statements, comments, or media created by and for users to interact with, share, and post. The laws exclude information that is uploaded to a platform solely for the purposes of cloud storage, file transmission, or file collaboration.

AB 587: Content Moderation Transparency

AB 587 requires extensive content moderation transparency. Social media platforms must (1) publicly post terms of service and (2) submit semiannual transparency reports to California's attorney general, which will be made publicly available on the attorney general's website.

Noteworthy Requirements

- **Terms of service.** Platforms must create and publicly post terms of service that describe violations of the terms and how users may report violating content. Platforms must also disclose the company's commitment to responding within a set time frame and provide contact information for questions.
- **Policies and enforcement data related to specific categories of content.** Transparency reports must include information on whether the platform maintains content policies on hate speech or racism, extremism or radicalism, disinformation or misinformation, harassment, and foreign political interference. The reports must also include data on how content was actioned under those policies.
- **Detailed descriptions of content moderation practices.** Platforms must disclose their content moderation practices, including enforcement methods, how the platform handles user reports, and, if applicable, languages in which the platform offers product features but not terms of service.

Timing and Enforcement

Platforms must submit their first transparency reports to the California attorney general by January 1, 2024, and cover content from the third quarter of 2023. The second report is due by April 1, 2024, containing data from the fourth quarter of 2023. After April 1, 2024, platforms must submit reports semiannually by October 1 and April 1 of each year. These semiannual reports must contain the information from the first two quarters of the current fiscal year and the last two quarters of the previous fiscal year, respectively.

Any platform that fails to post its terms of service, fails to submit reports to the attorney general, or materially omits or misrepresents any information will be in violation of this act. "Meaningful [civil] penalties sufficient to induce compliance with this act" may be awarded through an action brought by the attorney general or certain city attorneys, although damages may not exceed \$15,000 per violation per day. Courts will look to see if a platform made a reasonable, good-faith attempt to comply with the act before determining penalties.

AB 1628: Illegal Distribution of Controlled Substances

AB 1628 requires platforms to create and publicly post a policy statement that addresses the use of the platform for the illegal distribution of controlled substances.

Noteworthy Requirements

- **Content policy on illegal substances.** In addition to stating the platform's policy on the illegal distribution of controlled substances, the policy must include a description of the platform's practices in place to prevent the posting or sharing of such content. Platforms may exclude any information they believe may inhibit their ability to identify prohibited content or user activity or otherwise endanger users.
- **Reporting policies.** Platforms must share the link to their reporting mechanism for illegal or harmful content or behavior if one exists.
- **Education links.** Platforms must post a link to government public health resources for mental health and drug education.
- **Government requests.** Platforms must share their policies for responding to law enforcement requests such as subpoenas or warrants.

Timing and Enforcement

This law does not provide for any specific enforcement mechanisms. It is intended to be in effect from January 1, 2023, to January 1, 2028.

SB 1056: Violent Posts

SB 1056 requires platforms to "clearly and conspicuously" state whether they have mechanisms for reporting "violent posts" against users and nonusers. It also creates an avenue for court action for users mentioned in such posts.

Noteworthy Requirements

- **Reporting and removal mechanism for "violent posts."** A violent post is content that contains a true threat against a specific person and is not protected by the First Amendment. An unprotected "true threat" is one that explicitly intends to place the recipient in fear of actual harm. If a platform has a mechanism for reporting violent posts, it must state it clearly and conspicuously and provide a link to it. Platforms with such mechanisms must remove reported content within 48 hours or risk the court action described below.
- **Court action.** If a person is, or reasonably believes they are, the target of a violent post on a social media platform, they may bring an action in California court seeking removal of the violent content and any related violent content. Individuals may request these orders at any time if a platform does not have a reporting mechanism. Platforms with a reporting mechanism will be given 48 hours to remove the post before the court rules on the takedown request. Courts may also award court costs and reasonable attorneys' fees to prevailing plaintiffs.

Timing and Enforcement

SB 1056 will go into effect on January 1, 2023. As noted above, this law contemplates enforcement through private actions seeking courts to order the takedown of violent posts.

Providers in Scope: New York

The New York law, adopted in June 2022, requires social media networks that operate in New York to provide and maintain a mechanism for users to report hateful conduct. The law defines social media networks as service providers that operate internet platforms for profit-making purposes that are designed to let users share content with other users or to make it available to the public. Unlike the California laws, the New York law does not limit its application to networks that meet certain revenue or monthly user thresholds.

Noteworthy Requirements

- **Hateful conduct reporting mechanism.** The New York law requires social media networks to provide a hateful conduct reporting mechanism that is clear and accessible to users via the network's website and application. Users must receive a direct response from the social media network to each report submitted. The law broadly defines hateful conduct as using a social media network to vilify, humiliate, or incite violence against protected classes.
- **Hateful conduct policy.** Social media networks must have a clear and concise policy that addresses how the network will respond to and address hateful conduct reports.

Timing and Enforcement

This law will go into effect on December 3, 2022, and will be enforced by New York's attorney general. Social media networks that knowingly fail to comply with the law can be assessed a civil penalty of up to \$1,000 per day.

Next Steps for Online Service Providers

For online service providers that allow users to share content, the following are some steps to consider:

- Determine whether the service is in scope for one or more of these new social media laws. For California, this should include the consideration of whether the service's social interaction features constitute a substantial function of the service.
- If the service is in scope, assess the obligations and risks specific to the company and develop a compliance plan.

A compliance plan for these laws may benefit from the consideration of overlapping obligations arising from a number of global online safety laws. Companies that may be in scope should consult with experienced counsel to understand their obligations and how they may intersect with related global obligations.

© 2022 Perkins Coie LLP

Authors



Madeline McFee

Associate

MMcFee@perkinscoie.com [206.359.3056](tel:206.359.3056)



Naa Kai Koppoe

Associate

NKKoppoe@perkinscoie.com [310.788.3256](tel:310.788.3256)

Explore more in

[Privacy & Security](#) [Technology Transactions & Privacy Law](#) [Communications](#)

Related insights

Update

[**Ninth Circuit Rejects Mass-Arbitration Rules, Backs California Class Actions**](#)

Update

[**CFPB Finalizes Proposed Open Banking Rule on Personal Financial Data Rights**](#)