Comparing Social Media Content Regulation in the US and the EU: How the US Can Move Forward With Section 230 to Bolster Social Media Users' Freedom of Expression

TRENT SCHEURMAN*

TABLE OF CONTENTS

I. INTRODUCTION 414
II. BACKGROUND ON SOCIAL MEDIA, PLATFORM CONTENT REGULATION METHODS, AND FIRST AMENDMENT ISSUES 419
A. Social Media Users, Content Regulation, and Content-Bubbles 419
B. The Inverted Triangle of Social Media Content Regulation 425
C. The First Amendment, Viewpoint Discrimination, and the Hurdle of Social Media Companies Being Private Entities 427
III. APPLICABLE LAWS REGARDING SOCIAL MEDIA CONTENT REMOVAL IN THE US AND THE EU 431
A. US Section 230's Social Media Company Safe Harbor 431
B. The EU's Lesser Freedom of Expression, the E-Commerce Directive and Content Regulation 435
C. Major Key Differences In the US and the EU for Content Removal 440

* © Trent Scheurman. J.D. Candidate 2022, University of San Diego Law School.

IV. ANALYSIS OF SECTION 230'S DEFICIENCIES REGARDING SOCIAL MEDIA CONTENT REMOVAL PRACTICES 441
A. Social Media Companies' Unique Business Model Creates Overbroad Discretionary Authority in Content Removal 442
B. Platform's Lack of Transparency in Regulating Content 443
C. Viewpoint Discrimination 444
V. PROPOSED SOLUTIONS FOR DETERRING VIEWPOINT DISCRIMINATION UNDER 230(C)(2) 448
A. Revision of Statutory Language 449
B. Federal Guidelines for Independent Social Media Oversight Boards 452
VI. CONCLUSION 455

I. INTRODUCTION

In early January of 2021, President Donald Trump and some of his constituents, such as Stephen Bannon—a former White House chief strategist—were banned or suspended from their social media accounts by ten of the largest social media sites, including Facebook, YouTube, and Twitter.[1] The duration of these bans ranged from indefinite to permanent.[2] Facebook, YouTube, and Twitter were the leaders of the censorship campaign, reasoning that the accounts created a "risk of further incitement of violence" after accounts began circulating Trump's accusations of election fraud and his statements during the storming of the Capitol building in Washington D.C.[3]

Trump's presidential term was controversial, marked by polarization in party allegiance and scrutiny of the media. Specifically relevant to this Article is Trump highlighting social media platforms' abilities to block access to or remove user accounts and posts.[4] The storming of the Capitol

1. Hannah Denham, *These Are The Platforms That Have Banned Trump And His Allies*, THE WASH. POST (Jan. 14, 2021), washingtonpost.com/technology/2021/01/11/trump-banned-social-media [https://perma.cc/JR34-88B7]. For this Article, a social networking site ("social media website") is defined as a website that helps people communicate, socialize, and share information with other groups of people. *Social Networking Site*, CAMBRIDGE DICTIONARY, https://dictionary.cambridge.org/us/dictionary/english/social-networking-site [https://perma.cc/JQF6-X3EU]. Additionally, Facebook, as used in this Article, will refer to both the social media site and the company, as the company's renaming to Meta is still relatively new and uncertain during this Article's publication process. *See* Mike Isaac, *Facebook Renames Itself Meta*, N.Y. TIMES (Nov. 10, 2021), https://www.nytimes.com/2021/10/28/technology/facebook-meta-name-change.html [https://perma.cc/SQ76-BP6R].

2. Denham, *supra* note 1.

3. Denham, *supra* note 1.

4. *See* Jessica Guynn, *Biden and Section 230: New administration, same problems for Facebook, Google and Twitter as under Trump*, USA TODAY (Jan. 20, 2021),

building was not the first time the platforms' abilities to block access to or remove user accounts and posts came under scrutiny.[5] For instance, in June of 2020, conservative political party members were outraged over alleged censored campaign advertisements and other blocked political messages.[6] Then, in July of 2020, Facebook, YouTube, Twitter, and Squarespace censored doctors' advice about COVID-19 treatments and prevention methods declaring a video from the America's Frontline Doctors Summit spread misinformation related to COVID-19.[7] Despite the platforms' dismissive labeling of the potential cures and treatments discussed at the Summit, there is a legitimate public interest in hearing licensed medical professionals discuss options of treatment, and more broadly, other variations of opinion.

In both the banning of Trump's accounts and the blocking of the doctors' medical advice, the social platforms reasoned that the removal or blocking of the content served to prevent the spread of misinformation.[8]

---

https://www.usatoday.com/story/tech/2021/01/20/biden-trump-censorship-section-230-google-facebook-scrutiny/4238357001 [https://perma.cc/N5E7-KLGY].

5. *See* Wendy Davis, *Conservative Activists To Pursue 'Censorship' Battle Against Tech Companies*, MEDIA POST (June 29, 2020), https://www.mediapost.com/publications/article/353123/conservative-activists-to-pursue-censorship-batt.html [https://perma.cc/ZF7W-8XUX].

6. *See id.*

7. Caroline Warnock, *America's Frontline Doctors Summit COVID-19 Video Called 'False Information'*, HEAVY (July 28, 2020), https://heavy.com/news/2020/07/america-frontline-doctors-summit/ [https://perma.cc/E8RZ-MQ9S]; *see* also Stephanie Dwilson, *America's Frontline Doctors' Website Expired: Squarespace Site Down*, HEAVY (July 28, 2020), https://heavy.com/news/2020/07/americas-frontline-doctors-website-expired-down [https://perma.cc/2VXF-ET55]. Facebook, Twitter, YouTube, and Squarespace are all social media websites. Facebook is a site that uses a "Newsfeed" consisting of users posting pictures, videos, and text allowing other users to see, share, and respond to the post. Chaim Gartenberg, *What is Facebook? Just ask Mark Zuckerberg*, THE VERGE (Mar. 9, 2019), https://www.theverge.com/2019/3/8/18255269/facebook-mark-zuckerberg-definition-social-media-network-sharing-privacy [https://perma.cc/RUZ6-KS6E]. YouTube is a video-sharing platform that encourages users to make appealing videos and also watch other user videos. *What is YouTube*?, DIGITAL UNITE, https://www.digitalunite.com/technology-guides/tv-video/youtube/what-youtube [https://perma.cc/DTR6-3EK2]. Twitter is a micro- blogging site where users can post short 280-character posts called "Tweets" that communicate with other users. Paul Gill, *What Is Twitter & How Does It Work*?, LIFEWIRE (Aug. 29, 2021), https://www.lifewire.com/what-exactly-is-twitter-2483331 [https://perma.cc/5V8M-KYJ6]. Lastly, Squarespace is a website building platform, with blogging and hosting services, that lets businesses of all types create websites and network with each other. Sara Angeles, *How to Use Squarespace to Build a Website for Your Business*, BUS. NEWS DAILY (Jan. 29, 2019), https://www.businessnewsdaily.com/5484-how-to-use-squarespace.html [https://perma.cc/BU6X-2E5Y].

8. *See* Davis, *supra* note 5; *see also* Dwilson, *supra* note 7.

Regardless of the circumstances, such instances demonstrate the ability of social media organizations to broadly remove information across their user audience and render information unavailable to the public. These examples highlight the potential for viewpoint discrimination, which refers to singling out particular perspectives and suppressing those perspectives.[9]

For individualized viewpoint discrimination to occur, individuals must first avail themselves of the platforms' services. One must qualify as a social media platform user before facing potential viewpoint discrimination. To become a user of any of these major platforms, the user must agree to the respective platform's Terms of Service when creating an account. The Terms of Service, also known as "Terms of Use" or "Terms and Conditions," is a contract between the social media company and the user.[10] The social media user legally assents to the site's Terms of Service even without performing any affirmative act, such as clicking on an agreement through a hyperlink or radio button.[11] Users—often unknowingly—waive all meaningful rights, warranties, and remedies, while the platform asserts its interests to the limits of the law.[12] For example, Facebook's Terms of Service limits the "aggregate liability" arising out of or related to "these Terms" or the "Facebook Products" to not "exceed the greater of $100 or the amount you have paid [Facebook] in the past twelve months."[13]

Platforms' broad regulatory power over their own users enables platforms to discriminate against the users by inhibiting users that do not share the platforms' own viewpoints. As private companies, social platforms are entitled to self-regulate their businesses and form their own beneficial contracts. Because social media platforms are private companies, social

---

9. Boos v. Barry, 485 U.S. 312, 316–319 (1988) (offering an example of a viewpoint-based regulation found to be discriminatory on its face; statute was viewpoint discriminatory because it singled out one particular perspective—criticism of foreign governments—for suppression).

10. Michael L. Rustad & Thomas H. Koenig, *Wolves of the World Wide Web: Reforming Social Networks' Contracting* Practices, 49 WAKE FOREST L. REV. 1431, 1450–51 (2014).

11. *Id.* at 1451.

12. *Id.*

13. *Terms of Service*, FACEBOOK, https://www.facebook.com/terms.php [https://perma.cc/2TCP-26NC]. Additionally, if Facebook determines that a user "clearly, seriously or repeatedly breached" the "Terms or Policies," including the "Community Standards," Facebook "may suspend or permanently disable access" to the account. *Id.* Facebook "may also suspend or disable" the account if a user "repeatedly infringe[s] other people's intellectual property rights" or where "required to do so for legal reasons." *Id.* Facebook does prohibit forms of lawful speech on its website including "hate speech," defined as a direct attack against people on the basis of protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. Unlawful acts Facebook prohibits include violations of "intellectual property rights" and "fraud." *Facebook Community Standards*, META, https://transparency.fb.com/policies/community-standards/ [https://perma.cc/JZ6D-W5EZ].

media users face the continual threat of discriminatory post removal without any legal civil remedies.

But what happens when platforms remove important content regarding topics like politics, social movements, or healthcare? Or if the platforms funnel the content to the extreme, eliminating millions of people's access to minority viewpoints? The problem is the size, scope, and control of the platforms over users; the platforms connect billions of people daily, and have exclusive control over informing billions of people, selectively organizing what people see and can say on their timelines. As the platforms are centered around user expression via their users' posts, and exercise control over user expression, why not treat the platforms as a governmental structure in terms of how the platforms regulate posts? Or, alternatively, why not create some sort of independent unit to oversee uniform and fair review of user complaints regarding removal of content? These questions will be addressed and answered as this Article explores the issues and concerns surrounding the extent of platforms' censorship abilities, and how those abilities impact everyday users of the platforms.

This Article will compare 47 U.S.C. § 230 ("Section 230"), the United States law governing civil claims that prevent social media companies from being treated like the publishers of their own users' posts and the companies' abilities to remove user posts, with the European Union's ("EU") equivalent governing law, the E-commerce Directive.[14] The E-Commerce Directive will be used as an example of a governmental regulation that better prevents viewpoint discrimination, but at the cost of a lower standard of user expression. A lower standard of user expression means diminished rights in exercising free speech, as exemplified by the EU outlawing broader categories of speech than the US (Section III covers this point in detail). Then, this Article will demonstrate how the US may achieve the goal of decreasing discretionary power of platforms' content removal abilities, thereby minimizing viewpoint discrimination of lawful user-posted content, while preserving private governance of social media business practices.

Section II provides background on social media users, platform content regulation, and content removal practices. It continues with a discussion of the enormous amount of content social media platforms are responsible for monitoring and governing. Additionally, the relationships of social

14. 47 U.S.C. § 230(c) (2018); Council Directive 2000/31, 2000 O.J. (L 178) 1 (EC) [hereinafter Council Directive 2000/31].

media companies, governments, and users are explained in connection with social media content moderation. Lastly, Section II summarizes the First Amendment's boundaries on protection of speech and clarifies freedom of expression for US citizens only from government actors, leaving private platforms content removal practices currently out of the First Amendment's reach.

Section III lays out the social media content regulation laws governing both the US and the EU. Historically, the US's Section 230 has been referred to as the "26 words that created the internet" due to its thorough protection of private online platforms from third-party ("intermediary") liability arising from civil suits like defamation, and general allowance for platforms to leave up or take down content voluntarily.[15] Contrarily, the EU's E-commerce Directive offers platforms safe harbor from legal liability with two main requirements: the platform must (1) not have "actual knowledge of illegal activity," and (2) "act expeditiously to remove" illegal activity once actual knowledge is obtained.[16] This section concludes by illustrating the EU's approach to social media content regulation and reviewing its implications on viewpoint discrimination in social platform content moderation.

