

ANDREAS BORGEAS  
VICE CHAIR

MEMBERS

ANNA CABALLERO  
MARIA ELENA DURAZO  
LENA A. GONZALEZ  
ROBERT M. HERTZBERG  
BRIAN W. JONES  
JOHN LAIRD  
HENRY STERN  
ROBERT A. WIECKOWSKI  
SCOTT D. WIENER

# California Legislature

## Senate Committee on Judiciary

THOMAS J. UMBERG  
CHAIR



MARGIE ESTRADA CANIGLIA  
CHIEF COUNSEL

TIMOTHY S. GRIFFITHS  
CHRISTIAN A. KURPIEWSKI  
AMANDA MATTSON  
ALLISON MEREDITH  
COUNSEL

ERICA PORTER  
CLARE RECINOS  
COMMITTEE ASSISTANTS

STATE CAPITOL  
ROOM 2187  
SACRAMENTO, CA 95814  
TEL (916) 651-4113  
FAX (916) 403-7394

## State of Social Media Regulation: Misinformation, Exploitation, Harassment, and Radicalization

Tuesday, November 9, 2021

9:30 a.m.

Location: State Capitol, Room 4203

---

### BACKGROUND PAPER

#### **I. Introduction**

Social media platforms have existed for fewer than 25 years, yet are ubiquitous and affect nearly every facet of many people's lives. The size and scope of many social media platforms give them unprecedented power over what people read, what they believe, and how they vote; it is not an exaggeration to say that social media platforms are one of the driving forces behind the current state of the country.

While social media platforms provide many benefits, many have expressed concern about social media platforms' disproportionately negative influence on issues ranging from mental health to politics to crime. Social media platforms are, after all, businesses; what is best for individuals and for the country might not be best for the platforms' profits. Yet social media platforms have been given wide discretion to function without government regulation, in large part due to a federal law passed in 1995, long before social media and the technologies that enable it (e.g., location tracking, content algorithms) existed.

This hearing will investigate what measures the State of California can and should take to reduce the harms caused by social media platforms. In particular, the hearing will examine the state of the law — including the constitutional and federal barriers to state

regulation – and what steps are available to the state in order to better protect its residents.

## II. What is social media?

Although most people familiar with the internet are also familiar with the concept of social media, the terms “social media” and “social media platform” have proven difficult to define with specificity.<sup>1</sup> While the term “social media” is defined frequently in state law, the definitions appear in contexts where a broad definition is desirable (e.g., protecting online privacy) and include communication methods such as email and text messages;<sup>2</sup> such direct and private communication tools do not seem to fall within the popular understanding of social media.

Crafting a precise definition of social media for the purpose of regulation is a challenge this Committee has faced, but this background paper will not attempt to devise a definitive framework. Instead, this background paper will discuss “social media” as online platforms that more or less comport with the following criteria:

- The platform allows users to create a profile for purposes of signing into and using the service.
- The platform provides a landing site on which users may post comments and/or content generated from other sources; the user’s landing site may be public or limited to a set number of users, depending on the platform and the user’s preferences.
- The platform allows users to connect with other users in order to see the other users’ posts and content.
- The platform presents connected users’ content in a main “feed” – the content of which is often determined by algorithm – and also allows users to specifically seek out other users’ content, subject to those other users’ privacy settings.
- A main purpose of the platform is to facilitate social interaction and/or comments between users who post content and the users who view the content.

Examples of social media platforms include Facebook, YouTube, Instagram, Twitter, TikTok, and Snapchat. Excluded from this definition are platforms primarily dedicated to transmitting information or content without a social component (e.g., DropBox, Netflix), as are email and direct texting media.

---

<sup>1</sup> Nowhere in the United States Code are “social media” or “social media platform” defined – even in the statute establishing the federal Social Media Working Group. (See 6 USC § 195d.)

<sup>2</sup> See, e.g., Ed. Code, § 99120 (“As used in this chapter, ‘social media’ means an electronic service or account, or electronic content, including, but not limited to, videos or still photographs, blogs, video blogs, podcasts, instant and text messages, email, online services or accounts, or Internet Web site profiles or locations”).

### III. The existing legal framework governing speech on social media

Speech on social media is governed by two primary sources: the First Amendment of the United States Constitution and Section 230 of the Communications Decency Act, 47 U.S.C. § 230 (Section 230).

The First Amendment, as applied to the states through the Fourteenth Amendment, prohibits Congress or the states from passing any law “abridging the freedom of speech.”<sup>3</sup> “[A]s a general matter, the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content.”<sup>4</sup> However, while the amendment is written in absolute terms, the courts have created a handful of narrow exceptions to the First Amendment’s protections, including defamation,<sup>5</sup> obscenity,<sup>6</sup> “true threats,”<sup>7</sup> “fighting words,”<sup>8</sup> and incitement to imminent lawless action.<sup>9</sup> Expression on the internet is given the same measure of protection granted to in-person speech or statements published in a physical medium.<sup>10</sup> Accordingly, a social media user may generally post content and comments free from government regulation, but may incur civil or criminal liability if their comment falls within one of the First Amendment exceptions. At the same time, social media platforms themselves – as private businesses – are not subject to the constraints of the First Amendment and may limit or prohibit users’ speech on their sites as they see fit.<sup>11</sup>

Section 230 does not apply to the *users* of social media (or the internet generally), but rather applies to the *platforms themselves*. In the early 1990s, prior to the enactment of Section 230, two trial court orders – one in the United States District Court for the Southern District of New York, and New York state court – suggested that internet platforms could be held liable for allegedly defamatory statements made by the platforms’ users if the platforms engaged in any sort of content moderation (e.g., filtering out offensive material).<sup>12</sup> In response, two federal legislators and members of

---

<sup>3</sup> U.S. Const., 1st & 6th amends.

