SWORD OR SHIELD? § 230 OF THE CDA, RACIAL CAPITALISM, AND THE RISE OF FAR-RIGHT EXTREMISM ON SOCIAL MEDIA

MARYAM NAZIR

INTRODUCTION

In the early 1990s, less than 1% of the U.S. population had access to the internet.[1] Six Degrees, one of the internet's first prominent social media sites, garnered around 3.5 million users at its peak,[2] with sites like MySpace amassing 75.9 million users about a decade later.[3] Fast forward to 2024, 63.7% of the total global population are social media users,[4] with 7 out of every 10 Americans active on these platforms.[5] In less than a decade, Facebook grew from fewer than six million users to one billion, boasting nearly three billion by early 2023.[6] Social media has altered the very foundations of our political landscape. It exists at the apex of cultural influence, shaping our speech, opinions, and identities and revolutionizing our methods of communication, information sharing, and connectivity.

This abrupt rise in its popularity has also led to more sinister uses of social media—specifically by far-right extremist groups to propagate their ideologies. The Alternative Right, referred to as the "Alt-Right," consists of several "groups and individuals whose core belief is that 'White identity' is under attack."[7] Coined in 2008 by conservative pundit Paul Gottfried,[8] the group exploded in popularity during the 2016 presidential campaign as the subject of various debates and speeches.[9] The group's ideology originally formed on obscure forum platforms like 4chan and Reddit, eventually amassing mainstream popularity on Twitter (now X) and Facebook.[10] The Alt-Right relied heavily on social media platforms to circulate their

[1] SASI Group & Mark Newman, *Internet Use 1990* (2006) [https://perma.cc/E4XB-2QLU].

[2] Ervin Kalemi, *The First Social Media Platform: The History of Social Media*, PUBLER (Aug. 5, 2022), https://publer.io/blog/the-first-social-media-platform [https://perma.cc/E9BY-B2XZ].

[3] Jasper Jackson, *Time Inc Buys What Is Left of Myspace for Its User Data*, THE GUARDIAN (Feb. 11, 2016), https://www.theguardian.com/media/2016/feb/11/time-inc-buys-what-is-left-of-myspace-for-its-user-data [https://perma.cc/4DBF-6R5F].

[4] *Global Social Media Statistics*, DATAREPORTAL (2024), https://datareportal.com/social-media-users [https://perma.cc/7ADA-PMQF]; *Social Media Fact Sheet*, PEW RSCH. CTR. (Jan. 31, 2024), https://www.pewresearch.org/internet/fact-sheet/social-media [https://perma.cc/L7NF-L2DN].

[5] Tim Stumbles, *Facebook History Timeline*, OFF. TIMELINE (Feb. 1, 2018), https://www.officetimeline.com/blog/facebook-history-timeline [https://perma.cc/7NCD-LXA7].

[6] *Facebook Users, Stats, Data & Trends*, DATAREPORTAL (May 11, 2023), https://datareportal.com/essential-facebook-stats [https://perma.cc/CXW8-9HCR].

[7] *Alt-Right*, S. POVERTY L. CTR., https://www.splcenter.org/fighting-hate/extremist-files/ideology/alt-right [https://perma.cc/HLR9-AQ83].

[8] Gabriel Raeburn, *The Alt-Right and Social Media*, HERBERT D. KATZ CTR. (Nov. 30, 2020), https://katz.sas.upenn.edu/resources/blog/alt-right-and-social-media [perma.cc/2X77-XP3V].

[9] Josh Vandiver, *The Radical Roots of the Alt-Right*, GALE, https://www.gale.com/intl/essays/josh-vandiver-radical-roots-alt-right [https://perma.cc/KXK7-J8CF].

[10] Raeburn, *supra* note 8.

ideology,[11] with many attributing the popularity of meme culture at the time as the medium through which Alt-Right ideology repackaged itself into easily digestible and shareable graphics.[12] At the height of the movement, Alt-Right viewpoints gained massive virality on social media platforms,[13] allowing the movement's popular figures to catapult their views and identify other like-minded individuals.[14]

There are myriad features that contribute to the vast proliferation of Alt-Right ideology on social media—its culture of anonymity, its propensity for virality, or perhaps platforms' lack of incentives to effectively moderate the content users publish due to the immunity provision of § 230 of the Communications Decency Act ("CDA").[15]

The far-reaching immunity afforded to platforms by § 230 of the CDA allows social media sites to relinquish oversight when it comes to moderating content that appears on their sites, shielding internet service providers from liability for what users post.[16] In practice, this has led to an online environment where misinformation and hate speech run rampant,[17] where individuals conspired to riot at the U.S. Capitol, and where terrorist organizations recruited new members.[18] The CDA's § 230 immunity provision is criticized by both political parties, but for different reasons: Democrats scrutinize platforms for the lack of effective content moderation,[19] while Republicans fault platforms for over-moderation of content,[20] introducing laws like Texas House Bill 20[21] to afford individuals the ability to sue over platform decisions regarding bans and censorship.[22]

The immunity provision of § 230 of the CDA, written more than half a decade before the first social media site took off,[23] protects platforms from legal liability for content posted on their sites by users by shielding platforms from liability if they choose to moderate content, but never requiring them

[11] *Id.*

[12] Jordan McSwiney, Michael Vaughan, Annett Heft & Matthias Hoffmann, *Memes in Far-Right Digital Visual Culture*, GLOB. NETWORK ON EXTREMISM & TECH. (Aug. 20, 2021), https://gnet-research.org/2021/08/20/memes-in-far-right-digital-visual-culture [https://perma.cc/D283-RWDU].

[13] Aja Romano, *The Year Social Media Changed Everything*, VOX (Dec. 31, 2016), https://www.vox.com/2016/12/31/13869676/social-media-influence-alt-right [https://perma.cc/HU4M-VYGW].

[14] Id.

[15] 47 U.S.C. § 230(c)(1).

[16] *Section 230*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230 [https://perma.cc/9WTE-B3G3].

[17] *See generally* Koustubh "K.J." Bagchi, Elizabeth Banker & Ife Ogunleye, *Disrupting the Narrative: Diving Deeper into Section 230 Political Discourse*, 2023 PEPP. L. Rev. 1, 6–12 (Mar. 15, 2023).

[18] *Id.* at 3.

[19] *See, e.g.*, Shannon Bond, *Democrats Want to Hold Social Media Companies Responsible for Health Misinformation*, NPR (July 22, 2021), https://www.npr.org/2021/07/22/1019346177/democrats-want-to-hold-social-media-companies-responsible-for-health-misinformat [https://perma.cc/3LG8-5AG7].

[20] *See* Danielle Draper & Sabine Neschke, *Republican Midterm Agenda: Section 230, Censorship, and Big Tech*, BIPARTISAN POL'Y CTR. (Oct. 18, 2022), https://bipartisanpolicy.org/blog/republican-midterm-agenda-section-230 [https://perma.cc/T7CE-4UBF].

[21] Jesus Vidales, *Texas Social Media "Censorship" Law Goes into Effect After Federal Court Lifts Block*, TEX. TRIB. (Sept. 16, 2022), https://www.texastribune.org/2022/09/16/texas-social-media-law [https://perma.cc/7JZ7-TFYC].

[22] Ann E. Marimow & Cat Zakrzewski, *U.S. Supreme Court to Rule on Texas Law Limiting Social Media Content Removal*, TEX. TRIB. (Sept. 29, 2023), https://www.texastribune.org/2023/09/29/supreme-court-texas-social-media-law [https://perma.cc/Z39B-DKAD].

[23] 47 U.S.C. § 230(c)(1).

to do so.[24] This provision carries value—it is the reason social media sites have not drowned in the face of costly litigation whenever a user posts something that could be subject to civil liability.[25] However, as social media sites transformed from humble startups into omniscient media giants, the problems of this immunity provision became glaringly obvious.

This Note will explore § 230 of the CDA, particularly the "Good Samaritan" provision that protects social media platforms from liability arising from the content their users posts. Part I will highlight how racial capitalism—the theory that racial exploitation is a fundamental aspect of our capitalist system—interacts with the formation and continuation of these extremist groups online, providing them with the economic viability to continue spreading messages of hate.[26] Part II will explore the Good Samaritan provision of § 230 of the CDA, the initial motivations for enacting the immunity provision of § 230 of the CDA, and how this provision has allowed platforms to take a passive role in mitigating hateful content peddled on their sites.

Part III examines potential solutions to the issues arising from the immunity provision, focusing first on how platform transparency, despite its value, is ultimately ineffective at accomplishing a meaningful solution. It will then pivot to a discussion of the importance of independent research in achieving platform transparency and accountability, recognizing the legal obstacles faced by independent researchers that prevent substantive investigation into the practices of social media sites. Lastly, this Note will propose tangible steps to protect and encourage independent research into social media platforms—specifically, the enactment of a safe harbor provision within § 230 of the CDA that will allow independent researchers to conduct studies while also balancing concerns about protecting user data. This section will emphasize how the safe harbor provision can act as the incentive platforms need for the critical assessment and overhaul of current moderation practices, placing users on a more equal footing with the platforms that host them.

I. RACIAL CAPITALISM AND ONLINE EXTREMISM

A. THE ALT-RIGHT AS AN ECONOMIC MOVEMENT

The Alt-Right is infamous for its extreme right-wing ideologies and resurgence of American nationalist thought.[27] What often goes unnoticed, however, is the economic foundation that funds many of the movement's prominent figures. We often think of the Alt-Right solely as an extremist social movement, but it is important to recognize it also as an economic system—one in which its subscribers can profit from peddling its ideology to the larger public.

