William & Mary Business Law Review

Volume 15 (2023-2024)
Issue 2

Article 5

2-2024

Killing Two Birds with One Stone: Remedying Malicious Social Bot Behavior Via Section 230 Reform

Jackson Smith

KILLING TWO BIRDS WITH ONE STONE: REMEDYING MALICIOUS SOCIAL BOT BEHAVIOR VIA SECTION 230 REFORM

JACKSON SMITH*

ABSTRACT

As "interactive computer services" (social media sites) expanded over the past decade, so too did the prevalence of "social bots," software programs that mimic human behavior online. The capacity social bots have to exponentially amplify often-harmful content has led to calls for greater accountability from social media companies in the way they manage bot presence on their sites. In response, many social media companies and private researchers have developed bot-detection methodologies to better govern social bot activities. At the same time, the prevalence of harmful content on social media sites has led to calls to reform Section 230 of the Communications Decency Act of 1996, the law that largely immunizes social media sites from liability for third-party content on their platforms. Such reform proposals largely entail making Section 230 immunity contingent on social media companies following new requirements when moderating content. Social bots have been left out of these reform conversations, however. This Note suggests that including specific provisions regulating social bots within broader Section 230 reform will help remedy both outdated Section 230 provisions and malicious social bots' effects. Fusing characteristics from several Section 230 reform proposals with existing bot-governance technology will help establish a legal foundation for social media companies' new social bot management requirements. Two suggested requirements are: (1) interactive computer services must have some type of monitoring and

* JD Candidate, 2024, William & Mary Law School; Bachelor of Arts, 2021, Christopher Newport University. The Author would like to thank his family and friends for their love and support in all facets of his life, as well as the *William & Mary Business Law Review* Volume 15 staff for their assistance in preparing this Note for publication.

classification system that helps users determine the "bot-ness" of social media accounts; and (2) interactive computer services must provide an accessible medium for users to view the data that its monitoring and classification system produces. These requirements will help protect the validity of organic online exchanges and reduce the potential power of deceitful influence campaigns.

TABLE OF CONTENTS

INTRODUCTION

"Bots," specifically those present on the world's largest social media platforms (social bots) have had their fair share of media and academic coverage in the past several years.[1] Recent events like Russia's disinformation campaign during the 2016 Presidential Election,[2] the Saudi-backed information distortion effort in the aftermath of journalist Jamal Khashoggi's 2018 killing,[3] the divisive issues of the year 2020,[4] the emergence of the COVID-19 Pandemic,[5] Elon Musk's accusations against Twitter of dishonesty around bots on the site,[6] and several others, have

---

[1] *See, e.g.*, Siobhan Roberts, *Who's a Bot and Who's Not?*, N.Y. TIMES (July 16, 2020), https://www.nytimes.com/2020/06/16/science/social-media-bots-kazemi.html [https://perma.cc/K4UX-BDK8]; Pascal Podvin, *The Social Impact of Bad Bots and What To Do About Them*, FORBES (Dec. 4, 2020), https://www.forbes.com/sites/forbestechcouncil/2020/12/04/the-social-impact-of-bad-bots-and-what-to-do-about-them/?sh=364d399659e0 [https://perma.cc/YLQ8-A8YD]; Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. ECON. PERSPS., no. 2, Apr. 2017, at 211, 211–36; Andrew Leber & Alexei Abrahams, *A Storm of Tweets: Social Media Manipulation During the Gulf Crisis*, 53 REV. MIDDLE E. STUDS. 241, 241–48 (2019), https://www-jstor-org.proxy.wm.edu/stable/pdf/26896726.pdf [https://perma.cc/W4PA-FEMA].

[2] *See*, *e.g.*, Allcott & Gentzkow, *supra* note 1, at 211–12; Alexandre Bovet & Hernan A. Makse, *Influence of Fake News in Twitter During the 2016 US Presidential Election*, 10 NATURE COMM'N no. 1, Jan. 2019, at 1, 2.

[3] *See*, *e.g.*, Chris Bell & Allistair Coleman, *Khashoggi: Bots Feed Saudi Support After Disappearance*, BBC (Oct. 18, 2018), https://www.bbc.com/news/blogs-trending-45901584 [https://perma.cc/8AA4-R5ZC]; Andrew Leber & Alexei Abrahams, *Saudi Twitter Blew Up With Support for the Crown Prince. How Much of It Is Genuine?*, WASH. POST (Mar. 9, 2021), https://www.washingtonpost.com/politics/2021/03/09/saudi-twitter-blew-up-with-support-crown-prince-how-much-it-is-genuine/ [https://perma.cc/J6FE-UZB6].

[4] Emilio Ferrara et al., *Characterizing Social Media Manipulation in the 2020 U.S. Presidential Election*, 25 FIRST MONDAY no. 11 (2020), https://journals.uic.edu/ojs/index.php/fm/article/view/11431/9993 [https://perma.cc/G9QX-QVLP]; Elyse Samuels & Monica Akhtar, *Are 'Bots' Manipulating the 2020 Conversation? Here's What's Changed Since 2016*, WASH. POST (Nov. 20, 2019), https://www.washingtonpost.com/politics/2019/11/20/are-bots-manipulating-conversation-heres-whats-changed-since/ [https://perma.cc/5ULL-6BQY].

[5] Samuels & Akhtar, *supra* note 4.

[6] *Musk Countersuit Accuses Twitter of Fraud over 'Bot' Count*, AP NEWS (Aug. 5, 2022, 1:54 PM) [hereinafter *Musk Countersuit*], https://apnews.com/article/elon-musk-twitter-inc-technology-933f52cf58fea145e71f1112563951d4

familiarized the public with the controversies surrounding social media content moderation and bot activity.[7] Though online bots possess demonstrated power to rapidly amplify harmful or misleading content, impersonate individuals and entities, and influence online discussions,[8] public attention has focused on content moderation itself, not how the content is spread.[9] As a result of this attention, many legislators have proposed reforms to the legal protections large social media companies have in this arena.[10]

Section 230 of the Communications Decency Act of 1996 is the biggest target,[11] as it provides a liability shield to social media companies for the content and actions of third parties using their sites.[12] Until recently, this law was regarded as invaluable to the rapid expansion of the internet and all its wealth-creating applications.[13] Massive social media companies benefited hugely from the peace of mind that Section 230's liability protection brought them in the United States, where they would not have to worry about defensive legal spending or overburdensome monitoring

---

[https://perma.cc/V6DX-826F]; Clare Duffy & Brian Fung, *Elon Musk Commissioned This Bot Analysis in His fight with Twitter. Now it Shows He Could Face if He Takes Over the Platform*, CNN BUS. (Oct. 10, 2022), https://www.cnn.com/2022/10/10/tech/elon-musk-twitter-bot-analysis-cyabra/index.html [https://perma.cc/EP38-SGMG]; Marek N. Posard, *Elon Musk May Have a Point About Bots on Twitter*, RAND CORPORATION: THE RAND BLOG (Sept. 23, 2022), https://www.rand.org/blog/2022/09/elon-musk-may-have-a-point-about-bots-on-twitter.html [https://perma.cc/C7TM-NCSZ].

[7] *See* Matthew Hines, Comment, *I Smell A Bot: California's S.B. 1001, Free Speech, and the Future of Bot Regulation*, 57 HOUS. L. REV. 405, 410 (2019).

[8] *Id.* at 405.

[9] *See* discussion *infra* Section II.B.

[10] *See* Nina I. Brown, *Regulatory Goldilocks: Finding the Just and Right Fit for Content Moderation on Social Platforms*, 8 TEX. A&M L. REV. 451, 457 (2021).

[11] Quinta Jurecic, *The Politics of Section 230 Reform: Learning from FOSTA's Mistakes*, BROOKINGS INST. (Mar. 1, 2022), https://www.brookings.edu/research/the-politics-of-section-230-reform-learning-from-fostas-mistakes/ [https://perma.cc/Z428-PEVW]. In some instances, private plaintiffs have indirectly gone after Section 230 by trying to hold social media companies liable for third-party terrorist content. *See* Gonzalez v. Google LLC, 143 S. Ct. 1191, 1192 (2023) (remanding case due to failure to state a claim without answering Section 230 questions).

[12] *See* 47 U.S.C. § 230(c)(2).