<sup>4</sup> *Ashcroft v. American Civil Liberties Union* (2002) 535 U.S. 564, 573.

<sup>5</sup> *R.A.V. v. St. Paul* (1992) 505 U.S. 377, 383.

<sup>6</sup> *Ibid.*

<sup>7</sup> *Snyder v. Phelps* (2011) 562 U.S. 443, 452.

<sup>8</sup> *Cohen v. California* (1971) 403 U.S. 15, 20.

<sup>9</sup> *Virginia v. Black* (2003) 538 U.S. 343, 359.

<sup>10</sup> *Reno v. ACLU* (1997) 521 U.S. 844, 870.

<sup>11</sup> E.g., *Hudgens v. NLRB* (1976) 424 U.S. 507, 513. Some have argued that certain social media platforms are so essential to the freedom of expression that they should be treated as common carriers subject to the First Amendment; this issue is discussed in Part V, *infra*.

<sup>12</sup> See *Cubby, Inc. v. Compuserve, Inc.* (S.D.N.Y. 1991) 776 F.Supp. 135, 141; *Stratton Oakmont v. Prodigy Servs. Co.* (N.Y. Sup. Ct., May 26, 1995) 1995 N.Y. Misc. LEXIS 229, \*10-14. These opinions relied on case law developed in the context of other media, such as whether book stores and libraries could be held liable for distributing defamatory material when they had no reason to know the material was defamatory. (See *Cubby, Inc.*, 776 F. Supp. at p. 139; *Smith v. California* (1959) 361 U.S. 147, 152-153.)

the burgeoning internet industry crafted a law that would give internet platforms immunity from liability for users' statements, even if they might have reason to know that statements might be false, defamatory, or otherwise actionable.<sup>13</sup> The result – Section 230 – was relatively uncontroversial at the time, in part because of the relative novelty of the internet and in part because Section 230 was incorporated into a much more controversial internet regulation scheme that was the subject of greater debate.<sup>14</sup>

Section 230 begins with findings and a statement of policy that extol the value of the internet and the intention to let the internet develop without significant government regulation.<sup>15</sup> The crux of Section 230 is then laid out in two parts. The first 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.”<sup>16</sup> The second provides a safe harbor for content moderation, by stating that no provider or user shall be held liable because of good-faith efforts to restrict access to material that is “obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.”<sup>17</sup>

Together, these two provisions give platforms immunity from any civil or criminal liability that could be incurred by user statements, while explicitly authorizing platforms to engage in their own content moderation without risking that immunity. Section 230 specifies that it does not preempt federal criminal laws, but that “[n]o cause of action may be brought and no liability may be imposed under any State law that is inconsistent with this section.”<sup>18</sup> Since its passage, Congress has created one exemption to Section 230 to allow online platforms (including social media platforms) to be held liable if it was aware that the platform was being used to traffic children for sex.<sup>19</sup> Congress has also altered the liability with respect to hosting copyrighted material by allowing platforms to be held liable for users' copyright violations unless the platform blocks access to alleged infringing material upon receiving a notice of infringement.<sup>20</sup>

---

<sup>13</sup> Kosseff, *The Twenty-Six Words That Created The Internet* (2019) pp. 57-65.

<sup>14</sup> *Id.* at pp. 68-73. Section 230 was added to the Communications Decency Act of 1996 (title 5 of the Telecommunications Act of 1996, Pub. L. 104-104, 110 Stat. 56), which would have imposed criminal liability on internet platforms if they did not take steps to prevent minors from obtaining “obscene or indecent” material online. The Supreme Court invalidated the CDA, except for Section 230, on the basis that it violated the First Amendment. (*See Reno, supra*, 521 U.S. at p. 874.)

<sup>15</sup> 47 U.S.C. § 230(a) & (b).

<sup>16</sup> *Id.*, § 230(c)(1).

<sup>17</sup> *Id.*, § 230(c)(1) & (2).

<sup>18</sup> *Id.*, § 230(e)(1) & (3).

<sup>19</sup> *Id.*, § 230(e)(5). This exemption was enacted in 2018, as part of the Stop Enabling Sex Traffickers Act and the Allow States to Fight Online Sex Trafficking Act (SESTA/FOSTA) legislation package. (*See* P.L. 115-164, 113 Stat. 1253.) SESTA/FOSTA is discussed further in Part V.

<sup>20</sup> *See* 17 U.S.C. § 512, the Digital Millennium Copyright Act (DCMA).