Section IV discusses the deficiencies of Section 230 in its approach to platform content moderation. The analysis will continue with the three main problems arising from Section 230's current application, which allows social media companies: (1) overbroad discretionary authority, (2) the ability to operate with limited transparency, and (3) the ability to discriminate based on viewpoint. Additionally, the Article will explore the implementation and significance of the ground-breaking independent Facebook Oversight Board on providing an appellate process for wrongful censorship of posts.

Section V proposes two solutions to the three previously listed issues of Section 230(c) addressed in this Article. The first solution is statutory revision of Section 230(c)(2). There are two statutory revisions proposed in the first solution: (1) revision of the statute to grant immunity to social media platforms only if the platforms remove content that is illegal or otherwise unprotected by the First Amendment, and (2) introduction of a "bad faith" clause that removes platform immunity if the plaintiff can prove their lawful post was removed as a result of viewpoint discrimination. The second solution suggests federal statutes mandating large social media platforms create their own independent oversight boards. These Social Media Oversight Boards will be primarily based on Facebook's Oversight

15. Guynn, *supra* note 4.
16. Council Directive 2000/31, *supra* note 14, art. 14.

Board, with the new Oversight Boards' purpose being independent review of platform censorship practices through a board review process.

II. BACKGROUND ON SOCIAL MEDIA, PLATFORM CONTENT REGULATION METHODS, AND FIRST AMENDMENT ISSUES

This section will start by covering the monthly global reach of social media companies and how social media companies regulate online content. Next, this section will describe algorithms and their advantages and disadvantages as content-filtering tools used to sort content and will describe the inverted-triangle model relationship between governments, social media companies, and speakers. The section will end with an explanation of why First Amendment claims fail to beat Section 230's civil liability immunity.

A. Social Media Users, Content Regulation, and Content-Bubbles

Social media websites are currently used by a significant portion of the global population. As of July 2020, 4.574 billion people worldwide regularly use the internet, with 3.96 billion of those individuals also being social media users.[17] For this Article, an individual qualifies as a "social media user" if he or she accesses a social media website at least once per month.[18] Social media users utilize social media platforms for a multitude of purposes—socializing, working, networking, collaborating, revenue generating—all while on the couch, at the office, or on the go.[19]

Facebook, YouTube, and Twitter, three of the largest recognized global social media providers, respectively recorded 2.449 billion, 2.0 billion, and 340 million active monthly social media users worldwide as of January 2020.[20] Regionally, the United States reported 295 million active social

17. J. Clement, *Global Digital Population as of July 2020*, STATISTA (July 24, 2020), https://www.statista.com/statistics/617136/digitalpopulationworldwide/#:~:text=Almost%204.57%20billion%20people%20were,percent%20of%20the%20global%20population [https://perma.cc/6URL-RJPA].

18. *See* Simon Kemp, *Digital 2020: 3.8 Billion People Use Social Media*, WEARESOCIAL (Jan. 30, 2020), https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media [https://perma.cc/D4WC-PZRM].

19. *Id.*

20. *Id.*

media users in 2021.[21] Europe reported 325.70 million active social media users per month as of January 1, 2021.[22] Besides websites, social platforms also utilize mobile apps to appeal to more potential users through convenience and ease of use.[23]

In terms of content regulation, the social media platform serves as the gatekeeper to its own site, either passively accepting or actively rejecting the content its users post and the content other users view on the platform.[24] The high volume of users on the major sites contribute to the exponentially growing levels of individual content creation. To illustrate, Facebook has 350 million photos uploaded every day, or 4,000 photos uploaded per second.[25] YouTube has 500 hours of video uploaded every minute.[26] Twitter, on average, has around 6,000 tweets tweeted per second, or 500 million tweets per day.[27] The continuously increasing flood of content requires social media organizations to use complex measures to regulate the torrent of content posted on their platforms.[28]

The modern measures utilized by social media platforms in approaching content moderation and content removal are shared across the social media industry, and include human and algorithmic types of review based on platform community guidelines.[29] Human review of all posts by staff

21. *Number of social network users in the United States from 2017 to 2026,* STATISTA (Jan. 28, 2022), https://www.statista.com/statistics/278409/number-of-social-network-users-in-the-united-states/ [https://perma.cc/L6J8-YXZR].

22. Number includes users from Germany, the United Kingdom, France, Italy, Spain, Poland, Ukraine, Netherlands, and Belgium. *Number of monthly active mobile social media users in Europe as of January 2021, by country*, STATISTA (Jan 28, 2022), https://www.statista.com/statistics/299496/active-mobile-social-media-users-in-european-countries/ [https://perma.cc/YJ23-9U6M].

23. Paige Cooper, *All the Social Media Apps You Should Know in 2021*, HOOTSUITE (May 17, 2021), https://blog.hootsuite.com/best-social-media-apps-list [https://perma.cc/S3HT-AKMF].

24. David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 389 (2010).

25. *Facebook by the Numbers: Stats, Demographics & Fun Facts*, OMNICORE (Oct. 28, 2020), https://www.omnicoreagency.com/facebookstatistics/#:~:text=350%20million%20photos%20are%20uploaded,300%2C000%20users%20helping%20in%20translation [https://perma.cc/UJ8Z-FLCG].

26. Susan Wojcicki, *YouTube at 15: My Personal Journey and the Road Ahead*, YOUTUBE OFF. BLOG (Feb. 14, 2020), https://blog.youtube/news-and-events/youtube-at-15-my-personal-journey [https://perma.cc/JD7T-F35X].

27. *Twitter Usage Statistics*, INTERNET LIVE STATS, https://www.internetlive stats.com/twitter-statistics/ [https://perma.cc/47R4-VERY].

28. Abbey Stemler, *Regulation 2.0: The Marriage of New Governance and Lex Informatica*, 19 VAND. J. ENT. & TECH L. 87, 105–07 (2016).

29. Christopher Gao, *Social Media Censorship, Free Speech, and the Super Apps*, CAL. L. REV. BLOG (Oct. 2020), https://www.californialawreview.org/social-media-censorship-free-speech [https://perma.cc/M25N-E8VS].

members is impossible, given the size and exponentially increasing scope of the amount of posts.[30] Even with the specialized industry algorithms, companies struggle to differentiate between legitimate and illegitimate content.[31]

An algorithm is a set of instructions and rules created to give a computer the ability to perform a specific task or solve a specific problem.[32] Social media algorithms are proprietary information and individualized to each company, which varies how each platform undertakes the task of content-filtering. Content-filtering refers to software that screens and blocks online content that includes particular words or images.[33] The software includes blocked and buzzword lists created in two ways: human review and automated selection.[34] Human review consists of humans going through websites, searching for objectionable material and words commonly associated with such material (i.e. racial slurs, sexist phrases), and adding them to the software, while automated selection builds off those inputs, constantly updating itself with similar keywords.[35]

Algorithms have advantages and disadvantages in content-moderation. Algorithms are advantageous in showing relevant ads and suggesting similar content to users all based on the users' preferences and behaviors analyzed through machine-learning.[36] Also, social media algorithms help companies flag potentially problematic content for removal.[37] But algorithms are

30. The previous paragraph illustrates this point. Using Twitter tweets as an example, having a team of human reviewers attempting to review 500 million tweets per day is not cost effective. The number of employees needed to complete a single day's review of tweets could not be sustained long-term.

31. Mark Scott & Laura Kayali, *What Happened When Humans Stopped Managing Social Media Content*, POLITICO (Oct. 21, 2020, 5:56 PM), https://www.politico.eu/article/facebook-content-moderation-automation/ [https://perma.cc/VPU2-7B5X].

32. Veronica Appia, *What is an algorithm and how is it used by big tech?*, TORONTO.COM (Dec. 8, 2020), https://www.toronto.com/news-story/10283678-what-is-an-algorithm-and-how-is-it-used-by-big-tech-/ [https://perma.cc/U2XZ-ZAAP].

33. Norman Clark, *Content Filter*, BRITANNICA, https://www.britannica.com/technology/content-filter [https://perma.cc/Y36K-U2YV].

34. *Id.*

35. *Id.*

36. Appia, *supra* note 32.

37. *YouTube Community Guidelines enforcement: Videos removed, by source of first detection*, GOOGLE TRANSPARENCY REP., https://transparencyreport.google.com/youtube-policy/removals?hl=en [https://perma.cc/VW93-26EQ]. 9,321,948 videos were removed from YouTube from October 2020 through December 2020, with 8,800,082 of those removed videos being flagged automatically by algorithm. *Id.* Flags can come from the automated

disadvantageous as well. As with all things, a healthy level of skepticism is warranted when using algorithms. Many people are unaware of the underlying algorithms that work behind-the-scenes of social media.[38] If people think what they see on their feed is "news," instead of content that is curated specifically for them, as well as only engage with people that have similar beliefs, this creates a content bubble—an online space which only reinforces their beliefs.[39] These bubbles create higher levels of engagement amongst users interacting with the platform content, and the bubbles are directly related to profit-maximizing algorithms of the platforms.[40] The bubble leads to negative effects such as political polarization and limiting exposure to diverse views.[41]

Moreover, using algorithms repeatedly and pervasively over large populations of people may inappropriately associate risky or otherwise undesirable content with people's accounts, imposing unjustified burdens and hardships on populations, and reinforcing existing inequalities.[42] For example, in early 2019, news broke that YouTube was recommending explicit self-harm videos with titles like "my huge extreme self-harm scars" and search terms like "how to self-harm tutorial" to its users, many of

---

flagging systems, from members of the Trusted Flagger program (NGOs, government agencies, and individuals) or from users in the broader YouTube community. *Id.*

38. Emilie Robichaud, *How Social Media Algorithms Drive Political Polarization*, MEDIUM: THE STARTUP (Oct. 8, 2020), https://medium.com/swlh/how-persuasive-algorithms-drive-political-polarization-75819854c11d [https://perma.cc/AU6A-5KHW].

39. *Id.*

40. Michelle Hampson, *Smart Algorithm Bursts Social Networks' "Filter Bubbles"*, IEEE SPECTRUM (Jan. 21, 2021), https://spectrum.ieee.org/finally-a-means-for-bursting-social-media-bubbles [https://perma.cc/AA7R-5E9T].

41. Robichaud, *supra* note 38. *See also* Wendy Rose Gould, *Are you in a Social Media Bubble? Here Is How to Tell*, NBC NEWS (Oct. 21, 2019, 10:06 AM), https://www.nbcnews.com/better/lifestyle/problem-social-media-reinforcement-bubbles-what-you-can-do-about-ncna1063896 [https://perma.cc/4BTK-659L] ("Social media giants — including Google, Facebook and Twitter—use algorithms that are ever-changing and top secret, which ultimately create filter bubbles." "The reality is that all platforms now constantly feed us content that aligns with our own interests, friends and belief systems." "[Platforms] are able to take what we browse or post about and feed us back our own thoughts gathered from other social media followers as though we have hundreds and thousands of friends feeling the same way.").

42. Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1167 (2018) [hereinafter Balkin, *Free Speech in the Algorithmic Society*]. For this Article, an individual being treated as risky or otherwise undesirable means the individual is selected by the platforms' algorithm for being socially taboo (i.e., encouraging suicide or suicidal inclinations) and as a result those posts are removed or negatively impacted (i.e., removal of partial clips or sound bites). Unjustified burdens and hardships refer to individuals who rely on social media as their income source and suffer by having a video that the individuals invested time and potentially money into to be wrongly flagged and removed, preventing the individuals from receiving monetary benefits from the removal or censorship.

whom were young children.[43] YouTube responded to the crisis by deleting the original accounts encouraging self-harm as the accounts violated the community guidelines against suicide and self-injury, and suspended comments encouraging self-harm, but some felt this did not address the core problem: the platform's use of artificially intelligent algorithms built to maximize engagement metrics[44] above all else.[45] The reference algorithm treated the young children as having a tendency to harm themselves, enveloping the children in negative content bubbles, exposing young children to significant traumatic emotional material, all autonomously.[46] This is an extreme example, but it reinforces the real threat concerning platform algorithms trapping users in content-bubbles.

The current content moderation algorithms implemented by social media companies are imperfect tools for content regulation.[47] The algorithms are imperfect in the sense that these algorithms block more lawful speech than the social media companies wish, and are incapable of exercising human judgment.[48] For example, algorithms can identify certain flagged words or phrases (i.e., "gay," "crime," "abuse"),[49] but they are unable to determine whether the speech in fact defames others or invades an individual's privacy.[50] When it comes to handling internet defamation issues and other online reputation attacks—specifically in terms of removing the content

43. Daniyal Malik, *YouTube Faces Severe Criticism For Recommending Self Harm Videos Again*, DIGIT. INFO. WORLD (Feb. 7, 2019), https://www.digitalinformationworld.com/2019/02/youtube-recommending-self-harm-videos-in-search-results-criticized.html [https:// perma.cc/7VR5-6HR5].

