



---

3-2-2021

## Reimagining Content Moderation: Section 230 and the Path to Industry-Government Cooperation

Yeva Mikaelyan  
*LMU Loyola Law School, Los Angeles*

Follow this and additional works at: <https://digitalcommons.lmu.edu/elr>



Part of the [Entertainment, Arts, and Sports Law Commons](#)

---

### Recommended Citation

Yeva Mikaelyan, *Reimagining Content Moderation: Section 230 and the Path to Industry-Government Cooperation*, 41 Loy. L.A. Ent. L. Rev. 179 (2021).

Available at: <https://digitalcommons.lmu.edu/elr/vol41/iss2/3>

This Notes and Comments is brought to you for free and open access by the Law Reviews at Digital Commons @ Loyola Marymount University and Loyola Law School. It has been accepted for inclusion in Loyola of Los Angeles Entertainment Law Review by an authorized administrator of Digital Commons@Loyola Marymount University and Loyola Law School. For more information, please contact [digitalcommons@lmu.edu](mailto:digitalcommons@lmu.edu).

---

## Reimagining Content Moderation: Section 230 and the Path to Industry-Government Cooperation

### Cover Page Footnote

J.D. Candidate, 2021, LMU Loyola Law School, Los Angeles. The author would like to thank Professor Aaron Caplan for his incredible guidance and support throughout this process. Professor Caplan, thank you for constantly challenging and improving my ideas. The author would also like to thank Jameson Evans, Natalie Kalbakian, Fernanda Sanchez Jara, and the entire staff of Loyola of Los Angeles Entertainment Law Review for their hard work, feedback, and support on this Note. Finally, the author would like to thank Alena McLucas for taking one of our midnight conversations down the rabbit hole that spawned the idea for this Note

# REIMAGINING CONTENT MODERATION: SECTION 230 AND THE PATH TO INDUSTRY- GOVERNMENT COOPERATION

*Yeva Mikaelyan\**

In February 2020, the Ninth Circuit held that YouTube, as a private entity, does not have to provide First Amendment protections to its content creators. The holding was not surprising or groundbreaking, but the case served as catalyst in the discussion of how platforms should moderate content. This was further amplified when over the summer, Twitter started to add warnings under some of President Donald Trump’s tweets. In response, the President called to “REVOKE 230.”

“230” refers to Section 230 of the Communications Decency Act. At a high level, Section 230 allows platforms to moderate content at their discretion without fear of liability. But today, platforms dominate much of the national discourse and heavily influence politics and government. This Note explores the role of Section 230 in today’s social media-dominated world and critiques the recent push to regulate Section 230. This Note argues that the First Amendment and government intervention are ineffective and dangerous tools in Section 230 regulation. Instead, this Note proposes a collaborative industry-government approach that balances on the needs of social media platforms, government, users, and content creators.

---

\*J.D. Candidate, 2021, LMU Loyola Law School, Los Angeles. The author would like to thank Professor Aaron Caplan for his incredible guidance and support throughout this process. Professor Caplan, thank you for constantly challenging and improving my ideas. The author would also like to thank Jameson Evans, Natalie Kalbakian, Fernanda Sanchez Jara, and the entire staff of *Loyola of Los Angeles Entertainment Law Review* for their hard work, feedback, and support on this Note. Finally, the author would like to thank Alena McLucas for taking one of our midnight conversations down the rabbit hole that spawned the idea for this Note.

## I. INTRODUCTION

Early on the morning of May 29, 2020, the President of the United States, Donald Trump, posted a tweet.<sup>1</sup> It read: “REVOKE 230!”<sup>2</sup>

Though it was only two words, the Tweet spoke to an issue brewing for years in the internet space—whether the government should regulate content on social media platforms or whether platforms should continue to regulate themselves.

A few days before posting the tweet, President Trump tweeted that allowing mail-in voting would lead to mass voter fraud and lead to a “Rigged Election.”<sup>3</sup> Twitter responded by adding a warning below his tweet and flagging it as false and misleading.<sup>4</sup> Throughout the next few months, Twitter continued to add warning labels under President Trump’s tweets<sup>5</sup> and committed to flagging other false and misleading information on the platform.<sup>6</sup>

The ability of Twitter to flag and add warnings to his tweets was the exact reason that President Trump had called to revoke “230.” “230” refers to section 230 (“Section 230”) of the Communications Decency Act

---

1. Donald J. Trump (@realDonaldTrump), TWITTER (May 29, 2020, 8:15 AM), <https://twitter.com/realDonaldTrump/status/1266387743996870656> [<https://archive.is/iBmsK>].

2. *Id.*

3. Donald J. Trump (@realDonaldTrump), TWITTER (May 26, 2020, 5:17 AM), <https://twitter.com/realDonaldTrump/status/1265255835124539392> [<https://archive.is/XJ8eS>].

4. *Trump Makes Unsubstantiated Claim That Mail-In Ballots Will Lead to Voter Fraud*, TWITTER (May 26, 2020), <https://twitter.com/i/events/1265330601034256384> [<https://archive.is/YMrHI>].

5. During the national protests in the aftermath of George Floyd’s killing, President Trump tweeted that the military would intervene if the protests involved looting, tweeting “when the looting starts, the shooting starts.” Twitter added a content warning to the tweet, asserting that the tweet was “glorifying violence” but that it would not be taken down because “it may be in the public’s best interest for the Tweet to remain accessible.” Donald J. Trump (@realDonaldTrump), TWITTER (May 28, 2020, 9:53 AM), <https://twitter.com/realDonaldTrump/status/1266231100780744704> [<https://archive.is/Ghu6u>].

6. Lauren Feiner, *Twitter CEO Stands by Fact-Check on Trump’s Tweets as the White House Prepares an Executive Order on Social Media Bias*, CNBC (May 28, 2020, 10:22 AM), <https://www.cnbc.com/2020/05/28/twitter-ceo-stands-by-fact-check-on-trumps-tweets.html> [<https://perma.cc/3J48-BBUM>].

(“CDA”),<sup>7</sup> a law that was passed in 1996 because of growing concern that minors could easily access pornography on the internet.<sup>8</sup> Though the Supreme Court of the United States deemed many parts of the CDA unconstitutional, Section 230 remains in effect.<sup>9</sup> At a high level, Section 230 provides platforms with broad authority to remove user-generated content as they see fit and shields them from civil liability for most content moderation decisions.<sup>10</sup>

The ability of platforms to censor content at their discretion often causes tension<sup>11</sup> between regulators, platforms, content creators, and users.<sup>12</sup> In February 2020, these tensions flared when content creator Prager University opposed a content moderation decision by Google’s video hosting platform, YouTube.<sup>13</sup> Prager University is not a traditional university.<sup>14</sup> It is “a

---

7. 47 U.S.C. § 230(e) (2018).

8. William A. Sodeman, *Communications Decency Act*, ENCYCLOPEDIA BRITANNICA (Nov. 24, 2016), <https://www.britannica.com/topic/Communications-Decency-Act> [<https://perma.cc/8BPL-YH3Y>].

9. *Reno v. ACLU*, 521 U.S. 844, 844 (1997) (holding that the portions of the CDA banning “indecent” and “offensive” material are unconstitutional); *CDA 230: The Most Important Law Protecting Internet Speech*, ELEC. FRONTIER FOUND., <https://www.eff.org/issues/cda230/legislative-history> [<https://perma.cc/5EM2-WVWW>].

10. 47 U.S.C. § 230(c)(2); Section 230’s protections and limitations are described in greater detail in Part III.

11. *E.g.*, Michelle Castillo, *Why Some YouTube Stars Are Angry at the Platform*, CNBC (Sept. 7, 2016, 3:13 PM), <https://www.cnbc.com/2016/09/07/why-some-youtube-stars-are-angry-at-the-platform.html> [<https://perma.cc/UE7Y-PFG8>] (discussing that YouTube creators are angry at the platform for censoring sensitive content); Kevin Roose, *The President Versus the Mods*, N.Y. TIMES (Sept. 16, 2020), <https://www.nytimes.com/2020/05/29/technology/trump-twitter.html> [<https://perma.cc/Q3W4-DH6X>] (discussing Twitter users who are angry about Twitter’s warnings on some of President Trump’s tweets).

12. This article frequently uses the terms “content creator” and “user.” For the purposes of this article, a content creator is a person or group that creates and posts user-generated content, such as someone who posts a YouTube video or a tweet. A user is a person who consumes content, but does not necessarily create content, such as a person with a YouTube account that does not post videos. Content creators are often also users and vice versa. While most content creators post content casually, some people or groups are content creators for a living and thus have an additional financial stake in platform content moderation decisions.

13. *See generally* Prager Univ. v. Google LLC, 951 F.3d 991 (9th Cir. 2020).

14. *Id.* at 995.

nonprofit educational and media organization with a mission to ‘provide conservative viewpoints and perspective on public issues that it believes are often overlooked.’”<sup>15</sup> Prager University’s YouTube channel features videos speaking out against the Black Lives Matter movement,<sup>16</sup> condemning the anti-fascist political group Antifa,<sup>17</sup> and promoting content from conservative commentators like Ben Shapiro<sup>18</sup> and Candace Owens.<sup>19</sup>

Prager University sued YouTube because YouTube placed some of Prager University’s videos on Restricted Mode, which is a platform setting that a user can choose to enable “that makes unavailable certain age-inappropriate content.”<sup>20</sup> Only about 1.5-2% of YouTube’s users enable restricted mode.<sup>21</sup> To determine which videos are limited by Restricted Mode, YouTube uses algorithms to flag sensitive content such as material related to drug use, sex, and violence and makes the flagged content unavailable to those that enable Restricted Mode.<sup>22</sup> On Prager University’s channel specifically, YouTube restricted videos pertaining to subjects such as abortion and “radical” Islam.<sup>23</sup>

YouTube also disabled monetization—or “demonetized”—several of Prager University’s videos, which is a process that removes advertisements

---

15. *Id.*

16. PragerU, ‘Black Lives Matter’ Is Not Helping Blacks, YOUTUBE (June 11, 2020), <https://www.youtube.com/watch?v=zxLi-6RtmUQ> [https://perma.cc/3RZH-9DXN].

17. PragerU, *Antifa Declared A Terrorist Group*, YOUTUBE (June 4, 2020), <https://www.youtube.com/watch?v=DbA5MOnxFxo> [https://perma.cc/BB5X-8JBG].

18. PragerU, *New Release This Monday*, YOUTUBE (July 18, 2020), <https://www.youtube.com/watch?v=DuYvc-bPsgQ> [https://perma.cc/GZ9T-X68K].