Section 230 uses terminology generally applicable in defamation cases (e.g., “publisher,” “speaker”), but courts interpreting Section 230 did not limit its application to the defamation context. Instead, courts have applied Section 230 in a vast range of cases to immunize internet platforms from “virtually all suits arising from third-party content.”<sup>21</sup> Courts have even extended Section 230 immunity to situations where the platform’s moderator affirmatively solicited the information, selected the user’s statement for publication, and/or edited the content.<sup>22</sup>

Of course, Section 230 does not prevent a user from seeking redress from the actual *speaker* of a statement that is defamatory, threatening, or otherwise actionable. As a practical matter, however, this is rarely a realistic option. It can be difficult or impossible to determine the identity of an anonymous online speaker. Even if a plaintiff is able to obtain the speaker’s IP address from the social media platform via subpoena – subpoenas that are often met with challenges from the social media platform – the IP address often cannot be connected to a single user, leaving the plaintiff with no defendant from whom to seek recourse.<sup>23</sup> Moreover, individual speakers are often essentially judgment-proof, so even if the subject of defamation can find the speaker, the subject is unlikely to be able to obtain compensation for what can be life-shattering reputational, psychological, or other injuries.

Thus, counterintuitively, the breadth of Section 230 appears to grant speech on social media *greater* protections than speech elsewhere. Section 230 immunizes the social media platforms that enable abusers to post and publicize harassment, threats, and lies – indeed, allowing those abusers to spread hate and harassment to a wider audience than would be possible without the platform – while the nature of these platforms make it more difficult to locate the actual speaker than it would if the speaker were using a different medium. As discussed in the next section, the relative impunity with which speakers can spread false, hateful, or criminal information has caused social media platforms to be the source of significant harms.

---

<sup>21</sup> Kosseff, *supra*, fn. 13, at pp. 94-95; *see, e.g., Doe v. MySpace Inc.* (5th Cir. 2008) 528 F.3d 413, 421-422; *Carfano v. Metrospalsh.com, Inc.* (9th Cir. 2003) 339 F.3d 1119, 1125; *Zeran v. America Online, Inc.* (4th Cir. 1997) 129 F.3d 327, 333-334.

<sup>22</sup> *See, e.g., Jones v. Dirty World Entertainment Recordings LLC* (6th Cir. 2014) 755 F.3d 398, 415; *Batzel v. Smith* (9th Cir. 2003) 333 F.3d 1018, 1030-1031; *cf. Blumenthal v. Drudge* (D.D.C. 1998) 992 F.Supp. 44, 51-52.

<sup>23</sup> Kosseff, *supra*, fn. 13, at pp. 221-222.

#### IV. A brief overview of the problems with social media

While social media platforms undoubtedly provide societal benefits – by enabling friends and family to stay in touch, by allowing like-minded people to connect, and by amplifying the voices of those who have not been given space in traditional media – many argue that social media platforms are also responsible for great societal harm. The myriad concerns are documented extensively in multiple sources; this paper will address several of the most prominent issues in brief.

Many of the problems below are exacerbated by social media platforms' use of algorithms intended to keep users engaged with the platform – more user engagement generally translates to more profit. The use of algorithms has, however, resulted in harmful content being amplified – sometimes, as alleged by whistleblowers, with social media platforms' knowledge.<sup>24</sup> Social media companies also generally collect large amounts of data about their users, which raise both privacy concerns and can make it easier for bad actors to target individuals for, e.g., harassment or radicalization. Some of the recommendations for reforming social media do not target specific subject areas, but rather focus on obtaining more transparency into, and potentially limiting the use of, algorithms and data collection;<sup>25</sup> these are discussed further in Part V.

##### a. Misinformation and disinformation

Evidence suggests that social media platforms make it easier to spread misinformation (information the user does not know is false) and disinformation (information the user knows is false). The ease with which such false information can be shared – and can create an echo chamber that reinforces the “truth of the falsity” – has already led to dire consequences for our country's political and physical health.

An investigation by a select committee of the United States Congress determined that the 2016 U.S. presidential election was tainted by interference through a social media “information warfare campaign designed to spread disinformation and societal division in the United States.”<sup>26</sup> Although the interference was identified well before the next U.S. presidential election, evidence suggests that social media platforms were then used in 2020 to spread disinformation intended to prevent individuals from voting (often

---

<sup>24</sup> E.g., Zakrzewski, et al., *Facebook whistleblower Frances Haugen tells lawmakers that meaningful reform is necessary 'for our common good'*, Washington Post (Oct. 5, 2021), <https://www.washingtonpost.com/technology/2021/10/05/facebook-senate-hearing-frances-haugen/> [last visited Nov. 1, 2021].

<sup>25</sup> *Ibid.*

<sup>26</sup> Select Committee on Intelligence, *Russian Active Measures, Campaigns, and Interference in the 2016 U.S. Election*, United States Senate (Nov. 10, 2020), [https://www.intelligence.senate.gov/sites/default/files/documents/Report\\_Volume2.pdf](https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf) [last visited Oct. 26, 2021].

voters of color);<sup>27</sup> to spread lies about the integrity of the voting process;<sup>28</sup> and to perpetuate the dangerous myth that former President Donald J. Trump had actually won the election.<sup>29</sup> Social media platforms were then the planning grounds for the January 6, 2021, violent attack on the United States Capitol by Trump supporters.<sup>30</sup>