[13] *See* discussion *infra* Section II.A.

regimes.[14] Political and ideological fights in the past several years, however, have drastically reduced Section 230's popularity,[15] and several proposals now exist that would limit its liability shield or add more requirements for its application.[16] Many of these proposals seek to require social media companies to make more extensive disclosures about their content moderation techniques and improve their reporting to regulatory authorities.[17]

Largely missing from these Section 230 reform proposals are any specific requirements regulating bot activity on social media sites.[18] Despite this absence, there are laws and proposals for regulating online bots with requirements outside of Section 230's bounds.[19] California's 2019 "Bolstering Online Transparency Act" requires disclosures when bots communicate with humans in California.[20] Public and government scrutiny has also incentivized social media companies to undertake self-regulatory measures that try to identify and control the influence of online bots.[21]

While these different efforts may eventually bear fruit, a more coherent legal regime that combines their characteristics underneath the liability-shield incentive of Section 230 would likely do the most to remedy the controversies surrounding social media content moderation and bot activity.[22] Fusing characteristics from several Section 230 reform proposals with existing bot-governance technology will help establish a legal foundation for social media companies' new social bot management requirements.[23] Two suggested requirements are: (1) interactive

---

[14] *See* discussion *infra* Part III.

[15] *See* Jurecic, *supra* note 11.

[16] *Id.*

[17] *See* Brown, *supra* note 10, at 465, 470.

[18] *See* discussion *infra* Section II.B.

[19] *See* Barry Stricke, People v. Robots*: A Roadmap for Enforcing California's New Online Bot Disclosure Act*, 22 VAND. J. ENT. & TECH. L. 839, 839, 842 (2020).

[20] *See SB-1001 Bots: Disclosure,* CAL. LEGIS. INFO., https://leginfo.legislature.ca.gov/faces/billHistoryClient.xhtml?bill_id=201720180SB1001 [https://perma.cc/XN2U-9H6J].

[21] *See* Oliver Beatson et al.*, Automation on Twitter: Measuring the Effectiveness of Approaches to Bot Detection*, 41 SOC. SCI. COMPUT. REV. 181, 184 (Feb. 2023), https://doi.org/10.1177/08944393211034991 [https://perma.cc/V2EZ-C2WF].

[22] *See* discussion *infra* Part III.

[23] *See* discussion *infra* Part III.

computer services must have some type of monitoring and classification system that helps users determine the "bot-ness" of social media accounts and (2) interactive computer services must provide an accessible medium for users to view the data that its monitoring and classification system produces. These requirements will help protect the validity of organic online exchanges and reduce the potential power of deceitful influence campaigns.[24]

Section I.A will discuss general information about bots and social bots and highlight numerous examples of malicious social bot capabilities and the potential dangers associated with them.[25] Section I.B will discuss online bot regulation efforts, mainly, Twitter's attempts at self-regulation and California's novel law the "Bolstering Online Transparency Act."[26] Section II.A will introduce the general background of social media companies' liability shield under Section 230,[27] while Section II.B will discuss some of the recent proposals for legal reform of Section 230.[28] Part III will lay out a proposed legal foundation that fuses characteristics from several Section 230 reform proposals with existing bot-governance technology.[29]

I. ONLINE BOTS

A. The Rise of Social Bots on Social Media Platforms

Along with the massive growth of social media[30] has come a growing prevalence of bots.[31] Numerous instances in the past decade have brought media attention to the capabilities bots and

---

[24] *See* discussion *infra* Part III.

[25] *See* discussion *infra* Section I.A.

[26] *See* discussion *infra* Section I.B.

[27] *See* discussion *infra* Section II.A.

[28] *See* discussion *infra* Section II.B.

[29] *See* discussion *infra* Part III.

[30] Social Media Fact Sheet, PEW RES. CTR.: INTERNET & TECH. (June 12, 2019), https://www.pewresearch.org/internet/fact-sheet/social-media/ [https://perma.cc/B6P6-7HZD] (noting that by 2018, over 70% percent of U.S. adults had at least one social media account, with that figure rising to 88% for adults under the age of twenty-nine).

[31] *See* Onur Varol et al., *Online Human-Bot Interactions: Detection, Estimation, and Characterization*, ARXIV 280, 280 (2017), https://doi.org/10.48550/arXiv.1703.03107 [https://perma.cc/762L-PBHA].

their human controllers have to influence the online marketplace of ideas.[32] It is particularly bots' potential for malicious activity that has given them notoriety over the past decade.[33] As an ever-present and growing force on many of the world's top social media sites[34] (specifically on "X," formerly "Twitter"), bots should be a part of any conversation involving changes to the legal and regulatory regimes these sites operate under.[35] Attempting to regulate "bots" on social media sites, especially through the mechanism of Section 230 reform, requires a foundational understanding of exactly which "bots" a regulation aims to cover[36] and what malicious activities such bots can accomplish.[37]

"Bots" are generally defined as "software programmed to perform automated tasks in digital space"[38] and can take on a variety of forms and accomplish many different tasks.[39] For example: "web crawlers" comb through troves of online data to provide search results,[40] chatbots use conversational patterns to assist users in requested ways,[41] and others disseminate useful information through social media or direct posts.[42] "Social bots" refer to the type of bots most often found on social media sites and are arguably the bots most capable of malicious activities.[43]

---

[32] *See supra* notes 1–6 and accompanying text.

[33] *See* Varol et al., *supra* note 31, at 280; Victor Suarez-Lledo & Javier Alvarez-Galvez, *Assessing the Role of Social Bots During the COVID-19 Pandemic: Infodemic, Disagreement, and Criticism,* 24 J. MED. INTERNET RSCH. 1128, 1128 (2022), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9407159/ [https://perma.cc/7T4Y-BK2P].

[34] See Leber & Abrahams, *supra* note 1, at 241.

[35] *See generally* Ferrara et al., *supra* note 4 (demonstrating the impact of bots on democracy).

[36] Regulation is futile if such regulation does not clearly define what it is regulating. *See* Stricke, *supra* note 19, at 848–60.

[37] Hines, *supra* note 7, at 410.

[38] *Id.* at 408.

[39] *See* Varol et al., *supra* note 31, at 280; Niccolo Pescetelli et al., *Bots Influence Opinion Dynamics without Direct Human-Bot Interaction: The Mediating Role of Recommender Systems*, 7 APPLIED NETWORK SCI. no. 1 (Dec. 2022), https://appliednetsci.springeropen.com/articles/10.1007/s41109-022-00488-6 [https://perma.cc/BY56-WHW6].

[40] Hines, *supra* note 7, at 408.

[41] *Id.* at 409.

[42] Varol et al., *supra* note 31, at 280.

[43] Emilio Ferrara et al., *The Rise of Social Bots*, 59 COMMS. ACM. 96, 96 (2016) ("A 'social bot' is a computer algorithm that automatically produces

These bots "create an online account to complete tasks a human user does, posting content and feedback."[44] Social bots come in several different forms, but all have a degree of automated features programmed into them.[45] Hybrid bots, known as "cyborgs," are automated accounts which receive directions and input from humans through URLs, particular content, and spam messages.[46] "Full" social bots are autonomous (other than receiving human instructions through software), and the most sophisticated of them can successfully "mimic human behavior," which includes sending out content at different time intervals and replying to human accounts.[47] Social media users often have a difficult time differentiating between the most advanced bots' content and humans' organic content.[48] Social bots can prove greatly useful to human users when they are instructed to do beneficial tasks,[49] but they can be equally as harmful when used for malicious purposes.[50]

As put by one scholar, "bots are problematic for the same reason that they are useful—they amplify the power and efficiency of a single person."[51] This ability to amplify is made more effective when coupled with the nature of online discussion.[52] Online discussion is a forum ripe for "building consensus among individual users,"[53] which makes it easier for individuals using

---

content and interacts with humans on social media, trying to emulate and possibly alter their behavior.").

[44] Stricke, *supra* note 19, at 852.

[45] Nathaniel Persily, *Can Democracy Survive the Internet?*, 28 J. DEMOCRACY, no. 2, Apr. 2017, at 63, 70.

[46] *Id.*

[47] Matthew Hindman & Vlad Barash, *Disinformation, 'Fake News' and Influence Campaigns on Twitter*, KNIGHT FOUND. 1, 17 (Oct. 2018).

[48] *See* Shannon Liao, *Most Americans Say They Can't Tell the Difference Between a Social Media Bot and a Human*, VERGE (Oct. 15, 2016), https://www.theverge.com/2018/10/15/17980026/social-media-bot-human-difference-ai-study [https://perma.cc/23FS-KMNL].

[49] *See infra* notes 84–86.

[50] *See* discussion *infra* Section I.A.

[51] Hines, *supra* note 7, at 410 (citing Elisabeth Eaves, *The California Lawmaker Who Wants to Call a Bot a Bot*, BULL. ATOMIC SCIENTISTS (Aug. 23, 2018), https://thebulletin.org/2018/08/the-california-lawmaker-who-wants-to-call-a-bot-a-bot/ [https://perma.cc/Y4ZM-CVE4] ("The difference between a single individual attempting to stir controversy and what a bot can accomplish is one of scale . . . .")).