44. Engagement metrics measure how much and how often the audience accounts are interacting with other accounts. Examples of these metrics are likes, comments, and shares. Jenn Chen, *The most important social media metrics to track*, SPROUTSOCIAL (Mar. 26, 2021), https://sproutsocial.com/insights/social-media-metrics/ [https://perma.cc/V4SU-HZU2].

45. Allison Zakon, *Optimized for Addiction: Extending Product Liability Concepts to Defectively Designed Social Media Algorithms and Overcoming the Communications Decency Act*, 20 WIS. L. REV. 1107, 1110 (2020).

46. *See* Malik, *supra* note 43.

47. Ardia, *supra* note 24, at 390.

48. Ardia, *supra* note 24, at 390.

49. Lindsay Dodgson, *YouTubers have identified a long list of words that immediately get videos demonetized, and they include 'gay' and 'lesbian' but not 'straight' or 'heterosexual'* (Oct. 1, 2019), https://www.insider.com/youtubers-identify-title-words-that-get-videos-demonetized-experiment-2019-10 [https://perma.cc/UE5W-LLPV].

50. Ardia, *supra* note 24, at 390.

from the internet—the solutions are, in two words: fact dependent.[51] There is no perfect solution that can be applied to every instance of online defamation.[52] Such determinations require contextual analysis and additional fact gathering.[53] As a result, the process of assessing legal liability cannot be currently solved by algorithmic flagging of words or phrases.[54]

Regarding content-removal, when an algorithm does remove a post, a significant factor for the fate of the removed content is whether the post is controversial. Facebook, Twitter, and YouTube all have appeal processes for users appealing their removal of content.[55] Facebook is the first platform to create an external and independent appeal process for removed material. In November 2018, Facebook announced the construction of an independent oversight body, the Facebook Oversight Board ("FOB").[56] The FOB's purpose is to serve as an appellate review system for user content and to make content-moderation policy recommendations to Facebook.[57] It seeks to grant transparency to Facebook's internal deliberations in deciding how to design and implement the FOB.[58]

Facebook admits it does not "always get it right" when it decides to take something down, which is why they created the FOB.[59] The FOB independently reviews some of the most difficult and significant content decisions Facebook makes.[60] Facebook stipulates users can appeal removed posts to

---

51. Whitney Gibson, *Removing Internet Defamation From the Internet: Solutions are Fact-Dependent*, VORYS (Jan. 26, 2016), https://www.vorys.com/publications-1673.html [https://perma.cc/74EV-UQ7R].

52. *Id.*

53. Ardia, *supra* note 24, at 390.

54. Ardia, *supra* note 24, at 390–91.

55. Stan Horaczek, *Here's What You Can Do If Your Social Media Post Gets Taken Down*, POPULAR SCI. (July 20, 2018), https://www.popsci.com/appeal-social-media-post-takedowns [https://perma.cc/5RLX-YPC8]. Twitter follows a multi-level system depending on the severity of the violation and frequency of violations regarding their standards, including Tweet-Level Enforcement, Direct-Level Message Enforcement, or Account-Level Enforcement. *Our Range Of Enforcement Options*, TWITTER, https://help.twitter.com/en/rules-and-policies/enforcement-options [https://perma.cc/7TAV-HEV8]. YouTube only allows a single appeal on removals based on Community Guidelines, where a human reviews the removed post and determines whether to reinstate the post or keep the post removed. *Appeal Community Guidelines Actions*, YOUTUBE, https://support.google.com/youtube/answer/185111?hl=en [https://perma.cc/VFP4-G5NT]. The approaches are substantially similar across the US and the EU.

56. Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2418 (2020).

57. *Id.*

58. *Id.* at 2425–26.

59. *What is the Oversight Board?*, FACEBOOK, https://www.facebook.com/help/711867306096893?helpref=related&ref=related&source_cms_id=3463664531159 [https://perma.cc/PV2V-ACAE].

60. *Id.*

its new Oversight Board if Facebook has taken down the content and already upheld a human review finding that a post was a violation of the Community Standards.[61] Section IV will go into greater analysis concerning the significance of the FOB and what it means for creating other oversight boards to deter viewpoint discrimination by granting greater transparency.

It is wise for platforms to err on the side of blocking controversial posts to save face and protect revenue streams. From the platform's perspective, it costs less to remove expressive—yet problematic—posts rather than risking damage to the website's reputation.[62] Recalling the YouTube self-harm example, YouTube experienced severe negative pressure in the media and even though YouTube did not intend to promote self-harm to children, its underlying goals of profit and reputation remained constant, evidenced by YouTube instantly removing the accounts after the complaints by users, to salvage its reputation and ease its users' outrage.[63] Social media companies are for-profit businesses that prioritize profits and tenaciously protect their bottom line, which naturally leads to preemptively avoiding any issues by removing controversial content before profit margins suffer.[64]

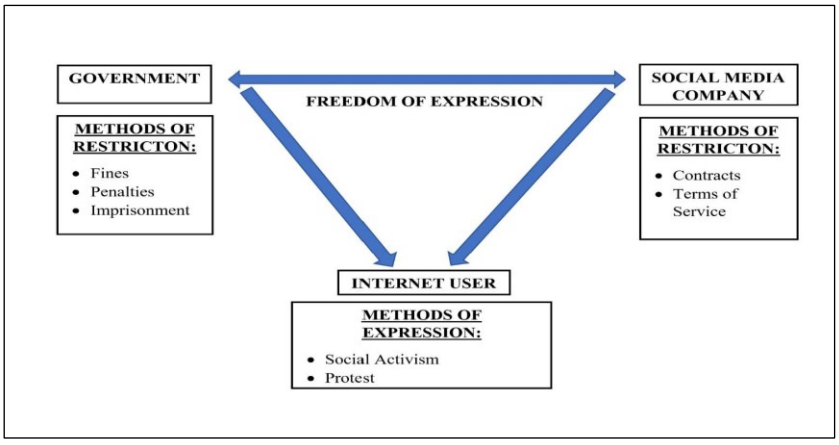
B. The Inverted Triangle of Social Media Content Regulation

Social media content moderation concerning speech involves balancing relationships between platforms, governments, and users, best understood when thought of as an inverted triangle.[65] In this inverted triangle, there

61. *Id.*
62. *See generally* Ardia, *supra* note 24, at 391.
63. Malik, *supra* note 43.
64. David L. Hudson, Jr., *Free Speech or Censorship? Social Media Litigation is a Hot Legal Battleground*, ABA J. (Apr. 1, 2019, 12:05 AM), https://www.abajournal.com/magazine/article/social-clashes-digital-free-speech [https://perma.cc/9GM3-H63K].
65. Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011, 2014-15 (2018) ("On one corner of the triangle are nation-states, states, municipalities, and supranational organizations like the European Union. On the second corner of the triangle are internet-infrastructure companies. These include social media companies . . . search engines, internet service providers (ISPs) . . . the internet infrastructure is important, if not crucial, to people's practical ability to speak. In most countries, this internet infrastructure, or important parts of it, are privately owned. On the third corner of the triangle, at the very bottom, we have speakers and legacy media, including mass-media organizations, protesters, civil-society organizations, hackers, and trolls. Although both states and infrastructure owners regulate their speech, they are sometimes able to influence states and infrastructure owners through social activism and protest.").

are the top two points: (1) the social media company, and (2) the relevant governmental body. Both points taper downward, exerting pressure on the last point—those in the user role like individuals, associations, and commercial companies.[66] The two top points hold power over the expression of the bottom.[67]

FIGURE 1



The above diagram illustrates the relationships at play in content regulation.[68] Social media companies regulate users' speech through private administrative practices like Terms of Service and privacy policies, while the government regulates speakers through fines, penalties, imprisonment, or other forms of punishment.[69] The top two points of the triangle connect through additional restrictions that governments implement on social media companies, vicariously impacting the expression of the speakers at the bottom of the triangle.[70]

An example of extreme government regulation on social platforms is demonstrated by China, which implements a strategy known as the "Great Firewall" to regulate internet access by criminalizing non-government

66. *Id.* at 2014–15.
67. *Id.*
68. Figure 1 is an original chart based entirely on illustrating Balkin's theory of the three aspects of free speech forming a triangle. *See id.*
69. *Id.* at 2015, 2021.
70. *Id.* at 1187–89.

approved material posted on available platforms.[71] China's internet users have less freedom of expression due to the government's stricter policies regarding what the speakers may post—the policies tighten the sides of the triangle and shrink the freedom of expression area within the triangle.[72] Thus, more restrictions on content lead to a smaller triangle of expression for speakers, while removing restrictions creates a larger triangle and more room for speakers to express themselves.[73]

Likewise, the more regulations social media platforms enforce, the less power the speakers have over expressing themselves. The inverted triangle, though a great starting visual, is not sufficient for understanding all complications of private companies censoring content that constitutes users' freedom of expression.[74]

C. The First Amendment, Viewpoint Discrimination, and the Hurdle of Social Media Companies Being Private Entities

The First Amendment provides that "Congress shall make no law . . . abridging the freedom of speech, or the press,"[75] however, First Amendment speech protection is generally not extended to private company self-regulation.[76] The major social media organizations—Facebook, YouTube, and Twitter—are private companies.[77] Just as the New York Times, CNN, or any other traditional media platform has no obligation to host a particular

71. Christopher Stevenson, *Breaching the Great Firewall: China's Internet Censorship and the Quest for Freedom of Expression in a Connected World*, 30 B.C. INT'L & COMP. L. REV. 531, 538 (2007).

72. Dennis Normal, *Science Suffers as China's Internet Censors Plug Holes in Great Firewall*, SCIENCE (Aug. 30, 2017), https://www.science.org/content/article/science-suffers-china-s-internet-censors-plug-holes-great-firewall [https://perma.cc/ARP9-57LN]. Elaborating on China's stricter policies regarding what the speakers may post, China reportedly has 50,000 internet police who monitor domestic social media sites, deleting posts deemed seditious or merely critical of the government. *Id.* Numerous research sites are also blocked, including Google Scholar which is important for scholarly, Google Docs and Dropbox, which allow scientists to share materials for organizing conferences and managing collaborations, and even the NASA Jet Propulsion Laboratory. *Id.*

73. Balkin, *Free Speech in the Algorithmic Society*, *supra* note 42, at 1187–89.

74. Hudson, *supra* note 64.

75. U.S. CONST. amend. I.

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

77. Nadine Strossen, *Does the First Amendment Apply to Social Media Companies*, TALKSONLAW, https://www.talksonlaw.com/briefs/does-the-first-amendment-require-social-media-platforms-to-grant-access-to-all-users [https://perma.cc/8D9V-35BR].

message, the same is true for social media organizations.[78] The terms "freedom of speech," "freedom of expression," and "freedom of the press" are treated equally in United States jurisprudence.[79] Therefore, freedom of speech and freedom of expression shall be used interchangeably and signify the same legal doctrines when referring to users' speech or how users went about expressing themselves.

The Supreme Court consistently interprets the First Amendment's protections to extend to individual and collective speech "in pursuit of a wide variety of political, social, economic, educational, religious, and cultural ends."[80] Therefore, speech is generally protected under the First Amendment unless it falls within one of the narrow categories of unprotected speech, which are obscenity, defamation, fraud, incitement, fighting words, true threats, speech integral to criminal conduct, and child pornography.[81] Whether courts apply strict scrutiny or a lower form of scrutiny depends on the character and context of the speech.[82] If the restriction is content-based, meaning the restriction discriminates against speech based on the substance of what it communicates, then strict scrutiny will be applied, and if the restriction is content-neutral, then a lower standard of scrutiny may be applicable.[83] At issue here is viewpoint discrimination, which is content-based in nature.

Political and ideological restrictions on speech receive strict scrutiny by the courts.[84] The Supreme Court considers political and ideological speech to be at the core of the First Amendment, including speech concerning "politics, nationalism, religion, or other matters of opinion."[85] Political

78. *Id.*

79. First Nat'l Bank v. Bellotti, 435 U.S. 765, 732, 799 (1978).

80. Roberts v. U.S. Jaycees, 468 U.S. 609, 622 (1984).

81. CONG. RESEARCH SERV., *The First Amendment: Categories of* Speech, (Jan. 16, 2019), https://fas.org/sgp/crs/misc/IF11072.pdf (noting the limited categories of speech that the government may regulate because of the speech's content, if the government regulates it evenhandedly); *see R.A.V. v. St. Paul*, 505 U.S. 377, 382–88 (1992) (identifying these unprotected speech categories as obscenity, defamation, fraud, incitement, fighting words, true threats, speech integral to criminal conduct, and child pornography).