Studies also indicate that misinformation poses a danger to public health. One study found that the more people rely on social media as their main news source, the more likely they are to believe misinformation about the COVID-19 pandemic.<sup>31</sup> Another found that a mere 12 people are responsible for 65 percent of the false and misleading claims about COVID-19 vaccines on Facebook, Instagram, and Twitter.<sup>32</sup>

Misinformation can also hinder emergency responses to natural disasters, when social media posts contain incorrect or out-of-date information.<sup>33</sup>

---

<sup>27</sup> Fessler, *Robocalls, Rumors And Emails: Last-Minute Election Disinformation Floods Voters*, NPR (Oct. 24, 2020), <https://www.npr.org/2020/10/24/927300432/robocalls-rumors-and-emails-last-minute-election-disinformation-floods-voters> [last visited Oct. 26, 2021]; Bond, *Black and Latino Voters Flooded With Disinformation in Election's Final Days*, NPR (Oct. 30, 2020), <https://www.npr.org/2020/10/30/929248146/black-and-latino-voters-flooded-with-disinformation-in-elections-final-days> [last visited Oct. 26, 2021].

<sup>28</sup> Frenkel, *How Misinformation 'Superspreaders' Seed False Election Theories*, New York Times (Nov. 23, 2020), <https://www.nytimes.com/2020/11/23/technology/election-misinformation-facebook-twitter.html> [last visited Oct. 26, 2021].

<sup>29</sup> Bump, *The chain between Trump's misinformation and violent anger remains unbroken*, Washington Post (May 12, 2021), <https://www.washingtonpost.com/politics/2021/05/12/chain-between-trumps-misinformation-violent-anger-remains-unbroken/> [last visited Oct. 26, 2021].

<sup>30</sup> Timberg, Dwoskin, & Albergotti, *Inside Facebook, Jan. 6 violence fueled anger, regret over missed warning signs*, Washington Post (Oct. 22, 2021) <https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/> [last visited Oct. 26, 2021]. Facebook has rejected its Oversight Board's recommendation that it study how its policies contributed to the violence on that day. (*Ibid.*)

<sup>31</sup> Su, *It doesn't take a village to fall for misinformation: Social media use, discussion heterogeneity preference, worry of the virus, faith in scientists, and COVID-19-related misinformation belief* (May 2021) *Telematics and Information*, Vol. 58, <https://www.sciencedirect.com/science/article/abs/pii/S0736585320302069?via%3Dihub> [last visited Oct. 26, 2021].

<sup>32</sup> Bond, *Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows*, NPR (May 14, 2021), <https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes> [last visited Oct. 26, 2021]. At a recent hearing before a subcommittee of the United States Senate Commerce Committee, a witness from YouTube testified that the platform has removed over a million videos related to COVID-19 misinformation and over 130,000 videos relating to COVID-19 vaccine misinformation since implementing anti-COVID-19 misinformation measures. (*See* U.S. Sen. Commerce Com., Subcom. on Consumer Protection, Product Safety, and Data Security (Oct. 26, 2021), testimony of Leslie Miller). However, YouTube did not suspend the accounts of many prominent anti-vaccine activists until September 2021. (De Vynck, *YouTube is banning prominent anti-vaccine activists and blocking all anti-vaccine content*, Washington Post (Sept. 29, 2021) <https://www.washingtonpost.com/technology/2021/09/29/youtube-ban-joseph-mercola/> [last visited Oct. 27, 2021].

<sup>33</sup> United States Department of Homeland Security, *Countering False Information on Social Media in Disasters and Emergencies* (March 2018),

## b. Harassment, bullying, and abuse

While social media brings people together, it also provides unprecedented avenues for harassment, bullying, and abuse. Research suggests that abuse and hate speech on social media are disproportionately directed at LGBTQ users<sup>34</sup> and women,<sup>35</sup> and particularly women of color; Amnesty International’s Troll Patrol reviewed millions of tweets on Twitter and found that women of color were 34 percent more likely than white women to be mentioned in such tweets, and that Black women were 84 percent more likely than white women to be mentioned in such tweets.<sup>36</sup> Forms of harassment include sexual harassment, swatting, doxing, physical threats (including rape threats), revenge porn, and stalking.<sup>37</sup> Such harassment is often intended to, and successful in, intimidating and silencing voices who threaten the status quo and challenge oppressive systems.<sup>38</sup> Thus, ironically, the “free speech” so prized on the internet turns into a zero sum game: by empowering harassers and abusers to engage in their vile tactics, there is less room for the women, people of color, and LGBTQ users.

## c. Mental health

Studies indicate that social media usage is correlated to increased negative mental health impacts.<sup>39</sup> In documents disclosed by whistleblower Frances Haugen, it was revealed that Facebook, the parent company of Instagram, had internal research showing that Instagram increases negative body image in around one-third of teen girls who use the platform, and that use of the platform can increase depression in teens.<sup>40</sup> Relatedly, reports show that Instagram, other social media platforms (such as Snapchat), and third-party content producers provide filters for photos that make users look thinner and, in some cases, paler; these filters create unrealistic beauty standards,

---

[https://www.dhs.gov/sites/default/files/publications/SMWG\\_Counteracting-False-Info-Social-Media-Disasters-Emergencies\\_Mar2018-508.pdf](https://www.dhs.gov/sites/default/files/publications/SMWG_Counteracting-False-Info-Social-Media-Disasters-Emergencies_Mar2018-508.pdf).