19. PragerU, *The Candace Owens Show: Dinesh D’Souza*, YOUTUBE (July 26, 2020), [https://www.youtube.com/watch?v=X5\\_DBg31ZWU](https://www.youtube.com/watch?v=X5_DBg31ZWU) [https://perma.cc/7QHR-LCJ8].

20. *Prager Univ.*, 951 F.3d at 996.

21. *Id.*

22. Johanna Wright, *An Update on Restricted Mode*, YOUTUBE OFFICIAL BLOG (Apr. 21, 2017), <https://blog.youtube/news-and-events/an-update-on-restricted-mode/> [https://perma.cc/7A58-JK4L].

23. *Editorial: YouTube has censorship problems. The 1st Amendment Isn’t the Tool to Fix Them*, L.A. TIMES (Feb. 27, 2020, 3:00 AM), <https://www.latimes.com/opinion/story/2020-02-27/prager-university-youtube-lawsuit-rejected-9th-circuit> [http://archive.vn/90KFA].

from a videos that would otherwise allow Prager University to profit financially from the video's views.<sup>24</sup> For a video to qualify for monetization, it has to follow YouTube's Community Guidelines, Terms of Service, copyright policies, and advertising policies.<sup>25</sup> YouTube's advertising guidelines are particularly elusive and restrict, among other things, content that is adult, hateful, demeaning, shocking, controversial, or sensitive.<sup>26</sup>

In its opinion in *Prager University v. Google*, the Ninth Circuit considered, among other things, whether YouTube infringed on Prager University's First Amendment right to free speech by restricting and demonetizing its videos.<sup>27</sup> Ultimately, the court held that because YouTube is a private company, it does not have to provide free speech protections to its content creators and may censor content on its platform at its discretion.<sup>28</sup> The court compared YouTube to a larger version of a private town bulletin board or coffee shop where the public could share ideas.<sup>29</sup> Although YouTube is certainly a private entity that hosts speech, the problem with this comparison is the extent to which YouTube offers a bigger platform. While a local coffee shop or bulletin board might reach a few hundred visitors, YouTube touts over 30 million visitors per day.<sup>30</sup>

The magnitude of YouTube cannot be ignored when discussing its place in society. YouTube, along with other major social networks like Facebook and Twitter, are massive platforms that, together, host billions of users.<sup>31</sup> In addition to providing entertainment, social media platforms provide

---

24. *Prager Univ.*, 951 F.3d at 996.

25. *YouTube Chanel Monetization Policies*, YOUTUBE (2020), <https://support.google.com/youtube/answer/1311392?hl=en> [<https://perma.cc/ZH9Q-AM6P>].

26. *Advertiser-Friendly Content Guidelines*, YOUTUBE, <https://support.google.com/youtube/answer/6162278> [<https://perma.cc/AL5H-K8EK>].

27. *Prager Univ.*, 951 F.3d at 996.

28. *Id.* at 996–97.

29. *Id.* at 995.

30. *Id.*

31. J. Clement, *Global Social Networks Ranked by Number of Users 2020*, STATISTA (Aug. 21, 2020), <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> [<https://perma.cc/R5NE-FXLV>].

Americans with information. In 2019, about 55% of American adults reported that they get their news from social media platforms such as Facebook and Twitter either “often” or “sometimes”—up by 8% from the year before.<sup>32</sup> Politicians can now directly post content for their supporters in a way that has never been done before; content creators can gain power by reaching followers around the world; and users can isolate themselves into virtual echo chambers by consuming only material that aligns with their views.<sup>33</sup> The ability of platforms to censor and promote content at their discretion gives them unprecedented power to influence opinions and discourse.

Though many Americans consider the current methods of content moderation problematic,<sup>34</sup> there is currently no consensus on a solution. Legal scholars have suggested various approaches such as treating companies as individual “governors” of their platforms,<sup>35</sup> comparing major platforms to nations with their own governments,<sup>36</sup> and modifying First Amendment doctrines to encompass regulation of platforms.<sup>37</sup>

This Note argues that the most optimal approach to content moderation focuses on the needs of content creators and users and involves both government and industry stakeholders. Part II explains content moderation and social media community standards in detail. Part III describes Section 230 protections and identifies two major issues with the current model of platform self-regulation: the constant threat of government interference and potential platform abandonment by users due to inconsistent and frustrating

---

32. Peter Suci, *More Americans Are Getting Their News from Social Media*, FORBES (Oct. 11, 2019, 10:35 AM), <https://www.forbes.com/sites/petersuci/2019/10/11/more-americans-are-getting-their-news-from-social-media/#4ab1c2ad3e17> [<https://perma.cc/323Z-KNGU>].

33. Sam Sanders, *Did Social Media Ruin Election 2016?*, NPR (Nov. 8, 2016, 5:00 AM), <https://www.npr.org/2016/11/08/500686320/did-social-media-ruin-election-2016> [<https://perma.cc/4AES-E62E>].

34. Suci, *supra* note 32 (describing a 2019 report in which “[88%] of Americans . . . recognized that social media companies now have at least some control over the mix of the news that people see each day [and] . . . 62% felt this was a problem”).

35. See generally Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1603 (2018).

36. See generally Anupam Chander, *Facebookistan*, 90 N.C. L. REV. 1807, 1815 (2012).

37. See generally Jonathan Peters, *The “Sovereigns of Cyberspace” and State Action: The First Amendment’s Application (or Lack Thereof) to Third-Party Platforms*, 32 BERKELEY TECH. L.J. 989 (2017).



content moderation decisions. Part IV explores the reasons that First Amendment free speech protections and government regulation are ineffective models for content moderation. Finally, this Note proposes a content moderation model which features cooperation between government and industry leaders and encompasses the values of transparency, accountability, and consistency.

## II. OVERVIEW OF CONTENT MODERATION

Why is it that we do not encounter violent videos, graphic images, or pornography more frequently on major social media platforms? The broad answer to this question is because platforms engage in content moderation. Content moderation is “the practice of flagging user-generated submissions based on a set of guidelines in order to determine whether the submission can be used or not in the related media.”<sup>38</sup> Section A of this Part explains the reasons that content moderation is in the best interest of users and platforms. Section B then describes how content moderation consists of a combination of human moderators, algorithms, and user reports and discusses the flaws inherent to each form.

### A. *Why Do Platforms Moderate Content?*

Platforms have both legal and social authority to moderate user-generated content. Section 230(c)(2) of the CDA provides platforms with the express legal authority to moderate content.<sup>39</sup> Beyond legal authorization, society has come to expect that platforms will moderate and filter certain content in order to protect users.<sup>40</sup> Platforms moderate content “out of a sense of corporate responsibility” that is based in “American free speech and

---

38. *What Is Content Moderation and Why Companies Need It*, BRIDGED (Sept. 3, 2019), <https://bridged.co/blog/what-is-content-moderation-why-companies-need-it/> [https://perma.cc/VE4J-UQWH].

39. 47 U.S.C. § 230(c)(2)(A)–(B) (2020) (stating that “[n]o provider or user of an interactive computer service shall be held liable on account of—(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).”).

40. See *What Is Content Moderation and Why Companies Need It*, *supra* note 38.

democratic values.”<sup>41</sup> Perhaps most importantly, a platform’s economic viability depends on content moderation, especially of violent or obscene content.<sup>42</sup> Imagine, for example, a world in which Facebook posed a constant risk of exposing users to videos of beheadings or pornographic images. Such a Facebook would be unusable and unacceptable to many users, likely resulting in those users abandoning the platform for its lack of moderation. Therefore, it is in a platform’s best interest to moderate content to ensure that the website is friendly to as many users as possible.

To properly moderate content, platforms today use sets of community guidelines to determine what sort of content is allowed on the platform.<sup>43</sup> If a post violates a platform’s community guidelines, the platform can moderate the content in various ways, such as issuing a warning over, removing, restricting, or demonetizing<sup>44</sup> the content.<sup>45</sup>

### B. *What Does Content Moderation Look Like?*

To conceptualize the monumental task of content moderation, it helps to understand that there are 317,000 Facebook statuses,<sup>46</sup> 347,000 tweets,<sup>47</sup>

---

41. Klonick, *supra* note 35, at 1625–26.

42. *Id.* at 1625.

43. *See id.* at 1631.

44. Demonetization is the process of disabling content’s ability to generate revenue through advertising income. James Johnson, *How to Survive YouTube Demonetization and Continue Making Money from Your Channel*, USCREEN (Aug. 8, 2019), <https://www.uscreen.tv/blog/how-to-survive-youtube-demonetization/> [https://perma.cc/YX88-8NXX].

45. *See generally* Jillian C. York & Corynne McSherry, *Content Moderation Is Broken. Let Us Count the Ways.*, ELEC. FRONTIER FOUND. (Apr. 29, 2019), <https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways> [https://perma.cc/7P6E-9ARB]; Christine Fisher, *YouTube Will Temporarily Increase Automated Content Moderation*, ENGADGET (Mar. 16, 2020), <https://www.engadget.com/2020-03-16-youtube-automated-content-moderation-coronavirus.html> [https://perma.cc/2JKT-RMW9].

46. *Facebook by the Numbers: Stats, Demographics and Fun Facts*, OMNICORE (Apr. 22, 2020), <https://www.omnicoreagency.com/facebook-statistics/> [https://perma.cc/EU89-A2FN].

47. *Twitter by the Numbers: Stats, Demographics and Fun Facts*, OMNICORE (Feb. 10, 2020), <https://www.omnicoreagency.com/twitter-statistics/> [https://perma.cc/P3AE-SQHU].

and 500 hours of YouTube video footage<sup>48</sup> posted every single minute. Social media platforms use a combination of human moderators, automated algorithms, and user reports to filter through the massive amount of content contained on their platforms.<sup>49</sup>

The need to moderate this material started with the rise of websites that featured large amounts of user-generated content, some of which could be inappropriate or undesirable for users to encounter.<sup>50</sup> The very first form of platform content moderation was performed by human employees.<sup>51</sup> Human content moderators still work in call center-like hubs around the world and review content reported by users.<sup>52</sup> Although these employees have handbooks—some up to 80 pages long<sup>53</sup>—that dictate what sort of content to remove, content moderation is an inevitably subjective and inconsistent art. Moderators can and do make exceptions for certain content based on its context.<sup>54</sup> For example, a platform could have a rule banning derogatory language, which may include the N-word. However, banning use of the N-word outright “could be completely insensitive to the African American community in the United States,” who may use the N-word in a non-derogatory manner.<sup>55</sup> Before making a content moderation decision, a content moderator would have to take into account more subjective factors such as who used the N-word and the context in which it was used. The process becomes thorny when content moderators, who often work in traumatic and mentally

---

48. *YouTube by the Numbers: Stats, Demographics and Fun Facts*, OMNICORE (Feb. 10, 2020), <https://www.omnicoreagency.com/youtube-statistics/> [<https://perma.cc/B67L-CRUL>].