[52] *Id.* at 410.

[53] *Id.* at 411.

the amplification capabilities of bots to present their viewpoint as widely held.[54]

Numerous studies and reports in the past several years have highlighted real-world examples of human controllers using the amplification power of bots to spread fake news and disinformation, attempt to influence online discussion, distort organic content, and more.[55] Social bots targeted conversations about events such as the 2016 U.S. Presidential Election, the murder of Saudi journalist Jamal Khashoggi, the 2020 COVID-19 Pandemic and U.S. Presidential Election, and recent happenings involving Elon Musk's takeover of Twitter.[56]

The 2016 U.S. Presidential Election was one of the first major events which brought social bots' capabilities into the spotlight.[57] Academic attention increased after U.S. intelligence agencies' assessments that "foreign agents used social media to influence the 2016 U.S. Presidential Election."[58] Many purveyors of misinformation on social media during the period simply latched onto the 2016 campaign's ideological passions for pecuniary gain.[59]

---

[54] *Id.* ("Because it is possible for a single independent producer to widely distribute content based on shares and trending tags, bots programmed to share specific content can be used 'to create a bandwagon effect, to build fake social media trends . . . and even to suppress the opinions of the opposition.'").

[55] *See* Chris Baraniuk, *How Twitter Bots Help Fuel Political Feuds*, SCI. AM. (Mar. 27, 2018), https://www.scientificamerican.com/article/how-twitter-bots-help-fuel-political-feuds/ [https://perma.cc/SK4B-F6G2].

[56] *See supra* notes 1–6.

[57] *See* Hines, *supra* note 7, at 410.

[58] *Id.*; Robert S. Mueller, III, REPORT ON THE INVESTIGATION INTO RUSSIAN INTERFERENCE IN THE 2016 PRESIDENTIAL ELECTION 15–16 (2019); *see* Hindman & Barash, *supra* note 47, at 33 (recounting Twitter's January, 2018 claim to have identified around 54,000 bot accounts linked to the Russian government); *see also* Jon Swaine, *Twitter Admits Far More Russian Bots Posted on Election Than It Had Disclosed*, GUARDIAN (Jan. 19, 2018, 7:46 PM), https://www.theguardian.com/technology/2018/jan/19/twitter-admits-far-more-russian-bots-posted-on-election-than-it-had-disclosed [https://perma.cc/P3LL-JYBU].

[59] *See* Allcott & Gentzkow, *supra* note 1, at 217 ("[N]ews articles that go viral on social media can draw significant advertising revenue when users click to the original site."). One prominent example of prolific misinformation authors motivated by ad revenue came out of North Macedonia, where "[m]ore than 100 sites posting fake news were run by teenagers in the small town of Veles, Macedonia . . . The teenagers of Veles . . . produced stories favoring both Trump and Clinton that earned them tens of thousands of dollars." *See id.*

Regardless of motive, much of the disinformation spread in 2016 relied on bots and botnets to reach wide audiences.[60] One 2018 study sought to understand the spread of fake news on Twitter by mapping out "more than 10 million tweets from 700,000 Twitter accounts that linked to more than 600 fake and conspiracy news outlets" in the months before and after November 2016.[61] Many bots coalesced into "botnets," networks of automated accounts which interact with each other's content, exposing the content (usually clickbait or misinformation) to as many human accounts as possible along the way.[62] Botnets and bot account clusters create dense webs of reciprocal engagement that increase the likelihood of a particular link or topic reaching the "trending" threshold.[63] The 2018 study's authors estimated that bots on Twitter created "two-thirds of Twitter links to popular sites"[64] (at the time), and the "dense core of misinformation accounts [identified in the study was] dominated by social bots."[65] Another study estimated that, in a period spanning from September 2016 to October 2016, "bots produced about a fifth of all tweets related to the upcoming election."[66] Bots were able to drive Twitter conversations, skew the results of polls, push topics to trend, and falsely inflate engagement numbers for certain accounts and their content.[67]

Though malicious social bot activity is problematic in liberal Western societies, its potential power is increased when authoritarian state actors (or those sympathetic to them) utilize it to influence domestic speech and temper international criticism.[68] In 2018, after Saudi agents murdered dissident journalist

---

[60] *See* Hindman & Barash, *supra* note 47, at 16.

[61] *Id.* at 3.

[62] *See id.* at 43.

[63] *See* Persily, *supra* note 45, at 70; *see also* Hindman & Barash, *supra* note 47, at 24.

[64] Hindman & Barash, *supra* note 47, at 18 (citing Stefan Wojcik et al., *Bots in the Twittersphere*, PEW RSCH. CTR. (Apr. 9, 2018), http://www.pewinternet.org/2018/04/09/bots-in-the-twittersphere/ [https://perma.cc/Z53P-K8P4].

[65] *Id.* at 4, 18 ("[M]achine learning models estimate that 33 percent of the 100 most-followed accounts in our postelection map—and 63 percent of a random sample of all accounts—are 'bots.'").

[66] Persily, *supra* note 45, at 70.

[67] *See id.*; *see also* Hindman & Barash, *supra* note 47, at 44 ("[M]any fake accounts exist to inflate numbers of followers, likes, and retweets.").

[68] *See* Hindman & Barash, *supra* note 47, at 44; *see* Leber & Abrahams, *supra* note 1, at 242–48; *see also* Samuel Woolley & Nicholas Monaco, *Amplify*

Jamal Khashoggi in the Saudi consulate in Istanbul, Turkey, Arabic hashtags pushing pro-Saudi government talking points began to trend on Twitter (especially in Saudi Arabia).[69] These hashtags included phrases such as: "[u]nfollow enemies of the nation," "[w]e all have trust in Mohammed bin Salman," and "[w]e have to stand by our leader," with several reaching top spots in global trends.[70] Ben Nimmo, a fellow at the Atlantic Council Digital Forensic Research Lab, analyzed the hashtag "unfollow enemies of the nation" and published his results in a Twitter thread.[71] In his thread, he highlighted how "just one account" was "driving the [online] traffic,"[72] accounting for over 103,000 mentions.[73] Looking at some of the other trending hashtags, Nimmo and journalists observed many telltale signs of bot activity such as sudden activation of dormant accounts, "identical or near-identical material" being posted between other "suspicious accounts,"[74] and a very high, over 96%, retweet proportion.[75]

The Khashoggi hashtags are just one example in a growing list of instances where governments (largely authoritarian) have used bots to influence domestic social media (Twitter) discussion.[76] "Networked authoritarianism"[77] has become a ubiquitous

---

the Party, Suppress the Opposition: Social Media, Bots, and Electoral Fraud, 4 GEO. L. TECH. REV. 447, 453–55 (2020) (highlighting instances where government and private actors in North Macedonia, Nigeria, India, Mexico, Russia, Italy, and the UK used social bots and botnets in attempts to boost their political viewpoints in online Twitter discussions).

[69] Bell & Coleman, *supra* note 3.

[70] *Id.*

[71] Ben Nimmo (@benimmo), TWITTER (Oct. 17, 2018, 4:31 AM), https://twitter.com/benimmo/status/1052477177638375425?s=20&t=faZGQ2SARxgX_mVXdcsIgQ [https://perma.cc/X62Y-PFNE].

[72] Ben Nimmo (@benimmo), TWITTER (Oct. 17, 2018, 4:37 AM), https://twitter.com/benimmo/status/1052478749923524608?s=20&t=faZGQ2SARxgX_mVXdcsIgQ [https://perma.cc/XHJ4-YXXQ].

[73] Ben Nimmo (@benimmo), *supra* note 71.

[74] *See* Bell & Coleman, *supra* note 3.

[75] Ben Nimmo (@benimmo), TWITTER (Oct. 17, 2018, 4:35 AM), https://twitter.com/benimmo/status/1052478178890055680?s=20&t=faZGQ2SARxgX_mVXdcsIgQ [https://perma.cc/8W22-H98J] ("Retweet proportion of 96.3% is off the charts. Either there's a ton of self-effacing user out there, or it's a coordinated retweet farm, or it's bots.").

[76] *See* Leber & Abrahams, *supra* note 1, at 242–48.