82. Identifying the category of speech at issue (e.g., commercial speech, obscenity) is an important step in determining what First Amendment standards, including what level of judicial scrutiny, a court might apply to the law. CONG. RESEARCH SERV., *supra* note 81, at 1. Regulations of protected speech generally receive strict or intermediate scrutiny, which are high bars for the government to meet. CONG. RESEARCH SERV., *supra* note 81, at 1. In contrast, the government typically has more leeway to regulate unprotected speech. CONG. RESEARCH SERV., *supra* note 81, at 1.

83. David L. Hudson Jr., *Content Based*, THE FIRST AMENDMENT ENCYCLOPEDIA, https://www.mtsu.edu/first-amendment/article/935/content-based [https://perma.cc/4V27-Z7SU].

84. *Id.*

85. W. Va. State Bd. of Educ. v. Barnette*,* 319 U.S. 624, 642 (1943).

speech can also take other forms beyond writing or words, such as symbolic acts.[86] A government regulation that implicates political, or ideological speech generally receives strict scrutiny in the courts, where the government must show that the law is "narrowly tailored" to achieve "a compelling government interest."[87]

A form of content-based discrimination at the level of strict scrutiny is viewpoint discrimination. Viewpoint discrimination is an "egregious" form of content-based discrimination.[88] When the government targets not subject matter, but particular views taken by speakers on a subject, the violation of the First Amendment is "all the more blatant."[89] The government must abstain from regulating speech when the specific motivating ideology or the opinion or perspective of the speaker is the rationale for the restriction.[90] The test for viewpoint discrimination is "whether—within the relevant subject category—the government has singled out a subset of messages for disfavor based on the views expressed."[91]

Further clarification of viewpoint discrimination is found in *Matal v. Tam*.[92] The Supreme Court in *Matal* explained that although viewpoint discriminatory conduct targeting offensive speech may be evenhanded and apply to multiple, allegedly offensive viewpoints or speakers, the First Amendment forbids any attempt to ban all offensive speech because "[g]iving offense is a viewpoint."[93] In its ruling, the Supreme Court emphasized that the First Amendment's prohibition against viewpoint discrimination

86. Texas v. Johnson, 491 U.S. 397, 412–14 (1989) (holding criminal conviction for burning the American flag violated the defendant's First Amendment rights because the defendant burning the flag qualified as political expression and the state lacked findings of that action breaching the peace during the demonstration).

87. Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1, 551 U.S. 701, 783 (2007).

88. Rosenberger v. Rector & Visitors of the Univ. of Va., 515 U.S. 819, 829 (1995).

89. *Id.*

90. *Id.*

91. Matal v. Tam, 137 S. Ct. 1744, 1750 (2017); *see also Boos*, 485 U.S. at 316 (offering an example of a viewpoint-based regulation found to be discriminatory on its face). In *Boos*, the Supreme Court struck down a District of Columbia statute that criminalized the display of any sign criticizing a foreign government within five hundred feet of its embassy. *Id.* On its face, this statute was viewpoint discriminatory because it singled out one particular perspective—criticism of foreign governments—for suppression. *Id.* at 319.

92. *Matal*, 137 S. Ct. at 1744.

93. *Id.* at 1763.

is central to an understanding of First Amendment protections.[94] The Supreme Court states that the law protects offensive, even "hateful" speech: "Speech that demeans on the basis of race, ethnicity, gender, religion, age, disability, or any other similar ground is hateful; but the proudest boast of our free speech jurisprudence is that we protect the freedom to express even 'the thought that we hate.'"[95]

However, the First Amendment does not extend to private companies prohibiting user expression.[96] Private companies are private actors, not governmental bodies, and the Supreme Court of the United States interprets the free-speech clause of the First Amendment to constrain only governmental actors, not private actors.[97] To draw the line between governmental and private actors, the Supreme Court applies the state-action doctrine.[98] Under the state-action doctrine, a private entity may only be considered a state actor when it "exercises a function traditionally exclusively reserved to the State."[99]

The Court finds private entities to qualify as state actors scarcely, an example being in a 1946 case named *Marsh v. Alabama*.[100] In *Marsh*, the Court held that a privately owned town made itself equivalent to public property, and could not stop a Jehovah's witness from passing out religious flyers on the streets of the town because the operation of the town constituted state action.[101] Therefore, defeating viewpoint discrimination perpetrated by platforms through the First Amendment's state-action doctrine is not likely to succeed without modification to the current governing law, 47 U.S.C. Section 230.[102]

---

94. *Kathleen* M. Hidy, *Social Media Use And Viewpoint Discrimination: A First Amendment Judicial Tightrope Walk With Rights And Risks Hanging In The Balance*, 102 MARQ. L. Rev. 1045, 1055 (2019).

95. *Matal*, 137 S. Ct. at 1764 (quoting Justice Holmes in United States v. Schwimmer, 279 U.S. 644, 655 (1929) (Holmes, J., dissenting)).

96. *Manhattan Cmty. Access Corp.*, 139 S. Ct. at 1926.

97. *Id.* (operation of a public access channel held insufficient to establish the role of an entity as a government actor).

98. *Id.*

99. *Id.* at 1928.

100. *See generally* Marsh v. Alabama, 326 U.S. 501, 506–09 (1946) (Court held that a privately owned town made itself equivalent to public property and could not stop a Jehovah's witness from passing out religious flyers because operation of the town constituted state action). However, the Supreme Court has limited this ruling. *See also Manhattan Cmty. Access Corp.*, 139 S. Ct. at 1921.

101. *Marsh*, 326 U.S. at 503, 509–10.

102. Hudson, *supra* note 64.

III. APPLICABLE LAWS REGARDING SOCIAL MEDIA CONTENT REMOVAL IN THE US AND THE EU

A. US Section 230's Social Media Company Safe Harbor

Online intermediary liability regarding user posted content in the United States is governed by the Communications Decency Act of 1996 ("CDA"), codified as 47 U.S.C. § 230 ("Section 230").[103] The relevant statutory language of Section 230 is the "Good Samaritan" provision, which is the authority surrounding civil liability of companies publishing and regulating third-party content on the internet.[104] There are two prongs to this well-known statute. The first prong, (c)(1), stipulates "no . . . interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[105] Essentially, (c)(1) prevents civil liability arising from posts social media companies allow on their websites by preventing the companies from being construed as the party creating (or "publishing") the post, despite the companies allowing circulation of the post.

The second prong, (c)(2), states no "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

---

103. *See* 47 U.S.C. § 230(b) for further policy considerations underlying implementing section 230 ("It is the policy of the United States—
> (1) to promote the continued development of the Internet and other interactive computer services and other interactive media;
> (2) to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation;
> (3) to encourage the development of technologies which maximize user control over what information is received by individuals, families, and schools who use the Internet and other interactive computer services;
> (4) to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to objectionable or inappropriate online material; and
> (5) to ensure vigorous enforcement of Federal criminal laws to deter and punish trafficking in obscenity, stalking, and harassment by means of computer.").

104. *Id.* § 230(c).

105. *Id.*

(B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described.[106]

Thus, (c)(2) sets forth immunity for "interactive computer service[s]" (social media companies) that shields these companies from civil liability regarding content regulation of "information content provider[s]" (users).[107] Simplified, (c)(2) enables platforms—or any company that meets the definition of "interactive computer service" under Section 230—to freely takedown or leave up any post, without civil liability, subject to two large exceptions.[108]

The two broadest categories of Section 230's exceptions are intellectual property claims and criminal prosecutions, which do not receive the immunity provided in section (c).[109] This is significant, for platforms that do not remove material that either violates intellectual property laws or material deemed criminal are exposed to liability for failure to remove those posts. Prior to Section 230's enactment in 1996, issues regarding liability centered around whether to treat interactive computer services as publishers or speakers.[110]

Then in 1997, one year after Section 230's enactment, the United States Supreme Court ruled on *Reno v. ACLU*.[111] In *Reno*, the Court reviewed sections of the CDA attempting to criminalize the communication via the internet of "indecent" and "patently offensive" content to any person under 18 years old.[112] The "patently offensive" and "indecent" material restrictions were struck down, failing the strict scrutiny standard, placing unacceptably

106. *Id.* (emphasis added to demonstrate broad language given to platforms in their content regulation capabilities).

107. *Id.* § 230(f)(3); Chi. Lawyers' Comm. for Civ. Rights Under the Law, Inc. v. Craigslist, Inc., 461 F.Supp.2d 681, 693 (N.D. Ill. 2006) ("Section 230(c)(1) . . . bars those causes of action that would require treating interactive computer services as publisher of third-party content"), aff'd, 519 F.3d 666, 672 (7th Cir. 2008).

108. 47 U.S.C. § 230(c); 47 U.S.C. § 230(f)(2) ("Interactive computer service" means "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.").

109. 47 U.S.C. § 230(e).

110. *See* Cubby, Inc. v. CompuServe, 776 F. Supp. 135, 135 (S.D.N.Y. 1991) (holding that Compuserve was a distributor and not liable for defamatory statements made by a third party because Compuserve had no reason to know of the defamatory statement at issue); *see also* Stratton Oakmont, Inc. v. Prodigy Servs. Co., 1995 WL 323710 at 5 (N.Y. Sup. Ct. 1995) (holding that Prodigy was subject to liability of defamatory statements made by a third party because Prodigy exercised some control over moderation of posts). *Stratton Oakmont, Inc.* disincentivized online computer services regulation of content because any regulation could open regulators to third-party liability as a publisher. *Stratton Oakmont, Inc.*, 1995 WL 323710 at 5; *see also* 47 U.S.C. § 230(b) (2018) (for policy purposes of Section 230's enactment).

111. Reno v. ACLU, 521 U.S. 844, 844 (1997).

112. *Id.* at 849.

heavy burdens on protected speech, and were insufficient for the "narrow tailoring" needed to justify the restrictions on freedom of speech.[113] The underlying rationale of the Court focused on the overbroad content-based restrictions of free speech.[114] Lastly, the Court declared its stance on internet censorship, stating, "encouraging freedom of expression in a democratic society outweighs any theoretical but unproven benefit of censorship."[115]

Reno demonstrates the Supreme Court's view on online material, granting online material the same degree of protection as public speech to facilitate the exchange of ideas.[116] Also, the Court is cautious of censorship, believing the ability to censor anything to individuals under eighteen years of age is untenable, given that "most Internet forums are open to all comers" and that even the strictest reading of the "indecent materials" requirement would "confer broad powers of censorship, in the form of a 'heckler's veto,' upon any opponent of indecent speech."[117]

Since 1997, Section 230 has been used as a sturdy shield for platforms to numerous claims. A model example is the D.C. Circuit case *Klayman v. Zuckerberg*, where the court held Section 230 prevented tort liability based on Facebook's decisions to allow or to remove content from its website.[118] Additionally, the court held a social networking website does not create or develop content by merely providing "a neutral means by which third parties can post information of their own independent choosing online."[119]

The current practical applications of Section 230 are once again exemplified in the Ninth Circuit case of *Barnes v. Yahoo!, Inc*.[120] The *Barnes* case centers around whether the CDA protects an "internet service provider" from suit where it attempted to remove harmful material—nude photographs of the plaintiff and ex-boyfriend defendant on a shared Yahoo public profile—from its website, but failed to do so.[121] The court held the CDA does protect against such suit, allowing Yahoo to invoke Section 230's immunity. The court reasoned that:

113. *Id.* at 882–83.
114. *Id.* at 880–82.
115. *Id.* at 885.
116. *Id.* at 884.
117. *Id.* at 880.
118. Klayman v. Zuckerberg, 753 F.3d 1354, 1354 (D.C. Cir. 2014).
119. *Id.* at 1358.
120. Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1101–02 (9th Cir. 2009).
121. *Id.* at 1098.