49. *What Is Content Moderation and Why Companies Need It*, *supra* note 38.

50. Minna Ruckenstein & Linda Lisa Maria Turunen, *Re-humanizing the Platform: Content Moderators and the Logic of Care*, 22.6 NEW MEDIA & SOC’Y 1026, <https://journals.sagepub.com/doi/full/10.1177/1461444819875990#> [<https://perma.cc/JPP2-MD22>].

51. York & McSherry, *supra* note 45.

52. Casey Newton, *The Trauma Floor: The Secret Lives of Facebook Moderators in America*, VERGE (Feb. 25, 2019, 8:00 AM), <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> [<https://archive.is/Qb9vb>].

53. Klonick, *supra* note 35, at 1634.

54. *Id.* at 1634–35.

55. *Id.* at 1634.

draining environments,<sup>56</sup> are put in the position of making snap decisions about sensitive content. Such decisions consistently invite individual biases regarding race, gender, and sexual orientation to play a role in content moderation decisions.<sup>57</sup>

Because of the large amount of content uploaded to platforms daily, human teams are not sufficient to moderate content on their own. Though it is unclear exactly what percentage of content is moderated by algorithms as opposed to humans, platforms want algorithms to play an increasingly prominent role in content moderation.<sup>58</sup> An example of when an algorithm is used in content moderation is in the period of time between when a video is uploaded and when it is posted.<sup>59</sup> During this period, algorithms scan the uploaded content against digital libraries and search for prohibited material such as graphic imagery or copyrighted matter.<sup>60</sup> However, algorithms cannot exercise the same level of nuanced discretion as human moderators which can result in “an even more secretive process in which false positives may never see the light of day.”<sup>61</sup> For example, algorithms can detect nudity with a high rate of accuracy, but may have trouble discerning between

---

56. Human content moderators spend their shifts reviewing “hate speech, . . . violent attacks, [and] graphic pornography.” Because of the nature of the work, some employees have reported developing severe mental health issues. Newton, *supra* note 52.

57. Daisy Soderberg-Rivkin, *When It Comes to Content Moderation, We've Been Focusing on the Wrong Type of Bias*, MORNING CONSULT (Dec. 5, 2019 5:00 AM), <https://morningconsult.com/opinions/when-it-comes-to-content-moderation-weve-been-focusing-on-the-wrong-type-of-bias/> [<https://perma.cc/D9J2-SLT2>] (giving several examples of content moderation decisions motivated by bias, including Facebook’s quick removal of a post about the fragility of white men as hate speech and YouTube’s restricting of LGBTQ+-themed videos at a high rate).

58. See, e.g., Klonick, *supra* note 35, at 1636; Shannon Bond, *Facebook, YouTube Warn of More Mistakes as Machines Replace Moderators*, NPR (Mar. 31, 2020, 5:06 AM), <https://www.npr.org/2020/03/31/820174744/facebook-youtube-warn-of-more-mistakes-as-machines-replace-moderators> [<https://perma.cc/ZDF4-KQ6W>] (explaining that major platforms sent many human content moderators home due to the COVID-19 pandemic and plan to rely more heavily on automated content moderation methods in the future).

59. Klonick, *supra* note 35, at 1636.

60. *Id.* at 1637.

61. York & McSherry, *supra* note 45.

“Renaissance art . . . and banned content such as sexual activity.”<sup>62</sup> The struggle of algorithms to recognize nuance has led to the removal of otherwise appropriate content, including YouTube news coverage of extremist attacks that algorithms failed to distinguish from the extremist attacks themselves.<sup>63</sup>

Finally, a platform’s users also help with the content moderation process.<sup>64</sup> Platforms generally have reporting tools that individual users can use to flag content that they believe should not be on the platform.<sup>65</sup> The flagged content is then reviewed by moderators who decide whether or not to remove the content.<sup>66</sup> However, platform users often flag content inappropriately.<sup>67</sup> Most flagged content does not violate community standards but instead “often reflect[s] . . . conflicts or disagreements of opinion.”<sup>68</sup> The reporting process is also administered unevenly because it can result in flagging more visible or disliked people at a higher rate, even if similar content posted others with lower profiles is permitted.<sup>69</sup>

No single method of content moderation is sufficient on its own. Ultimately, whether content moderation remains controlled by platforms, becomes regulated by the government, or is implemented by another method, the combination of human content moderators, algorithms, and user reports will inevitably play a role in ensuring that websites remain safe and usable for users.

---

62. Ben Dickson, *Human Help Wanted: Why AI Is Terrible at Content Moderation*, PC MAG. (July 10, 2019), <https://www.pcmag.com/opinions/human-help-wanted-why-ai-is-terrible-at-content-moderation> [<https://perma.cc/7VFF-EA8K>].

63. *Id.*

64. *What Is Content Moderation and Why Companies Need It*, *supra* note 38.

65. *Id.*

66. Newton, *supra* note 52.

67. See Klonick, *supra* note 35, at 1638.

68. *Id.*

69. York & McSherry, *supra* note 45.

### III. THE PROBLEM WITH PLATFORM SELF-REGULATION

Section 230 largely allows platforms to moderate user-generated content at their discretion.<sup>70</sup> However, the autonomy that platforms enjoy in this realm is under threat. In 2020, the Department of Justice (“DOJ”) declared that “the time is ripe” to amend Section 230.<sup>71</sup> Congress heeded the call and proposed several bills that seek to roll back Section 230 protections.<sup>72</sup> In addition to legal threats, creators and users have grown increasingly frustrated with community guidelines that are “opaque” and change frequently, creating uncertainty as to whether they are violating platform standards.<sup>73</sup> Section A of this Part details Section 230’s protections and limitations in the content moderation arena. Section B discusses current government attempts to limit Section 230 protections. Finally, Section C explains how the growing frustration from creators and users regarding platform content moderation decisions could eventually lead to the decline or abandonment of platforms. The increasing attempts at government regulation, combined with creator and user frustration at platform decisions, makes total content moderation autonomy an increasingly unlikely model for the future.

#### *A. How Does Section 230 Protect Platforms?*

Section 230 provides platforms with two protections that allow them to regulate user-generated content without the fear of civil liability.

First, Section 230(c)(1) sets forth 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>74</sup> This

---

70. 47 U.S.C. § 230(c)(1) (2018) (providing that platforms shall not be treated as publishers of user-generated content); 47 U.S.C. § 230(c)(2) (2018) (protecting platforms from civil liability for steps taken to moderate most content).

71. *Section 230—Nurturing Innovation or Fostering Unaccountability?*, U.S. DEPT. JUST. (June 2020), <https://www.justice.gov/file/1286331/download> [<https://perma.cc/T9AK-PMFX>].

72. *See, e.g.*, S. 3983, 116th Cong. (2020); H.R. 4027, 116th Cong. (2020).

73. Klonick, *supra* note 35, at 1630. Klonick discusses the evolution of content moderation. In the early days of social media, platforms had small, vague lists of things that were not allowed, and much of the moderation was based on “gut feelings.” Over time, major platforms developed large, international teams that use more robust sets of guidelines to moderate content.

74. 47 U.S.C. § 230(c)(1).

provision applies to any “interactive computer service,”<sup>75</sup> such as a social media platform<sup>76</sup> or the comment section of a newspaper’s website.<sup>77</sup> The protection under Section 230(c)(1) means that platforms are not treated as the publishers of user-generated content.<sup>78</sup> Therefore, if a YouTube creator posts a defamatory video, YouTube will not be liable for defamation because the creator, not YouTube, is considered the publisher of the content.<sup>79</sup>

Second, Section 230(c)(2) sets forth that providers of interactive computer services will not face civil liability for “any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.”<sup>80</sup> This provision provides social media platforms with broad authority to set and enforce their own content moderation rules.<sup>81</sup> Section 230(c)(1) and Section 230(c)(2) thus work together to provide platforms with the freedom to set their own community guidelines, to shield them from civil liability for most actions taken to moderate content, and to shield them from liability for content posted by their users.

---

75. An interactive computer service is “any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server . . .”. 47 U.S.C. § 230(f)(2) (2018).

76. See *Fed. Agency of News LLC v. Facebook, Inc.*, 395 F. Supp. 3d 1295, 1305 (N.D. Cal. 2019) (asserting that Facebook clearly meets the definition of an interactive computer service); *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 330 (4th Cir. 1997) (holding that AOL is not liable for tortious content posted by a user because AOL is an interactive computer service and not the publisher of the material).

77. Makena Kelly, *Big Tech’s Liability Shield Under Fire Yet Again from Republicans*, VERGE (July 28, 2019, 1:00 PM), <https://www.theverge.com/2019/7/28/8933871/big-techs-liability-shield-under-fire-yet-again-from-republicans> [<http://archive.today/h3rge>].

78. *Zeran*, 129 F.3d at 330 (explaining that “[Section 230] precludes courts from entertaining claims that would place a computer service provider in a publisher’s role.”).

79. *Id.*

80. 47 U.S.C. § 230(c)(2)(A).

81. Russell Brandom, *Senate Republicans Want to Make It Easier to Sue Tech Companies for Bias*, VERGE (June 17, 2020, 9:46 AM), <https://www.theverge.com/2020/6/17/21294032/section-230-hawley-rubio-conservative-bias-lawsuit-good-faith> [<http://archive.today/QmhYc>].

The protections under Section 230(c), however, are not absolute. Section 230(e) sets forth limitations to Section 230(c).<sup>82</sup> Section 230(e) explains that interactive computer services could face liability under criminal law, intellectual property law, state law, communications privacy law, and sex trafficking law for their content moderation decisions.<sup>83</sup>

The most controversial Section 230(e) exception to date is the provision against sex trafficking,<sup>84</sup> which Congress passed as a package in 2018 through the Allow States and Victims to Fight Online Sex Trafficking Act (“FOSTA”)<sup>85</sup> and Stop Enabling Sex Traffickers Act (“SESTA”).<sup>86</sup> The House bill clarified that “[Section 230(c)] does not prohibit the enforcement against providers and users of interactive computer services of Federal and State criminal and civil law relating to sexual exploitation of children or sex trafficking . . . .”<sup>87</sup> It also amended the federal criminal code to define illegal “participation in a venture” that engages in sex trafficking as “knowingly assisting, supporting, or facilitating [a sex trafficking violation].”<sup>88</sup> SESTA largely echoed FOSTA and set forth that “[Section 230(c)(2)] shall not be construed to impair or limit . . . any claim in a civil action brought under [federal sex trafficking law] . . . or . . . any charge in a criminal prosecution brought under State law if the conduct underlying the charge constitutes a

---

82. See 47 U.S.C. § 230(e).

83. *Id.*

84. “Nothing in this section (other than subsection (c)(2)(A)) shall be construed to impair or limit—(A) any claim in a civil action brought under section 1595 of Title 18, if the conduct underlying the claim constitutes a violation of section 1591 of that title; (B) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 1591 of Title 18; or (C) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 2421A of Title 18, and promotion or facilitation of prostitution is illegal in the jurisdiction where the defendant’s promotion or facilitation of prostitution was targeted.” 47 U.S.C. § 230(e)(5).

85. Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No. 115-164, 132 Stat. 1253 (codified as amended at 18 U.S.C. § 2421A and 47 U.S.C. § 230 (2018)).

86. S. 1693, 115th Cong. (2017).

87. H.R. 1865, 115th Cong. (2017).

88. H.R. 1865, 115th Cong. § 5 (2017).



violation of [federal sex trafficking law].”<sup>89</sup> Ultimately, FOSTA and SESTA worked together to create an exception under Section 230(e) for sex trafficking,<sup>90</sup> but did little to specify exactly what terms such as “facilitating” sex trafficking meant.

FOSTA-SESTA was criticized by groups such as the American Civil Liberties Union (“ACLU”) and the Sex Workers Outreach Project (“SWOP”) as a “disguised internet censorship bill” intended to weaken Section 230 protections and control the content allowed by interactive service providers.<sup>91</sup> The seemingly innocuous goal of preventing sex trafficking was extremely consequential for platforms, users, and creators. To avoid potential liability created by FOSTA-SESTA, many platforms created stricter community guidelines around any sexual or sex-adjacent content.<sup>92</sup> The stricter enforcement by platforms resulted in unfortunate consequences, such as encouraging consensual online sex workers off of the internet and to the street<sup>93</sup> and resulting in the deletion of content that was posted to help victims of sex trafficking.<sup>94</sup>

---

89. S. 1693, 115th Cong. § 3 (2017).

90. *See supra* note 84.

91. *SWOP-USA Stands in Opposition of Disguised Internet Censorship Bill SESTA*, S. 1963, SWOP-USA (Aug. 11, 2017), <https://swopusa.org/blog/2017/08/11/call-to-actionpress-release-swop-usa-stands-in-direct-opposition-of-disguised-internet-censorship-bill-sesta-s-1963-call-your-state-representatives-and-tell-them-to-fight/> [<https://perma.cc/T4TX-CRUK>].

92. Anthony Cuthbertson, *Tumblr Porn Ban: One-Fifth of Users Have Deserted Site Since It Removed Adult Content*, INDEPENDENT (Mar. 11, 2019, 4:06 PM), <https://www.independent.co.uk/life-style/gadgets-and-tech/news/tumblr-porn-ban-nsfw-verizon-yahoo-adult-content-a8817546.html> [<https://perma.cc/UT8W-V4YU>].

93. Mike Masnick, *The Human Cost of FOSTA*, TECHDIRT (May 7, 2019, 9:39 AM), <https://www.techdirt.com/articles/20190503/13180842135/human-cost-fosta.shtml> [<https://perma.cc/B78E-LAUP>].

94. *Id.*

## B. Government Attempts to Regulate Section 230

Today, Section 230 protections for social media platforms face threats from both Congress and executive branch.<sup>95</sup> Subsections (1) and (2) focus on two recent government efforts to scale back Section 230 protections for social media platforms. The first is former President Trump’s executive order<sup>96</sup> directing the Federal Communications Commission (“FCC”) to draft regulations that clarify, among other things, when interactive computer services can be held liable for content moderation decisions not taken in good faith.<sup>97</sup> The second is a bipartisan bill called the Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2020 (“EARN IT Act”), which seeks to expand Section 230(e)’s to include an exception for violations of “child sexual exploitation law.”<sup>98</sup>

### 1. President Trump’s “Executive Order on Preventing Online Censorship”

President Trump’s May 26, 2020 tweet stating that the use of mail-in ballots for the 2020 presidential election would lead to widespread voter fraud<sup>99</sup> catapulted the discussion about the future of Section 230 into public discourse. In response to the tweet, Twitter added a link below the tweet titled: “Get the facts about mail-in ballots” and characterized President Trump’s claims as “unsubstantiated.”<sup>100</sup> Upset at Twitter’s response,

---

95. Casey Newton, *Everything You Need to Know About Section 230*, VERGE (May 28, 2020, 3:46 PM), <https://www.theverge.com/21273768/section-230-explained-internet-speech-law-definition-guide-free-moderation> [<https://archive.is/Pt3uG>].

96. Exec. Order No. 13925, 85 Fed. Reg. 34079, § 8(a)(i) (June 2, 2020).

97. Section 230(c)(2)(A) protects platforms from civil liability for “any action voluntarily taken *in good faith* to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.” (emphasis added); 47 U.S.C. § 230(c)(2)(A) (2018). In his executive order, President Trump asks the FCC to clarify when content moderation decisions could be considered not “in good faith.”

98. EARN IT Act of 2020, S. 3398, 116th Cong. (2020).

99. Donald J. Trump (@realDonaldTrump), TWITTER (May 26, 2020, 5:17 AM), <https://twitter.com/realDonaldTrump/status/1265255835124539392> [<https://archive.is/XJ8eS>].

100. *Id.*; *Trump Makes Unsubstantiated Claim That Mail-In Ballots Will Lead to Voter Fraud*, *supra* note 4.

President Trump issued an executive order titled “Executive Order on Preventing Online Censorship.”<sup>101</sup>

The executive order asserts that it is “un-American” and “dangerous” to allow a small number of companies to “pick the speech that Americans may access and convey on the internet.”<sup>102</sup> It then directs the Federal Communications Commission (“FCC”) to draft regulations that: (1) “clarify and determine the circumstances under which a provider of an interactive computer service that restricts access to content in a manner not specifically protected by [Section 230(c)(2)(A)] may also not be able to claim protection under [Section 230(c)(1);]” (2) specify when a company’s conduct could violate Section 230’s “good faith” provision; and (3) propose any other regulation that “may be appropriate to expand [this] policy.”<sup>103</sup> The executive order also calls for an investigation of whether social media companies engage in discrimination based on the political viewpoint of a content creator.<sup>104</sup>

Although the President does not have the authority to change an existing law by issuing an executive order,<sup>105</sup> the order could prove impactful in one of two ways. First, courts have determined that where a law is unclear, federal agencies may step in and “fill the gap.”<sup>106</sup> A statute may provide this gap filling authority either explicitly through Congress or may provide implicit authorization by not defining key terms.<sup>107</sup> When the relevant federal

---

101. Exec. Order No. 13925, 85 Fed. Reg. at 34079.

102. *Id.*

103. *Id.* at 34081.

104. *Id.* at 34081–82.

105. *Youngstown Sheet & Tube Co. et al. v. Sawyer*, 343 U.S. 579, 637 (1952) (Jackson, J., concurring).

106. *E.g.*, *Chevron U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 843–44 (1984) (stating that “if the statute is silent or ambiguous with respect to the specific issue, the question for the court is whether the agency’s answer is based on a permissible construction of the statute” which can be either explicit or implicit); *Norfolk S. Ry. Co. v. Geodis Logistics, LLC*, No. 1:19-CV-03341-JPB, 2020 WL 4938362, at \*3 (N.D. Ga. July 7, 2020) (explaining that “because the statute is silent as to who is liable for demurrage, the [Surface Transformation Board] had the authority to ‘fill the gap’ and issue a regulation regarding the issue”).

107. *S.F. Baykeeper v. Cargill Salt Div.*, 481 F.3d 700, 704 (9th Cir. 2007) (explaining that “[b]y not defining further the meaning of ‘waters of the United States,’ Congress implicitly

agency interprets a statute, “the court should defer to the agency’s interpretation as long as it is reasonable.”<sup>108</sup> Thus with regards to Section 230, the FCC may be able to formulate guidelines on what sort of content moderation decisions do and do not constitute “good faith” pursuant to Section 230(c)(2).<sup>109</sup> Second, the executive order signaled the Trump administration’s approval for Congress to amend Section 230, an initiative that is seemingly also supported by President Joe Biden’s administration.<sup>110</sup> Congress could heed the call given that there are already active bills intended to roll back Section 230 protections. One such bill would require interactive computer services to undertake a duty of good faith to keep their Section 230 protections and would impose monetary penalties for breach.<sup>111</sup> Another bill seeks to restrict an interactive computer service’s ability to moderate content the company finds “objectionable.”<sup>112</sup> Even the DOJ has proposed that “the time is ripe” to amend Section 230 altogether.<sup>113</sup> Thus, even though President Trump’s executive order does not have the force of law, the executive order, combined with President Biden’s stance on Section 230 nonetheless signals bipartisan approval of the initiatives to roll back Section 230 protections.

---

delegated policy-making authority to the EPA and the Corps, the agencies charged with the CWA’s administration.”).

108. *Chao v. Symms Fruit Ranch, Inc.*, 242 F.3d 894, 897 (9th Cir. 2001).

109. *See supra* note 97. Section 230(c)(2)(A) does not currently specify exactly what sort of conduct constitutes good faith.

110. Joseph Fawbush, *After Executive Order Condemning Online Censorship, Should Social Media Companies Fear Liability?*, FINDLAW (June 1, 2020, 2:30 PM), [https://blogs.findlaw.com/technologist/2020/06/after-executive-order-condemning-online-censorship-should-social-media-companies-fear-liability.html?DCMP=pro\\_special:nwl:z:2020june:feature](https://blogs.findlaw.com/technologist/2020/06/after-executive-order-condemning-online-censorship-should-social-media-companies-fear-liability.html?DCMP=pro_special:nwl:z:2020june:feature) [<https://perma.cc/CEU8-HM3E>]. President Biden has stated that he thinks Section 230 should be revoked and his current nominee for Secretary of Commerce, Gina Raimondo, believes that Section 230 should be reformed. Of note, the Biden administration, has not taken action in that direction as of the publication of this Note. Makena Kelly, *Biden’s Commerce Nominee Backs Changes to Section 230*, VERGE (Jan. 26, 2021, 1:40 PM), <https://www.theverge.com/2021/1/26/22250746/biden-gina-raimondo-commerce-secretary-section-230> [<https://archive.is/tCKsdj>].