<sup>34</sup> Diaz, *Social Media Hate Speech, Harassment ‘Significant’ Problem for LGBTQ Users: Report*, NPR (May 10, 2021), <https://www.npr.org/2021/05/10/995328226/social-media-hate-speech-harassment-significant-problem-for-lgbtq-users-report> [last visited Oct. 28, 2021].

<sup>35</sup> E.g., Nadim & Fladmoe, *Silencing Women? Gender and Online harassment*, *Social Science Computer Review* 2021, Vol. 39(2) (2021), at pp. 246-247 (2021).

<sup>36</sup> See Amnesty International, *Troll Patrol Findings*, <https://decoders.amnesty.org/projects/troll-patrol/findings> [last visited Oct. 28, 2021].

<sup>37</sup> ADL, *Online Hate and Harassment: The American Experience 2020* (2020), at p. 8, <https://www.adl.org/online-hate-2020> [last visited Oct. 28, 2021].

<sup>38</sup> E.g., Vilks, et al., *No Excuse for Abuse: What Social Media Companies Can Do Now to Combat Online Harassment and Empower Users*, PEN America (2020) <https://pen.org/report/no-excuse-for-abuse/> [last visited Oct. 28, 2021].

<sup>39</sup> E.g., Bouygues, *Social Media Is a Public Health Crisis. Let’s Treat It Like One.*, *U.S News and World Report* (Jul. 20, 2021), <https://www.usnews.com/news/health-news/articles/2021-07-20/social-media-is-a-public-health-crisis> [last visited Oct. 26, 2021].

<sup>40</sup> See Wells, Horwitz, & Seetharaman, *Facebook Knows Instagram Is Toxic For Teen Girls, Company Documents Show*, *Wall Street Journal* (Sept. 14, 2021).



perpetuate racism and colorism, and damage users' self-esteem.<sup>41</sup> Journalists have also reported that Facebook programmed content-boosting algorithms to prioritize posts that produce anger or contain toxic content.<sup>42</sup>

The potential breadth of the negative health consequences of social media use, along with social media platforms' knowledge of those consequences and failure to protect their users, has led some to compare social media platforms to Big Tobacco.<sup>43</sup>

#### d. Radicalization and crime

The ability to remain anonymous online makes the internet as a whole a petri dish for crime and radicalization into terrorist and hate groups. Evidence suggests that social media platforms are often aware that their platforms are being used for illegal activities and hate group recruitment, but fail to act. For example, the Wall Street Journal reports that Facebook was aware that its platform was being used by Mexican drug cartels to recruit, train, and pay hit men, by Middle Eastern human traffickers to facilitate their trafficking, and to incite violence against the Rohingya population in Myanmar, and was aware that its mitigation attempts were insufficient.<sup>44</sup> An administrator of the U.S. Drug Enforcement Administration has specifically called out Snapchat and TikTok for taking insufficient steps to prevent drug trafficking on their platforms, particularly of pills laced with fentanyl.<sup>45</sup>

Reports also indicate that social media platforms are valuable recruiting tools for terrorist and extremist groups, both domestically and abroad. One 2016 study showed that both the Islamic State (ISIS) and American white nationalist and Nazi sympathizers used Twitter as a key recruitment tool, leading to significant growth.<sup>46</sup> The "alt-right" community has also exploded in size due to online engagement, usually by adolescent

---

<sup>41</sup> Ryan-Mosley, *How digital beauty filters perpetuate colorism*, Technology Review (Aug. 15, 2021), <https://www.technologyreview.com/2021/08/15/1031804/digital-beauty-filters-photoshop-photo-editing-colorism-racism/> [last visited Oct. 26, 2021].

<sup>42</sup> Merrill & Oremus, *Five points for anger, one for a "like"; How Facebook's formula fostered rage and misinformation*, Washington Post (Oct. 26, 2021), <https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/> [last visited Oct. 26, 2021].

<sup>43</sup> E.g., Stelter, *Wall Street Journal's Facebook Files series prompts comparisons to Big Tobacco*, CNN.com (Sept. 18, 2021), <https://www.cnn.com/2021/09/18/media/reliable-sources-facebook-files/index.html> [last visited Oct. 26, 2021].

<sup>44</sup> Scheck, Purnell, & Horwitz, *Facebook Employees Flag Drug Cartels and Human Traffickers. The Company's Response Is Weak, Documents Show.*, Wall Street Journal (Sept. 16, 2021).

<sup>45</sup> Lerman & De Vynck, *Snapchat, TikTok, Instagram face pressure to stop illegal drug sales as overdose deaths soar*, Washington Post (Sept. 28, 2021), <https://www.washingtonpost.com/technology/2021/09/28/tiktok-snapchat-fentanyl/> [last visited Oct. 26, 2021].

<sup>46</sup> Berger, *Nazis v. ISIS on Twitter: A Comparative Study of White Nationalist and ISIS Online Social Media Networks*, George Washington University Program on Extremism (Sept. 2016), at pp. 3-4.

or young men, who are wooed with memes and videos that slowly introduce them to white supremacist, anti-Semitic, and misogynistic ideologies.<sup>47</sup> The QAnon conspiracy, one of the driving forces behind the January 6 attack on the Capitol, was an online phenomenon; despite warning signs that the conspiracy would turn violent, Twitter waited until after the attack to remove QAnon-connected profiles.<sup>48</sup>

## V. Proposed and attempted solutions

Calls to regulate or reform social media have been made more or less since the invention of social media. This section discusses a range of measures that are under discussion some approaches that have been attempted in the past (with varying success).