[24] Michael D. Smith & Marshall W. Van Alstyne, *It's Time to Update Section 230*, HARV. BUS. REV. (Aug. 12, 2021), https://hbr.org/2021/08/its-time-to-update-section-230 [https://perma.cc/6RTW-DGXU].

[25] Jason Kelley, *Section 230 Is Good, Actually*, ELEC. FRONTIER FOUND. (Dec. 3, 2020), https://www.eff.org/deeplinks/2020/12/section-230-good-actually [https://perma.cc/QK7B-6UJF].

[26] Raeburn, *supra* note 8.

[27] Vandivier, *supra* note 9.

Those spearheading the movement have created an economy premised upon the bigoted rhetoric of the Alt-Right, and in this economy, you can find dating apps, social media platforms, and even online marketplaces all dedicated to the promotion of Alt-Right thought and its subscribers.[28] These funds operate not only to line the pockets of Alt-Right pundits but also to ensure the viability of Alt-Right thought and right-wing extremism in American politics. Patriot Mobile, a right-wing Christian conservative wireless cellphone company, raised over six hundred thousand dollars to elect eleven candidates, winning all the seats of four school boards in Texas.[29] The acquisition of these school board seats is quintessential for both the political sustainability of right-wing ideology and the recruitment of members from newer generations. Following the elections, one school board implemented a policy calling for the removal of nearly forty books from library shelves, and another implemented stringent policies restricting the ways teachers can discuss the topics of race and gender.

The immense profitability of Alt-Right ideology is intriguing, considering White radicalism is not even close to being a popular social movement in the United States. A study on the demography of the Alt-Right showed that only 6% of White American survey respondents agreed with all three attitudes of White identity politics, suggesting the actual amount of individuals who subscribe to Alt-Right thought is particularly low.[30] The question of how such a statistically insignificant portion of the United States population can fund an industry coined as a "parallel economy"[31] to that of the general economy may be answered using the critical race theory of racial capitalism.

B. THE THEORY OF RACIAL CAPITALISM

Cedric Robinson coined the term "racial capitalism" in *Black Marxism: The Making of the Black Radical Tradition* to refer to the idea that "racialized exploitation and capital accumulation are mutually constitutive."[32] Robinson posited that capitalism and racism are two mutually dependent ideas that rely on the success and continuation of the other.[33] The subjugation of individuals in our capitalist system parallels that of individuals subjugated in our racialized society, and this is no coincidence. Robinson notes how these two oppressive forces often interplay when impacting certain demographics, emphasizing how racial oppression ensures the structural integrity of our capitalist system.[34] The theory of racial capitalism argues that race is used as a basis to justify economic exploitation.[35] Jodi Melamed writes in *Racial*

[28] *Conservative Americans are Building a Parallel Economy*, THE ECONOMIST (June 1, 2023), https://www.economist.com/united-states/2023/06/01/conservative-americans-are-building-a-parallel-economy [https://perma.cc/SD9A-5MF8].

[29] Mike Hixenbaugh, *How a Far-Right, Christian Cellphone Company 'Took Over' Four Texas School Boards*, NBC NEWS (Aug. 25, 2022), https://www.nbcnews.com/news/us-news/-christian-cell-company-patriot-mobile-took-four-texas-school-boards-rcna44583 [https://perma.cc/T8KB-BBXW].

[30] George Hawley, *The Demography of the Alt-Right*, INST. FOR FAM. STUD. (Aug. 9, 2018), https://ifstudies.org/blog/the-demography-of-the-alt-right [https://perma.cc/LM68-P6BA].

[31] *See generally* THE ECONOMIST, *supra* note 28.

[32] Whitney N. Laster Pirtle, Racial Capitalism: A Fundamental Cause of Novel Coronavirus (COVID-19) Pandemic Inequities in the United States, 47 HEALTH EDUC. & BEHAV. 504, 504 (2020).

[33] *Id.*

[34] Cedric J. Robinson, Black Marxism: The Making of the Black Radical Tradition 315–16 (1983).

[35] Jodi Melamed, *Racial Capitalism*, 1 CRITICAL ETHNIC STUD. 76, 77 (2015).

Capitalism that "[c]apital can only be capital when it is accumulating, and it can only accumulate by producing and moving through relations of severe inequality among human groups."[36] We can observe the effects of racial capitalism in real time—simply look at the myriad of low-income, non-White communities in "Third World" countries being exploited for resources by Western empires. For Western countries to be shielded from the consequences of capitalism's most extreme failures, Third World countries must bear the brunt, consistently boasting the highest rates of child labor,[37] extreme poverty,[38] and climate change-induced catastrophes.[39]

Racial capitalism is premised on the idea that "whiteness" carries inherent value that is extracted through the oppression of non-White identities.[40] Cheryl Harris, in *Whiteness as Property*, theorized that America's legal and economic structures shaped whiteness as a racially privileged identity that retains its inherent value through the subjugation of non-White people.[41] The value of whiteness was codified in law, dictating who was allowed to vote, attend schools, buy property, work, and own businesses.[42] These status-based rights, dependent on a White identity, allowed for the accumulation of capital at the expense of non-White people for years.[43] This accumulation of capital became inextricable from whiteness—transforming a once abstract identity trait into a vehicle for attaining economic power.[44] It is virtually impossible to separate whiteness from the economic privileges afforded to it, and these economic privileges would cease to exist were it not for the economic disadvantages faced by people of color.[45]

By assigning value to whiteness, racial capitalism provides the Alt-Right with a basis for the proliferation of its ideologies. As we enter a new anthropological epoch in which non-White people are predicted to become the majority in the United States,[46] the Alt-Right has popularized its ideology by emphasizing that the loss of majority status equates to the devaluation of whiteness and the loss of power associated with it.[47] To the Alt-Right, becoming a minority demographic means no longer being able to subjugate non-White identities—and without subjugation, the value of whiteness loses

[36] Id.

[37] *Child Labour*, UNICEF (July 2024), https://data.unicef.org/topic/child-protection/child-labour [https://perma.cc/5WX4-9PLV].

[38] *Global Extreme Poverty*, QERY (Apr. 2024), https://qery.no/global-poverty [https://perma.cc/7JJA-3W58].

[39] Jihad Azour & Abebe Aemro Selassie, *Africa's Fragile States are Greatest Climate Change Casualties*, IMF BLOG (Aug. 30, 2023), https://www.imf.org/en/Blogs/Articles/2023/08/30/africas-fragile-states-are-greatest-climate-change-casualties [https://perma.cc/RC7R-UMXY].

[40] Melamed, *supra* note 35.

[41] *See generally* Cheryl I. Harris, *Whiteness as Property*, 106 HARV. L. REV. 1709 (1993).

[42] *Id.* at 1728.

[43] Steve Maas, *Exploring the 160 Years of the Black-White Wealth Gap*, NAT'L BUREAU OF ECON. RSCH.: THE DIGEST (Aug. 1, 2022), https://www.nber.org/digest/202208/exploring-160-years-black-white-wealth-gap [https://perma.cc/5ZLJ-DPTE].

[44] Harris, *supra* note 41, at 1718.

[45] *Id*. at 1721.

[46] William H. Frey, *The Nation is Diversifying Even Faster Than Predicted, According to New Census Data*, BROOKINGS (July 1, 2020), https://www.brookings.edu/articles/new-census-data-shows-the-nation-is-diversifying-even-faster-than-predicted [https://perma.cc/792F-BAG9].

[47] Mark R. Reiff, *How Important is White Fear?*, AEON (June 28, 2021), https://aeon.co/essays/the-demographic-shift-isnt-driving-white-people-to-the-right [https://perma.cc/B7UV-5VYM].

its economic basis.[48] It is precisely the protection of this value that motivates the Alt-Right movement,[49] accomplished through the resurgence and widespread use of racist rhetoric designed to reinforce both the subjugated position of minorities and White identity as the status quo.[50]

C. THE ECONOMIC SUCCESS OF THE ALT-RIGHT

The economic foundation of the Alt-Right, explained by racial capitalism, is evident in the recent mainstream success of many Alt-Right pundits. Alex Jones, for example, host of the radio show *InfoWars* and known supporter of White nationalism,[51] accrued a net worth of $270 million from his various companies premised on the propagation of his Alt-Right ideologies.[52] Richard Spencer, a prominent figurehead of the Alt-Right movement, had a net worth of $185 million as of mid-2018.[53] Although most advertisers shy away from outwardly endorsing Alt-Right viewpoints or figureheads, they are keenly aware of the amount of engagement these viewpoints garner and target social media users based on the knowledge that these viewpoints will continue to attract engagement due to weak content moderation policies.[54] Researchers from New York University's Center for Cybersecurity found that far-right content generated the highest percentage of engagement among all political posts and far-right accounts notorious for spreading misinformation are "more successful than other kinds of accounts" at acquiring user engagement.[55]

---

[48] Harris, *supra* note 41.

[49] Vandiver, *supra* note 9.

[50] Monica Duffy Toft, *White Right? How Demographics is Changing US Politics*, THE CONVERSATION (Jan. 7, 2019), https://theconversation.com/white-right-how-demographics-is-changing-us-politics-107872 [https://perma.cc/363X-CD2J].

[51] *Alex Jones*, WIKIPEDIA, https://en.wikipedia.org/w/index.php?title=Alex_Jones&oldid=1189591634 [https://perma.cc/7MK5-PQ27].