111. S. 3983, 116th Cong. (2020); Brandom, *supra* note 81.

112. H.R. 4027, 116th Cong. (2019); Kelly, *supra* note 77.

113. *Section 230—Nurturing Innovation or Fostering Unaccountability*, *supra* note 71.

## 2. EARN IT Act

The EARN IT Act is perhaps the biggest threat to Section 230 protections today. The bipartisan bill introduced by Senators Lindsey Graham and Richard Blumenthal seeks to expand Section 230(e) by amending it to include an exception for violations of federal and state “child sexual exploitation law.”<sup>114</sup> In the same way that FOSTA-SESTA added a sex trafficking exception to Section 230(e), the EARN IT Act seeks to add an exception for child sexual exploitation law. Ultimately, the EARN IT Act would set forth that interactive service providers could be held liable for civil and criminal prosecution for the “advertisement, promotion, presentation, distribution, or solicitation of child sexual abuse material.”<sup>115</sup>

The EARN IT Act would also create a National Commission on Online Child Sexual Exploitation Prevention (“Commission”) to construct a set of “recommended best practices that providers of interactive computer services may choose to implement to prevent, reduce, and respond to the online sexual exploitation of children, including the enticement, grooming, sex trafficking, and sexual abuse of children and the proliferation of online child sexual abuse material.”<sup>116</sup> The Commission would consist of sixteen members<sup>117</sup> comprised of law enforcement, online child crime experts or survivors, and technology experts.<sup>118</sup> The original version of the bill created a safe harbor from liability for interactive computer services that followed the

---

114. “Section 230(e) of the Communications Act of 1934 (47 U.S.C. 230(e) is amended by adding at the end the following: ‘(6) NO EFFECT ON CHILD SEXUAL EXPLOITATION LAW.—Nothing in this section (other than subsection (c)(2)(A)) shall be construed to impair or limit— ‘(A) any claim in a civil action brought against a provider of an interactive computer service under section 2255 of title 18, United States Code, if the conduct underlying the claim constitutes a violation of section 2252 or section 2252A of that title; ‘(B) any charge in a criminal prosecution brought against a provider of an interactive computer service under State law regarding the advertisement, promotion, presentation, distribution, or solicitation of child sexual abuse material, as defined in section 2256(8) of title 18, United States Code; or ‘(C) any claim in a civil action brought against a provider of an interactive computer service under State law regarding the advertisement, promotion, presentation, distribution, or solicitation of child sexual abuse material, as defined in section 2256(8) of title 18, United States Code.’ EARN IT Act of 2020, S. 3398, 116th Cong. (2020).

115. *Id.* § 5(B).

116. *Id.* § 3.

117. *Id.* § 3(c)(1)(A).

118. *Id.* § 3(c)(2).

Commission's recommendations.<sup>119</sup> The safe harbor provision created undue pressure for interactive computer services to follow their recommendations in order to avoid liability.<sup>120</sup> Although the bill has been amended to remove the safe harbor provision, many still worry that platforms will inevitably face pressure to follow the Commission's recommendations or risk liability for child sexual exploitation.<sup>121</sup>

Critics of the bill are concerned that if the EARN IT Act passes, platforms will have to take restrictive content moderation steps to comply with potentially conflicting and expansive child exploitation laws in all fifty states, as they did with FOSTA-SESTA.<sup>122</sup> The passage of the EARN IT Act will similarly result in stricter rules that will "force platforms to over-censor speech out of an abundance of caution due to fear of endless litigation."<sup>123</sup>

The DOJ and Congress are eager to play a bigger role in content moderation.<sup>124</sup> Instead of amending Section 230(c) in its entirety, Congress has proposed bills to slowly chip away at its protections.<sup>125</sup> Like FOSTA-SESTA before it, the EARN IT Act seeks to add yet another exception to Section 230(e), giving Congress more power over platform content moderation. The EARN IT Act is thus an attempt by Congress to heed the DOJ's call that "the time is ripe" to roll back Section 230 protections.<sup>126</sup>

---

119. *Id.* § 6(a)(6)(B).

120. Michael S. Horikawa & Jon Schreiber, *EARN IT Act Amendments Could Shift Section 230 Protection from DOJ Guideline Compliance to Post Hoc Enforcement Regime*, JD SUPRA: PILLSBURY (Aug. 3, 2020), <https://www.jdsupra.com/legalnews/earn-it-act-amendments-could-shift-48919/> [<https://perma.cc/NKD5-3Z4D>].

121. *Id.*

122. *Id.*

123. Corey Silverstein, *The EARN IT Act: Are You Paying Attention Yet?*, XBIZ (Aug. 23, 2020), <https://www.xbiz.com/features/253832/the-earn-it-act-are-you-paying-attention-yet> [<https://perma.cc/VD39-CJQD>]; Horikawa & Schreiber, *supra* note 120.

124. *Department of Justice's Review of Section 230 of the Communications Decency Act of 1996*, *supra* note 64.

125. *E.g.*, EARN IT Act of 2020, S. 3398, 116th Cong. (2020); H.R. 4027, 116th Cong. (2019).

126. *Section 230—Nurturing Innovation or Fostering Unaccountability*, *supra* note 71.

### C. Creator and User Frustration with Content Regulation Methods

The possibility of government intervention in content moderation is not the only threat to platforms. Another is the threat of usage decline by increasingly confused, frustrated, and distrustful creators and users.<sup>127</sup> Subsections (1) and (2) below focus on two areas of platform abandonment: (1) a brief overview of the reasons behind previous platform abandonment and (2) what platform abandonment could mean for major social media platforms today.

#### 1. A History of Platform Abandonment

It is difficult to imagine a world where Facebook, Twitter, and YouTube do not play a role in everyday life. In 2010, it must have been equally difficult to imagine a world without renting a weekend movie from Blockbuster. Ten years prior, it would be similarly difficult to imagine a world without malls for back-to-school shopping. And yet, both Blockbuster<sup>128</sup> and malls,<sup>129</sup> once staples of daily life, today seem like relics of a different era. The idea that today's social media platforms will last forever is therefore misguided, despite their explosive growth in the 21<sup>st</sup> century.<sup>130</sup>

The first notable instance of a fallen powerhouse in the social media realm was the fall of MySpace in the late 2000s. MySpace, which at one point touted more daily visitors than Google,<sup>131</sup> declined sharply in the late

---

127. Audrey Schomer, *The Content Moderation Report: Social platforms Are Facing A Massive Content Crisis—Here's Why We're Thinking Regulation is Coming and What it Will Look Like*, BUS. INSIDER (Nov. 14, 2019, 7:03 PM), <https://www.businessinsider.com/content-moderation-report-2019-11> [<https://perma.cc/Q3RU-SM7S>].

128. Greg Satell, *A Look Back At Why Blockbuster Really Failed and Why It Didn't Have To*, FORBES (Sept. 5, 2014, 11:38 PM), <https://www.forbes.com/sites/gregsatell/2014/09/05/a-look-back-at-why-blockbuster-really-failed-and-why-it-didnt-have-to/#27fc20961d64> [<https://perma.cc/E745-E9AJ>].

129. Mary Hanbury, *50 Haunting Photos of Abandoned Shopping Malls Across America*, BUS. INSIDER (Nov. 8, 2019, 3:49 AM), <https://www.businessinsider.com/american-retail-apocalypse-in-photos-2018-1> [<https://perma.cc/BU5E-J7KQ>].

130. Esteban Ortiz-Ospina, *The Rise of Social Media*, OUR WORLD IN DATA (Sept. 18, 2019), <https://ourworldindata.org/rise-of-social-media> [<https://perma.cc/T3UE-J7VP>].

131. *Id.*

2000s.<sup>132</sup> The once simple platform became clunky, ad-littered, and difficult to use.<sup>133</sup> The fall of MySpace demonstrated that a burdensome user experience can be fatal even to the largest and most popular platform.<sup>134</sup>

Content moderation decisions are also decisions in user experience.<sup>135</sup> This was apparent with the fall of the social networking and blogging platform Tumblr. In an attempt to comply with FOSTA-SESTA in 2018, Tumblr created an algorithm to find and delete adult content.<sup>136</sup> The algorithm targeted not only illegal content, but legal content that fell under “photos, videos, or GIFs that show real-life human genitals or female-presenting nipples, and any content . . . that depict sex acts.”<sup>137</sup> Before implementation of the algorithm, Tumblr featured a large base of consensual sex workers, blogs that posted sexual imagery, and blogs that discussed sexual health and safety.<sup>138</sup> Once the stricter guidelines were implemented and blogs featuring sexual content were removed, Tumblr lost a fifth of its users.<sup>139</sup>

---

132. Matthew Garrahan, *The Rise and Fall of MySpace*, FIN. TIMES (Dec. 4, 2009), <https://www.ft.com/content/fd9ffd9c-dee5-11de-adff-00144feab49a> [<http://archive.today/tooaF>].

133. *Id.*

134. *Id.*

135. See generally *Content Moderation and the User Experience*, HAA SITES AT THE UNIV. OF PITT. (Mar. 31, 2020), <https://sites.haa.pitt.edu/digitalcriticalmethods/2020/03/31/content-moderation-and-the-user-experience/> [<https://perma.cc/7H7T-WVC6>].

136. *What Tumblr’s Ban on ‘Adult Content’ Actually Did*, ELEC. FRONTIER FOUND., <https://www.eff.org/tossedout/tumblr-ban-adult-content> [<https://perma.cc/VHS2-RUG5>].

137. Cuthbertson, *supra* note 92.

138. *Id.*

139. *Id.*



## 2. Potential Platform Abandonment Today

Today, as Facebook<sup>140</sup> and Twitter<sup>141</sup> lose users for the first time and YouTube<sup>142</sup> creators report less income from ads, decisions on how to moderate content could be highly impactful for the major social media platforms. On YouTube, creators frequently criticize the platform's constantly changing and confusing content-moderation standards. The standards force creators to guess whether their content will remain on the website, lose monetization, or remain untouched.<sup>143</sup>

Conservative-leaning content creators like Prager University are not the only ones claiming that the platform is biased against their content.<sup>144</sup> LGBTQ+ content creators also claim that their videos are demonetized or restricted just for having LGBTQ+-related words in the titles.<sup>145</sup> Even creators that make videos about sensitive topics are targeted. For example, YouTube creator Phillip DeFranco posted a video covering the police killing of George Floyd separately from his usual daily upload for fear that the sensitive content would subject the entire video to demonetization or

---

140. Abrar Al-Heeti, *Facebook Lost 15 Million US Users in the Past Two Years, Report Says*, CNET (Mar. 6, 2019, 1:58 PM), <https://www.cnet.com/news/facebook-lost-15-million-users-in-the-past-two-years-report-says/> [<https://perma.cc/L8X7-4JSU>].

