DEALING WITH DISINFORMATION: EVALUATING THE CASE FOR CDA 230 AMENDMENT

Tim Hwang[1]

Recent revelations surrounding Russian interference in the 2016 US presidential election and the role that "fake news" may have played in shaping voter preferences have sparked a broad conversation among researchers, policymakers, technologists, and others on how to combat the spread and influence of disinformation online. Emerging from this conversation has been a number of proposals that seek to pass legislation or promulgate regulations that would make it more difficult for disinformation to flow through the web.

To that end, these interventions will confront the long-standing legal protections provided by Section 230 of the Communications Decency Act of 1996 (CDA 230), a key legal provision which broadly shields platforms from legal liability for the actions of third-party users of their services. For the past two decades, this provision has been seen as major driver in the growth of online services, and a cornerstone supporting free expression on the web. Simultaneously, CDA 230 has also been argued to inhibit platform responsiveness to the harms posed by harassment, defamation, sex trafficking, and a host of other activities online. The present-day debates on how to address "fake news" will join the legacy of efforts to reform or eliminate the shield provided by CDA 230.

This paper seeks to address three questions given this historical background. First, would modifications to CDA 230 pave the way to an effective response to the challenges posed by disinformation online? Second, if so, should such modifications be made? Finally, how should such modifications be crafted?

* * *

[1] Fellow, Knight-Stanford Project on Democracy and the Internet

Electronic copy available at: https://ssrn.com/abstract=3089442

* * *

INTRODUCTION

Recent revelations surrounding Russian interference in the 2016 US presidential election and the role that "fake news" may have played in shaping voter preferences have sparked a broad conversation among researchers, policymakers, technologists, and others on how to combat the spread and influence of disinformation online. Emerging from this conversation has been a number of proposals that seek to pass legislation or promulgate regulations that would make it more difficult for disinformation to flow through the web.

Given the fragmented nature of information *creation* across the web, many of these legal proposals rely on the central role that online platforms such as Google, Facebook, and Twitter play in shaping the *distribution* of information throughout the web. By creating incentives - or penalties - encouraging platforms to take a more proactive role in removing or combatting disinformation, these interventions seek to leverage the unique

Electronic copy available at: https://ssrn.com/abstract=3089442

position of these companies as potentially the most effective "least cost avoiders" in addressing the challenge posed by disinformation.

To that end, these interventions will confront the long-standing legal protections provided by Section 230 of the Communications Decency Act of 1996 (CDA 230), a key legal provision which broadly shields platforms from legal liability for the actions of third-party users of their services. For the past two decades, this provision has been seen as major driver in the growth of online services, and a cornerstone supporting free expression on the web. Simultaneously, CDA 230 has also been argued to inhibit platform responsiveness to the harms posed by harassment, defamation, child pornography, and a host of other activities online. The present-day debates on how to address "fake news" will join the legacy of efforts to reform or eliminate the shield provided by CDA 230.

This paper seeks to address three questions given this historical background. First, would modifications to CDA 230 pave the way to an effective response to the challenges posed by disinformation online? Second, if so, should such modifications be made? Finally, how should such modifications be crafted?

Part I frames the challenge posed by disinformation online, specifically the threat posed by the confluence of politically motivated actors, financially motivated media outlets, and online “troll” culture. Part II examines the legislative history and caselaw surrounding CDA 230, and evaluates its impact on contending with the challenge of online disinformation. Part III takes up the challenge of whether or not CDA 230 should be modified in light of this analysis, concluding that partial amendment focused on the techniques leveraged by disinformation campaigns is warranted.