[52] Lisa Backus, *Alex Jones' Net Worth as High as $270M, Expert Says at Trial, Before Sandy Hook Parents Awarded $45.2M*, NEWS TIMES (Aug. 7, 2022), https://www.newstimes.com/news/article/Alex-Jones-net-worth-Sandy-Hook-damages-17352929.php [https://perma.cc/7DYA-7J3L].

[53] Daniel Wanburg, *Who is White Supremacist Richard Spencer? His Bio: Wife, Net Worth, Parents, Height, Children*, BIOGRAPHY TRIB. (June 28, 2024), https://biographytribune.com/who-is-white-supremacist-richard-spencer-his-bio-wife-net-worth-parents-height-children [https://perma.cc/8GGU-6634].

[54] *See generally Social Media Platforms are Monetizing Searches for Hate Groups and Extremists*, ANTI-DEFAMATION LEAGUE (Sept. 27, 2023), https://www.adl.org/resources/press-release/social-media-platforms-are-monetizing-searches-hate-groups-and-extremists [https://perma.cc/J5TL-R9JX].

[55] *See generally* Michel Martin & Will Jarvis, *Far-Right Misinformation is Thriving on Facebook. A New Study Shows Just How Much*, NPR (Mar. 6, 2021), https://www.npr.org/2021/03/06/974394783/far-right-misinformation-is-thriving-on-facebook-a-new-study-shows-just-how-much [https://perma.cc/HXH9-MAFN]; Laura Edelson, Minh-Kha Nguyen, Ian Goldstein, Oana Goga, Tobias Lauinger & Damon McCoy, *Far-Right News Sources on Facebook More Engaging*, MEDIUM (Mar. 3, 2021), https://medium.com/cybersecurity-for-democracy/far-right-news-sources-on-facebook-more-engaging-e04a01efae90 [https://perma.cc/W64M-UMUY].

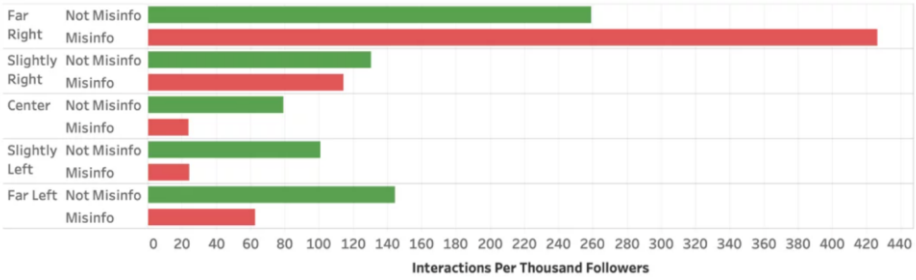


Fig. 1. Research from New York University's Center for Cybersecurity analyzing eight million posts from news and information sources.[56]

This propensity for user engagement does two things: it creates an environment where platforms allow advertisers to target their ads to users who subscribe to alt-right ideologies,[57] and it provides Alt-Right creators the ability to fund their bigotry through the amassment of a staunch fan base.[58] In either case, the profit extracted depends on the spewing of racist ideology that actively harms minority communities.[59] In 2018, *The Intercept* created a paid promotion specifically targeting users who expressed an interest in pages related to "white genocide conspiracy theory," which was approved by Facebook's advertising division.[60] A year earlier, ProPublica had a Facebook promotion approved that targeted users who expressed an interest in topics like "Jew hater" or "How to burn Jews."[61] Though neither platform nor advertiser is directly lining the pockets of these Alt-Right figureheads, their profit is chained to the continued engagement of extremist users and the proliferation of their bigoted ideologies to even more users.

When it comes to profit creation for Alt-Right creators themselves, most are self-funded.[62] Because of companies' hesitancy to outwardly monetize the content of Alt-Right creators, these individuals must rely on the creation of a loyal fan base to finance their various ventures.[63]

The talking points of these figures serve to recruit a high number of devoted, like-minded fans, further boosting their own engagement and catapulting Alt-Right ideology into the mainstream. These viewpoints appeal

[56] Martin & Jarvis, *supra* note 55.

[57] Jeremy B. Merrill, *Targeting Trump Fans, QAnon Ad Slips Through Facebook's Filters*, THE MARKUP (Nov. 17, 2020), https://themarkup.org/election-2020/2020/11/17/targeting-trump-fans-qanon-ad-slips-through-facebooks-filters [https://perma.cc/QR5Q-5J2E].

[58] See, e.g., id.

[59] *From Alt Right to Alt Lite: Naming the Hate*, ANTI-DEFAMATION LEAGUE (July 12, 2017), https://www.adl.org/resources/backgrounder/alt-right-alt-lite-naming-hate [https://perma.cc/54L5-54JD].

[60] Sam Biddle, *Facebook Allowed Advertisers to Target Users Interested in "White Genocide" – Even in Wake of Pittsburgh Massacre*, THE INTERCEPT (Nov. 2, 2018), https://theintercept.com/2018/11/02/facebook-ads-white-supremacy-pittsburgh-shooting [https://perma.cc/F68G-TUM2].

[61] Julia Angwin, Madeleine Varner & Ariana Tobin, *Facebook Enabled Advertisers to Reach 'Jew Haters'*, PROPUBLICA (Sept. 14, 2017), https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters [https://perma.cc/6MQ5-YSDS].

[62] *Funding Hate: How White Supremacists Raise Their Money*, ANTI-DEFAMATION LEAGUE (Dec. 1, 2017), https://www.adl.org/resources/report/funding-hate-how-white-supremacists-raise-their-money [https://perma.cc/6MD5-TN3H].

[63] Id.

to the identities of individuals most susceptible to Alt-Right ideology. Journalist Robert Evans collected comments and chat logs from seventy-five fascist activities to study how fascist activists were radicalized, uncovering that the vast majority of White nationalist and fascist organizations were overwhelmingly male.[64] Analysts from the Center of Extremism identified two hundred individuals from over thirty-five states who attended a White supremacist rally and concluded that the overwhelming majority of attendees were young males.[65] Figure 2 illustrates that these individuals are also less likely to have a college degree and more likely to earn less than $75,000 per year.[66]

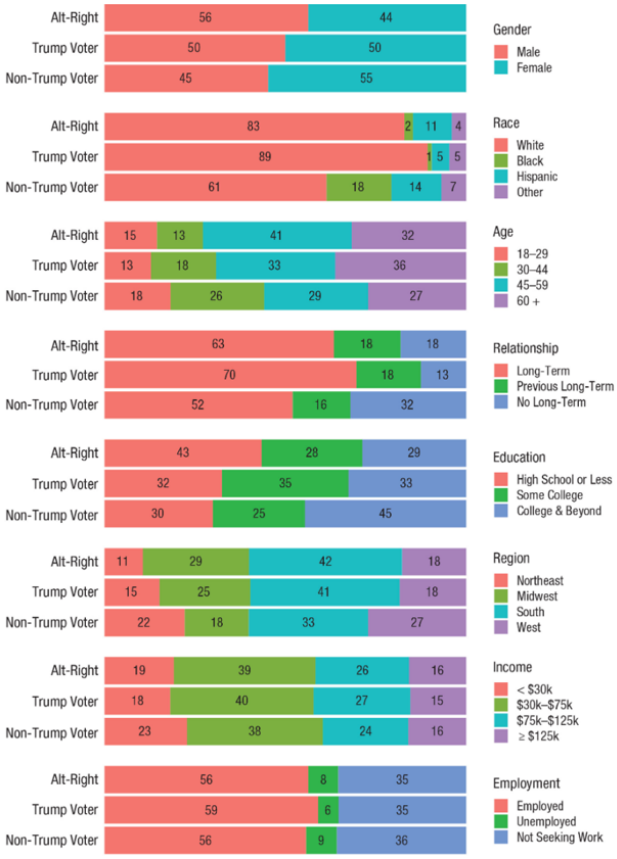


Fig. 2. Study surveying the demographic breakdown of the Alt-Right, Trump voting, and non-Trump voting samples.[67]

[64] Robert Evans, *From Memes to Infowars: How 75 Fascist Activists Were "Red-Pilled"*, BELLINGCAT (Oct. 11, 2018), https://www.bellingcat.com/news/americas/2018/10/11/memes-infowars-75-fascist-activists-red-pilled [https://perma.cc/HWN3-TQVF].

[65] *Have Hate, Will Travel: The Demographics of Unite the Right*, ANTI-DEFAMATION LEAGUE (Oct. 8, 2017), https://www.adl.org/resources/blog/have-hate-will-travel-demographics-unite-right [https://perma.cc/NNY5-LD9G].

[66] Patrick S. Forscher & Nour S. Kteily, *A Psychological Profile of the Alt-Right*, PERSP. PSYCH. SCI. 90, 102 (2020).

[67] Id.

It is no wonder that the majority of talking points from prominent Alt-Right figures like Alex Jones focus on topics such as xenophobia, global elitism, and misogyny.[68] Alt-Right figureheads capitalize on aspects of their audience's identities closely intertwined with power and value, formulating their viewpoints around perceived attacks on these identities.[69] For the most susceptible victims of Alt-Right ideology, these identity aspects boil down to being a White, heterosexual male—aspects that hold value due to the subjugation of other groups.[70] These value-laden identities, reinforced by racial capitalism, are preyed upon by both Alt-Right figures and platforms alike to garner profit—and it works.