141. Kalev Leetaru, *A Fading Twitter Changes Its User Metrics Once Again*, FORBES (Apr. 23, 2019, 10:02 PM), <https://www.forbes.com/sites/kalevleetaru/2019/04/23/a-fading-twitter-changes-its-user-metrics-once-again/#42c25b0c7a31> [<https://perma.cc/FX6Q-TWVQ>].

142. Geoff Weiss, *A Lot of YouTube Creators Just Disclosed Their Declining AdSense Rates Amid the Coronavirus Pandemic*, TUBEFILTER (Apr. 16, 2020), <https://www.tubefilter.com/2020/04/16/creators-disclose-declining-adsense-rates-coronavirus/> [<https://perma.cc/PWP2-UARV>].

143. Louise Matsakis, *YouTube Doesn't Know Where Its Own Line Is*, WIRED (Mar. 2, 2018, 11:41 AM), <https://www.wired.com/story/youtube-content-moderation-inconsistent/> [<https://perma.cc/2YDT-5Z3R>].

144. Prager Univ. v. Google LLC, 951 F.3d 991, 999 (9th Cir. 2020).

145. Aja Romano, *A Group of YouTubers is Trying to Prove the Site Systematically Demonetizes Queer Content*, VOX (Oct. 10, 2019, 9:40 AM), <https://www.vox.com/culture/2019/10/10/20893258/youtube-lgbtq-censorship-demonetization-nerd-city-algorithm-report> [<https://archive.is/qXiQl>] (describing a study of LGBTQ+-related video titles that were deemed “fit for monetization” found that 33 out of the 100 videos tested were demonetized).

restriction.<sup>146</sup> The separate video discussing George Floyd was demonetized.<sup>147</sup> Similarly, creator Jessica Kent, who makes educational videos about her former incarceration and drug addiction, has started to bleep out words related to drug use in her videos, despite the fact that her content is educational and meant to deter viewers from drug use and illegal activity.<sup>148</sup> Though these may seem like small annoyances, a consistent stream of small annoyances could deter content creators from discussing and addressing sensitive topics on their channels. Further, YouTube's inconsistent and vague monetization and restriction requirements make it so that content creators often make guesses at what it would take to continue to make a living off of their content. Even with the measures taken to prevent demonetization, creators still remain unclear as to whether the methods will work. This sort of confusion is slowly driving some content creators away from YouTube.<sup>149</sup> As with MySpace and Tumblr, if content creators leave YouTube, they are likely to take their fans to new platforms along with them.

Similarly, Twitter's content moderation decisions have also pushed some creators and users to other platforms. For example, after Twitter placed the content warning on President Trump's tweet about mail-in ballots, many of his ardent supporters abandoned Twitter for Parler, a Twitter-like platform which claims to engage in minimal content regulation only driven by "the *spirit* of the First Amendment to the United States Constitution."<sup>150</sup> Parler, which has quickly grown to almost twelve million users,<sup>151</sup> rapidly turned

---

146. Phillip DeFranco, *Disgusting! We Need To Talk About What Happened To George Floyd ... What We Know & What Happens Next*, YOUTUBE (May 27, 2020), [https://www.youtube.com/watch?v=JGecrIMn\\_Ds](https://www.youtube.com/watch?v=JGecrIMn_Ds) [<https://perma.cc/PH5Q-6HXZ>].

147. *Id.*

148. Jessica Kent, *Top Ten Things I Used to Do*, YOUTUBE (July 27, 2020), <https://www.youtube.com/watch?v=Z7Uk-siZL6o> [<https://perma.cc/TMG2-J3EK>] (explaining, in a video about the negative behaviors she exhibited during her drug addiction, that she must bleep out names of certain drugs because "YouTube is kicking [her] ass" by demonetizing videos that discuss drug use).

149. Julia Alexander, *YouTubers Look to New Platforms After Viewer Suppression, Demonetization Issues*, POLYGON (Apr. 17, 2018, 6:30 PM), <https://www.polygon.com/2018/4/17/17246948/defranco-patreon-casey-neistat-youtube-ceo-demonetized> [<https://archive.is/kw5vs>].

150. *Parler Community Guidelines*, PARLER (Aug. 21, 2020), <https://legal.parler.com/documents/guidelines.pdf> [<https://perma.cc/T3T3-SDYX>].

151. In early 2021, after the pro-Trump riot at the United States Capitol, President Trump was banned from Twitter and Parler's user base grew from about three million to twelve million.

into a conservative echo chamber. Upon joining the platform, users are immediately prompted to follow Breitbart, the Daily Caller, and Eric Trump—all avid supporters of Parler.<sup>152</sup> Although Parler's user base of 12 million pales in comparison to Twitter's 159 million active users,<sup>153</sup> it would be a mistake for today's social media giants to ignore the threat of emerging platforms to their success. Large platforms like Twitter, Facebook, and YouTube should be alarmed by statistics showing that 80% of Americans distrust their content moderation methodologies<sup>154</sup> because that distrust chips away at their content moderation autonomy and fuels additional minimization of Section 230 protections.

#### IV. WHAT IS THE SOLUTION?

If full platform self-governance is unlikely to remain, what is the alternative? Part IV analyzes several alternative methods to regulate content moderation. Section A explains why the First Amendment's free speech protections could not prevent platforms from censoring speech. Section B describes how low technological literacy and political self-interest are likely to result in draconian content moderation decisions if the task is entirely left to the government. Finally, Section C poses a solution based on government-industry cooperation with a focus on transparency, consistency, and accountability.

---

After being taken offline by Amazon's cloud provider, Parler sued Amazon, but the court denied Parler's request to force Amazon to allow it back online. As of the publication of this Note, Parler is back online using an independent server. Rachel Lerman & Natasha Tiku, *Parler is Offline, but Violent Posts Scraped by Hackers Will Haunt Users*, WASH. POST (Jan. 12, 2021, 12:04 PM), <https://www.washingtonpost.com/technology/2021/01/12/parler-data-downloaded/> [<https://perma.cc/Q5X9-LV6F>]. Igor Bonifacic, *Parler Is Back Online After Amazon Kicked It Off the Internet*, YAHOO FINANCE (Feb. 15, 2021), <https://finance.yahoo.com/news/parler-back-online-173810043.html> [<https://perma.cc/29HT-JXDJ>].

152. Mark Sullivan, *I Joined Parler, The Right-Wing Echo Chamber's New Favorite Alt-Twitter*, FAST COMPANY (June 27, 2020), <https://www.fastcompany.com/90522049/i-joined-parler-the-right-wing-echo-chambers-new-favorite-alt-twitter> [<https://perma.cc/6XFB-UYMZ>].

153. *Twitter by the Numbers: Stats, Demographics and Fun Facts*, *supra* note 47.

154. Sarah Brennan, *80% of Americans Distrust Tech Companies' Content Moderation*, IT PRO (June 17, 2020), <https://www.itpro.com/marketing-comms/social-media/356109/80-of-americans-trust-tech-companies-content-moderation> [<https://perma.cc/79XU-LN2G>].

### A. *The First Amendment Does Not Protect the Public's Ability to Say Anything Online*

The First Amendment to the Constitution of the United States says that “Congress shall make no law . . . abridging the freedom of speech.”<sup>155</sup> The First Amendment, especially the right to free speech, is perhaps the most cherished right in the United States.<sup>156</sup> For Americans, free speech is more than a right—it is a value,<sup>157</sup> a “central tenet of our American way of life,”<sup>158</sup> and something that feels almost dystopian when denied. Because of this, Americans often have visceral reactions when they are not able to say the things they want to say, evoking their right to free speech “every time [someone is] banned from Twitter.”<sup>159</sup> Americans value the right to free speech so much that they often expect it in all situations, forgetting that the Constitutional right to free speech actually has several limitations.<sup>160</sup> Most people know the popular analogy that “free speech would not protect a man in falsely shouting fire in a theatre.”<sup>161</sup> But there are other narrow categories of speech that the First Amendment does not protect, including speech that incites lawless action, fighting words, true threats, and libel.<sup>162</sup> The limitation that is often forgotten, however, is that the Constitutional right to free speech—subject to all limitations, of course—is only guaranteed from *the*

---

155. U.S. CONST. amend. I.

156. Alex Cook, *Americans Say Freedom of Speech is the Most Important Constitutional Right, According to FindLaw.com Survey for Law Day, May 1*, FINDLAW (Apr. 30, 2015, 5:20), <https://www.pnewsire.com/news-releases/americans-say-freedom-of-speech-is-the-most-important-constitutional-right-according-to-findlawcom-survey-for-law-day-may-1-300074847.html> [<https://perma.cc/2E52-KQLP>].

157. Logan Chipkin, *Free Speech is a Value, Not a Right*, QUILLETTE (July 6, 2020), <https://quillette.com/2020/07/06/free-speech-is-a-value-not-a-right/> [<https://perma.cc/8G7W-PXMY>].

158. AJ Willingham, *The First Amendment Doesn't Guarantee You the Rights You Think It Does*, CNN (Sept. 6, 2018, 7:36 PM), <https://www.cnn.com/2017/04/27/politics/first-amendment-explainer-trnd/index.html> [<https://perma.cc/QL8Y-VNPC>].

159. *Id.*

160. *Id.*

161. *Schenck v. United States*, 249 U.S. 47, 52 (1919).

162. DAVID HUDSON, *THE FIRST AMENDMENT: FREEDOM OF SPEECH* § 3:1 (2012).

*government.*<sup>163</sup> Private entities are not obligated to guarantee or allow free speech in most situations.<sup>164</sup>

There are some narrow exceptions for when private entities may have to provide certain rights that normally only the government is obligated to provide.<sup>165</sup> Courts determine when this exception applies by using the judge-made state action doctrine.<sup>166</sup> The state action doctrine creates certain obligations for a private entity when it exercises “powers traditionally exclusively reserved to the State.”<sup>167</sup> The appellant in *Marsh v. Alabama*,<sup>168</sup> Grace Marsh, was arrested by authorities in a town that was owned and operated by a private company for distributing religious literature in the town.<sup>169</sup> The Supreme Court of the United States found that because the town operated just like a regular town, it was obligated to provide to the public the same rights that a government-operated town would have to provide.<sup>170</sup> Thus, the First Amendment allowed Marsh to distribute religious literature even in a company town.<sup>171</sup>

The state action doctrine, however, has its limits. Recently in *Manhattan Community Access Corp. v. Halleck*, the Supreme Court refused to extend the state action doctrine to a private nonprofit corporation that ran a city’s public access channel.<sup>172</sup> A public access channel is a publicly-owned television channel that is for public use but is generally administered by a

---

163. *Manhattan Cmty. Access Corp. v. Halleck*, 139 S. Ct. 1921, 1926 (2019).