> What matters is not the name of the cause of action—defamation versus negligence versus intentional infliction of emotional distress—what matters is whether the cause of action inherently requires the court to treat the defendant as the "publisher or speaker" of content provided by another. To put it another way, courts must ask whether the duty that the plaintiff alleges the defendant violated derives from the defendant's status or conduct as a "publisher or speaker." If it does, section 230(c)(1) precludes liability.[122]

The court also stated "(c)(2) also protects Internet service providers from liability not for publishing or speaking, but rather for actions taken to restrict access to obscene or otherwise objectionable content."[123] The phrase "otherwise objectionable content" serves as the statutory catch-all enabling censorship power, while the private company status—combined with terms and conditions contracts—adds additional strength to the social media platforms' right to censor content.[124]

Section 230 has been successfully invoked in cases involving negligence, deceptive trade practices, unfair competition, false advertising, common-law privacy torts, tortious interference with contract or business relations, intentional infliction of emotional distress, and dozens of other legal doctrines.[125] Yet, the future of Section 230 is uncertain.[126] Emphasizing the antiquated nature of the legislation, Section 230 was enacted in 1996, with the only update to the statute being added in 2018 regarding the safe harbor not covering websites "that promote and facilitate prostitution."[127] For reference, Facebook, YouTube, and Twitter opened their platforms to the public in 2004, 2005, and 2006.[128] Since 1996, social media's influence over society has exponentially grown—as well as the amount of content the platforms restrict—with no meaningful update to the governing statute.[129]

---

122. *Id.* at 1101-02 (quoting 47 U.S.C. § 230).

123. *Id.* at 1105.

124. *See* 47 U.S.C. § 230(c)(2).

125. *See* Eric Goldman, *How Section 230 Enhances the First Amendment*, AM. CONST. SOC'Y (July 28, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3662475/ [perma.cc/CV98-H4R2].

126. *See* Exec. Order No. 13925, 85 Fed. Reg. 34,079 (May 28, 2020). The order is titled "Preventing Online Censorship," and the order specifically states that Facebook, YouTube, and Twitter are "engaging in selective censorship that is harming our national discourse" and instructs the Secretary of Commerce, in consultation with the Attorney General, and acting through the National Telecommunications and Information Administration, to file a petition for rulemaking with the Federal Communications Commission requesting that the FCC expeditiously propose regulations to changed Section 230(c). *Id.*

127. *See generally* 47 U.S.C. § 230 notes ("[S]ection 230 . . . was never intended to provide legal protection to websites that unlawfully promote and facilitate prostitution and websites that facilitate traffickers in advertising the sale of unlawful sex acts with sex trafficking victims.").

128. *The History of Social Media: A Timeline*, PHRASEE (Aug. 6, 2018), https://phrasee.co/the-history-of-social-media-a-timeline/ [perma.cc/34KS-4KX7].

129. *See generally* Kemp, *supra* note 18.

Hence, social media platforms' overbroad discretion over choice of removal of user posts must be scrutinized.

B. The EU's Lesser Freedom of Expression, the E-Commerce Directive and Content Regulation

The EU has less expression available for platform users than the US. In the EU, the governing legislation on freedom of expression is the EU Charter of Fundamental Rights, which establishes every individual has the right to freedom of expression.[130] However, this right to expression is limited by "formalities, conditions, restrictions or penalties as are prescribed by law."[131] These "formalities" and "conditions" refer to the individual governance of the Member States of the EU.[132] Additionally, there is a lower standard level of expression in the EU than the US. For example, the EU outlaws hate speech, while the US jurisprudence protects it.[133]

In application, the lower standard of expression impacts platforms under the EU's domain. For example, in 2017, Germany, a Member State of the EU and a country with some of the strictest laws on what is acceptable speech, passed the Network Enforcement Act ("NetzDG").[134] The NetzDG requires social media platforms like Facebook to quickly take down illegal material (i.e., comments inciting hatred against national, religious, ethnic or racial groups) or face large fines.[135] Thus, the EU provides less protection for expression of social media users than the United States.[136]

The EU has its own laws concerning social media content regulation and liability. The EU's governing law concerning online intermediary liability

130. Charter of Fundamental Rights of the European Union, art. 21, Oct. 26, 2012, 2012 O.J. (C 326/02) [hereinafter Charter of Fundamental Rights of the European Union].

131. European Convention for the Protection of Human Rights and Fundamental Freedoms, art. 10, Nov. 4, 1950, 213 U.N.T.S. 221.

132. *Id.*

133. *Matal*, 137 S. Ct. at 1764; Online Platforms' Moderation of Illegal Content Online Law, Practices and Options for Reform, EUR. PARL. DOC. (PE 652.718) (2020).

134. Netzdurchsetzunggesetz, [NetzDG][Network Enforcement Act], Sept. 1, 2017, BGBL I at 3352 (Ger.), https://germanlawarchive.iuscomp.org/?p=1245 [perma.cc/3L4Q-XXC2].

135. Janosch Delcker, *Germany's Balancing Act: Fighting Online Hate While Protecting Free* Speech, POLITICO (Oct. 1, 2020), https://www.politico.eu/article/germany-hate-speech-internet-netzdg-controversial-legislation/ [https://perma.cc/9GRC-E8KU].

136. *See* Kitsuron Sangsuvan, *Balancing Freedom of Speech on the Internet Under International Law*, 39 N.C.J. INT'L L. & COM. REG. 701, 716 (2014).

is the E-commerce Directive, enacted in 2000.[137] The E-commerce Directive grants safe harbor for three types of online intermediaries who host or transmit content provided by third parties.[138] The three types of online intermediaries covered under the safe harbor are (1) Mere Conduit Service Providers, (2) Caching Providers, and (3) Hosting Providers.[139] Social media companies like Facebook, YouTube, and Twitter are considered hosting providers under the E-commerce Directive.[140]

Hosting providers must meet two requirements to qualify for safe harbor.[141] First, hosting providers cannot be held liable without "actual knowledge" of illegal activity or information when unaware of facts or circumstances from which the illegal activity or information is apparent.[142] What constitutes "illegal activity," however, has not been precisely defined in the E-commerce Directive.[143] Second, hosting providers must also "act expeditiously to remove" (take down) or to "disable access" (block) illegal activity or information of which they have obtained actual knowledge.[144] The main tools used by EU online platforms to identify illegal content online are "notice-and-takedown" flags used by users, keywords or filters, and AI machine learning models.[145] All large online platforms allow users to appeal against their decisions on the moderation of illegal content online through a "counter-notice" procedure.[146]Additionally, Article 15 of the E-commerce Directive enables Member States of the EU to impose additional obligations on internet service providers to report "alleged illegal activities," such as the previously discussed NetzDG legislation.[147]

Although "illegal activity" has not been precisely defined in the E-Commerce Directive, the EU consistently finds four types of content illegal: (1) child sexual abuse material; (2) racist and xenophobic hate speech; (3) terrorist content; and (4) content infringing on intellectual property rights.[148] Online intermediaries can be subject to injunctive relief

---

137. Council Directive 2000/31, *supra* note 14.
138. Council Directive 2000/31, *supra* note 14.
139. Council Directive 2000/31, *supra* note 14, art. 12–14.
140. Council Directive 2000/31, *supra* note 14, art. 14.
141. Council Directive 2000/31, *supra* note 14, art. 14.
142. Council Directive 2000/31, *supra* note 14, art. 14.
143. Council Directive 2000/31, *supra* note 14, art. 14.
144. Council Directive 2000/31, *supra* note 14, art. 14.
145. Online Platforms' Moderation of Illegal Content Online: Law, Practices and Options for Reform, EUR. PARL. DOC. PE 652.718 (2020).
146. *Id.* The counter-notice is accessible through a link on the blocked-content and allows for users to manually type their requests for reinstatement of the post and the contents of the post. *Id.*
147. Council Directive 2000/31, *supra* note 14, art. 15.
148. Online Platforms' Moderation of Illegal Content Online Law, Practices and Options for Reform, EUR. PARL. DOC. PE 652.718, 16–17 (2020).

when they are found to be in breach of any piece of specific legislation regarding the previously listed areas, such as copyright law.[149]

An illustrative case of the application of the E-Commerce Directive is Glawischnig-Piesczek v. Facebook Ireland Ltd.[150] Ms. Glawischnig-Piesczek was a member of the Nationalrat (National Council, Austria), chair of the parliamentary party "die Grünen" (The Greens) and federal spokesperson for that party.[151] On April 3, 2016, a Facebook user shared on their own personal page an article from an Austrian online news magazine titled "Greens: Minimum income for refugees should stay," which included a photograph of Glawischnig-Piesczek.[152] The user also published, in connection with that article, comments which the Supreme Court of Austria found to be harmful to the reputation of Glawischnig-Piesczek by insulting and defaming her.[153]

The comments called Glawischnig-Piesczek "miese Volksverräterin" (lousy traitor), "korrupten Trampel" (corrupt bumpkin), and her party a "Faschistenpartei" (fascist party).[154] Glawischnig-Piesczek requested that Facebook Ireland delete the comments, but Facebook Ireland refused.[155] This led her to bring an action before the Commercial Court of Vienna in Austria which ordered Facebook Ireland to cease from publishing and disseminating photographs showing Glawischnig-Piesczek if the accompanying text contained the assertions—verbatim or using words with equivalent meaning—of the comments at issue.[156] Facebook Ireland then appealed to the Supreme Court of Austria.[157]

The Supreme Court of Austria referred the question to the Court of Justice of the European Union ("CJEU"), asking whether Article 15(1) of the Directive required hosting providers to remove not only illegal information within the meaning of Article 14(1)(a) of the Directive, but

149. Reform of the EU Liability Regime For Online Intermediaries, EUR. PARL. DOC. PE 649.404, 3 (2020).
150. Case C-18/18, Glawischnig-Piesczek v. Facebook Ireland Ltd., ECLI:EU:C:2019:821.
151. *Id.* ¶ 10.
152. *Id.* ¶ 12.
153. *Id.*
154. *Glawischnig-Piesczek v. Facebook Ireland Limited*, GLOB. FREEDOM OF EXPRESSION COLUM. UNIV., https://globalfreedomofexpression.columbia.edu/cases/glawischnig-piesczek-v-facebook-ireland-limited/ [https://perma.cc/3BXN-J8LG].
155. *Id.*
156. *Id.*
157. *Id.*

also other identically worded, or equivalent meaning information.[158] The CJEU held the Directive did not preclude a Member state from ordering a hosting provider to remove information found to be unlawful, as well as information that is identical or equivalent to such unlawful information posted by any user.[159] The CJEU's reasoning included referencing Article 14(1) of the Directive, which exempts information service providers from liability if they have (1) no knowledge of any illegal activity or information, and once obtaining "actual knowledge" of illegal activity, (2) act "expeditiously to remove or disable access to the information" immediately.[160] Facebook Ireland was notified of the illegal content, but failed to "expeditiously" remove or disable access to the defamatory content, precluding Facebook Ireland from claiming immunity under the E-Commerce Directive.[161]

The EU allows Member States and their respective courts to establish procedures to remove or disable illegal content and require platforms "to terminate or prevent infringement."[162] Glawischnig-Piesczek extended this allowance to "information with an equivalent meaning," as long as the host is not required to "carry out an independent assessment of that content," and only employs automated search tools for the "elements specified in the injunction."[163] As a result, EU law currently allows, but does not mandate, the removal of content that is identical, or equivalent to the initial illegal content already brought to the attention of the social platform.[164]

In evaluating the EU's approach to platform liability for content removal for possibilities to diminish viewpoint discrimination, the E-Commerce Directive does not directly address viewpoint discrimination, nor does it allow a cause of action for social media users against social media companies that remove lawful content. However, the E-Commerce Directive does incentivize over-removal of illegal posts. Encouraging over-removal increases censorship over every viewpoint and prevents viewpoints that are different than the viewpoints of the company, its advertisers, or the majority of its users.[165]

---

158. *Id.*; Council Directive 2000/31, *supra* note 14, art. 13 ("(a) the provider does not have actual knowledge of illegal activity or information and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or information is apparent").

159. Case C-18/18, Glawischnig-Piesczek v. Facebook Ireland Ltd., ECLI:EU:C: 2019:821, ¶ 53.

160. *Id.* ¶ 4.

161. *Id.* ¶ 27.

162. *Id.* ¶ 4.

163. *Id.* ¶¶ 39, 41, 45–46.

164. Luc von Danwitz, *The Contribution Of Eu Law To The Regulation Of Online Speech*, 27 MICH. TECH. L. REV. 167, 211 (2020).