II. THE GOOD SAMARITAN PROVISION OF CDA §230

Section 230 of the CDA works as a legislative backbone, reinforcing the proliferation of Alt-Right ideology and the passivity of platforms in adequately moderating extremist content. Section 230(c), specifically the "Good Samaritan" provision, shields internet service providers from liability for the moderation of the content their users post—allowing internet service providers to take a passive role in overseeing user content.[71] Written in 1996, the CDA has failed to keep up with the evolving influence that major social media sites have on our political and social landscape.[72] The consequences of this provision extend far beyond our screens—numerous terrorist events have been linked to hate speech manifestos posted on social media platforms. The El Paso, Christchurch, and Poway Synagogue mass shootings all occurred within a six-month period in 2019 and involved hate speech manifestos posted on prominent social media sites.[73]

With platforms' far-reaching immunity in regard to the content that their users post, there exists no adequate incentive for platforms to effectively crack down on instances of hate speech, especially considering how racial capitalism renders these instances immensely profitable.

A. HISTORY

Section 230 of the CDA was drafted in response to two cases: *Cubby Inc. v. CompuServe, Inc.* and *Stratton Oakmont, Inc. v. Prodigy Services Co*.[74] Both cases involved defamatory comments posted on the sites and were interpreted quite differently. While the court in *Cubby Inc.* held that CompuServe would not be held liable because the site did not moderate the

[68] *Alex Jones*, S. POVERTY L. CTR., https://www.splcenter.org/fighting-hate/extremist-files/individual/alex-jones [https://perma.cc/NF4Z-CENH].

[69] *See generally Alt-Right, supra* note 7.

[70] *See generally* Harris, *supra* note 41, at 1718, 1721.

[71] *Section 230*, WIKIPEDIA, https://en.wikipedia.org/w/index.php?title=Section_230&oldid=1189080618 [https://perma.cc/77RT-SRZV].

[72] *Id.*

[73] Drew Harwell, *Three Mass Shootings This Year Began with a Hateful Screed on 8chan. Its Founder Calls it a Terrorist Refuge in Plain Sight*, WASH. POST (Aug. 4, 2019), https://www.washingtonpost.com/technology/2019/08/04/three-mass-shootings-this-year-began-with-hateful-screed-chan-its-founder-calls-it-terrorist-refuge-plain-sight [https://perma.cc/ST36-J4YD].

[74] *Section 230: Legislative History*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230/legislative-history [https://perma.cc/E36X-F3XE].

content on its forums, the court in *Stratton* determined that because Prodigy moderated its content, it should be liable for all of its users' posts.[75] These conflicting decisions, if left without further legislative guidance, would essentially disincentivize websites from moderating content for fear of potential liability and create an environment where virtually no websites would engage in content moderation. Fearful of the effect this decision would have on the economic viability of internet sites at the time, legislators introduced § 230 as an amendment to the CDA to ensure internet sites hosting content posted by users would not be "treated as the publisher or speaker of any information provided by another information content provider."[76] The amendment shielded websites from liability regarding decisions made in attempts to moderate or censor content posted by their users.[77] Section 230 of the CDA was essential to the survival of internet sites like Wikipedia, WordPress, and Yelp[78]—without it, these websites would be significantly more costly to operate and subject to near-constant litigation.

B. THE TEXT OF § 230 OF THE CDA

The "Good Samaritan" provision of § 230 of the CDA, which provides internet service providers with legal immunity, is broken up into two primary sections: (c)(1) and (c)(2). Section (c)(1) states that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[79] The term "information content provider" encompasses "any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service."[80] The text of this provision contains ambiguities that affect how this section is interpreted—namely, what degree of involvement by a platform in third-party content is required for that platform to be "treated as the publisher or speaker"[81] of that content. Construed broadly, this provision could prohibit all lawsuits against platforms involving harm from user-generated content. If interpreted narrowly, however, platforms could face liability if, for example, they knowingly host harmful content or promote it in some way.[82]

The first case to interpret this provision of § 230 was *Zeran v. AOL*, which was decided by the Fourth Circuit Court of Appeals in 1997. Kenneth Zeran brought the action against AOL, an online messaging site, for unreasonably delaying the removal of "defamatory messages posted by an unidentified third party," among other claims.[83] Zeran argued that § 230(c)(1) does not afford immunity to service providers "who possess notice of defamatory material posted through their services,"[84] thus leaving

[75] Id.
[76] 47 U.S.C. § 230(c)(1).
[77] Id.
[78] *Section 230 Successes*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230/successes [https://perma.cc/8D5T-PS22].
[79] 47 U.S.C. § 230(c)(1).
[80] 47 U.S.C. § 230(f)(3).
[81] 47 U.S.C. § 230(c)(1).
[82] Alan Z. Rozenshtein, *Interpreting the Ambiguities of Section 230*, BROOKINGS (Oct. 26, 2023), https://www.brookings.edu/articles/interpreting-the-ambiguities-of-section-230 [https://perma.cc/AB46-MGX8].
[83] Zeran v. Am. Online, Inc., 129 F.3d 327, 328 (4th Cir. 1997).
[84] Id.

AOL liable because he had provided the site with sufficient notice of the defamatory messages.[85] Zeran claimed that once internet providers acquire knowledge of defamatory statements, their status becomes more akin to that of *distributors* rather than publishers of content, thereby making them subject to liability because distributors are not protected by the immunity provisions of § 230.[86]

The court held that § 230(c)(1) does not necessarily distinguish distributors as separate entities from publishers; rather, it construes distributors as encompassed under the larger umbrella of the publisher category, thus protected by § 230 immunity.[87] The court defines a publisher as including those who "intentionally mak[e] the contents available to others, sometimes without knowing all of the contents,"[88] thereby embracing a broad interpretation of § 230(c)(1). Courts in the years since have adopted this broad interpretation of § 230, and it has essentially been the only means of interpretation used.[89] The ambiguity of § 230, leading to either a broad or narrow interpretation, is not being properly grappled with in courts; lower courts are instead opting to follow the approach established in *Zeran* nearly three decades ago, which presents concerns, especially considering the stark landscape shift that has occurred on the internet in the past few years alone.

Section 230(c)(2) grants immunity to providers and users of an interactive computer service who decide to moderate content they deem "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable."[90] It is important to note that the text of § 230(c)(2) leaves the question of objectivity up to the discretion of platforms themselves, rather than requiring an inquiry into whether the content is truly objectionable.[91] The question of what is encompassed by the "otherwise objectionable" portion of § 230(c)(2) has been debated in court, with the Ninth Circuit opting to interpret this category as more general, rather than as a catch-all category for content that does not necessarily fit into the previously enumerated categories.[92] The ambiguity in the text, though seemingly trivial, has implications that may grant companies far too much deference in deciding what constitutes "otherwise objectionable" content, such as political content. However, many argue that this interpretation does not align with Congress's intentions and that "otherwise objectionable" refers to "material that Congress itself found objectionable in the Communications Decency Act of 1996," rather than a category companies may use to immunize any actions they deem objectionable.[93]

Section 230 presents numerous ambiguities that have yet to be resolved. What is clear from courts' attempts to clarify these ambiguities is that major leeway is given to platforms under this immunity provision. In most cases,

[85] *Id.* at 331.
[86] Id.
[87] *Id.* at 332.
[88] Id.
[89] Rozenshtein, *supra* note 82.
[90] Adam Candeub & Eugene Volokh, *Interpreting 47 U.S.C. § 230(c)(2)*, 1 J. FREE SPEECH L. 175, 176 (2021).
[91] 47 U.S.C. § 230(c)(2).
[92] Enigma Software Grp. USA, v. Malwarebytes, Inc., 946 F.3d 1040, 1051–52 (9th Cir. 2019).
[93] Candeub & Volokh, *supra* note 90, at 189.

these ambiguities are interpreted in a way that affords internet service providers the greatest chance of immunization. Initially, this was important to the survival and continuation of the internet, especially social media sites; however, as their power has grown, it may be prudent to revisit these ambiguities with a different perspective.

C. FROM ONLINE HATE SPEECH TO REAL WORLD HARMS

In the years since § 230 was drafted, social media sites have mutated into corporate giants—providing news, dictating culture, influencing beliefs, and impacting our political institutions.[94] This is a far cry from what the original drafters of § 230 envisioned the internet would become, and the flaws of this amendment have manifested in violent and harmful ways. In 2019, three mass shootings began with hate speech manifestos posted on the social media forum 8chan.[95] The message board had become home to several White nationalist extremist groups,[96] and once these terrorist acts were carried out, their manifestos became the subject of far-reaching virality.[97]

8chan and similar forum sites hosting extremist user content are shielded by § 230 of the CDA, a main reason why many of these sites continue to host extremist content to this day.[98] Though 8chan was taken down due to public outcry, its counterparts, like 4chan, still persevere.[99] 4chan, a forum site similar to 8chan, has been the origin of some of the largest instances of cyberbullying, conspiracy theories, and hoaxes on the internet.[100] 4chan's terms of service are extremely limited—users are banned from violating U.S. law, posting other people's personal information, impersonating site administrators, and using bots.[101] When it comes to content moderation, 4chan is not transparent about its practices, claiming "there is no way for an end user to accurately judge the amount of moderation taking place at any given point in time" and "no way of knowing which [threads] have been removed by the moderation team."[102]

---

[94] *See generally* Esteban Ortiz-Ospina, *The Rise of Social Media*, OUR WORLD IN DATA (Sept. 18, 2019), https://ourworldindata.org/rise-of-social-media [https://perma.cc/92ZQ-EYJ9]; Giles Crouch, *The Effects of Social Media and Global Culture*, MEDIUM (Mar. 13, 2022), https://gilescrouch.medium.com/the-effects-of-social-media-and-global-culture-febd6858b6c [https://perma.cc/8K6A-93QU]; Elizabeth Germino, *How Social Media Has Changed the U.S. Congress*, CBS NEWS: 60 MINUTES OVERTIME (Nov. 6, 2022), https://www.cbsnews.com/news/social-media-u-s-congress-60-minutes-2022-11-06 [https://perma.cc/97AN-7RXH].