### a. Increased social media platform self-regulation

Section 230 has been described as “an implicit contract between Congress and the technology community: if online platforms develop responsible and reasonable moderation procedures, Congress will grant them extraordinary legal immunity.”<sup>49</sup> While many social media platforms do engage in content moderation to benefit the common welfare—for example, by removing some COVID-19-related misinformation—many believe that social media platforms are not living up to their end of the bargain.<sup>50</sup> A recent Congressional Research Services Report discussed the issue of content moderation and specifically the spread of misinformation and the role that social media platforms play in worsening the issue:

Two features of social media platforms—the user networks and the algorithmic filtering used to manage content—can contribute to the spread of misinformation. Users can build their own social networks, which affect the content that they see, including the types of misinformation they may be exposed to. Most social media operators use algorithms to sort and prioritize the content placed on their sites. These

---

<sup>47</sup> E.g., Schroeder, *Racists Are Recruiting. Watch Your White Sons*, New York Times (Oct. 12, 2019), <https://www.nytimes.com/2019/10/12/opinion/sunday/white-supremacist-recruitment.html> [last visited Oct. 28, 2021].

<sup>48</sup> Bond, *Unwelcome on Facebook and YouTube, QAnon Followers Flock to Fringe Sites*, NPR (Jan. 31, 2021), <https://www.npr.org/2021/01/31/962104747/unwelcome-on-facebook-twitter-qanon-followers-flock-to-fringe-sites> [last visited Oct. 28, 2021].

<sup>49</sup> Kosseff, *supra*, fn. 13, at p. 250.

<sup>50</sup> See, e.g., Scott & Lerman, *Biden clarifies comments about Facebook ‘killing people’ with vaccine information*, Washington Post (July 19, 2021), <https://www.washingtonpost.com/politics/2021/07/19/biden-facebook-misinformation/> [last visited Oct. 27, 2021]; Schechner, Horwitz, & Glazer, *How Facebook Hobbled Mark Zuckerberg’s Bid to Get America Vaccinated*, Wall Street Journal (Sept. 17, 2021); Horwitz, *Facebook Says Its Rules Apply to All. Company Documents Reveal a Secret Elite That’s Exempt*, Wall Street Journal (Sept. 13, 2021).

algorithms are generally built to increase user engagement, such as clicking links or commenting on posts. In particular, social media operators that rely on advertising placed next to user-generated content as their primary source of revenue have incentives to increase user engagement. These operators may be able to increase their revenue by serving more ads to users and potentially charging higher fees to advertisers. Thus, algorithms may amplify certain content, which can include misinformation, if it captures users' attention.<sup>51</sup>

The role that content moderation, or the lack thereof, has in alleviating or exacerbating these issues has been a source of much debate, mostly centered on social media platforms' use of algorithms. As disclosed by Haugen and other sources, many social media platforms' algorithms will recommend harmful, divisive, or false content even where a user was not looking for it.<sup>52</sup>

These algorithms and other content moderation tools remain, first and foremost, under the control of the social media platforms themselves. At a number of Congressional hearings, representatives from various social media platforms have testified that their platforms' self-regulation efforts have been, or can be altered to be, sufficient to avoid the need for further governmental regulation.<sup>53</sup>

#### b. Federal action

As discussed above, Section 230 gives social media platforms enormous power to decide which voices to boost, bury, or block. And unlike the state and federal governments, social media platforms are not constrained by the First Amendment — meaning they can prioritize or silence favored and disfavored voices without any legal recourse from the silenced. As a result of this power to shape democratic discourse, it has been posited that social media platforms “should be thought of the New Governors of online speech,” who may or may not be motivated to moderate content consistently with the United States' interests.<sup>54</sup>

---

<sup>51</sup> Jason A. Gallo & Clare Y. Cho, *Social Media: Misinformation and Content Moderation Issues for Congress* (January 27, 2021) Congressional Research Service, <https://crsreports.congress.gov/product/pdf/R/R46662>.

<sup>52</sup> E.g., Oremus, *Lawmakers' latest idea to fix Facebook: Regulate the algorithm*, Washington Post (Oct. 12, 2021), <https://www.washingtonpost.com/technology/2021/10/12/congress-regulate-facebook-algorithm/> [last visited Oct. 28, 2021].

<sup>53</sup> E.g., Lerman & Lima, *TikTok, Snap, YouTube defend how they protect kids online in congressional hearing*, Washington Post (Oct. 26, 2021), <https://www.washingtonpost.com/technology/2021/10/26/tiktok-snapchat-youtube-congress-hearing/> [last visited Oct. 27, 2021].

<sup>54</sup> Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598, 15603 (Apr. 2018). Mark Zuckerberg has even said that “ ‘Facebook is more like a government than a traditional company.’ ” (*Id.* at p. 1599.)