PART I: THE DISINFORMATION CHALLENGE

“Fake news” has become a commonplace term for characterizing the prevalence of false or inaccurate stories circulating online, considered a symptom of the poor state of information quality throughout media and society generally. These stories were widely distributed during the 2016 US presidential election, with one survey suggesting that close to one in five US adults saw headlines claiming (falsely) that the Pope had endorsed then-candidate Donald Trump, and that protestors had been paid $3,500 to

disrupt a Trump rally.[2] These stories were also considered credible, with 64% and 79% of respondents reporting that they believed the stories to be "very or somewhat accurate", respectively.[3]

However, as has been observed by others, the use of "fake news" as a conceptual frame is problematic on a number of levels.[4] In of itself, the spreading of false information under the pretense of truth is, of course, not a novel phenomenon, either online or in channels of communications more generally.[5] Moreover, there are numerous problems with defining the contours of the "fake news" phenomena. Should it apply only to the outright fabrication of events and assertions about reality? Or does it also include the partial presentation of information or an unfair characterization of events?

A sharper framing of the nature of the purported threat is necessary to evaluate the case for modifying or eliminating CDA 230. This paper focuses on three major developments that have been drivers motivating the post-2016 discussion around online disinformation and its regulatory response. This includes (1) the active spreading of disinformation by governments and state-owned media, (2) financially motivated actors pushing disinformation for the purposes of obtaining advertising revenue, and (3) the activities of online "troll" communities as a vector for spreading disinformation. This paper narrows in specifically on campaigns of *political* disinformation, false information targeted to shape perceptions around some

---

[2] Craig Silverman Singer-Vine Jeremy, Most Americans Who See Fake News Believe It, New Survey Says BuzzFeed, https://www.buzzfeed.com/craigsilverman/fake-news-survey (last visited Dec 16, 2017).

[3] *Id.*

[4] *See, e.g.,* Will Oremus, *Stop Calling Everything "Fake News,"* Slate, 2016, http://www.slate.com/articles/technology/technology/2016/12/stop_calling_everything_fake_news.html; Margaret Sullivan, *It's time to retire the tainted term "fake news,"* Washington Post, https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5_story.html (last visited Nov 20, 2017); Rasmus Kleis Nielsen and Lucas Graves, *What do ordinary people think fake news is? Poor journalism and political propaganda*, Columbia Journalism Review, https://www.cjr.org/analysis/fake-news-study.php (last visited Nov 20, 2017).

[5] *See, e.g.,* Carla Mulford, *Benjamin Franklin's Savage Eloquence: Hoaxes from the Press at Passy, 1782*, 152 Proceedings of the American Philosophical Society 490–530 (2008) (political hoaxes in newspapers); ERIK BARNOUW, A TOWER IN BABEL, 168-72 (1966) (medical hoaxes in radio); Drunk Driving on the Internet, Museum of Hoaxes, http://hoaxes.org/af_database/permalink/drunk_driving_on_the_internet (last visited Nov 20, 2017) (regulatory hoaxes on the Internet).

aspect of political discourse, rather than efforts which spread inaccurate stories on other topics such as corporate acquisitions or celebrity deaths.[6]

As they have been less of a primary focus in the regulatory debates around what to do with "fake news", this paper also excludes some other types of falsity. It does not cover the inadvertent spread of false or misleading information through the web which do not result from a coordinated effort. Similarly, this paper focuses on activities primarily targeted at disseminating disinformation - distinguishing it from cases in which a campaign simply attempts to amplify a point of view, or spread awareness of a fact.

There is no doubt these categories are by necessity blurry at the edges: a disinformation campaign may leverage an existing misconception spreading organically, or an effort to bring attention to a certain point of view may strategically frame the truth or even shade into falsehood. Issues frequently bleed across the fuzzy boundary between "political" and "non-political" discourse, such as debunked theories around the dangers of vaccines promoted by "antivax" activists. However, through this rough framework, this section seeks to offer background context around some of the activities that have provided the impetus for recent calls for legislative and regulatory action on online disinformation.