[95] Harwell, *supra* note 73.

[96] Julia Carrie Wong, *8chan: The Far-Right Website Linked to the Rise in Hate Crimes*, THE GUARDIAN (Aug. 4, 2019), https://www.theguardian.com/technology/2019/aug/04/mass-shootings-el-paso-texas-dayton-ohio-8chan-far-right-website [https://perma.cc/MW57-SJFA].

[97] Thomas Plant, *America Needs a New Protocol for Handling Shooter Manifestos*, FOUND. FOR DEF. OF DEMOCRACIES (May 19, 2023), https://www.fdd.org/analysis/2023/05/19/america-needs-a-new-protocol-for-handling-shooter-manifestos [https://perma.cc/NXJ3-USEP].

[98] *See 8chan*, WIKIPEDIA, https://en.wikipedia.org/w/index.php?title=8chan&oldid=1188720198 [https://perma.cc/9NGU-NCJ3].

[99] Caitlin Dewey, *Absolutely Everything You Need to Know to Understand 4chan, the Internet's Own Bogeyman*, WASH. POST (Sept. 26, 2014), https://www.washingtonpost.com/news/the-intersect/wp/2014/09/25/absolutely-everything-you-need-to-know-to-understand-4chan-the-internets-own-bogeyman [https://perma.cc/M9XP-C86Q].

[100] Id.

[101] Id.

[102] *Frequently Asked Questions*, 4CHAN, https://www.4chan.org/faq#whomod [https://perma.cc/W6K4-QXBR].

When it comes to the content moderation policies of more-mainstream social media platforms, although more tangible efforts are made to strengthen moderation practices and transparency,[103] issues still persist. Because the immunity provision of § 230 of the CDA does not dictate guidelines for content moderation,[104] platforms design their moderation practices with very little legal or public accountability.[105] Content moderation policies are routinely applied inconsistently, resulting in the over-moderation of marginalized groups and the lack of moderation of dominant groups.[106] When it comes to moderating content categorized as "terrorist speech," for example, many social media platforms rely on national terrorism designations, such as the U.S. Department of State's list of foreign terrorist organizations, which disproportionately designates Muslim and Middle Eastern-dominated demographics as "terrorist organizations."[107] In practice, this has led to instances of Facebook suspending dozens of accounts belonging to Syrian journalists and Palestinian activists for erroneously falling under the category of "terrorist speech."[108]

On the other side, platforms' moderation of Alt-Right content consistently falls flat. Facebook's own terms of service categorizes members of the Alt-Right as the lowest tier of priority under its "dangerous organizations and individuals" category.[109] Documents leaked from Facebook revealed a variation in content moderation practices across different countries.[110] While American users are offered a myriad of resources to protect themselves,[111] users in countries like Ethiopia are not even afforded community standards translated into their official language.[112] Facebook lacked classifiers for misinformation in Myanmar, Pakistan, and Ethiopia—even lacking classifiers for hate speech in Ethiopia "in the midst of a bloody civil conflict."[113]

Platforms' responses to Alt-Right hate speech are shockingly disproportionate in vigor and swiftness when compared to responses to non-hate-speech content posted by users of color. Facebook's Chief Executive Officer Mark Zuckerberg shielded figures, like Alex Jones, from being permanently banned under the company's own community standards by creating an entirely new classification for Jones under its policies.[114] As

---

[103] *See generally Facebook Community Standards*, META, https://transparency.fb.com/policies/community-standards [https://perma.cc/CDS9-PUT8]; *The X Rules*, X CORP., https://help.x.com/en/rules-and-policies/x-rules [https://perma.cc/R7V4-4UUD].

[104] 47 U.S.C. § 230(c)(1).

[105] *See generally* 47 U.S.C. § 230(c)(1).

[106] Angel Díaz & Laura Hecht-Felella, *Double Standards in Social Media Content Moderation*, BRENNAN CTR. FOR JUST. 23 (Aug. 4, 2021), https://www.brennancenter.org/sites/default/files/2021-08/Double_Standards_Content_Moderation.pdf [https://perma.cc/2CWL-HNJM].

[107] *Id.* at 5.

[108] Id.

[109] *Dangerous Organizations and Individuals*, META, https://transparency.fb.com/policies/community-standards/dangerous-individuals-organizations [https://perma.cc/CTK2-L7ES].

[110] Casey Newton, *The Tier List: How Facebook Decides Which Countries Need Protection*, VERGE (Oct. 25, 2021), https://www.theverge.com/22743753/facebook-tier-list-countries-leaked-documents-content-moderation [https://perma.cc/5CD6-ZPZR].

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

[112] Id.

[113] Id.

[114] Ryan Mac & Craig Silverman, *"Mark Changed the Rules": How Facebook Went Easy on Alex Jones and Other Right-Wing Figures*, BUZZFEED NEWS (Feb. 22, 2021),

members of QAnon, another far-right extremist group, began peddling their ideologies on Twitter, the platform failed to enforce its existing standards against hate speech or harassment.[115] Instead, the platform limited appearances of the conspiracy theory in its trends, recommendations, and search results, until public pressure from the 2021 capitol riot forced the site to begin permanently suspending thousands of QAnon accounts.[116]

The effects of inconsistent moderation are not only being felt domestically through the propagation of groups like the Alt-Right, but it is also fundamentally affecting the lives of social media users abroad. Considering the use of social media as a tool for activism and political organization,[117] these platforms—backed by § 230 of the CDA—have had a hand in influencing major geopolitical conflicts around the globe. The 2017 Rohingya crisis in Myanmar provided a shocking example of this.

In 2017, Myanmar forces were involved in a horrific ethnic cleansing campaign against the Rohingya people, which involved state-sponsored discrimination, displacement, rape, and killing.[118] Amnesty International compiled a report based on the research of Rohingya activists and refugees[119] that revealed how Facebook's algorithm contributed to the conflict by creating an echo chamber of hate speech rhetoric against the Rohingya people.[120] The Institute for War and Peace reported findings from a hate speech monitoring project in Myanmar, revealing that the majority of hate speech disseminating anti-Muslim or anti-Rohingya sentiment originated from Facebook.[121] The popularity of extremist anti-Rohingya groups in Myanmar "were among the most prominent Facebook users in Myanmar from 2015 onwards."[122]

The Independent International Fact-Finding Mission on Myanmar documented over one hundred fifty examples of social media accounts, pages, and groups that spread hate speech against the Rohingya people, many of which contained death threats, dehumanizing language, and promises of violence.[123] During this time, Meta failed to enforce its own policies on hate speech—an internal document written by a Meta employee revealed that the company has "evidence from a variety of sources that hate speech, divisive political speech, and misinformation on Facebook . . . are affecting societies

https://www.buzzfeednews.com/article/ryanmac/mark-zuckerberg-joel-kaplan-facebook-alex-jones [https://perma.cc/3T95-265K].

[115] Díaz & Hecht-Felella, *supra* note 106, at 7.

[116] Id.

[117] *See, e.g.*, Anna Rees, *The Role of Digital Tools in the World of Online Activism*, RESET: DIGITAL FOR GOOD (Feb. 6, 2024), https://en.reset.org/digital-and-online-activism [https://perma.cc/XHT3-X45H].

[118] *Myanmar: Facebook's Systems Promoted Violence Against Rohingya; Meta Owes Reparations – New Report*, AMNESTY INT'L (Sept. 29, 2022) [hereinafter *Myanmar*], https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report [https://perma.cc/78QD-ZDBB].

[119] *The Social Atrocity Meta and the Right to Remedy for the Rohingya*, AMNESTY INT'L 12 (Sept. 29, 2022) [hereinafter *The Social Atrocity*], https://www.amnesty.org/en/documents/ASA16/5933/2022/en [https://perma.cc/7T2Z-953X].

[120] *Myanmar, supra* note 118.

[121] *The Social Atrocity, supra* note 119, at 29.

[122] *Id.* at 27.

[123] *Id.* at 28.

around the world."[124] Since the conflict, Meta has partially recognized its role in the genocide but has taken no substantive action to prevent another instance of this happening again.[125] Section 230 of the CDA's broad protections effectively allowed Facebook to steer this genocide, evading any accountability until years after the genocide was over.[126] The legislation's provisions provide no requirement for social media companies to conduct their content moderation processes transparently,[127] which explains why Facebook's role in the genocide was uncovered nearly five years following the actual atrocities.[128] Had there been an efficient means of accessing the internal mechanisms by which Facebook conducts its moderation practices and enforces policies, perhaps the magnitude of this atrocity could have been curbed.

The protection of platforms via § 230 of the CDA carries a serious propensity for real-world harm, as demonstrated by instances both domestically and abroad. As social media sites amass more users and their influence continues to grow, this problem will continue to fester in the coming years. Social media platforms' ever-expanding role in our political landscape[129] has increased the risk of platforms prioritizing political and economic concerns over the safety of its users.[130] The lack of insight into how platforms conduct themselves internally has only exacerbated these risks. Achieving increased platform accountability requires a critical reevaluation of § 230 of the CDA.