In recent weeks, revelations about Facebook have spurred an increased interest in the problems of social media and discussions about federal legislation to limit social media platforms' power.<sup>55</sup> There are several bills pending in both houses of Congress that would regulate or limit social media by, for example, creating a federal agency to regulate data-collection practices,<sup>56</sup> requiring social media platforms to give users the option to use the platform without algorithms that use data not expressly provided by the user,<sup>57</sup> and requiring social media platforms to obtain user consent before collecting and selling user data.<sup>58</sup>

Congress could also create additional exemptions to Section 230, as it did with SESTA/FOSTA, which allows a social media platform to be held liable if it is aware that child sex trafficking is being conducted on the platform,<sup>59</sup> and the DMCA, which allows platforms to be liable for users' copyright violations if they do not promptly remove the material after being notified of the alleged violation.<sup>60</sup> Both former President Trump and President Biden have expressed interest in repealing or revising Section 230.<sup>61</sup>

Amending Section 230, however, is not without its downsides, as demonstrated by SESTA/FOSTA itself. While SESTA/FOSTA was intended to make it easier to prosecute child sex trafficking, a report from the Government Accountability Office (GAO) found that the post-SESTA/FOSTA landscape "heightens already-existing challenges law enforcement face in gathering tips and evidence" about child sex trafficking.<sup>62</sup> In brief, the GAO reported that SESTA/FOSTA caused many legitimate and U.S.-based platforms to shut down pages that were commonly used by sex traffickers as well as others for legitimate purposes, but the sex traffickers simply moved to other platforms, many of which are located overseas, making it more difficult to gather evidence to prosecute the traffickers *or* the platforms.<sup>63</sup> The DMCA has also been abused to remove legitimate speech, such as scientific and religious disagreement.<sup>64</sup>

---

<sup>55</sup> E.g., Lima, *Lawmakers hammer Facebook for hiding how its products hurt kids*, Washington Post (Sept. 30, 2021), <https://www.washingtonpost.com/technology/2021/09/30/facebook-instagram-congress-hearing-antigone-davis/> [last visited Oct. 27, 2021].

<sup>56</sup> S.2134 (117th Congress, 2021-2022).

<sup>57</sup> S.2024 (117th Congress, 2021-2022).

<sup>58</sup> S.1477 (117th Congress, 2021-2022).

<sup>59</sup> Kosseff, *supra*, fn. 13, at pp. 269-270.

<sup>60</sup> 17 U.S.C. § 512.

<sup>61</sup> E.g., Guynn, *Donald Trump and Joe Biden vs. Facebook and Twitter: Why Section 230 could get repealed in 2021*, USA Today (Jan. 5, 2021), <https://www.usatoday.com/story/tech/2021/01/04/trump-biden-pelosi-section-230-repeal-facebook-twitter-google/4132529001/> [last visited Oct. 27, 2021].

<sup>62</sup> United States Government Accountability Office, Report to Congressional Committees, *Sex Trafficking: Online Platforms and Federal Prosecutions*, No. 21-385 (June 2021), at p. 20.

<sup>63</sup> *Id.* at pp. 20-25.

<sup>64</sup> Keller, *The Right Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation*, 33 Berkeley Tech. L.J. 297, 308-309 (2018).

Many free speech advocates warn that limiting Section 230 would result in a similar restriction of other types of speech. They argue that social media platforms, under the threat of legal liability, would always err on the side of removing or prohibiting content, which would disproportionately harm and silence marginalized people whose voices were historically not given space in mainstream press outlets.<sup>65</sup> If, for example, a social media platform could face legal and monetary liability for users' defamatory statements, the platform would have less incentive to keep up any post flagged as defamatory, even if it is not. Under such a system, a person could prevent the spread of criticism about them by registering complaints with the platform.<sup>66</sup> Such concerns may induce Congress to go in a different direction, as it has in many of the pending social media regulation bills.

Finally, there has been a push to employ antitrust laws in order to limit the reach of individual social media companies. Although a federal judge dismissed an antitrust suit filed by the Federal Trade Commission and 46 states in 2020 against Facebook – on the ground that the plaintiffs waited too long to sue after Facebook consolidated its market share by purchasing competitors – the plaintiffs have filed an amended complaint.<sup>67</sup> Federal legislators have also introduced a package of five bills targeted at constraining the technology industry's consolidation, including consolidation by Facebook and other social media companies.<sup>68</sup>

### c. State laws

At least two states have passed legislation to regulate the content of social media. In 2021, both Florida and Texas passed legislation not to protect users of social media from harassment or misinformation, but to prohibit what they perceived as unfair moderation of certain viewpoints, namely posts advocating conservative positions.

Florida's SB 7072 prohibited a social media platform from deplatforming a candidate for office while that user was actively running for office,<sup>69</sup> in direct response to Facebook and Twitter's bans of former President Trump following the January 6, 2021,

---

<sup>65</sup> See Common Cause, et al., Letter from 70+ human rights and social justice groups opposing repeal of or overbroad changes to Section 230 (Jan. 27, 2021), <https://www.fightforthefuture.org/news/2021-01-27-letter-from-70-human-rights-and-social-justice/> [last visited Oct. 27, 2021].

<sup>66</sup> Such a system would make the internet in the U.S. look more like the internet in Europe, which does not have a Section 230 analog. (Kosseff, *supra*, fn. 13, at pp. 147-148.)