Disinformation from State Actors

Perhaps the primary trigger of calls for a regulatory response to the challenges posed by disinformation threats online has been confirmation by the intelligence community that Russian state actors engaged in an active effort to shape discourse around the 2016 US presidential election.[7]

The 2016 Russian campaign was a multi-faceted effort aimed at undermining trust in targeted political figures. This included conspiracy theories such as "Pizzagate", which spread the notion that Democratic

---

[6] *See, e.g.,* Timothy B. Lee, *Dow Jones posts fake story claiming Google was buying Apple,* Ars Technica (2017), https://arstechnica.com/tech-policy/2017/10/dow-jones-posts-fake-story-claiming-google-was-buying-apple/ (last visited Nov 20, 2017); *Musician Started Bon Jovi Death Hoax*, Rolling Stone, http://www.rollingstone.com/music/news/musician-started-bon-jovi-death-hoax-20111228 (last visited Nov 20, 2017).

[7] NAT'L INTELLIGENCE COUNCIL, OFFICE OF THE DIR. OF NAT'L INTELLIGENCE, ICA 2017-01D, ASSESSING RUSSIAN ACTIVITIES AND INTENTIONS IN RECENT US ELECTIONS (2017).

nominee Hilary Clinton and Clinton campaign chairman John Podesta were members of an underground sex trafficking ring.[8] Beyond efforts to spread disinformation, the effort also included attempts to exacerbate political polarization, in one case stoking racial controversy around law enforcement between activist "Black Lives Matter" and "Blue Lives Matter" groups.[9]

The campaign also operated through a range of different channels. State-owned media outlets such as Sputnik and Russia Today were leveraged to create and disseminate disinformation widely.[10] These more obvious channels operated alongside more subtle "grassroots" infiltration of online communities and the purchase of targeted advertising across various social media platforms.[11] Beyond the spread of disinformation, the campaign also engaged in hacking targeted at compromising private information held by political parties and candidates on both sides of the electoral race.[12]

While the 2016 Russian campaign has been a widely-discussed example of state-driven online disinformation, the use of these techniques is not new. Researchers have tracked similar online disinformation campaigns launched by Russia to influence political discourse throughout Central and Eastern Europe, as well as the Middle East, in recent years.[13]

---

[8] *See* Amanda Robb, *Pizzagate: Anatomy of a Fake News Scandal*, Rolling Stone, http://www.rollingstone.com/politics/news/pizzagate-anatomy-of-a-fake-news-scandal-w511904 (last visited Nov 20, 2017).

[9] *See* Yamiche Alcindor, *Black Lawmakers Pressure Facebook Over Racially Divisive Russian Ads, The New York Times,* September 28, 2017, https://www.nytimes.com/2017/09/28/us/politics/facebook-russia-race-congressional-black-caucus.html (last visited Nov 20, 2017); Deepa Seetharaman, *Russian-Backed Facebook Accounts Staged Events Around Divisive Issues*, Wall Street Journal, October 30, 2017, https://www.wsj.com/articles/russian-backed-facebook-accounts-organized-events-on-all-sides-of-polarizing-issues-1509355801 (last visited Nov 14, 2017).

[10] *See* Nat'l Intelligence Council, *supra* note 4, at 3.

[11] *See id.* at 3-4.

[12] *See* NAT'L CYBERSEC. AND COMMC'NS INTEGRATION CTR., DEP'T OF HOMELAND SEC., JAR-16-20296A, GRIZZLY STEPPE - RUSSIAN MALICIOUS CYBER ACTIVITY (2016).

[13] *See* Elina Lange-Ionatamishvili, Sanda Svetoka & Kenneth Geers, *Strategic Communication and Social Media in the Russia Ukraine Conflict*, Cyber War in Perspective: Russian Aggression against Ukraine, Tallinn: NATO CCD COE Publications (2015); Patrick Wintour, *Russian hackers to blame for sparking Qatar crisis, FBI inquiry finds*, The Guardian, June 7, 2017, http://www.theguardian.com/world/2017/jun/07/russian-hackers-qatar-crisis-fbi-inquiry-saudi-arabia-uae (last visited Nov 21, 2017); Adrian Chen, *The Agency*, The New York Times, June 2, 2015, https://www.nytimes.com/2015/06/07/magazine/the-agency.html (last visited Oct 11, 2017).