III. SOLUTIONS TO §230 OF THE CDA

A. THE INADEQUACIES OF PLATFORM TRANSPARENCY

Section 230 of the CDA's lack of guidelines regarding how internet platforms should moderate content has created a black box of platform moderation practices and enforcement of terms of service. Platforms have no duty to release information regarding the exact methods used to moderate content, target ads, or curate what their users see due to the legal immunity provision of § 230.[131] Without transparency, there is no accountability between users and platforms, thus allowing platforms to conduct their practices in a vacuum.[132] Because there are no effective means to uncover

[124] Maktoob Staff, *Facebook Promoted Violence Against Rohingya; Meta Owes Reparation: Amnesty*, MAKTOOBMEDIA (Sept. 29, 2022), https://maktoobmedia.com/world/facebook-promoted-violence-against-rohingya-meta-owes-reparation-amnesty [https://perma.cc/W4EA-YHK9].

[125] Natasha Lomas, *Meta Urged to Pay Reparations for Facebook's Role in Rohingya Genocide*, TECHCRUNCH (Sept. 29, 2022), https://techcrunch.com/2022/09/29/amnesty-report-facebook-rohingya-reparations [https://perma.cc/3MSC-FJD5].

[126] *See generally* ANTI-DEFAMATION LEAGUE, *supra* note 65; Forscher & Kteily, *supra* note 66.

[127] Smith & Alstyne, *supra* note 24.

[128] Id.

[129] *See generally* Mason Walker & Katerina Eva Matsa, *News Consumption Across Social Media in 2021*, PEW RSCH. CTR. (Sept. 20, 2021), https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021 [https://perma.cc/ZQ4U-8PNW].

[130] Luke Barr, *Social Media Companies Prioritizing Profit Over Harmful Content: Senate Report*, ABC NEWS (Nov. 16, 2022), https://abcnews.go.com/Politics/social-media-companies-prioritizing-profit-harmful-content-senate/story?id=93360057 [https://perma.cc/CXZ8-HREY].

[131] 47 U.S.C. § 230(c)(1).

[132] *See, e.g.*, Svea Windwehr & Jillian C. York, *Facebook's Most Recent Transparency Report Demonstrates the Pitfalls of Automated Content Moderation*, ELEC. FRONTIER FOUND. (Oct. 8, 2020),

the moderation practices utilized by platforms, the effects of poor content moderation observed today most likely represent a small fraction of the actual consequences.

Current transparency practices by platforms are inept at addressing the issues resulting from poor content moderation. A coalition of organizations, advocates, and experts formulated the Santa Clara Principles on Transparency and Accountability in Content Moderation, a list of principles outlining how platforms can attain meaningful transparency and accountability regarding their practices, which has since been updated.[133] The list was endorsed by twelve major companies, including Apple, Facebook, Google, and Twitter.[134] Figure 3 displays a comparative analysis study of transparency reporting that examined the reports of Google, Microsoft, Apple, Meta, TikTok, and Twitter based on the Santa Clara Principles, revealing that none of these companies fully adhered to any of the principles outlined.[135]

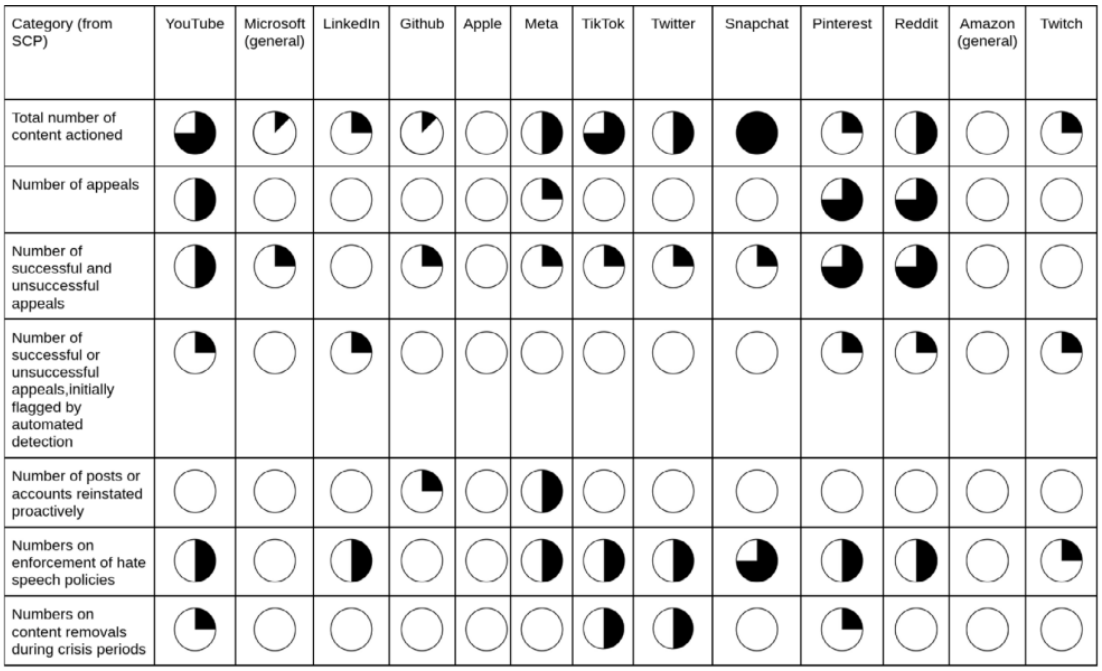


Fig. 3. Black circles represent full compliance to the Santa Clara Principles, while white circles indicate that no data on that particular category was included in the report.[136]

https://www.eff.org/deeplinks/2020/10/facebooks-most-recent-transparency-report-demonstrates-pitfalls-automated-content [https://perma.cc/K9AK-Q8LD].

[133] *The Santa Clara Principles on Transparency and Accountability in Content Moderation,* THE SANTA CLARA PRINCIPLES, https://santaclaraprinciples.org [https://perma.cc/6XSL-X9TR].

[134] Id.

[135] Aleksandra Urman & Mykola Makhortykh, *How Transparent Are Transparency Reports? Comparative Analysis of Transparency Reporting Across Online Platforms*, 47 TELECOMM. POL'Y 5 (2023), https://doi.org/10.1016/j.telpol.2022.102477 [https://perma.cc/4T89-UHVW].

[136] Id.

At one point, the study analyzed the degree to which platforms complied with the principle of releasing "the numbers on content removals during crisis periods" (Fig. 3, Row 8). Of the thirteen companies analyzed, none fully complied with the standard.[137] The highest level of compliance with this standard was only 50%, with nine companies not complying at all.[138] In fact, among all seven principles, there was only a single instance of full compliance.[139] The average rate of compliance with the principles was 20%, and the category with the highest compliance rate still only boasted a 30% average rate of compliance.[140] The study emphasized how current transparency practices by platforms fall flat despite the corporations' public endorsement of the principles. Without independent agencies or legislative guidelines incentivizing corporations to follow these transparency guidelines, companies self-enforce—revealing data that bolster their image and concealing data that do not.

This is not to say that transparency as a potential solution is ineffective. Current transparency practices taken by platforms, despite their insufficiency, still produce immense benefits for both users and the general public. A study on how transparency in moderation decisions affects user behavior on a popular social media platform, Reddit, revealed that "when moderated users are provided explanations, their subsequent post removal rate decreases," indicating how transparency regarding moderation decisions can improve the posting behavior of users on platforms.[141] CrowdTangle, a transparency tool purchased by Facebook to monitor trends and provide insights into the platform, has been used to dismantle the false claim of anti-conservative bias by platforms,[142] track the spread of misinformation relating to the COVID-19 vaccine, [143] and analyze Facebook's own moderation efforts of certain groups.[144] The tool has since been shut down by Facebook, leaving academics, journalists, and oversight organizations void of tools with which to conduct valuable research.[145] TikTok's global transparency report in 2020 also provided valuable insights on how the platform was combatting instances of misinformation related to COVID-19 and the U.S. presidential elections.[146]

The benefits of increased transparency by social media platforms are clear—not only does it facilitate trust between a user and their host platform, but it also provides an avenue for the general public to hold platforms

[137] Id.
[138] Id.
[139] Id.
[140] Id.
[141] Shagun Jhaver, Amy Bruckman & Eric Gilbert, Does Transparency in Moderation Really Matter?: User Behavior After Content Removal Explanations on Reddit, PROC. ACM ON HUM.-COMPUT. INTERACTION, Nov. 2019, at 21.
[142] *See generally* Paul M. Barrett & J. Grant Sims, False Accusation: The Unfounded Claim That Social Media Companies Censor Conservatives 4–5 (NYU Stern Ctr. for Bus. & Hum. Rts. 2021).
[143] John Albert, *Facebook's Gutting of CrowdTangle: A Step Backward for Platform Transparency*, ALGORITHM WATCH (Aug. 3, 2022), https://algorithmwatch.org/en/crowdtangle-platform-transparency [https://perma.cc/JW4S-M5UZ].
[144] Id.
[145] Id.
[146] Spandana Singh & Leila Doty, *The Transparency Report Tracking Tool: How Internet Platforms Are Reporting on the Enforcement of Their Content Rules*, NEW AM. (Dec. 9, 2021), https://www.newamerica.org/oti/reports/transparency-report-tracking-tool [https://perma.cc/CF86-4VWF].

accountable for the decisions they make—decisions with the capacity for real-world harm. Current practices, however, are ineffective due to the broad immunity provisions of § 230 of the CDA, providing platforms with no legislative incentive or guidance on what should be released to the public. Current insights released by platforms lack independent verification[147] and are self-selected data points—preventing organizations and researchers from using this data effectively in empirical studies or reports. There is also no way to ensure the data being released is accurate or comprehensive.[148] The solution to § 230 of the CDA, then, may not lie in the hands of platforms, but rather in those of independent researchers.