164. *Hudgens v. NLRB*, 424 U.S. 507, 1033 (1976).

165. *See Manhattan Cmty. Access*, 139 S. Ct. at 1926.

166. *Id.*

167. *Jackson v. Metro. Edison Co.*, 419 U.S. 345, 352 (1974).

168. *Marsh v. Alabama*, 326 U.S. 501, 502 (1946).

169. *Id.*

170. *Id.* at 508.

171. *See id.* at 508–09.

172. *Manhattan Cmty. Access Corp. v. Halleck*, 139 S. Ct. 1921, 1926 (2019).

private network.<sup>173</sup> The Court held that the private network “merely hosting speech by others is not a traditional, exclusive public function.”<sup>174</sup> Therefore, the corporation that owned the public access channel was permitted to exercise editorial discretion and censor content that was critical of it.<sup>175</sup>

Because of the narrow scope of the state action doctrine, by the time the Ninth Circuit had to decide *Prager University*,<sup>176</sup> the decision was clear—YouTube, by hosting videos, did not perform a function that was traditionally reserved for government entities.<sup>177</sup> Prager University attempted to argue that YouTube became a state actor when it hosted public speech.<sup>178</sup> However, the Ninth Circuit easily analogized the argument to the one that failed in *Halleck*, and held that the state action doctrine did not mandate YouTube to provide First Amendment protections to its users.<sup>179</sup> It was not enough that hosting videos on a platform is a function that the government has performed at some point.<sup>180</sup> For YouTube to become a state actor, video hosting would have to be a function traditionally exclusive to the government.<sup>181</sup> Simply hosting public speech did not “transform” YouTube into a public forum.<sup>182</sup> Like YouTube, Facebook and Twitter are private entities that host public speech. Therefore, courts are not likely to require them to provide First Amendment protections either.

---

173. *Public, Educational, and Governmental Access Channels* (“PEG Channels”), FCC (Dec. 9, 2015) <https://www.fcc.gov/media/public-educational-and-governmental-access-channels-peg-channels> [<https://perma.cc/CM9Q-8Y55>].

174. *Manhattan Cmty. Access*, 139 S. Ct. at 1930.

175. *Id.* at 1930–31.

176. *Prager Univ. v. Google LLC*, 951 F.3d 991, 997 n.3 (9th Cir. 2020) (citing to six different cases in other Circuits that refused to find that social media platforms, including Facebook and YouTube, were public forums or state actors).

177. *Id.* at 998.

178. *Id.* at 997–98.

179. *Id.* at 998–99.

180. *Id.* at 997.

181. *Id.*

182. *Id.*

The right to free speech is a valuable right at the cornerstone of American culture. However, arguing that the First Amendment protects the ability to say absolutely anything online is an easy argument for courts to strike down. Precedent establishes that social media companies do not provide a government service.<sup>183</sup> Thus, the right to free speech outlined in the First Amendment is not the correct tool to use to solve the visceral reaction Americans have toward internet censorship.

*B. Government Regulation of Content Moderation is Ineffective at Best and Harmful at Worst*

In 2018, Google's Chief Executive Officer, Sundar Pichai, testified in front of Congress for three and a half hours.<sup>184</sup> The hearing was meant to address Google's alleged anti-conservative bias.<sup>185</sup> Some of the many strange questions Pichai had to answer were questions about why a photograph of President Trump appears as a top result in a Google image search for the word "idiot" and whether Google "knew" if someone moved from one end of a room to the other.<sup>186</sup> Facebook Chief Executive Officer Mark Zuckerberg also answered similar questions in a congressional hearing earlier that year.<sup>187</sup> Media coverage of the hearings mostly consisted of jokes about the low technical literacy of members of Congress and upheld a general consensus that the government is not well-equipped to regulate large technology companies.<sup>188</sup>

---

183. *See id.*

184. Minda Zetlin, *Google CEO Sundar Pichai Spent 3 1/2 Hours Before Congress. Here are the Strangest Things They Asked*, INC. (Dec. 13, 2018), <https://www.inc.com/minda-zetlin/google-ceo-sundar-pichai-congress-representatives-hearings-funny-stupid-questions.html> [http://archive.today/krcor].

185. *Id.*

186. *Id.*

187. Emily Stewart, *Lawmakers Seem Confused About What Facebook Does – And How to Fix It*, VOX (Apr. 10, 2018, 7:50 PM), <https://www.vox.com/policy-and-politics/2018/4/10/17222062/mark-zuckerbewestrg-testimony-graham-facebook-regulations> [http://archive.today/cRWtw].

188. *E.g.*, Margaret Sullivan, *Members of Congress Can't Possibly Regulate Facebook. They Don't Understand It*, WASH. POST (Apr. 10, 2018, 3:42 PM), [https://www.washingtonpost.com/lifestyle/style/members-of-congress-cant-possibly-regulate-facebook-they-dont-understand-it/2018/04/10/27fa163e-3cd1-11e8-8d53-eba0ed2371cc\\_story.html](https://www.washingtonpost.com/lifestyle/style/members-of-congress-cant-possibly-regulate-facebook-they-dont-understand-it/2018/04/10/27fa163e-3cd1-11e8-8d53-eba0ed2371cc_story.html) [https://perma.cc/F9NT-D4WX]; Cecelia Kang et al., *Knowledge Gap Hinders Ability of Congress to Regulate*

Of course, the argument that the government cannot regulate an industry simply because it is large and complex is not only inaccurate, but dangerous. Many large and complex industries are regulated, such as the food industry, the pharmaceutical industry, and the financial industry. To uphold a standard that complex industries are beyond regulation could create an environment where technology companies engage in bad faith practices and take advantage of their users with impunity.

Nevertheless, the hearings demonstrate that technological literacy for the top lawmakers in the country is low. Given that in less than two decades, platforms went from inception to use by 72% of all Americans, it would be a mistake to undertake regulation of social media platforms without understanding their nuances.<sup>189</sup> In addition, regulators, platforms, content creators, and users *all* have an interest in what content moderation looks like. Regulators and politicians have an especially strong interest, given that Americans increasingly use social media to connect with political figures.<sup>190</sup>

A combination of low technical literacy and a strong personal stake in content moderation is a recipe for oppressive legislation that stifles speech. Congressional attempts to regulate Section 230 thus far do little to instill faith in the government's ability to regulate platforms without such negative consequences. For example, FOSTA-SESTA failed to protect victims of sex trafficking or reduce the incidence of sex work as intended.<sup>191</sup> Instead, it led to platforms creating stricter community guidelines and taking down content meant to help victims of sex trafficking, thus making it more difficult for sex

---

*Silicon Valley*, N.Y. TIMES (Apr. 12, 2018), <https://www.nytimes.com/2018/04/12/business/congress-facebook-regulation.html> [<https://perma.cc/Z3L4-WV9T>].

189. *Social Media Fact Sheet*, PEW RSCH. CTR. (June 12, 2019), <https://www.pewresearch.org/internet/fact-sheet/social-media/> [<https://perma.cc/94AM-9V3A>].

190. Monica Anderson, *More Americans Are Using Social Media to Connect With Politicians*, PEW RES. CTR. (May 19, 2015), <https://www.pewresearch.org/fact-tank/2015/05/19/more-americans-are-using-social-media-to-connect-with-politicians/> [<https://perma.cc/9RVN-BXQ2>].

191. Karol Markowicz, *Congress' awful anti-sex-trafficking law has only put sex workers in danger and wasted taxpayer money*, BUS. INSIDER (July 14, 2019, 12:38 PM), <https://www.businessinsider.com/fosta-sesta-anti-sex-trafficking-law-has-been-failure-opinion-2019-7> [<https://perma.cc/RBU4-SRYV>]; Karen Gullo & David Greene, *With FOSTA Already Leading to Censorship, Plaintiffs Are Seeking Reinstatement of Their Lawsuit Challenging the Law's Constitutionality*, ELEC. FRONTIER FOUND. (Mar. 1, 2019), <https://www.eff.org/deeplinks/2019/02/fosta-already-leading-censorship-we-are-seeking-reinstatement-our-lawsuit> [<https://perma.cc/68BR-EYKS>].



workers to “share information or warn each other away from violent clients.”<sup>192</sup>

Like FOSTA-SESTA, the EARN IT Act raises content moderation freedom concerns under the guise of preventing child sexual exploitation.<sup>193</sup> The First Amendment provides interactive computer services with editorial authority over the content posted on their platforms.<sup>194</sup> The EARN IT Act threatens editorial authority by pressuring platforms to create content moderation standards based on suggestions by the EARN IT Act’s Commission that go beyond simply requiring platforms to ban unlawful content.<sup>195</sup>

The EARN IT Act’s potential problems with vagueness and overbreadth also pose dangers to the First Amendment rights of interactive computer services.<sup>196</sup> Vagueness in law refers to specificity—“[a] statute regulating behavior must be specific enough to allow ordinary people to understand what conduct is prohibited and to prevent arbitrary or discriminatory enforcement.”<sup>197</sup> Overbreadth, on the other hand, refers to law that is overly broad in scope.<sup>198</sup> The EARN IT Act imposes “vague and expansive

---

192. Markowicz, *supra* at 191; Gullo & Greene, *supra* note 191.

193. Letter from Sophia Cope & Aaron Mackey, Staff Attorneys, Electronic Frontier Foundation to Senate Judiciary Comm. (Mar. 9, 2020), <https://www.eff.org/document/eff-earn-it-act-first-amendment-letter-sjc> [<https://perma.cc/9393-GL6D>].

194. *Id.*; Miami Herald Publ’g Co., Inc. v. Tornillo, 418 U.S. 241, 258 (1974).

195. The Electronic Frontier Foundation, a non-profit organization that works to protect online free speech and privacy, wrote a letter to Congress detailing its concerns that the law is not narrowly tailored enough to pass constitutional scrutiny. The letter identified as concerns “[t]he ‘best practices,’ [that] will govern how online service providers must prevent, identify, disrupt, and report child sexual exploitation; how they must work with ‘non-profit organizations and other providers of interactive computer services to preserve, remove from view, and report child sexual exploitation;’ how they must implement ‘a standard rating and categorization system to identify the type and severity of child sexual abuse material;’ and how they must employ ‘age rating and age gating systems to reduce child sexual exploitation.’” Letter from Sophia Cope & Aaron Mackey to Senate Judiciary Comm., *supra* note 193.