[77] *See id.* at 245 ("Here, state-led action (by some combination of real users and automated accounts) promotes a particular hashtag as a 'trending topic'

phenomenon in many of the world's largest authoritarian nations like China (within its own domestic social networks), Russia, and several of the Gulf States.[78]

Notwithstanding the uproar that emerged after bots' malicious activities on social media sites came to light in the years after the 2016 election, and despite some sites' actions to stem malicious social bot activity,[79] such activity has continued in the years after.[80]

More recently, bad actors turned to social bots to try to distort online discussion around the 2020 COVID-19 Pandemic, prominent social issues of that year, and the 2020 U.S. Presidential Election.[81] A study encompassing over 240,000,000 election-related tweets between June and September of 2020 resulted in

---

for the country in question or retweets posts by other accounts to make them seem more influential than they actually are."); *see also id.* at 242 ("[R]egimes in the Middle East and North Africa . . . have now gone on the offensive, exploiting Twitter and other forms of social media as vectors for political propaganda by which to manufacture the perception of support for themselves and their policies while dividing and discouraging their opposition.").

[78] Lorenzo Franceschi-Bicchierai, *How "Mr. Hashtag" Helped Saudi Arabia Spy on Dissidents*, MOTHERBOARD (Oct. 29, 2018), https://motherboard.vice.com/en_us/article/kzjmze/saud-al-qahtanisaudi-arabia-hacking-team [https://perma.cc/RF9K-GYKU]; *see* Leber & Abrahams, *supra* note 1, at 246–48. Bots retweeted "favorable remarks about Saudi Arabia" made by President Donald Trump when he visited the Kingdom in 2017. *See id.* at 246. A conversation on Twitter between Bahraini activists was "suddenly inundated with sectarian hate speech in a coordinated flooding attack by numerous bot accounts." *See id.* at 247–48; *see also Kuwait: Teacher Faces Jail over Twitter Comments,* HUM. RTS. WATCH (July 20, 2013), https://www.hrw.org/news/2013/07/20/kuwait-teacher-faces-jail-over-twitter-comments [https://perma.cc/3469-RP2K].

[79] "Following the 2016 election, several internet platforms changed their policies concerning information on their sites to address perceived shortcomings of the communications environment. Google, Facebook, and Twitter each enacted new rules for news and other communication on their platforms, based on complaints related to the 2016 presidential campaign." Persily, *supra* note 45, at 72–73.

[80] *See* discussion *infra* Section I.A.

[81] *See, e.g.*, Suarez-Lledo & Alvarez-Galvez, *supra* note 33, at 1128; Joshua Uyheng et al., *Bots Amplify and Redirect Hate Speech in Online Discourse About Racism During the COVID-19 Pandemic*, 8 SOC. MEDIA + SOC'Y, no. 3, July–Sept. 2022, at 1, https://journals-sagepub-com.proxy.wm.edu/doi/10.1177/20563051221104749; Menghan Zhang et al., *Social Bots' Involvement in the COVID-19 Vaccine Discussions on Twitter*, 19 INT'L J. OF ENV'T RSCH. AND PUB. HEALTH no. 3, 1651 (2022), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8835429/_[https://perma.cc/K9GS-53VG].

findings similar to those made in numerous studies focused on 2016 election-related tweets.[82] Among the findings were that social bots "generated spikes of conversations around real-world political events," that they "exacerbate the consumption of content produced by users with their same political views," and that they (again) provided evidence of "coordinated efforts carried out by Russia, China, and other countries."[83]

Many of the most controversial social topics were intertwined with 2020's election discussion and were also targets of social bot influence attempts.[84] Social bots helped direct and redirect online debates about racism,[85] spread pandemic-related conspiracies,[86] and provide opinions about the efficacy of COVID-19 vaccines.[87]

One of the most recent instances involving social bots that garnered widespread media attention was Elon Musk's high-profile takeover of Twitter.[88] Here, concern was that the harm that social bots can cause would directly influence a business deal consequential to the future of social media.[89] Musk tried to pull out of the $44 billion deal after accusing Twitter of not being transparent about the number of social bots present on its platform.[90] Musk alleged Twitter's users were far smaller in

---

[82] *See* Ferrara et al., *supra* note 4.

[83] *Id.*

[84] *See id.*; *see also* Zhang et al., *supra* note 81, at 1651; *see also* Uyheng et al., *supra* note 81.

[85] *See* Uyheng et al., *supra* note 81, at 12 ("[w]e empirically characterized bot-fueled amplification and redirection of hate from focusing on Asian and Chinese populations in March to targeting political discourse surrounding US political figures in August.").

[86] *See* Suarez-Lledo & Alvarez-Galvez, *supra* note 33, at 1129 (finding that bots contributed to what was characterized as an "infodemic" of abundant health-related information, both reliable and false, increasing the difficulty of finding reliable material).

[87] *See* Zhang et al., *supra* note 81, at 1652.

[88] *See, e.g.*, *Musk Countersuit*, *supra* note 6; Duffy & Fung, *supra* note 6; Posard, *supra* note 6.

[89] *See* Duffy & Fung, *supra* note 6. Twitter, however, accused Musk of using bot concern as a pretext to back out of the deal due to buyer's remorse. *See id.*

[90] *See Musk Countersuit*, *supra* note 6 ("Musk offered to buy Twitter earlier this year, then tried to back out of the deal by claiming the social platform was infested with a larger number [sic] of 'spam bots' and fake accounts than

number than Twitter led on (because of a higher actual proportion of spam bots being counted as "monetized" users), the smaller user count produced less ad revenue, and the smaller user count made the company worth less than Twitter claimed.[91] Since his Twitter takeover, Musk has routinely used polls on the site to make Twitter-related decisions, but has had to confront (or promote) the suspicion that these polls' results may be vulnerable to social bot manipulation.[92]

The numerous examples of social bots' potential to influence or distort public opinion online are distinct from the actual effects social bots produce.[93] Many scholars continuously debate whether bots and the disinformation and distortion they spread concretely changes public opinion (or election results);[94] but, a multitude of researchers agree at the very least, social bots help drive societal polarization through their influence in online discussions.[95] This polarization may be a paramount objective of those directing social bots, especially among state actors engaged

---

Twitter had disclosed."); *see also* Duffy & Fung, *supra* note 6 ("[A] study commissioned by Musk . . . found spam and bot accounts make up an estimated 11% of Twitter's total user base . . . .") ("Twitter has for years said that bots make up less than 5% of its monetizable daily active users."). Musk stated that clearing Twitter of harmful social bots was one of the main reasons he wished to buy the social media site. *Id.*

[91] See *Musk Countersuit*, *supra* note 6.

[92] Davey Alba & Bloomberg, *The People Have Spoken—or Maybe Not: Elon Musk's Use of Twitter Polls for Key Decisions Invites Manipulation*, FORTUNE (Dec. 22, 2022, 2:26 PM), https://fortune.com/2022/12/22/elon-musk-twitter-poll-manipulation-bots-acquisition/ [https://perma.cc/5TH3-P2WL].

[93] *See generally* Chang-Feng Chen, Social Bots' Role in Climate Change Discussion on Twitter: Measuring Standpoints, Topics, and Interaction Strategies, 12 ADVANCES IN CLIMATE CHANGE RSCH., 913, 921 (2021).

[94] *See* Allcott & Gentzkow, *supra* note 1 (arguing that the "ideological isolation" present in society prevents many people from being receptive to changing their most strongly held political beliefs).

[95] *See id.* (finding in their study that online consumers of fake news overwhelmingly consumed fake news that supported their already-held beliefs, filtering out any opposing opinions); *see also* Hines, *supra* note 7, at 411 (sharing how the Internet is an effective space for building consensus among individual users); Persily, *supra* note 45, at 72 ("[T]he Internet's unprecedented ability to facilitate the targeted delivery of relevant information, marketing, and even friendship also leads to the bubbles, filters, and echo chambers that shelter people from information that might challenge sent to them by campaigns, partisan media, or social networks.").

in "information warfare."[96] The fact that such bots and their human controllers have the capability and determination to influence wide-ranging subjects online, as demonstrated in this Section, continues to be a worrying threat.[97]

B. Attempts at Regulating Online Bots

Until recently, "bot regulation" was an untested concept.[98] California became one of the first jurisdictions in the world to try to specifically regulate bots when it enacted the "Bolstering Online Transparency Act" (the "Bot Act").[99] The Bot Act focuses primarily on requiring online bots to disclose the fact that they are bots (though in narrowly defined circumstances), hoping to protect internet users from manipulation.[100]

The Bot Act regulates bots beyond just those present on social media sites but is written with restraint to not discourage many of the beneficial uses of bots.[101] California legislators drew from many of the same concerns this Note discussed earlier as reasons for crafting the Bot Act, and they specifically emphasized the potential for bot users to perpetuate fraud in online commercial contexts.[102] The Bot Act's scope is clear from its text:

> It shall be unlawful for any person to use a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a

---

[96] *See* Hindman & Barash, *supra* note 47, at 11 ("Many observers have especially noted the evolution of Russian information warfare doctrine, along with its 'deep roots in long-standing Soviet practice.'").