165. Rikke Frank Jørgensen & Lumi Zuleta, *Private Governance of Freedom of Expression on Social Media Platforms: EU Content Regulation Through The Lens Of*

Moreover, additional requirements are placed on social media companies by the Audio-Visual Media Services Directive ("AVMSD") to "take appropriate measures" to remove illegal content.[166] Such "appropriate measures" include platforms having mechanisms for users to flag non-compliant content, effective procedures for user complaints, providing effective media literacy tools, and raising users' awareness of those tools.[167]

The AVMSD's 2018 amendment also notes "social media services" should be treated as "audiovisual media services" due to both industries competing for the "same audiences and revenues," and the "considerable impact" in the possibility of users shaping and influencing opinions of other users.[168] The 2018 revision further detailed duties of social media companies, holding the companies responsible for the duty to protect the general public from the four types of online content illegal under EU law, elaborating on the definition of hate speech being illegal if based on the EU Charter of Fundamental Rights—sex, race, color, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age, or sexual orientation.[169]

Lastly, the revision assigns the responsibility to protect minors from content which may impair their physical, mental, or moral development.[170] These additional duties of care are covered by social media companies under the AVMSD to the extent that they meet the definition of a video-sharing platform service.[171]

Human Rights Standards, 41 NORDICOM REV. 1, 59 (2020), https://www.researchgate.net/publication/339682228_Private_Governance_of_Freedom_of_Expression_on_Social_Media_Platforms [https://perma.cc/PLA8-NV5W]

166. Council Directive 2010/13, art. 4(2)(b)-(3)(a), 2010 O.J. (L 95) 1, 14-15 (EC); Council Directive 2018/1808, ¶ 4, 2018 O.J. (L 303) 69 (EC) [hereinafter Council Directive 2018/1808].

167. Council Directive 2018/1808, *supra* note 166, ¶ 28.

168. Council Directive 2018/1808, *supra* note 166, ¶ 4.

169. Council Directive 2018/1808, *supra* note 166, ¶¶ 13, 47; Charter of Fundamental Rights of the European Union, *supra* note 130, art. 21.

170. Online Platforms' Moderation of Illegal Content Online: Law, Practices and Options for Reform, EUR. PARL. DOC. PE 652.718, 24 (2020).

171. Council Directive 2018/1808, *supra* note 166, ¶ 5 (Defining a video-sharing platform as "a service offering programmes, user-generated videos, or both, to the general public, for which the video-sharing platform provider does not have editorial responsibility, in order to inform, entertain or educate, using electronic communications networks, and the organisation of which is determined by the video-sharing platform provider, including by use of automatic means or algorithms, in particular by displaying, tagging and sequencing.").

C. Major Key Differences In the US and the EU for Content Removal

Before proceeding with the Section 230 analysis, it is important to note the major relevant differences in US and EU law pertaining to user expression and social media content removal practices. First, the EU offers less protection over freedom of expression than the US.[172] Second, the EU's social media censorship remedies are based on criminalization, not civil remedies.[173] Nevertheless, the EU serves as Europe's baseline for online content moderation, which helps highlight and assess the US's strengths and weaknesses in its own current statutory framework. Additionally, the two approaches share similarities.

A hypothetical scenario will best serve to illustrate the different approaches to online intermediary liability concerning regulation of user content used by both the US and the EU. Suppose a social media user posts a string of content defaming a polarizing political figure on Facebook, YouTube, and Twitter. The hypothetical post, in part, states "This French bureaucrat is lining his pockets with baguettes paid for by his citizens' taxes." The social media companies remove this post via their own algorithms for violating respective guidelines of community standards. Some users, upon viewing the post, believe the post constituted defamation, and other users felt it deserved a Pulitzer Prize for displaying a brilliant political message.

Under US law, Section 230 would prevent any civil liability for the social media companies when removing the posts.[174] The liability lies with the publisher (poster) of the content and not the intermediary platform.[175] The user posting on the platform only allows the user to be held responsible for the post, making the post able to be used against that respective user in civil and criminal cases while leaving the social media platform out of the scope of liability.[176] There is no successful claim of liability against the social media companies and their respective censorship. Even if Facebook, YouTube, and Twitter did not censor the material, the platforms would not be liable for allowing the content to remain available on their sites, as (c)(1) blocks any interpretation of the companies being publishers or speakers of the information.[177]

Under EU law, social media companies (hosting providers) are not liable (1) without "actual knowledge" of illegal activity or information when

---

172. Kalev Leetaru, *The EU Will Be The End Of Free Speech Online*, FORBES (June 6, 2019), https://www.forbes.com/sites/kalevleetaru/2019/06/06/the-eu-will-be-the-end-of-free-speech-online/?sh=132b5ce056a8 [https://perma.cc/8FSS-CNDJ].

173. *See* Council Directive 2000/31, *supra* note 14, art. 4.

174. 47 U.S.C. § 230(c)(2).

175. *Barnes*, 570 F.3d at 1101-02 (quoting 47 U.S.C. § 230).

176. 47 U.S.C. § 230(c).

177. *Barnes*, 570 F.3d at 1101-02 (quoting 47 U.S.C. § 230).

unaware of facts or circumstances from which the illegal activity or information is apparent and (2) must also "act expeditiously to remove" ("take down") or to "disable access" ("block") illegal activity or information of which they have obtained actual knowledge.[178] The content, explicitly referencing French culture negatively ("lining his pockets with baguettes"), due to its borderline xenophobic nature, would likely be deemed illegal content.[179] Thus, Article 14 of the Directive requires the companies to take appropriate measures to take it down once notified of its illegality, or be fined as a result of their inaction.[180]

In this case, Section 230 protects user expression to a greater degree than the E-Commerce Directive. The US law grants immunity of civil liability to the social media intermediary regardless of censorship or not, while the EU demands the statement, which is illegal, be removed by the social media intermediary. The contrast highlights the discretionary power offered to the social media companies in the US, which enables viewpoint discrimination upon its users.

Increasing civil liability for intermediary social media platforms in the US would provide the benefit of incentivizing platforms to moderate more accurately, and to develop more efficient content-filtering approaches. Adversely, increased liability would also provide the incentive of moderating strictly and strain companies by requiring more monetary resources towards review of content restrictions. However, when balancing these two different sides of the scale, the weight of allowing individuals to express themselves despite disapproval of the majority or established societal norms must win out, even if "the idea itself [is] offensive or disagreeable."[181] A society that silences disagreement will not be able to grow if it shelters itself from adversity.

IV. ANALYSIS OF SECTION 230'S DEFICIENCIES REGARDING SOCIAL MEDIA CONTENT REMOVAL PRACTICES

This section will start with a discussion of Section 230's allowance of platforms' overbroad discretionary authority in the removal of their content. Next, the section will move on to issues surrounding lack of transparency with the content removal process. Then, the analysis will turn to viewpoint

178. Council Directive 2000/31, *supra* note 14, art. 14.
179. Charter of Fundamental Rights of the European Union, *supra* note 130, art. 21.
180. Council Directive 2000/31, *supra* note 14, art. 14.
181. *Texas v. Johnson*, 491 U.S. at 414.

discrimination, examining how platforms are able to viewpoint discriminate and the political and social ramifications of such behavior.

A. Social Media Companies' Unique Business Model Creates Overbroad Discretionary Authority in Content Removal

Social media companies are unique in that their profit originates by providing a platform for the online speech of others, which entangles their commercial interests and their governing roles. Under Section 230, overbroad discretionary power of online intermediaries is exercised by curating online content; ranking or giving priority to some content, while diverting attention away from other types of information, leading to content-bubbles.[182] These content-bubbles help foster viewpoint discrimination by reaffirming individual users' beliefs without any chance for minority opinions to be introduced.[183] But, platforms need some degree of discretionary power to effectively run their sites. Therefore, a balance must be set between discretionary autonomy of the platforms to run their own business and the freedom of speech issues that arise from unrestricted discretion.

Discretionary power is the root of a laissez-faire economy, allowing companies to adjust to market supply and demand based on their own decision-making.[184] However, online intermediaries are unique in that their profit originates by providing a platform for the online speech of others—their own users—which complicates their commercial interests and their governing roles.[185] Permitting social media companies exclusive discretion over consumer complaints while simultaneously allowing the companies to control those same consumers is ill-advised. Letting those companies adjudicate the content they distribute may trap intermediaries in a conflict of interest.[186] The conflict of interest is being placed in a position to grab one's cake (user posts), and eat it (profits from user posts), and then blame someone else that the cake is gone (blame user if the posts are removed because the posts did not meet guidelines). The platforms review their own mistakes and have no official consequences for simply ignoring mistakes made and allowing posts to stay blocked.[187]

---

182. Niva Elkin-Koren & Maayan Perel, *Guarding the Guardians: Content Moderation by Online Intermediaries and the Rule of Law*, OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIAB. 669, 671 (2020).

183. Gould, *supra* note 41.

184. Mark H. Moore, *Public Values In An Era Of Privatization: Introduction*, 116 HARV. L. REV. 1212, 1223 (2003).

185. Elkin-Koren & Perel, *supra* note 182, at 671.

186. Elkin-Koren & Perel, *supra* note 182, at 671.

187. *See* Will Oremus, Facebook keeps researching its own harms — and burying the findings, THE WASH. POST (Sept. 16, 2021), https://www.washingtonpost.com/technology/2021/09/16/facebook-files-internal-research-harms/ [https://perma.cc/7QTP-BFZD].

The vital element of discretion is the ability to abuse it. In terms of Section 230, a website does not open itself to civil claims when it merely provides "a neutral means by which third parties can post information of their own independent choosing online."[188] From the platforms' perspective, the incentive to censor user posts that offend management and sources of revenues, like other users, far outweighs adhering to the optional First Amendment standard of freedom of expression.

B. Platform's Lack Transparency in Regulating Content

Platforms lack transparency in their exact content-removal processes and procedures. The platforms, as private businesses, must protect their proprietary property—content-filtering systems, algorithms, and other intellectual property—inherently preventing a fully transparent business model. Transparency, for its purpose here, is defined as "openness, communication, and accountability."[189] Transparency builds user trust in the company and helps the exchange of honest dealings. Facebook, YouTube, and Twitter do voluntarily post transparency reports of varying degrees.[190] However, voluntary transparency alone remains a suboptimal solution for mitigating the problem of oppressive content regulation for two main reasons.[191] First, these reports are not held to any uniform standard, and secondly, the reports come in varying degrees of disclosure.[192] Yet, the overall main trend for social media transparency is trending upward.

For example, Facebook took the next step in its attempt to provide transparency without completely divulging trade secrets. Specifically, Facebook announced construction of the Facebook Oversight Board ("FOB") in November of 2018.[193] Facebook's incentives for undergoing such a large endeavor include building goodwill and trust with its users, maintaining

188. *Klayman v. Zuckerberg*, 753 F.3d at 1354.

189. Brian Carter, *The Illusion Of Transparency In Social Media*, SEARCH ENGINE PEOPLE (Feb. 12, 2009) https://www.searchenginepeople.com/blog/the-illusion-of-transparency-in-social-media.html [https://perma.cc/6JZK-HPMA].

190. *Transparency Reports*, META TRANSPARENCY CENTER, https://transparency.fb.com/data/ [https://perma.cc/9EWS-3J6J]; *Google Transparency Report*, GOOGLE, https://transparencyreport.google.com/youtube-policy/removals?hl=en [https://perma.cc/M2TB-RNB3]; *Twitter Transparency*, TWITTER, https://transparency.twitter.com/ [https://perma.cc/7DMR-4BME].

191. Michal Lavi, *Evil Nudges*, 21 VAND. J. ENT. & TECH. L. 1, 11 (2018).

192. *See* Elkin-Koren & Perel, *supra* note 182, at 671.

193. Klonick, *supra* note 56, at 2418.

viewership to help advertising revenue, or more cynically, establishing a convenient scapegoat for diverting outside social and legal pressure.[194] It is a mutually beneficial move for the company and its users in developing trust. To succeed, jurisdictional, intellectual, and financial independence from Facebook must be established.[195] Facebook's legal division developed a trust agreement and created a beneficiary for the trust—the Oversight Board—and an LLC to handle the operation, both being independent entities.[196]

The FOB's first 5 rulings each tell Facebook what to do with a single piece of content at issue; all but one are unanimous.[197] Four rulings overturned Facebook's original decisions to remove posts, and only one ruling agreed with Facebook's removal of the post at issue.[198] Where does this board get its authority? In every case, the board first assesses Facebook's decisions against Facebook's own standards and then against international human rights law.[199] But in all its rulings, the FOB came to the same conclusion under both its own as well as international norms, and in no case did the FOB confront the question of what happens if Facebook's rules conflict with international human rights law.[200] Overall, increasing transparency through oversight boards may not cure viewpoint discrimination, but increased transparency will decrease viewpoint discrimination's likelihood because of independent review by external sources.