196. Aaron H. Caplan, *Free Speech and Civil Harassment Orders*, 64 HASTINGS L.J. 781, 781 (2013) (explaining that vagueness and overbreadth in statutes pose concerns to First Amendment protections).

197. *Id.* at 810.

198. *Id.* at 817.

liability for user-generated content.”<sup>199</sup> The original version of the EARN IT Act required platforms to “earn” their Section 230 protections by taking steps to prevent child sexual exploitation based on a set of undefined “best practices” and creating a safe harbor for interactive computer services that followed the Commission’s “best practices.”<sup>200</sup> The latest version of the EARN IT Act removes the safe harbor and does not specify whether platforms will have to “earn” their Section 230 protection.<sup>201</sup> It does, however, create liability for interactive computer services pursuant to child sexual exploitation laws from all 50 states, creating “potential offenses that will encompass a wide variety of state laws that apply different legal standards to the same conduct.”<sup>202</sup> This sort of broad liability will likely result in the same type of blanket censorship caused by FOSTA-SESTA and may “condition the granting of a governmental privilege [of editorial authority] on individuals or entities doing things that amount to a violation of their First Amendment rights.”<sup>203</sup>

Thus, full government control of platform content moderation could be hindered by lack of understanding or vague and overly broad requirements. As FOSTA-SESTA and the EARN IT Act demonstrate, such a model is ineffective at best and harmful at worst.

### *C. The Solution is Government-Industry Cooperation with a Focus on Transparency, Consistency, and Accountability*

The current method of industry self-regulation in the content moderation arena is unlikely to last. Users grow frustrated with vague and inconsistent community guidelines and Congress and the DOJ have started to slowly chip away at Section 230 protections for platforms. But increasing

---

199. Emma Llansó, *Amendments to EARN IT Act Can't Fix the Bill's Fundamental Flaws*, CTR. FOR DEMOCRACY & TECH. (July 1, 2020), <https://cdt.org/insights/amendments-to-earn-it-act-cant-fix-the-bills-fundamental-flaws/> [<https://perma.cc/CC7L-CTNX>].

200. Letter from Sophia Cope & Aaron Mackey to Senate Judiciary Comm., *supra* note 193.

201. *See* EARN IT Act of 2020, S. 3398, 116th Cong. (2020).

202. *Id.*; Llansó, *supra* note 199.

203. Sophia Cope, Aaron Mackey, & Andrew Crocker, *The EARN IT Act Violates the Constitution*, ELEC. FRONTIER FOUND. (Mar. 31, 2020), <https://www EFF.ORG/deepinks/2020/03/earn-it-act-violates-constitution> [<https://perma.cc/NHN9-6BNT>].

criminal liability for platforms is not the best method if we value free speech and open dialogue on platforms.

The advent of social media platforms reimagined what social interaction and connection could look like. An effective content moderation model must also reimagine possible solutions. This Note asserts that the best method of content moderation is a committee (“Committee”) that brings together regulators, platforms, content creators, users, and third-party technology experts to propose a set of content moderation best practices that are voluntary to undertake that are voluntary to undertake. This Committee should be different from the EARN IT Act’s Commission, which is mainly comprised of law enforcement and seeks to create guidelines for platforms to avoid criminal liability.<sup>204</sup>

The role of the Committee proposed here is to create clear content moderation standards, regularly update and publish the standards, and publicly rate platforms based on their adherence to the standards.<sup>205</sup> The Committee should create a website where the best practices are published and where platforms and other websites that choose to adopt the best practices may be reported. Although platforms may adopt the standards on a voluntary basis, adoption with minimal changes will create clear content moderation expectations for users and creators.

The guiding principles of the standards must address what current content moderation guidelines lack: transparency, consistency, and accountability. Transparency will address the uncertainty creators experience when they upload content. Community guidelines today are too often described as a “black box,” where creators are uncertain as to whether their content will remain posted or whether it will get removed, demonetized, or restricted.<sup>206</sup> Although platforms may argue that releasing their proprietary algorithms and rules could expose their trade secrets, the issue could be remedied if “the

---

204. S. 3398, 116th Cong. § 3(c)(2) (2020).

205. Notably, Facebook has already created a similar committee, called the Oversight Board, to propose content moderation standards for Facebook. The committee proposed by this article takes the concept a step further by creating consistent standards across platforms so that users can expect certain baseline content moderation rules across the many platforms that they use. Nick Clegg, *Welcoming the Oversight Board*, FACEBOOK (May 6, 2020) <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/> [<https://perma.cc/7CNP-HMF7>].

206. Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353, 1377 (2018).

public disclosures were general rather than granular in nature.”<sup>207</sup> With standards clearly disclosed, creators could better understand what content is permissible.

Consistency will address standards that change frequently and abruptly. President Trump’s executive order and tweet to “REVOKE 230”<sup>208</sup> was likely due to Twitter’s inconsistent enforcement standards. Twitter had never flagged a post of his as potentially misleading before, and, suddenly, it decided to start flagging when President Trump’s rhetoric contained misinformation in a way it had never done before.<sup>209</sup> Therefore, consistency may quell government attempts to regulate Section 230, because such attempts generally focus on creating more predictability in the sort of content platforms do and do not allow. Further, since each platform currently creates its own standards, what is acceptable on one platform may be unacceptable on another. Though there is value in some variety from platform to platform, a consistent set of basic standards that serve as a floor would help create the predictability that is lacking across platforms. To address issues with consistency, the Committee must create and update standards on a regular schedule instead of changing them suddenly.

Finally, accountability will address the lack of consequences to platforms for inconsistent and opaque community guidelines without threatening their Section 230 protections. As previously described, the Committee must have a system in which it publicly ranks platforms and websites based on their adherence to the published best practices. This method seeks to incentivize platforms to comply by rewarding them with a high score for transparent and consistent content moderation standards without taking away their rights to moderate content as they please under Section 230.<sup>210</sup> In addition, the ranking system will serve as an important resource for content creators when deciding on which platforms to publish content. Platforms that have

---

207. *Id.* at 1384.

208. Donald J. Trump (@realDonaldTrump), TWITTER (May 29, 2020, 8:15 AM), <https://twitter.com/realDonaldTrump/status/1266387743996870656> [https://archive.is/iBmsK].

209. Nicholas Reimann, *Twitter Fact-Checks President Trump for the First Time*, FORBES (May 26, 2020, 5:56 PM), <https://www.forbes.com/sites/nicholasreimann/2020/05/26/twitter-fact-checks-president-trump-for-the-first-time/#752786f82905> [https://perma.cc/42D8-LGR8].

210. 47 U.S.C. § 230(c)(1) (providing that platforms shall not be treated as publishers of user-generated content); 47 U.S.C. § 230(c)(2) (protecting platforms from civil liability for steps taken to moderate most content).

more transparent and consistent rules are more appealing to creators that derive income from platforms and have the power to push users from one platform to another.<sup>211</sup>

This proposed committee would not be the first government-industry joint initiative in the technology space. A similar framework was used to create the cybersecurity framework of the National Institute of Standards and Technology (“NIST”).<sup>212</sup> “The [NIST] [f]ramework is voluntary guidance, based on existing standards, guidelines, and practices for organizations to better manage and reduce cybersecurity risk.”<sup>213</sup> The framework was first developed in 2014 and was involved collaboration from “industry, academia, and government stakeholders” and used “workshops, extensive outreach and consultation, and a public comment process” to create the final product.<sup>214</sup> In 2015, 30% of United States companies used the NIST framework, a number projected to increase to 50% by 2020.<sup>215</sup> The NIST framework provides all stakeholders involved with clear guidance on cybersecurity issues and gives both regulators and industry a hand in framing best practices.

Today, regulators, platforms, users, and creators all want more control over what content moderation looks like. What if, instead, content moderation involved input from all of these groups? The Committee proposed in this Note provides regulators with a set of clear best practices they can encourage, provides platforms with a seat at the table to decide how to moderate their own content, and provides users and creators with predictable and solid rules to rely on when deciding how and where to publish content.

---

211. Alexander, *supra* note 149.

212. *Questions and Answers*, NAT’L INST. OF STANDARDS AND TECH., <https://www.nist.gov/cyberframework/frequently-asked-questions/framework-basics#framework> [<https://perma.cc/63MF-WH93>].

213. *Id.*

214. *Id.*

215. *Cybersecurity Framework*, NAT’L INST. OF STANDARDS AND TECH., <https://www.nist.gov/industry-impacts/cybersecurity-framework> [<https://perma.cc/VG5J-CA3E>].

## V. CONCLUSION

In the backdrop of a global pandemic<sup>216</sup> and mass civil unrest,<sup>217</sup> President Trump propelled the issue of content moderation into public discourse in 2020.<sup>218</sup> The current content moderation method of complete platform self-regulation, however, has created unrest and dissatisfaction for regulators, platforms, content creators, and users alike.<sup>219</sup>

Popular methods proposed instead of self-regulation are expanding the First Amendment's free speech protections and full control by the government. However, both methods have inherent failings. First Amendment free speech protections consistently fail in courts because platforms are private entities with the power to regulate their own content.<sup>220</sup> Government-run solutions, on the other hand, have constitutional concerns and are often ill-informed and oppressive.<sup>221</sup> However, a third method is available—government-industry cooperation. Because social media is now a part of the daily lives of most Americans, the public has a stake in what content moderation looks like. A system of joint governance gives all interested parties a seat at the table and ensures that the needs of all parties are taken into account.

---

216. *Rolling Updates on Coronavirus Disease (COVID-19)*, WORLD HEALTH ORG., <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen> [<https://perma.cc/HPT7-9PPK>].

217. Larry Buchanan et al., *Black Lives Matter May Be the Largest Movement in U.S. History*, N.Y. TIMES (July 3, 2020), <https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html> [<https://perma.cc/X3BD-3W37>].

218. *See generally* Donald J. Trump (@realDonaldTrump), TWITTER (May 29, 2020, 8:15 AM), <https://twitter.com/realDonaldTrump/status/1266387743996870656> [<https://archive.is/iBmsK>].

219. Langvardt, *supra* note 206.

220. *E.g.*, Prager Univ. v. Google LLC, 951 F.3d 991, 995 (9th Cir. 2020).

221. Letter from Sophia Cope & Aaron Mackey to Senate Judiciary Comm., *supra* note 193; *Elrod v. Burns*, 427 U.S. 347, 347 (1976).