B. INDEPENDENT RESEARCH: THE PREMISE OF REMEDYING § 230 OF THE CDA

Though increased transparency certainly paves the way for solving the issues posed by the black box practices of social media platforms, it is unlikely that platforms themselves will reach the requisite degree of transparency necessary to accomplish this goal.[149] The best way to capture a glimpse into this void lies with independent researchers. Most of the important information already acquired, revealing the hidden practices of social media platforms, has come from the work of independent researchers.[150] This work is essential both in compelling platforms to ensure their practices are unproblematic and in giving users insight into how platforms conduct themselves.

Independent platform research is quintessential to obtaining insights into social media platforms that would otherwise remain hidden. For example, a group of independent researchers from Facebook's Data Transparency Advisory Group released a report on the platform's community standards, giving users insights into specific areas where Facebook was lacking in enforcement of these policies.[151] Similarly, another independent research organization, Upturn, used publicly available information from Facebook to reveal how the platform was "violating U.S. law by creating discriminatory effects in housing advertising."[152] Despite the limited information available to researchers, much has been revealed about the practices hidden by platforms. What stands in the way of more substantive investigation, however, are a variety of legal roadblocks that prevent researchers from obtaining access to platform data and subject them to legal liability if they do obtain access.

---

[147] *See, e.g.*, *Platform Transparency Reports – Just How Transparent?*, ANTI-DEFAMATION LEAGUE (Feb. 14, 2024), https://www.adl.org/resources/blog/platform-transparency-reports-just-how-transparent [https://perma.cc/GVW2-EUGS].

[148] See generally id.

[149] *See generally* Urman & Makhortykh, *supra* note 135; ANTI-DEFAMATION LEAGUE, *supra* note 147.

[150] Daphne Keller & Paddy Leerssen, *Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation, in* SOCIAL MEDIA AND DEMOCRACY THE STATE OF THE FIELD, PROSPECTS FOR REFORM 220, 237 (Nathaniel Persily & Joshua A. Tucker eds., 2020).

[151] Mark MacCarthy, Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry 1–4 (Transatlantic Working Group Paper, SSRN No. 3615726, 2020).

[152] *Id.* at 14 (citation omitted).

Many of these legal roadblocks took root in the Cambridge Analytica scandal,[153] which fundamentally changed the way data was shared by social media companies. The scandal involved a university researcher who legally obtained over fifty million Facebook profiles, which were then given to a political consulting company to further the political campaigns of Ted Cruz and Donald Trump.[154] Facebook was eventually forced to pay a five billion dollar fine to the Federal Trade Commission and shut down many automated programmer interfaces ("APIs") as a result.[155]

API access is the primary way researchers can access data available on platforms and serves as a more holistic alternative to self-reports from users or reliance on data freely given by platforms.[156] APIs allow researchers to download data from platforms efficiently without having to open a myriad of webpages.[157] Following Facebook's API lockdown, thirty internet academics signed an open letter condemning the decision, claiming it would detrimentally alter current and future research insights into platforms' practices, which are already tightly constrained due to the lack of legislation compelling platforms to make this data available.[158] Without access to API data, hundreds of important publications and research into platforms would cease to exist—publications ranging from the role of Twitter in the Hong Kong occupy protests to cyberstalking and opinion polarization.[159] APIs are quintessential in acquiring the data necessary to capture a glimpse into the practices of social media platforms. They serve as an important link between the black box practices of social media platforms and the need for platform transparency and accountability.

Beyond limited access to data, U.S. law has made it increasingly difficult for researchers to access the data concealed by social media companies. The Computer Fraud and Abuse Act ("CFAA"), for example, has been interpreted to ban independent researchers from obtaining information via the use of "scraping."[160] Social media scraping is a type of automatic data extraction similar to API's, whereby researchers can analyze and collect large amounts of information available on social media sites.[161] Scraping gives researchers access to data that would otherwise require far too much time and resources

---

[153] Nathaniel Persily & Joshua A. Tucker, *How to Fix Social Media? Start with Independent Research.*, BROOKINGS (Dec. 1, 2021), https://www.brookings.edu/articles/how-to-fix-social-media-start-with-independent-research [https://perma.cc/9M3Y-FD3X].

[154] *See generally Facebook—Cambridge Analytica Data Scandal*, WIKIPEDIA, https://en.wikipedia.org/w/index.php?title=Facebook%E2%80%93Cambridge_Analytica_data_scandal&oldid=1187694481 [https://perma.cc/RZ3C-4H5D]; Sara Fischer & David McCabe, *Cambridge Analytica Data Scandal Highlights Chaos at Facebook*, AXIOS (Mar. 18, 2018), https://www.axios.com/2018/03/18/cambridge-analytica-scandal-highlights-chaos-at-facebook [https://perma.cc/C74Y-F4BF].

[155] *Facebook–Cambridge Analytica Data Scandal, supra* note 154.

[156] Johannes Breuer, Zoltan Kmetty, Mario Haim & Sebastian Stier, User-Centric Approaches for Collecting Facebook Data in the 'Post-API Age': Experiences from Two Studies and Recommendations for Future Research, INFO., COMMC'N & SOC'Y, July 8, 2022, at 2.

[157] Persily & Tucker, *supra,* note 153.

[158] Axel Bruns, *Facebook Shuts the Gate After the Horse Has Bolted, and Hurts Real Research in the Process*, MEDIUM (Apr. 25, 2018), https://medium.com/@Snurb/facebook-research-data-18662cf2cacb [https://perma.cc/24WQ-RQ8J].

[159] Anja Bechmann, Open Letter from Anja Bechmann on Publications That Could Not Have Existed Without Access to API Data, GOOGLE DOC (Apr. 15, 2018).

[160] MacCarthy, *supra* note 151, at 13.

[161] Saheed Opeyemi, *What Is Social Media Scraping? (and Why You Should Do It)*, SCRAPINGROBOT (Apr. 11, 2021), https://scrapingrobot.com/blog/social-media-scraping [https://perma.cc/63UJ-NFCC].

to acquire manually, providing another potential glimpse into the practices being concealed by platforms.[162] Although recent court cases have alluded to keeping scraping legal, the CFAA is applied inconsistently, suggesting the modicum of legal protection afforded to scrapers rests on shaky ground.[163] Additionally, the CFAA gives platforms unlimited veto power over independent researchers' ability to request and collect important data.[164] This leads to both an unclear application in the courts and unfettered power for platforms to dictate whether researchers are able to use the data from their sites.

C. IMPLEMENTATION OF A REMEDY TO § 230 OF THE CDA: A SAFE HARBOR PROVISION FOR INDEPENDENT RESEARCHERS

The solution to many of the issues resulting from § 230 of the CDA lies in the hands of independent researchers, particularly in their ability to access data concealed by big platforms. Given the current legal roadblocks faced by independent researchers, effective reforms to § 230 of the CDA should include a safe harbor provision for social media sites that benefit from § 230 immunity, giving independent researchers access to data with proper safeguards to protect the data privacy of users. The immunity afforded to platforms should be contingent upon providing independent researchers with this safe harbor. Legislators have proposed several different bills with provisions aimed at protecting independent researchers' access to data from social media companies.[165] These bills address matters including which groups are designated as researchers, whether restrictions should exist on the type of research conducted, and privacy efforts to prevent the misuse of user data.

The 117th Congress introduced over twenty-five bills to either completely repeal or amend § 230 of the CDA.[166] While Democrats aim their proposals at increasing content moderation, specifically surrounding hate speech and misinformation, Republicans hope to stop platforms from moderating content entirely, especially in the wake of platforms banning figures like Donald Trump.[167] Some proposals specifically target speech in advertisements, which avoids many of the First Amendment challenges that other proposals face[168] but does not address the brunt of § 230's

[162] Id.

[163] Andrew Sellars, Twenty Years of Web Scraping and the Computer Fraud and Abuse Act, 24 B.U. J. SCI. & TECH. L. 372, 412 (2018).

[164] Alex Abdo, Ramya Krishnan, Stephanie Krent, Evan Welber Falcón & Andrew Keane Woods, *A Safe Harbor for Platform Research*, KNIGHT FIRST AMEND. INST. AT COLUM. UNIV. (Jan. 19, 2022), https://knightcolumbia.org/content/a-safe-harbor-for-platform-research [https://perma.cc/VB5Q-XRUR].

[165] Caitlin Vogus, *Independent Researcher Access to Social Media Data: Comparing Legislative Proposals*, CTR. FOR DEMOCRACY & TECH. (Apr. 21, 2022), https://cdt.org/insights/independent-researcher-access-to-social-media-data-comparing-legislative-proposals [https://perma.cc/2JFC-H4JK].

[166] Mike Brand, *Section 230 and Disinformation*, MCCAIN INST. (Aug. 14, 2023), https://www.mccaininstitute.org/resources/blog/section-230-and-disinformation [https://perma.cc/4C5D-3SEF].

[167] Id.