C. Viewpoint Discrimination

Section 230's "Good Samaritan" safe harbor provision enables viewpoint discrimination.[201] Viewpoint discrimination refers to the silencing or muffling the expression of disfavored viewpoints.[202] Social media platforms have the ability to abuse their regulatory discretion over content, and this must be considered moving forward. The balance of selectively removing material without overly inhibiting users' expression is delicate and Section 230 does

194. Klonick, *supra* note 56, at 2426–27.
195. Klonick, *supra* note 56, at 2426–27.
196. Brent Harris, *Establishing Structure and Governance for an Independent Oversight Board*, FACEBOOK NEWSROOM (Sept. 17, 2019), https://about.fb.com/news/2019/09/oversight-board-structure [https://perma.cc/N6ER-LCV9].
197. Evelyn Douek, *The Facebook Oversight Board's First Decisions: Ambitious, and Perhaps Impractical*, LAWFARE (Jan. 28, 2021), https://www.lawfareblog.com/facebook-oversight-boards-first-decisions-ambitious-and-perhaps-impractical [https://perma.cc/3WK5-N3D9].
198. *Id.*
199. *Id.*
200. *Id.*
201. 47 U.S.C. § 230(c) (2018).
202. *Matal*, 137 S. Ct. at 1749, 1750.

not give discretionary guidance on how to balance removal of material and expression of users.

Currently, online intermediaries are effectively performing three roles at the same time: they act (1) like a legislature, in defining what constitutes legitimate content on their platform, (2) like a judge who determine the legitimacy of content in particular instances, and (3) like an administrative agency who acts on these adjudications to block illegitimate content.[203] A user may post something, the system (most likely an algorithm) will remove it, and then the user is expected to lodge a complaint to the same administration that deemed removal of the post necessary with no independent external review.[204] The proprietary content-filtering systems used to regulate the content effectively blend law enforcement and discretionary powers, signaling a transformation in the traditional system of governance by law.[205] Such practices reflect an institutional shift in lawmaking power, from state to private companies, and a fundamental transformation of the nature of law enforcement.[206]

The American public is wary of this blend of law enforcement and discretionary power. A survey conducted in June of 2020 found that roughly three-quarters of US adults say it is very (37%) or somewhat (36%) likely that social media sites intentionally censor political viewpoints that they find objectionable.[207] Roughly three-quarters of Americans (73%) think it is very or somewhat likely that social media sites intentionally censor political viewpoints the platforms find objectionable.[208] Splitting these statistics into party lines, 90% of Republicans surveyed felt that social media sites intentionally censored political viewpoints that the sites found objectionable

203. Elkin-Koren & Perel, *supra* note 182, at 669–70.

204. *See Appeal An Account Suspension Or Locked Account*, TWITTER, https://help.twitter.com/forms/general?subtopic=suspended [https://perma.cc/E468-PNG8] (for example, Twitter allows appeals of post removals, which are reviewed by Twitter staff).

205. Elkin-Koren & Perel, *supra* note 182, at 672.

206. Elkin-Koren & Perel, *supra* note 182, at 672.

207. Emily A. Vogels et al., *Most Americans Think Social Media Sites Censor Political Viewpoints*, PEW RSCH. CTR. (Aug. 19, 2020), https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints/ [https://perma.cc/85WV-52U6]. For the analysis, Pew Research Center surveyed 4,708 U.S. adults from June 16 to 22, 2020. Everyone who took part is a member of the Center's American Trends Panel (ATP), an online survey panel that is recruited through national, random sampling of residential addresses. *Id.* Nearly all U.S. adults have a chance of selection. *Id.* The survey is weighted to be representative of the U.S. adult population by gender, race, ethnicity, partisan affiliation, education and other categories. *Id.*

208. *Id.*

—with 60% saying this is very likely the case.[209] By comparison, only 59% of Democrats believed censoring based on political views was very (19%) or somewhat (40%) likely.[210]

A counter argument to this Article's stance is the argument that the market will adapt to the needs and demands of the supply; if people are unhappy with a platform, they will use another competitor in the same or similar service. If there are no competitors, then the market will meet the rising demand, eventually producing a competitor. Yet, this argument does not consider multiple factors. The first, is the similarity in political views of the social media companies' CEOs.[211] Second, is the barriers to entry of the market. If someone is censored from a social platform, one needs to have access to the necessary capital, and underlying skills or connections to someone with sufficient skills in coding to enter the market as an unregulated social media platform.

There is a way to attack Section 230 besides the First Amendment—through the legislature. Politically, there are two sides to the Section 230 debate: conservatives bemoan of big tech companies asserting liberal bias against conservative speech and seek to prevent the platforms' discretion in removing content, while the liberal side complains of too little action over hate speech, fake news, and other problematic content, and desires obligations to remove such content.[212] To simplify the positions, liberals want the large social platforms to ramp up censorship, and conservatives want the platforms to significantly reduce the current level of censorship.

The tremendous legal immunity for the platforms has not gone unnoticed. Numerous political figures are attempting to reform, or even to dispose of, Section 230.[213] In May of 2020, former President Trump signed Executive order 13925, effectively demanding Section 230 be revised.[214] There are now an influx of bills circulating in Congress. The list includes (1) EARN IT Act of 2020, (2) Stopping Big Tech's Censorship Act, (3) Platform Accountability and Consumer Transparency Act, (4) Limiting Section 230 Immunity to Good Samaritans Act, and (5) the Online Freedom and

209. *Id.*
210. *Id.*
211. Ninety-eight percent (98%) of political contributions from internet companies this Presidential cycle (2020) went to Democrat campaigns. *See* Ari Levy, *Here's The Final Tally Of Where Tech Billionaires Donated For The 2020 Election*, CNBC (Nov. 2, 2020, 8:42 PM), https://www.cnbc.com/2020/11/02/tech-billionaire-2020-election-donations-final-tally.html [https://perma.cc/A8L6-XLJY].
212. Michael A. Cheah, *Section 230 and the Twitter Presidency*, 115 NW. U. L. REV. 192, 194 (2020).
213. Exec. Order No. 13925, 85 Fed. Reg. 34079 (May 28, 2020).
214. *Id.*

Viewpoint Diversity Act.[215] The Department of Justice even proposed four prongs to focus reforming discussions on Section 230.[216]

One of the most extreme proposed acts, the Stopping Big Tech's Censorship Act, completely removes immunity under Section 230.[217] The Stopping Big Tech's Censorship Act gives big tech companies the ability to earn immunity through external audits.[218] Big tech companies would have to prove to the Federal Trade Commission ("FTC") by clear and convincing evidence that their algorithms and content-removal practices are politically neutral.[219] The bill only applies to companies with more than 30 million active monthly users in the US, more than 300 million active monthly users worldwide, or who have more than $500 million in global annual revenue.[220]

---

215. EARN IT Act of 2020, S. 3398, 116th Cong. (2020); Stopping Big Tech's Censorship Act, S. 4062, 116th Cong. (2020); Platform Accountability and Consumer Transparency Act, S. 4066, 116th Cong. (2020); Limiting Section 230 Immunity to Good Samaritans Act, S. 3983, 116th Cong. (2020); and Online Freedom and Viewpoint Diversity Act, S. 4534, 116th Cong. (2020). *See also* Protecting Constitutional Rights from Online Platform Censorship Act, H.R. 83, 117th Cong. (2021); CASE-IT Act, H.R. 285, 117th Cong. (2021); and See Something, Say Something Online Act of 2021, S. 27, 117th Cong. (2021).

216. U.S. Dep't of Just., 20-556, *Department of Justice's Review of Section 230 of the Communications Decency Act of 1996* (2020) ("(1) Incentivizing online platforms to address illicit content; (2) clarifying federal government enforcement capabilities to address unlawful content; (3) promoting competition; and (4) promoting open discourse and greater transparency.").

217. Stopping Big Tech's Censorship Act, S.4062, 116th Cong. (2020); *see also Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies,* JOSH HAWLEY (June 19, 2019), https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies [https://perma.cc/66V3-RW4Y] [hereinafter JOSH HAWLEY].

218. Stopping Big Tech's Censorship Act, S.4062, 116th Cong. (2020); *see also Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies*, JOSH HAWLEY (June 19, 2019), https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies [https://perma.cc/66V3-RW4Y].

219. Stopping Big Tech's Censorship Act, S.4062, 116th Cong. (2020); *see also Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies,* JOSH HAWLEY (June 19, 2019), https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies [https://perma.cc/66V3-RW4Y].

220. Stopping Big Tech's Censorship Act, S.4062, 116th Cong. (2020); *see also Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies*, JOSH HAWLEY (June 19, 2019), https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies [https://perma.cc/66V3-RW4Y].

Big tech companies would be responsible for the cost of conducting audits, and reapplying for immunity every two years.[221]

Lastly, there is support for simply leaving Section 230 unchanged.[222] The underlying reason for maintaining Section 230 originates from the statute enabling the internet to grow exponentially into the unique asset it is today, and also that Section 230 should not be tinkered with until broken.[223] Together, Section 230 and the First Amendment have contributed to the Internet's emergence as one of the most remarkable speech venues in human history.[224] Wikipedia's crowdsourced encyclopedia, consumer review websites like Yelp, and user-uploaded video sites like YouTube were simply not possible in the offline world.[225] These user generated content services give Internet users an unprecedented ability to express themselves to a global audience and also created many private benefits including new jobs and wealth.[226] Those supporting Section 230, as it stands, focus on Section 230 increasing litigation efficiency via its immunity defense, making litigation more predictable than First Amendment claims, and Section 230's ability to cover claims that do not receive First Amendment protection.[227]

Ultimately, to protect citizens from viewpoint discrimination through abuses of online censorship on otherwise lawful posts, the US must introduce civil intermediary liability for social media providers under section 230(c).[228] This goal of decreasing discretionary power of social media companies over content regulation without encroaching on private governance of social media business practices is possible.

V. PROPOSED SOLUTIONS FOR DETERRING VIEWPOINT DISCRIMINATION UNDER SECTION 230(C)(2)

Section 230(c)(1) has proven effective in protecting online platforms, allowing the platforms to act with impunity and grow into the titans of industry they are today, however, several solutions are feasible to update Section 230(c)(2) to highly discourage abuses of biased social media content

221. Stopping Big Tech's Censorship Act, S.4062, 116th Cong. (2020); *see also Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies,* JOSH HAWLEY (June 19, 2019), https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies [https://perma.cc/66V3-RW4Y].

222. Goldman, *supra* note 125.

223. Goldman, *supra* note 125.

224. Goldman, *supra* note 125.

225. Goldman, *supra* note 125.

226. Goldman, *supra* note 125.

227. Goldman, *supra* note 125.

228. *See generally* 47 U.S.C. § 230.

removal.[229] The current problems with Section 230(c)(2), as discussed above, are overbroad discretionary authority, limited transparency, and viewpoint discrimination.[230] The solutions to these problems can be found in two possibilities. The first possibility is revision of the statutory language of Section 230(c)(2). The second possibility is introduction of mandatory federal censorship guidelines for each large-scale social media platform.

A. Revision of Statutory Language

First, the statutory language of (c)(2) may be restructured into granting social media companies immunity from liability for removing posts only if the posts contained illegal content or otherwise unprotected speech by the First Amendment, treating the platforms equivalently to the government for purposes of speech regulation.[231] The authority for enforcement comes from the federal statute, not the First Amendment. Thus, the social media platforms may be sued civilly for violations of the statute if they remove expression protected by the First Amendment.

For example, California's Leonard Law, passed in 1992, grants students at private universities free speech rights they can assert against their own institution.[232] The statute intended to transplant constitutional free speech rights students have off campus and apply it on campus.[233] However, the Leonard Law protections are weaker than those provided by the First Amendment.[234] Specifically, institutional violations of the Leonard Law

229. *See id.* § 230(c)(2).

230. *See* Elkin-Koren & Perel, *supra* note 182, at 674–77; *see also Matal*, 137 S. Ct. at 1750.

231. *See generally* 47 U.S.C. § 230(c)(2) ("No 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 *only if the posts contain illegal content or otherwise unprotected speech by the First Amendment* . . . 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.") (emphasis added).

232. Cal. Educ. Code § 94367(a) (West 2009) ("No private postsecondary educational institution shall make or enforce a rule subjecting a student to disciplinary sanctions *solely* on the basis of conduct that is speech or other communication that, when engaged in outside the campus or facility of a private postsecondary institution, is protected from governmental restriction by the First Amendment to the United States Constitution or Section 2 of Article 1 of the California Constitution") (emphasis added).