[97] *See generally* discussion *supra* Section I.A.

[98] *See* Stricke, *supra* note 19, at 841.

[99] *Id.* at 842–43.

[100] *See* Hines, *supra* note 7, at 407, 412.

[101] *See* Stricke, *supra* note 19, at 848.

[102] *See* Hines, *supra* note 7, at 412. An author of the Bot Act, California State Senator Robert Hertzberg, identified the three examples of "businesses . . . us[ing] bot accounts to pad follower or subscriber accounts," companies and individuals using bots to "misrepresent the scope of their organic reach," and "bot-boosted" fake news generating ad revenue as some of the few specific instances the Act seeks to remedy. *Id.* In each example, the bot users seek material gain through bot-enabled misrepresentations. *Id.*

> purchase or sale of goods or services in a commercial transaction or to influence a vote in an election. A person using a bot shall not be liable under this section if the person discloses that it is a bot.[103]

Further, the Bot Act defines a "bot" as "an automated online account where all or substantially all of the actions or posts of that account are not the result of a person."[104] The Act defines "online" as "appearing on any public-facing Internet Website, Web application, or digital application, including a social network or publication."[105] A later provision specifies that the Bot Act "does not impose a duty" on "online platforms."[106] For an Internet site (including a social media site) to qualify as an "online platform," it must have "10,000,000 or more unique monthly United States visitors or users for a majority of months during the preceding 12 months."[107] As seen, the Act's legal demands target bot *users* (not platforms simply hosting bots) and attach a disclosure requirement to make certain bot communications legally permissible.[108]

These characteristics have spawned questions about the Bot Act's enforceability and possible First Amendment complications.[109] The Bot Act's enforceability remains an untested concept at this point,[110] but any action brought to punish violations of the Act would likely be brought via "civil actions under California's Unfair Competition Law and/or tort of fraudulent deceit."[111] Some have criticized the Act's supposed lack of government enforcement delegation as well as its inattention to the large sites that bots operate on, with one author contending that "a future bot law must not only go after users, but must also go after social

---

[103] CAL. BUS. & PROF. CODE § 17941(a) (West).

[104] *Id.* § 17940(a).

[105] *Id.* § 17940(b).

[106] *Id.* § 17942(c).

[107] *Id.* § 17940(c).

[108] *See* Stricke, *supra* note 19, at 842.

[109] *See id.* at 887; Hines *supra* note 7, at 415–16.

[110] *See generally* Renee Diresta, *A New Law Makes Bots Identify Themselves—That's the Problem*, WIRED (July 24, 2019, 9:00 AM), https://www.wired.com/story/law-makes-bots-identify-themselves/ (last visited Jan. 4, 2024).

[111] *See* Stricke, *supra* note 19, at 861 ("The UCL prohibits 'unfair competition' . . . Plaintiffs may 'borrow' violations of other laws to stand as unlawful practices 'independently actionable' under the UCL, provided the borrowed statutes are 'pursuant to business activity.'").

networks for not taking steps to prevent the [bot] behavior themselves."[112] One possible solution for plugging enforcement holes in a future (federal) act is delegating rulemaking authority to an administrative agency, either new or existing.[113] Relatedly, the discussion around potential First Amendment concerns is quite complicated, but the issue most pertinent to this Note would be whether required bot disclosure and/or labelling triggers First Amendment coverage.[114]

Though the Bot Act is unique in the legal world for its targeted regulation of bots, large social media companies, particularly Twitter, have launched self-implemented mechanisms for policing social bots.[115] In addition, private researchers have developed *numerous* methodologies for trying to identify social bots.[116] Parag Agrawal, the CEO of Twitter before Elon Musk's purchase of the company, discussed in a May 2022 Twitter thread some of the methods Twitter uses (used) to detect spam and bots.[117] He characterized the effort as a "dynamic" one based on reviewing "both public and private data (e.g., IP address, phone number, geolocation, cline/browser signatures, what the account does when it's active . . .) to make a determination on each account."[118] Twitter also recently launched an "automated" label for "good" bots to self-identify themselves with.[119]

Though efficient and prompt social bot monitoring has proven difficult, the vast array of detection methods currently existing could be used as a foundation for some minimum types

---

[112] *See* Hines, *supra* note 7, at 434–35.

[113] *See id.* at 435 ("[t]he federal government could create an agency to oversee internet communities . . . giving it the power to promulgate rules for websites that allow for the use of bots.").

[114] *See id.* at 427; *see also* discussion *infra* Part III.

[115] *See* Beatson et al., *supra* note 21, at 184.

[116] *See id.* at 184–86; *see also* discussion *supra* Section I.A.

[117] Parag Agrawal (@paraga), TWITTER (May 16, 2022, 12:26 PM), https://twitter.com/paraga/status/1526237578843672576?s=20 [https://perma.cc/8U9J-YGZL].

[118] Parag Agrawal (@paraga), TWITTER (May 16, 2022, 12:26 PM), https://twitter.com/paraga/status/1526237587085463553?s=20 [https://perma.cc/2526-9FBP].

[119] Ayrshare, *Twitter Launches "Automated" Label for Bots* (Feb. 17, 2022), https://www.ayrshare.com/twitter-launches-automated-label-for-bots/ [https://perma.cc/9WWM-6CJX].

of disclosure or identification regulation.[120] What is needed to make such regulation stick is legal reinforcement.[121]

II. SECTION 230

A. Background of Section 230

While bots proliferated in step with social media companies' reach in the past decade,[122] the legal foundation that enabled social media's growth in the United States goes back further.[123] During the 1990s, as the internet's potential was only beginning to become clear, many members of Congress sought to create a friendly legal foundation for the internet to continue to grow into the future.[124] Congress sought to accomplish this goal through the provisions of Section 230 of the Communications Act of 1996,[125] which among other things, bestows liability shields upon "interactive computer services" both for the content of third parties using their sites and for their efforts to moderate content.[126] Through these provisions, legislators aimed to achieve the goal of an open internet free from government content regulation.[127] Considering the rationale behind Section 230 is essential to understanding the litany of criticism and reform attempts that have sprung up over the past several years.[128] Additionally, the incredible value Section 230 gives to social media companies demonstrates why its reform would be an effective mechanism to regulate social bots.[129]

The positions justifying Section 230 grew out of reaction to a court case and preceding legislative efforts.[130] The original

---

[120] *See* discussion *infra* Part III.

[121] *See* discussion *infra* Part III.

[122] *See* discussion *supra* Section I.A.

[123] *See* discussion *infra* Section II.A.

[124] *See* 47 U.S.C. § 230(b)(1)–(2).

[125] 47 U.S.C § 230 (amending the Communications Act of 1934 as a part of the Communications Decency Act of 1996).

[126] *See* 47 U.S.C. § 230(c)(1)–(2).

[127] *See* 141 CONG. REC. H8470 (daily ed. Aug. 4, 1995) (statement of Rep. Christopher Cox).

[128] *See* discussion *infra* Part II.

[129] *See* discussion *infra* Part II.

[130] *See* discussion *infra* Section II.A.

authors of Section 230, then-representatives Christopher Cox and Ron Wyden, introduced what would become Section 230 as "Section 104" of a different House bill.[131] A New York trial court case result that Representative Cox derided as "backward" spurred on Section 230's drafting.[132]

Stratton Oakmont, Inc. v. Prodigy Services Co.[133] involved a plaintiff (Stratton) seeking to hold Prodigy Services Co. (Prodigy), an online service provider, liable for alleged defamatory statements a third party wrote on the defendant's online "bulletin boards."[134] The plaintiff argued Prodigy was a "publisher"[135] much like a newspaper, opening it up to defamation liability, as "one who repeats or otherwise republishes a libel is subject to liability as if he had originally published it."[136] Prodigy argued they instead should be likened to a "book store" or "library," which "may be liable for defamatory statements of others only if they knew or had reason to know of the defamatory statement at issue."[137] The court agreed with the plaintiff, pointing to Prodigy's policy to "continually monitor incoming transmissions and . . . spend time censoring notes" as justification for Prodigy being labeled a "publisher."[138] The court then concluded Prodigy was liable (as a principal) for the defamatory content a third party posted on its site.[139]

Anger at *Stratton Oakmont*'s result prompted legislators to introduce Section 104 (of the Telecommunications Act of 1996),[140]

---

[131] *See* Valerie C. Brannon & Eric N. Holmes, Cong. Rsch. Serv., R46751, *Section 230: An Overview*, 5 (2021) ("Representatives Cox and Wyden offered the provision that would become section 230 as section 104 of House Bill 1555 (1995).").