<sup>67</sup> See Federal Trade Commission, Press Release, *FTC Alleges Facebook Resorted to Illegal Buy-or-Bury Scheme to Crush Competition After String of Failed Attempts to Innovate* (Aug. 19, 2021), <https://www.ftc.gov/news-events/press-releases/2021/08/ftc-alleges-facebook-resorted-illegal-buy-or-bury-scheme-crush> [last visited Oct. 27, 2021].

<sup>68</sup> See Kang, *Lawmakers Aim at Big Tech, Push Sweeping Overhaul of Antitrust*, NY Times (Jun. 11, 2021), <https://www.nytimes.com/2021/06/11/technology/big-tech-antitrust-bills.html> [last visited Oct. 21, 2021].

<sup>69</sup> SB 7072 (Fla. 2021).

attack on the United States Capitol.<sup>70</sup> The law’s prohibition was predicated on the Legislature’s finding that “[s]ocial media platforms hold a unique place in preserving first amendment [*sic*] protections for all Floridians and should be treated similarly to common carriers”<sup>71</sup>; under this common carrier analogy, social media platforms would be subject to certain First Amendment requirements, thus curtailing the platforms’ right to restrict content or block users unless the users’ speech fell into one of the narrow exceptions to the First Amendment.<sup>72</sup> A federal court enjoined enforcement of the law in June 2021, rejecting the common carrier analogy and finding that the plaintiffs were likely to prevail on their claims that the law violated the First Amendment.<sup>73</sup> Florida appealed the ruling, and the case is pending before the United States Court of Appeals for the Eleventh Circuit.<sup>74</sup>

Texas’s law relies on the same “common carrier” approach as the now-blocked Florida law, but prohibits an even greater range of social media content-moderation activities.<sup>75</sup> The bill prohibits a social media platform from “censoring” a user based on the user’s viewpoint or the viewpoint of another person.<sup>76</sup> Like the Florida law, the Texas law was adopted in response to the state’s Republicans’ belief that tech companies have an anti-conservative bias and that social media companies are attempting to “ ‘silence conservative ideas [and] religious beliefs.’ ”<sup>77</sup> Litigation is currently pending to invalidate the Texas law as contrary to the First Amendment.<sup>78</sup>

---

<sup>70</sup> McCabe, *Florida, in a First, Will Fine Social Media Companies That Ban Candidates*, New York Times (May 24, 2021), <https://www.nytimes.com/2021/05/24/technology/florida-twitter-facebook-ban-politicians.html> [last visited Oct. 26, 2021].

<sup>71</sup> SB 7072 (Fla. 2021).

<sup>72</sup> *E.g., Biden v. Knight First Amendment Inst.* (2021) 141 S. Ct. 1220, 1224-1227 (Thomas, J., concurring) (arguing that digital platforms, should be treated and regulated as common carriers subject to the First Amendment).

<sup>73</sup> Zakrewski, *Federal judge blocks Florida law that would penalize social media companies*, Washington Post (Jun. 30, 2021), available at <https://www.washingtonpost.com/technology/2021/06/30/florida-social-media-law-trump/> [last visited Oct. 26, 2021]. The court found that, among other things, the law likely violated the First Amendment because its exception for “social-media providers under common ownership with a large Florida theme park” could not survive strict scrutiny. (*Netchoice, LLC v. Ashley Brooke Moody*, Case No. 4:21cv220-RH-MAF, Order, Docket No. 113 (Jun. 30, 2021).)

<sup>74</sup> Aspuru, *Florida awaits appeal decision on new social media law; legal experts say it probably won’t survive*, WUFT.org (Oct. 1, 2021), <https://www.wuft.org/news/2021/10/01/florida-awaits-appeals-decision-on-new-social-media-law-legal-experts-say-it-probably-wont-survive/> [last visited Oct. 26, 2021].

<sup>75</sup> See H.B. 20 (Tex. 2021).

<sup>76</sup> *Ibid.* The bill also requires social media platforms to make certain disclosures with respect to their content moderation polices, and requires social media platforms to provide a written notice and the opportunity to appeal to any user whose content is removed for violation the platform’s acceptable use policy.

<sup>77</sup> Pollard, *Texas sued over bill stopping social media companies from banning users for political views*, Texas Tribune (Sept. 22, 2021), available at <https://www.texastribune.org/2021/09/21/texas-social-media-law/> [last visited Oct. 26, 2021].

<sup>78</sup> *Ibid.*

## **VI. Issues to consider going forward**

California has already taken steps to protect residents from certain online threats, such as the Consumer Privacy Act of 2018 and criminalizing revenge porn. These measures do not directly regulate social media, but give Californians the tools to minimize or seek recourse for some of the worst elements of social media. To the extent the Legislature seeks to regulate social media platforms as such, considerations could include:

- Whether any measure is consistent with the First Amendment of the United States Constitution and the State Constitution's guarantee of free expression;
- Whether any measure is preempted by Section 230 or other federal law;
- Whether there are elements of social media platform activity other than direct content moderation that could be regulated, such as transparency and data protection;
- Whether a measure is overbroad and will chill legitimate speech;
- Whether a measure might inadvertently drive bad actors off of U.S.-based sites without slowing their behavior, making it more difficult for law enforcement to gather evidence.