233. *Id.*

234. David Urban, *Free Speech Rights at Private Colleges and Universities*, CAL. PUB. AGENCY & LAB. BLOG (Aug. 14, 2018), https://www.calpublicagencylaboremploymentblog.com/first-amendment/free-speech-rights-at-private-colleges-and-universities/

provide students only declaratory and injunctive remedies with the option of attorney's fees—not damages.[235]

The significance of a private entity's censorship power over speech being statutorily regulated is relevant here.[236] Additionally, a compromise of limiting the available remedy of the revised (c)(2) to the same Leonard Law declaratory and injunctive relief with attorney's fees option would still grant the users their day in court while keeping damages reasonable for social media companies.[237] University of California Berkeley Law School Dean Erwin Chemerinsky, in an online discussion regarding social media and Section 230, discussed the Leonard Law and Section 230.[238] Chemerinsky stated "there needs to be a thorough rethinking of Section 230," adding that he would favor a law that forbids platforms from engaging in viewpoint discrimination.[239] Chemerinsky recognizes that the current Section 230, as it stands now, is susceptible to viewpoint discrimination and even advocated for the same type of preventative viewpoint discrimination measure taken in the Leonard law.[240]

This revised section (c)(2) will solve the discretionary authority issue of the social media companies by allowing review of the reasoning for banning user content by a judicial body. Review by the court system weakens the platforms' autonomous governance over censoring and does not infringe on (c)(1)'s purpose of blocking any sort of liability regarding social media companies as publishing user content. The issue is the potential for viewpoint discrimination based on the discretion of the platforms, not who should be held accountable for the post. Addressing transparency, courts will provide public records (redacted to the extent necessary to protect any proprietary or personal information) of what sort of content is protected and what is illegal. Concerning propriety filter-system knowledge, courts

---

#:~:text=In%20California%2C%20a%201992%20statute,in%20some%20way%20on%20campus [https://perma.cc/EX5U-9N8J]

235. Cal. Educ. Code § 94367(b) (2009).

236. *See First Amendment on Private Campuses*, HARV. CIV. RTS. – CIV. LIBERTIES L. REV. (Dec. 1, 2015), https://harvardcrcl.org/first-amendment-on-private-campuses/ [https://perma.cc/PH7P-YPT7].

237. Although not statutorily prohibited, federal courts have been reluctant to resolve constitutional controversies by means of declaratory judgments. The primary basis for these denials tends to be that the claimed injury was too "remote or speculative." RUSSELL WEAVER ET AL., REMEDIES 203–04 (5th ed. 2019).

238. *See* Nicholas Iovino, *Twitter's Trump Ban Sets Dangerous Precedent for Free Speech, Legal Scholar Warns*, COURTHOUSE NEWS SERV. (Jan. 14, 2021), https://www.courthousenews.com/twitters-trump-ban-sets-dangerous-precedent-for-free-speech-legal-scholar-warns/ [https://perma.cc/8JP2-ZY9U].

239. *Id.*

240. *Id.* (statement of Erwin Chemerinsky) ("I advocated for the Leonard Law because I think the values of free speech should be protected by private school administrations as it would be by public school administrations.").

may utilize in camera proceedings to protect any proprietary information the social platforms do not wish exposed to the public and its competitors.

The (c)(2) revision will additionally resolve the viewpoint discrimination issue by granting civil remedy options to users who prove the social media company knowingly censored the user's content on the grounds of viewpoint discrimination. The test for viewpoint discrimination is whether—within the relevant subject category—the government has singled out a subset of messages for disfavor based on the views expressed.[241] If the social media companies are to take the place of government actors via the revision, then this test would apply when determining if a platform removes a post under the First Amendment standard.

Alternatively, the introduction of a "bad faith" clause to deter broad removal of minority ideals may be utilized in (c)(2).[242] For example, inserting a clause that prevents an online platform from removing user content the provider knew did not violate freedom of expression will result in the good faith safe harbor no longer applying, and reinstatement of civil liability.[243] This "bad faith" clause puts a duty of care on the platform, but will not overly-burden content moderation for the platforms because it places the burden on the plaintiff to prove that the company knowingly reviewed the user's content and then removed it due to viewpoint discrimination.[244] If the removal occurred by algorithm, the platform is given a grace period of ten days to enable the post after user appeal. If the platform does not rectify the wrongful censorship in the set time period of ten days, the bad faith clause is triggered. Upon being triggered, the bad faith may result in

241. *Matal*, 137 S. Ct. at 1750.

242. *Bad faith*, L. DICTIONARY, https://dictionary.law.com/Default.aspx?selected=21 [https://perma.cc/TLD5-7ERS] (defining bad faith as "intentional dishonest act by not fulfilling legal or contractual obligations, misleading another, entering into an agreement without the intention or means to fulfill it, or violating basic standards of honesty in dealing with others.").

243. An alternative statute with a bad faith clause inserted into it may read as follows: No 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; *and (C) no provider of an interactive computer service may utilize sections (c)(1) or (c)(2) as defenses in legal proceedings if the provider is found to have engaged in "bad faith" removal of content.* 47 U.S.C. § 230 (c)(2) (emphasis added).

244. 47 U.S.C. § 230.

fines, attorneys' fees, sanctions, default judgment, or anything else within judicial discretion to punish the bad faith offender.

B. Federal Guidelines for Independent Social Media Oversight Boards

The second possibility in solving the issues of Section 230(c)(2) is the introduction of federally mandated Social Media Oversight Boards ("Oversight Boards"). This proposal suggests the United States federal government enacts statutes that require each large social media platform establish and maintain an independent oversight board to review decisions regarding removal of content. The Oversight Boards follow a similar structure to the Facebook Oversight Board ("FOB"), along with the FOB's respective charter.[245]

The social media companies that will have the responsibility of implementing these Oversight Boards meet any of the three following size or revenue requirements: (1) more than thirty million active monthly users in the US, or (2) more than 300 million active monthly users worldwide, or (3) companies who have more than $500 million in global annual revenue. This solution takes the FOB, and replicates it for each large social media platform, requiring all designated social media platforms establish and maintain an independent oversight board.

Prior to the Facebook Oversight Board's creation, the Sarbanes-Oxley Act ("SOX Act") created the Public Company Accounting Oversight Board ("PCAOB").[246] Large-scale auditing companies prior to the SOX Act were self-regulated—just as the content regulation of social media companies exists now.[247] The lack of regulation ended with immense financial issues

245. *Oversight Board Charter*, FACEBOOK (Sept. 2019), https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf; *see generally* 15 U.S.C. § 7211 (2010) (example of an oversight board).

246. 15 U.S.C. § 7201. The SOX Act created the PCAOB, which has four primary duties: "(1) Register public accounting firms that prepare audit reports for issuers, brokers, and dealers; (2) establish or adopt auditing and related attestation, quality control, ethics, and independence standards; (3) inspect registered firms' audits and quality control systems; and (4) investigate and discipline registered public accounting firms and their associated persons for violations of specified laws, rules, or professional standards." "The Securities and Exchanges Commission ("SEC") has oversight authority over the PCAOB, including the approval of the Board's rules, standards, and budget." *About*, PUB. CO. ACCT. OVERSIGHT BD., https://pcaobus.org/about#:~:text=TTh%20PCAOB%20is%20a%20nonnonpro,accurate%2C%20and%20indeindepen%20audit%20reports.&text=Inspect%20registered%20firms'%20audits%20and%20quality%20concont%20systems [https://perma.cc/TVY8-NMDA].

247. Andriy Blokhin, *The Impact of the Sarbanes-Oxley Act of 2002*, INVESTOPEDIA (June 27, 2019), https://www.investopedia.com/ask/answers/052815/what-impact-did-sarbanesoxley-act-have-corporate-governance-unitedstates.asp#:~:text=After%20a%20prolonged%20period%20of,public%20companies%20to%20defraud%20investors [https://perma.cc/PJ4W-HCHX].

and demonstrated public company auditing needed a government fence post to keep it on the straight and narrow.[248] Here, there is a distinction as to what the PCAOB and Oversight Boards are overseeing, as freedom of expression is a more fluid concept than whether a set of financial statements hides or exhibits fraud. However, established and independent evaluation of content removal is needed to prevent similar public injustice by the discretionary bottleneck the social media companies hold over regulating posts.[249]

The Oversight Boards follow the FOB's lead, assigning authority to the members of these Oversight Boards with jurisdictional, intellectual, and financial independence of the board members from the social media companies.[250] The reviews by the Oversights Boards are to be "neutral," "independent judgments" that are rendered "impartially" with no influential pressure from the respective platform.[251] However, the adjudicating members of the Oversight Boards still need a firm grasp of the relevant laws and how to apply those laws, while also being familiar with the social media industry. Furthermore, the FOB Charter specifies that each review panel for each case will contain "at least one member from the region" where the case arose.[252] Conflicts of interest would need to be thoroughly researched to prevent the possibility of inequitable rulings.

Regarding the governance relationship between the proposed Oversight Boards and the necessary legal structure, the FOB's structure is once again a viable precedent. Between the FOB, legal Trust, and Facebook, the Charter states that the FOB is to "review content and issue reasoned, public decisions" and "provide advisory opinions on Facebook's content policies."[253] The Trust's responsibilities are to fund FOB's budget and appoint and remove members of FOB, while Facebook's responsibilities are to "commit to the board's independent oversight on content decisions and the implementation of those decisions," to fund the Trust, to appoint trustees, and to "contract for services" with the FOB.[254] If this legal structure is replicated across the required platforms, it will allow each Oversight Board

248. *Id.*
249. *Id.*
250. Klonick, *supra* note 56.
251. *Oversight Board Charter*, FACEBOOK, art. 1 (Sept. 2019), https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf.
252. *See id.* art. 1, § 2.
253. *Id.* art. 3, § 2.
254. *Id.* art. 5, § 1.

to be self-sustaining and maintain autonomy in its decision-making processes.

As for adopting uniform standards of assessment among the Oversight Boards, it is possible to allow the Oversight Boards, based on their platforms, to individualize the assessment standards. The FOB uses Facebook's set of values to guide its content policies and decisions.[255] The FOB reviews content enforcement decisions and determines whether they were consistent with Facebook's content policies and values.[256] For each decision, any prior board decisions have precedential value and are to be viewed as highly persuasive when the facts, applicable policies, or other factors are substantially similar, and when reviewing decisions, the board pays particular attention to the impact of removing content in light of human rights norms protecting free expression.[257] Consistent and reliable protection against viewpoint discrimination is feasible with this model.

The Oversight Boards proposed will apply only to users based within the US as to avoid jurisdictional issues for users based outside the US and allow uniform application of US legal standards to all reviews. Additionally, the Oversight Boards shall follow the FOB's specification that each review panel for each case will contain "at least one member from the region" where the case arose.[258] Foreign cases will add more expenses and consume more time than strictly keeping the Oversight Boards domestic.[259] Once the Oversight Boards are established and functioning adequately, then expanding into foreign jurisdictions may be possible.

In terms of end results, the Oversight Boards decide whether to allow or remove the content properly brought to it for review.[260] The board can also uphold or reverse a designation that led to an enforcement outcome, such as deciding that content depicts graphic violence, and should therefore display a warning screen.[261] The financial burden is put on the social media companies as a business expense. However, only the social media companies meeting the statutory requirements of sufficient scope or revenue shoulder the additional burden and will have sufficient notice to plan accordingly.

Given the current political and health-adverse climate, implementation of these Oversight Boards may take several years to realize, and FOB is still a fledgling in terms of a case study. Yet, it is quite a promising solution. First, platforms' overbroad discretion will be put in check through independent

255. *Id.* art. 5.
256. *Id.* art. 2, § 2.
257. *Id.*
258. *Id.* art. 2, § 2.
259. *Id.* art. 3, § 2.
260. *Id.* art. 5, § 3.
261. *Id.* art. 3, § 5.

review reports and established case precedent. Second, transparency would be addressed through the detailed public reports available after every decision.[262] Third, viewpoint discrimination will be significantly diminished by the independent, neutral review process. These solutions do not disincentive regulation of intellectual property violations or criminal content, and simultaneously put in measures to combat involuntary or biased censorship of lawful expression.

VI. CONCLUSION

Social media companies are here to stay. Their massive growth combined with their societal influence opens up worlds of possibilities: both good and bad. As technology advances, so must the methods used to ensure individual expression. Achieving the goal of decreasing discretionary power of social media company censorship to viewpoint discrimination of lawful user-posted content, while preserving private governance of social media business practices, is achievable. This goal is possible through two methods: (1) restructuring (c)(2) to grant social media companies immunity from liability for removing posts only if the posts contain illegal content or otherwise unprotected speech by the First Amendment, and (2) through Social Media Oversight Boards. Section 230 sheltered the internet companies during their adolescence, but now the internet companies can fend for themselves. It is time for the new era of user focused protection to begin.

262. *Id.* art. 3, § 5.