[132] *Id.* at 7 (quoting 141 CONG. REC. H8470 (daily ed. Aug. 4, 1995) (statement of Rep. Christopher Cox)).

[133] No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

[134] *Id.* at *2.

[135] *Id.*

[136] *Id.* at *3.

[137] *Id.*

[138] *Id.* at *5 ("[i]t is Prodigy's own policies, technology, and staffing decisions which have altered the scenario and mandated the finding that it is a publisher.").

[139] *Id.* at *7.

[140] *See* Brannon & Holmes, *supra* note 131, at 1. Section 230 was enacted as part of the Communications Decency Act (CDA) of 1996 (a common name

which would later be redrafted as Section 230.[141] Many in Congress derided the result of *Stratton Oakmont* as penalizing a private party for helping moderate content on the internet, something many in government preferred private parties do as opposed to federal authorities.[142] The conference report on Section 104 itself singled out *Stratton Oakmont* as a reason for the provision's proposal.[143] Congress intended the provision to "provide 'Good Samaritan' protections from civil liability for providers or users of an 'interactive computer service' for actions to restrict or to enable restriction of access to objectionable online material."[144]

Section 104 was reborn as Section 230 of the Communications Decency Act when Congress enacted the latter as a part of the Telecommunications Act of 1996 (including the Communications Decency Act of 1996).[145] Congress stated in its conference report concerning the Telecommunications Act that Congress intended to "modernize the existing protections against obscene, lewd, indecent, or harassing uses of a telephone."[146] Section 230 further applied this intent to the Internet.[147] By enacting the Communications Decency Act with Section 230 contained therein,

---

for Title V of the Telecommunications Act of 1996), codified as part of the Communications Act of 1934 at 47 U.S.C. § 230. *Id.*

[141] *See* S. REP. NO. 104-230, at 194 (1996); Brannon & Holmes, *supra* note 131, at 1.

[142] *See, e.g.*, 141 CONG. REC. H8470, *supra* note 127 ("[W]e do not wish to have a Federal Computer Commission with an army of bureaucrats regulating the Internet") (statement of Rep. Ron Wyden) ("[t]he Internet is the shining star of the information age, and Government censors must not be allowed to spoil its promise.").

[143] S. REP. NO. 104-230, *supra* note 141, at 194 ("One of the specific purposes of this section is to overrule *Stratton-Oakmont v. Prodigy* and any other similar decisions which have treated such providers and users as publishers or speakers of the content that is not their own because they have restricted access to objectionable material.").

[144] *Id.*

[145] Brannon & Holmes, *supra* note 131.

[146] *Id.* at 1–2 (quoting S. REP. NO. 104-230, at 59 (1995)) ("The decency provisions increase the penalties for obscene, indecent, harassing or other wrongful uses of telecommunications facilities; protect privacy; protect families from uninvited and unwanted cable programming which is unsuitable for children and give cable operators authority to refuse to transmit programs or portions of programs on public or leased access channels which contain obscenity, indecency, or nudity.").

[147] *See* S. REP. NO. 104-230, *supra* note 141.

Congress achieved two aforementioned policy goals summarized as "promot[ing] the free exchange of information and ideas over the Internet and encourage[ing] voluntary monitoring for offensive or obscene material."[148] Subsection (b) of Section 230 likewise elaborates on the several policy goals justifying the provision.[149] Though the Supreme Court later struck down parts of the Communications Decency Act as unconstitutional,[150] Section 230 remains today with few amendments.[151]

The so-called "heart of Section 230"[152] is in subsection (c).[153] Subsection (c)(1) addresses the Congressionally disdained result from *Stratton Oakmont* by specifying that "[n]o . . . interactive computer service[154] shall be treated as the publisher or speaker of any information provided by another information content provider."[155] Subsection (c)(2) ensures that "service providers may not

---

[148] Carafano v. Metrosplash.com, Inc., 339 F.3d 1119, 1122 (9th Cir. 2003).

[149] 47 U.S.C. § 230(b) ("It is the policy of the United States (1) to promote the continued development of the Internet . . . (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 . . . ."); *see* Brown, *supra* note 10, at 462 ("The goal of Congress' decision in enacting section 230 was that Internet companies would be encouraged to develop platforms that relied almost entirely on user-generated content without fear of liability for the content users posted."). Without Section 230, the "'potential liability that would arise from allowing users to freely exchange information with one another, at this [large] scale would have been astronomical,' and could very well have prevented investors from supporting platforms." *Id.*

[150] *See* Reno v. ACLU, 521 U.S. 844, 882 (1997) (holding the Act's bans on transmitting obscene material to be content-based and overbroad restrictions in violation of the First Amendment).

[151] Pub. L. No. 105-277, § 1404, 112 Stat. 2681-739 (1998); Brannon & Holmes, *supra* note 131, at 1.

[152] Brannon & Holmes, *supra* note 131, at 2.

[153] 47 U.S.C. § 230(c).

[154] The statute defines the term of art, "interactive computer service" as follows:

> The term '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.

47 U.S.C. § 230(f).

[155] 47 U.S.C. § 230(c)(1).

be held liable for voluntarily acting to restrict access to objectionable material."[156] Together, these two subparts of subsection (c) give interactive computer service providers both a liability shield and a license to moderate third-party content.[157]

Several court rulings since Section 230's enactment have expanded the provision's liability shield, which was already broad on its face.[158] *Zeran v. America Online, Inc.* was one of the earliest challenges to Section 230's liability shield, and its result demonstrated that courts were willing to err on applying the shield's protection liberally rather than cautiously.[159]

In *Zeran*, an unknown person posted an advertisement on the defendant's site (AOL) showing shirts "celebrating" the 1995 Oklahoma City Bombing and encouraging viewers to call the plaintiff (Zeran) on his home phone to purchase the shirts.[160] Zeran began receiving a large volume of angry phone calls from individuals who had seen the ad online.[161] These calls got worse after news spread to other forms of media.[162] Zeran called AOL, who removed the ad.[163] When individuals put the ad back up several more times, Zeran brought a defamation suit against AOL, claiming that Section 230(c)(1) did not apply because AOL was acting as a "distributor," not a "publisher" (the word contained in Section 230(c)(1)).[164] A "distributor" is likened to a

---

[156] Brannon & Holmes, *supra* note 131 (citing 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 that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).

Id.

[157] Brannon & Homes, *supra* note 131, at 3.

[158] *See* discussion *supra* Section II.A.

[159] *See* Zeran v. Am. Online, Inc., 129 F.3d 327 (4th Cir. 1997).

[160] *Id.* at 329.

[161] *Id.*

[162] *Id.*

[163] *Id.*

[164] *Id.* at 331.

"traditional news vendor or bookseller," but must have some actual knowledge of the defamatory material it distributes to be liable.[165] Zeran followed up with the contention that because he had informed AOL of the first defamatory ad, AOL had sufficient knowledge to be liable as a distributor.[166] The Fourth Circuit rejected Zeran's argument and concluded that holding AOL liable upon notice would be antithetical to the purposes of Section 230(c)(1).[167] The holding in *Zeran* "has informed the approach of a vast number of courts interpreting Section 230(c)(1)" since the late 1990s.[168] Several other case holdings since *Zeran* demonstrate just how difficult it is to break Section 230's liability shield.[169] In such cases, Section 230 shielded social media companies from the plaintiffs' allegations of complicity in acts like terrorism and sex trafficking.[170] Section 230's powerful liability shield on its face, coupled with considerable judicial generosity, has been an invaluable factor in the growth of interactive computer services for over two decades now.[171] This fact is especially true for the interactive computer services that malicious social

---

[165] *Id.* at 333.

[166] *Id.* The liability Zeran argued that AOL was subject to is better known at "notice-based distributor liability." Brannon & Holmes, *supra* note 131, at 11.

[167] *Zeran*, 129 F.3d at 333 ("liability upon notice reinforces service providers' incentives to restrict speech and abstain from self-regulation.").

[168] Brannon & Holmes, *supra* note 131, at 11 ("As one commentator has noted, 'the rule of Zeran [barring distributor liability] has been uniformly applied by every federal circuit court to consider it and by numerous state courts."); Ian C. Ballon, *Zeran v. AOL and Its Inconsistent Legacy*, LAW JOURNAL NEWSLETTERS (Dec. 2017), https://www.law.com/therecorder/2017/11/10/2017ballon-essay/ [https://perma.cc/XP5H-9S8Q].