[168] Chris Riley & David Morar, *Legislative Efforts and Policy Frameworks Within the Section 230 Debate*, BROOKINGS (Sept. 21, 2021), https://www.brookings.edu/articles/legislative-efforts-and-policy-frameworks-within-the-section-230-debate [https://perma.cc/JGP7-BGAL].

ineffectiveness regarding hate speech and lack of platform transparency.[169] Other proposals call for the creation of standards for platforms to follow; one by Danielle Citron and Benjamin Wittes calls for adding a provision requiring reasonableness on the part of platforms' practices, allowing courts to define this standard through individual cases.[170] A major criticism of this proposal is the potential First Amendment challenges, as the reform is dangerously close to allowing government restriction of corporate speech.[171]

More viable reforms to § 230 of the CDA focus on ensuring independent researchers have access to data and arming them with legal immunity, conditioned upon satisfying certain requirements. These reforms balance the importance of making data available and protecting user privacy. Congress has introduced four bills addressing independent researchers' access to data: the Platform Accountability and Transparency Act, Social Media Data Act, Digital Services Oversight and Safety Act, and Kids Online Safety Act. All the Acts include provisions defining the qualifications of researchers, with some broadly defining a researcher as one affiliated with an institution or nonprofit, and others limiting the definition solely to academic researchers.[172] Regarding the type of data researchers would be able to access, one bill specifies data must relate to research interests "regarding harms to the safety and well being [sic] of minors," [173] while others give government agencies, like the Federal Trade Commission ("FTC") and National Science Foundation, the authority to determine which data is relevant to independent researchers.[174] The proposals vary on the types of research to be covered, with only one containing no restrictions on research subjects.[175] Several safeguards are proposed to protect confidential data from misuse: one bill assigns data to different tiers based on how sensitive the information is and how broadly it should be accessed by researchers, while another proposes that the FTC's Platform Accountability and Transparency Office establish privacy criteria and safeguards.[176] Further, most of the proposals include some form of conditional immunity provision for independent researchers.[177]

An ideal and effective reform to § 230 of the CDA should balance protecting independent researchers and safeguarding user data privacy. The reform should provide independent researchers with legal immunity as long as they meet certain requirements. Researchers should be qualified either through a higher education institution or a nonprofit whose mission is related to the researcher's research subject. Additionally, the purpose of the research should be to uncover information that will generally benefit users of social media platforms, and the research should be non-commercial. Researchers must take reasonable measures to protect the privacy of user data, and

[169] Aaron Mackey, India McKinney & Jason Kelley, *The SAFE Tech Act Wouldn't Make the Internet Safer for Users*, ELEC. FRONTIER FOUND. (Feb. 25, 2021), https://www.eff.org/deeplinks/2021/02/safe-tech-act-wouldnt-make-internet-safer-users [https://perma.cc/NY8U-YEQJ].
[170] Riley & Morar, s*upra* note 168.
[171] Id.
[172] Vogus, *supra* note 165.
[173] Id.
[174] Id.
[175] Id.
[176] Id.
[177] Id.

information acquired must not be used outside the direct furtherance of the research subject matter.[178] The methods of investigation protected under this provision will encompass collecting data through automated means, users, or the creation of research accounts.[179] Information covered under the provision should include any publicly available information on social media sites, information about advertisements, and other information that does not significantly burden user privacy.[180]

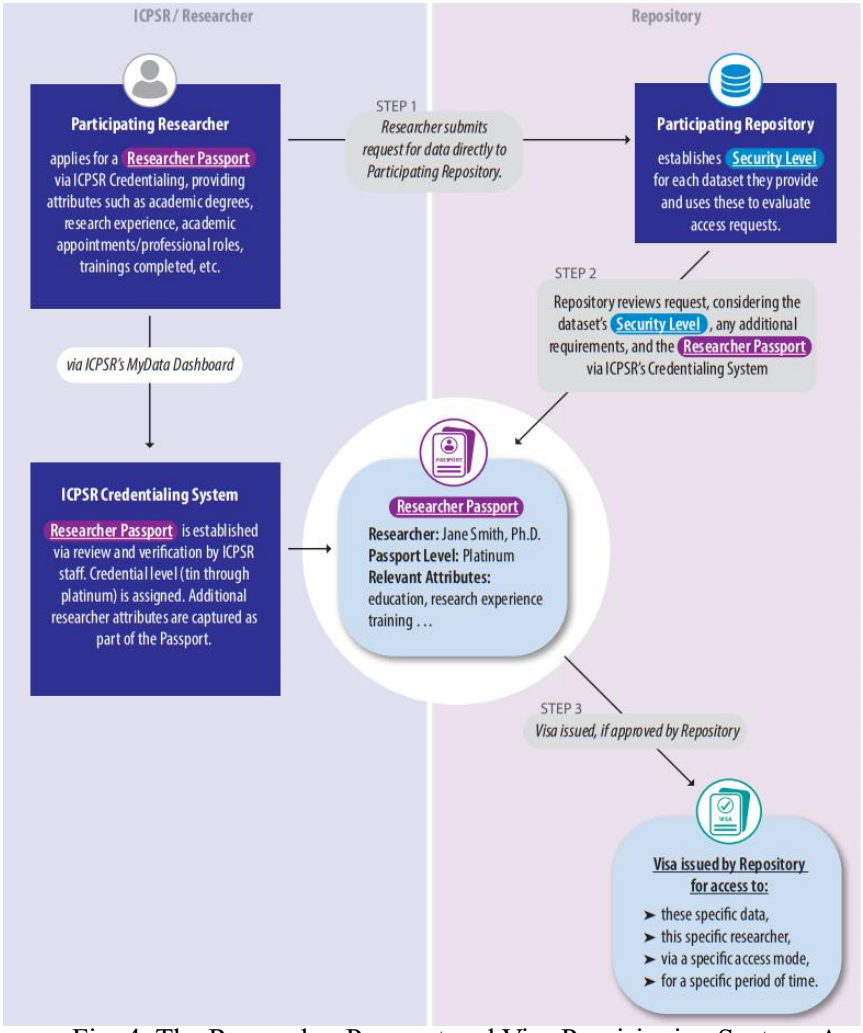


Fig. 4. The Researcher Passport and Visa Provisioning System: A Roadmap.[181]

[178] Abdo, Krishnan, Krent, Falcón & Woods, *supra* note 164.
[179] Id.
[180] Id.
[181] Margaret C. Levenstein, Allison R.B. Tyler & Johanna Davidson Bleckman, The Researcher Passport: Improving Data Access and Confidentiality Protection ICPSR's Strategy for a Community-Normed System of Digital Identities of Access 2 (2018).

Tiered access to data will ameliorate many privacy concerns regarding the misuse of user data. Researchers from the University of Michigan's Institute for Social Research proposed the idea of a "researcher passport," whereby researchers provide relevant credentials, including their research experience, professional roles, and training, to the company from which data is being requested.[182] These credentials may be independently verified to ensure accuracy.[183] Companies would then categorize the security level of data sets based on the degree of sensitive information and evaluate which tier of data may be released to each tier of researchers. If approved, companies can issue "visas" granting researchers access to the specified tier of data.[184] Companies should be required to conduct decisions in good faith to prevent instances of arbitrary denials.

Racial capitalism has shown that, without proper incentives, companies will act in ways that align with profit maximization—and for social media companies, this maximization of profit is inextricably tied to the proliferation of hate speech and Alt-Right viewpoints. It is clear that incentives must come from both legislation and the general public—and the safe harbor provision provides both. The safe harbor provision will arm independent researchers and watchdog organizations with important information on how social media companies moderate content and enforce their terms of service. This information will allow researchers to communicate their findings to the general public, which carries the collective power to demand more from platforms. Although Alt-Right viewpoints and hate speech will continue to exist, social media companies' marketability and profitability will be increasingly reliant on public perception of their conduct. With access to more information, users will wield more power in deciding what they will and will not accept from platforms.

CONCLUSION

The increase in the use and influence of social media platforms over the past decade is unparalleled. Social media has become an integral part of society, both domestically and abroad. The rise of hateful acts from extremist groups like the Alt-Right has transformed how we regard online extremist movements and should equally transform how they are regulated online. Racial capitalism underpins the explosion of hateful rhetoric from groups like the Alt-Right observed on social media, emphasizing how the maintenance of a racist status quo is imperative to the continuation of capitalism and instillment of the values so inherent to whiteness.

Section 230 of the CDA has only strengthened platforms' ability to conduct themselves in unethical ways. Without a legislative incentive to moderate hate speech, platforms fail to prioritize effective content moderation, allowing hate speech and extremist groups like the Alt-Right to run rampant. Lacking transparency into platforms' moderation practices, there is no accountability to the government or members of the general public. This has led to the inconsistent application of community standard

[182] Id.
[183] *Id.* at 25.
[184] *Id.* at 20–24.

policies, wrongfully targeting minority groups while shielding Alt-Right extremists from consequences. The lack of platform accountability can also be traced to the incitement of major geopolitical conflicts, underscoring platforms' failure to satisfy moral and ethical obligations to prevent harm, choosing instead the path that maximizes profit.

The solution to these issues lies in the work of independent researchers to provide the accountability and transparency that platforms currently lack. Transparency has the power to equip users with information, giving platforms incentives to change. Legal roadblocks, data privacy concerns, and platforms' economic interests in keeping their practices covert have made it increasingly difficult for independent researchers to collect data and conduct studies. Viable solutions to § 230 of the CDA must center on both protecting researchers' ability to collect data and balancing important data privacy concerns that users have.

To combat the issues with § 230 of the CDA, a safe harbor provision should be included for independent researchers in the crusade for increased platform transparency. The implementation of this provision will allow for more accurate and extensive studies regarding the practices of platforms—hopefully giving the general public more comprehensive insight into how platforms conduct themselves. The revelation of any wrongful practices to the public should provide platforms adequate incentive to critically examine their methods to avoid poor public perception. Overall, the implementation of this provision has the potential to drastically improve both user experience on social media and limit the propensity of real-world harm. This provision will hopefully stifle the spread of Alt-Right rhetoric online and lead to more unified and cohesive environments both online and beyond.