[169] *See*, *e.g.*, Klayman v. Zuckerberg, 753 F.3d 1354, 1355 (D.C. Cir. 2014) (shielding Facebook from liability for allegedly taking too long to remove "Third Palestinian Intifada" page from its website); Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 18–24 (1st Cir. 2016) (using Section 230(c)(1) to dismiss claim brought against Backpage.com under state and federal anti-sex-trafficking laws); Doe v. MySpace, Inc., 528 F.3d 413, 420 (5th Cir. 2008) (denying claim of negligence liability for site being used as medium for adult to meet and abuse underage girl); Force v. Facebook, Inc., 934 F.3d 53, 65–68 (2d Cir. 2019) (using Section 230(c)(1) to reject claim by terrorism victims that Facebook was liable for content supporting terrorism).

[170] *See*, *e.g.*, *Klayman,* 753 F.3d at 1355; *Backpage.com, LLC*, 817 F.3d at 18–24; *MySpace, Inc.*, 528 F.3d at 420; *Force*, 934 F.3d at 65–68.

[171] *See* Brown, *supra* note 10, at 463.

bot behavior affects the most,[172] and this fact underscores why such services would have an incentive to follow any new legal requirements necessary to keep Section 230's liability protections.[173]

B. Proposals to Amend Section 230

Legislators and commentators from both sides of the aisle have suggested proposals to amend or repeal Section 230 in recent years.[174] The reasons for such proposals revolve almost exclusively around the issue of content moderation (or an alleged lack of it).[175] Many of the issues related to social media content regulation are similar to issues involving malicious social bot regulation.[176] Canvassing recent proposals to modify Section 230 provides a useful starting point when putting together a viable foundation for regulating social bots.[177]

Legislators calling for Section 230 changes are largely divided between those accusing platforms of having biased content moderation policies and those alleging platforms of having too lenient content moderation policies.[178] Despite political divisions between the two groups, a theme of leveraging Section 230 protections in exchange for platform actions underlies many of the

---

[172] *See* Varol et al., *supra* note 31, at 280.

[173] *See* Brown, *supra* note 10, at 463.

[174] *See id.* at 465.

[175] *See id.* ("Section 230's breadth has made it an easy target for those hungry for change in the way social platforms curate—or do not curate—their platforms."); *see also* Michelle Roter, *With Great Power Comes Great Responsibility: Imposing a "Duty to Take Down" Terrorist Incitement on Social Media*, 45 HOFSTRA L. REV. 1379 (2017); Edward Lee, *Moderating Content Moderation: a Framework for Nonpartisanship in Online Governance*, 70 AM. U. L. REV. 913 (2021); Patricia Spiccia, *The Best Things in Life Are Not Free: Why Immunity Under Section 230 of the Communications Decency Act Should Be Earned and Not Freely Given*, 48 VAL. U. L. REV. 369 (2013).

[176] These issues include mandates on social media companies to do something (like disclose a process they use to moderate content (or potentially regulate bots)), as well as types of enforcement mechanisms and First Amendment considerations. *See* Stricke, *supra* note 19, at 845–46, 887–89; Hines, *supra* note 7, at 407–08, 416–24; Brown, *supra* note 10, at 464–65, 480–81, 485–88, 489–94.

[177] *See* Brown, *supra* note 10, at 464.

[178] *See id.*

reform proposals, regardless of their specific goals.[179] Legislators could use the same theme when regulating social bots.[180]

Various legal proposals have emerged, with the goals of supposed "viewpoint neutrality" at the forefront.[181] Former President Donald Trump attempted through a May 2020 Executive Order to limit Section 230's shield in order to force social media companies to adhere to "viewpoint neutrality."[182] The Order tasked the Federal Communications Commission with proposing regulations that clarify a narrow interpretation of Section 230; the Order asked the Federal Trade Commission to enforce social platforms' own conditions and directed the Department of Justice to identify "viewpoint-based speech restrictions" on social platforms that could serve as grounds to reduce government ad-spending on such platforms (as punishment).[183] The Biden administration later revoked this Executive Order,[184] but the Order demonstrated the existence of an appetite for imposing regulation on social media companies in exchange for continued Section 230 protection.[185]

Missouri Senator Josh Hawley has introduced several legislative proposals also intent on creating new rules for social media companies' content moderation practices.[186] His proposed "Ending Support for Internet Censorship Act" would give a social platform Section 230 immunity *only after* it received certification (by supermajority vote) from the Federal Trade Commission that the platform does not engage in biased moderation.[187] Social platforms would have to "prove . . . by clear and convincing evidence that their algorithms . . . are politically neutral."[188]

---

[179] *See id.* at 465, 470.

[180] *See* discussion *infra* Part III.

[181] *See* Brown, *supra* note 10, at 465.

[182] *See id.* at 466. Trump's move was "easily interpreted as retaliatory: it came days after Twitter decided to add a fact check label to two of the President's arguably false tweets about mail-in voting." *Id.*

[183] *See id.*

[184] Exec. Order No. 14029, 86 Fed. Reg. 27025 (May 19, 2021).

[185] *See* Brown, *supra* note 10, at 466.

[186] *See id.* at 467–69.

[187] Zoe Bedell & John Major, *What's Next for Section 230? A Roundup of Proposals*, LAWFARE (July 29, 2020, 9:01 AM), https://www.lawfaremedia.org/article/whats-next-section-230-roundup-proposals [https://perma.cc/62ZB-JRLW].

[188] *Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies* (June 19, 2019) [hereinafter *Senator Hawley Introduces*

They would also have to reapply every two years for immunity.[189] To protect industry competition, these provisions would only apply to companies "with more than 30 million active monthly users in the U.S., more than 300 million active monthly users worldwide, or who have more than $500 million in global annual revenue."[190]

Hawley and fellow Senator Marco Rubio also brought forward the "Limiting Section 230 Immunity to Good Samaritans Act," which would "make a social network's immunity under Section 230 contingent upon the network's contractual commitment to using 'good faith practices' when making content moderation decisions."[191] Other legislators have drafted additional proposals envisioning social platform compliance with new rules in exchange for Section 230 immunity.[192]

A bipartisan bill entitled the "Procedural Accountability and Consumer Transparency (PACT) Act" provides some of the most detailed proposals for Section 230 reform to date.[193] The PACT Act would mandate a "notice-and-takedown regime" for unlawful content, requiring social platforms to remove illegal content "deemed unlawful by a court within 24 hours," or risk losing Section 230 protections.[194] Interactive computer services would also need to "publish an acceptable use policy . . . in a location that is easily acceptable to the user."[195] One of the Act's most consequential provisions is its requirement for a "biannual transparency report" that discloses certain data from social platforms every six months.[196] The disclosures would include "the total number of unique monthly visitors to the [site]," the "number of instances" where illegal content was flagged and the site took action (or did not), a "descriptive summary of the kinds of

---

Legislation], https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies [https://perma.cc/C2JT-C976].

[189] *Id.*

[190] *Id.*

[191] Brown, *supra* note 10, at 468.

[192] *See id.* at 469.

[193] *See* Bedell & Major, *supra* note 187.

[194] *See id.*

[195] PACT Act, S.797, 117th Cong. (2021) [hereinafter PACT Act], https://www.congress.gov/bill/117th-congress/senate-bill/797/text [https://perma.cc/VU93-KW6Q].

[196] *Id.* § 5(d).

tools . . . [used in] enforcing the acceptable use policy," and other miscellaneous information.[197] The bill would also require disclosures about the processing of complaints and content removal.[198]

Combining factors of proposed Section 230 reforms with elements of existing strategies for tackling social bot proliferation would create a regulatory regime that uses legal enforcement to tackle social bots' harmful behaviors.[199]

III. Developing a Broad Regulatory Framework for Social Bots by Including New Requirements for Social Media Companies in Section 230 Reform

Given the array of harms social bots can cause, Congress should require interactive computer services to have a basic framework for monitoring and publicly disclosing bot and bot-like activity in exchange for Section 230's continued protections. The broad starting point here is legislative action revising Section 230 in which an enforcer would be required to further granularize the technical aspects of bot and bot-like activity.[200] Fusing characteristics from several Section 230 reform proposals with existing bot-governance technology helps establish the bases of social media companies' new social bot management requirements.[201]

The first legal requirement draws from social bot monitoring and identification methods that are already used by social media companies and researchers.[202] Interactive computer services would need to have *some* minimum system for policing and classifying social bots. An example would be a system like the one the former Twitter CEO, Parag Agrawal, described.[203] This system reviews "both public and private data (e.g., IP address, phone number, geolocation, cline/browser signatures, what the account does when it's active . . .) to make a determination [of bot-ness] on each account."[204] Other systems use "automated" or

---

[197] *Id.* § 5(d)(2)(A)–(G).
[198] *Id.* § 5(b)–(c).
[199] *See* discussion *infra* Part III.
[200] *See* discussion *infra* Part III.
[201] *See* discussion *infra* Part III.
[202] *See* discussion *supra* Section I.B.
[203] *See* discussion *supra* Section I.B.
[204] Parag Agrawal (@paraga), *supra* note 118.

"investigative" approaches for determining an account's "bot-ness."[205] Almost all social media companies likely already meet such a requirement.[206]

However, a secondary requirement would be that these companies provide an accessible way for users to benefit from and access (at least some of) the data these monitoring systems produce. One form this access could take borrows from the Bot Act and Twitter's automation label feature.[207] Instead of putting the onus on social bot users to self-identify bots (like in the Bot Act), the law could require social media companies to by default display[208] a "score" or "scale" on-site profiles that shows the probability that an account is a social bot or engages in "bot-like" behavior.[209] Though updating the display may now be unfeasible in real time,[210] a default "zero" label combined with a timestamped reporting system where users could submit accounts (their own or others) to the company for a preliminary score is a good starting point to providing accessible data.[211] The social media site could apply the same reporting process to viral links, designating to users the account such links originated from.[212] Such a labelling system, however, might result in instances

---

[205] *See* Beatson et al., *supra* note 21, at 184. "Automated" approaches focus on machine learning that focuses on areas like "an account's network . . . user information . . . user friends and how they interact with the specified account . . . and sentiment analysis." *Id.* at 185. "Investigative" approaches rely more on manual investigations, though still look for similar markers that automated approaches do. *Id.* at 186.

[206] *See id.* at 184.

[207] *See* discussion *supra* Section I.B.

[208] This requirement is similar to and could be fused with the PACT Act's proposed requirement that companies publish "acceptable use policies . . . in a location that is easily acceptable to the user." *See* PACT Act, *supra* note 195.

[209] Avoiding specific speech regulation could avoid First Amendment issues that the Bot Act potentially faces. *See* Hines, *supra* note 7, at 427. Emphasizing that a bot-focused regulation is "concerned with the 'how' of online communication, not the 'what,'" strengthens the argument that a regulation is content-neutral and thus permissible under the First Amendment. *See id.* at 426–27.

[210] *See* Beatson et al., *supra* note 21, at 181.

[211] *See, e.g.*, *id.* at 185.

[212] *See* Jamie Williams, *Cavalier Bot Regulation and the First Amendment's Threat Model*, KNIGHT FIRST AMEND. INST. COLUM. UNIV. (Aug. 21, 2019),

of mislabeling, which could inadvertently suppress speech and require processes for users to appeal unwanted labels.[213] An alternative to explicit "labelling" would be making more raw data on account behavior available (without impermissibly violating privacy) and letting users interpret specific data points for themselves, thereby avoiding most speech concerns.[214] Such data could include points on some of the telltale signs of bot and botnet behavior including: repetitious acts, mimicking, and suspect timing intervals.[215] In any iteration of the proposals above, users' access to pertinent account data would help them better decide whether their interactions are with an authentic account and whether the information they view reflects organic discussions.[216] Such transparency would diminish the negative effects of bot-driven disinformation campaigns and other malicious acts.[217]

Even so, the growing capabilities of social bots and generative AI to mimic human behavior demonstrate that transparency is only one piece in remedying malicious bot behavior.[218] Using data gleaned from the actions of accounts to determine the veracity of such accounts' content is much less helpful when the data shows nothing apparently malicious.[219] The first requirement to have a minimum bot policing system would fill in some of the gaps here, while the subject of Section 230 content moderation is a separate, but necessary, compliment to social bot regulation.[220]

When it comes to enforcement and compliance with the minimum requirements discussed above, lawmakers could take several possible routes.[221] President Trump's short-lived 2020 Executive Order envisioned enforcement duties going to the FTC,

---

https://knightcolumbia.org/content/cavalier-bot-regulation-and-the-first-amendments-threat-model [https://perma.cc/NVK3-SP7J].

[213] *See id.*

[214] *See id.*

[215] *See* discussion *supra* Section I.A.

[216] *See* Andrew Hutchinson, *Twitter Launches New Bot Labels to Identify Bot Accounts In-Stream*, SOCIALMEDIATODAY (Feb. 16, 2022), https://www.socialmediatoday.com/news/twitter-launches-new-bot-labels-to-identify-bot-accounts-in-stream/618971/ [https://perma.cc/SK4V-ERQP].

[217] *See id.*

[218] *See* discussion *supra* Section I.A.

[219] *See* Hindman & Barash, *supra* note 47, at 17.

[220] *See* Brown, *supra* note 10, at 456–58.

[221] *See* discussion *infra* Part III.

as it is the agency that most closely regulates issues like those on social media sites.[222] Due to the FCC's familiarity with Section 230, proximity to social media regulation issues, and its enforcement capabilities, an explicit enforcement delegation to that agency would work here.[223] To ensure continuing adherence, a social media company should require a supermajority of the FTC Board to certify every few years that the company is in compliance with the new policies in order to maintain Section 230 immunity, like the proposal in one of Senator Hawley's bills.[224] To increase transparency even further, the regulation should require disclosures that take a shape similar to those the PACT Act proposes, where social media companies must produce biannual "transparency reports" that in this case would disclose facts about their bot-governance regimes and whatever specifics a federal agency later requires.[225] As with most instances of novel government regulation, concerns exist about potential overregulation.[226] Many of these concerns, however, are more specific to reforming Section 230's liability shield and not to bot regulation.[227] Active user thresholds would do much to remedy such concerns.[228]

To protect market competition and prevent undue barriers to entry, the legislative reform should include site-specific thresholds that excuse compliance unless a social media company meets

---

[222] Brown, *supra* note 10, at 466. The constitutionality of directing the FCC to reinterpret Section 230 was unclear and highly controversial. *See id.*

[223] *See* Devin Coldewey, *Who Regulates Social Media?*, TECHCRUNCH (Oct. 19, 2020), https://techcrunch.com/2020/10/19/who-regulates-social-media/ [https://perma.cc/6XGK-NNUU]. Delegation to the FCC (to directly regulate social media company actions in the proposed way) would need to be explicitly stated in a new statute, as the agency currently lacks an explicit designation. *See id.* Other commentators have suggested tapping the Federal Trade Commission for the job, though that agency regulates matters more general to business itself, rather than matters specific to social media companies. *See id.* Another proposal is to delegate enforcement to an industry level (nongovernmental) "self-regulatory council," which would avoid the bureaucratic headache involved with government regulation but would give amending Section 230 unclear legal significance. *See* Brown, *supra* note 10, at 489.

[224] Bedell & Major, *supra* note 187.

[225] PACT Act, *supra* note 195.

[226] *See* Williams, *supra* note 212.

[227] *See* Brown, *supra* note 10, at 457.

[228] *See* discussion *infra* Part III.

them.[229] As previously mentioned, the Bot Act only applies to sites with "10,000,000 or more unique monthly United States visitors or users for a majority of months during the preceding 12 months,"[230] while one of Senator Hawley's proposals would require companies to have "more than 30 million active monthly users in the U.S., more than 300 million active monthly users worldwide, or . . . more than $500 million in global annual revenue."[231] Only the largest American social media companies with ample revenue would meet the minimum thresholds to fall under these new legal requirements.[232]

Building on the broad legal foundation this Section lays out, a specialized entity tasked with enforcing this foundation will be able to further specify the technical criteria indicating an acceptable monitoring and disclosure system.[233] This Note provides a starting point for further academic debate on legal strategies to tackle malicious social bot usage via Section 230 reform.

CONCLUSION

While Section 230 reform proposals have largely focused on changing the requirements of interactive computer services' content moderation, Congress and regulatory agencies have a chance to remedy the issues associated with malicious social bot behavior via bot-specific Section 230 reform. Bot-specific regulation via Section 230 should start with a framework that makes Section 230's liability protections contingent on interactive computer services having (1) a system in place that polices and classifies social bots and (2) a way for users to access the material data that such a system produces. While these requirements alone cannot solve all the issues associated with malicious bot behavior, they would provide social media users with minimum tools to assess the authenticity of their online interactions.

---

[229] *See* discussion *infra* Part III.

[230] CAL. BUS. & PROF. CODE § 17940(c) (West).

[231] *Senator Hawley Introduces Legislation*, *supra* note 188.

[232] *See id.*

[233] *See* Brown, *supra* note 10, at 493–94 (demonstrating how a government agency can apply enforcement guidelines).