Nor are these campaigns specific to Russia. Researchers have found that social media has been leveraged for political disinformation purposes in a range of different contexts in recent years. Incidents include efforts seen in Mexico, Brazil, Canada, and China, to name a few.[14] These campaigns have been launched by state actors as in the Russian case, but have also been launched by a range of independent groups, as well.[15]

Financial Incentives for Disinformation

Recognition that politically motivated actors engaged in efforts to influence the 2016 election has emerged alongside a growing number of commentators highlighting the financial incentives driving the creation and dissemination of disinformation. Online advertising, in particular, has been seen as motivator to create false, but highly sharable content that drives monetizable pageviews to content online.

In the context of the 2016 US presidential election, businesses both within the country and abroad engaged in the creation of sites spreading disinformation through the web. Media outlets included such sites as "The Denver Guardian", which spread a range of conspiracy theories, such as one story connecting Clinton to the murder of an FBI agent investigating her use of a private e-mail server, shared millions of times across Facebook.[16] While the site was designed with the appearance of a local paper in Colorado, it was in actuality operated by a Los Angeles based entrepreneur who also ran a collection of other sites profiting from the sharing of disinformation.[17]

---

[14] *See* Samuel C. Woolley & Philip N. Howard, Computational propaganda worldwide: Executive summary (2017), *available at* http://comprop.oii.ox.ac.uk/publishing/working-papers/computational-propaganda-worldwide-executive-summary/; Klint Finley Klint Finley Security, Pro-Government Twitter Bots Try to Hush Mexican Activists WIRED, https://www.wired.com/2015/08/pro-government-twitter-bots-try-hush-mexican-activists/ (last visited Nov 21, 2017).

[15] *See, e.g.,* Jordan Robertson et al., *How to Hack an Election*, Bloomberg.com, , https://www.bloomberg.com/features/2016-how-to-hack-an-election/ (last visited Oct 11, 2017) (paid election manipulation for hire in Latin America).

[16] Jestin Coler, *We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned,* NPR.org, http://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs (last visited Oct 11, 2017).

[17] *Id.*

Outside the US, journalists have uncovered groups of entrepreneurs in Macedonia and elsewhere profiting by selling advertisements running alongside disinformation catering to right-wing readers online.[18] Stories included "news" of Pope Francis endorsing then-candidate Trump, and fabricated reports of the candidate slapping a protestor at a campaign rally.[19] These sites sometimes acted as an amplifier rather than an originator of disinformation, copying content from other sites online and promoting them through swarms of fake accounts on social media platforms like Twitter and Facebook.[20]

Discussion around financial incentives for disinformation have not been limited to discussing the outlets producing and promoting this content online. Since many of the most prominent online platforms such as Google and Facebook are themselves reliant on advertising, critics and researchers have also underscored that the companies hosting this activity may have perverse incentives to harbor it given that disinformation content is often widely shared and viewed.[21] For their part, these platforms have disputed this notion in numerous public statements and have taken action to restrict distributing advertising against disinformation.[22]

"Trolling Culture" as a Disinformation Source

Beyond the activities of politically and financially motivated actors, the participation of grassroots online "troll" culture has also been a force in

---

[18] *See* Samantha Subramanian, *Inside the Macedonian Fake-News Complex*, Wired, https://www.wired.com/2017/02/veles-macedonia-fake-news/ (last visited Oct 11, 2017); Dan Tynan, *How Facebook powers money machines for obscure political "news" sites*, The Guardian, August 24, 2016, http://www.theguardian.com/technology/2016/aug/24/facebook-clickbait-political-news-sites-us-election-trump (last visited Nov 21, 2017).

[19] Subramanian, *supra* note 19.

[20] *Id.*

[21] *See, e.g.,* Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 Journal of Economic Perspectives 211–236 (2017); Nicholas Thompson, *Our Minds Have Been Hijacked By Our Phones. Tristan Harris Wants to Rescue Them*, Wired, https://www.wired.com/story/our-minds-have-been-hijacked-by-our-phones-tristan-harris-wants-to-rescue-them/ (last visited Nov 21, 2017).

[22] *See* Nick Wingfield, Mike Isaac & Katie Benner, *Google and Facebook Take Aim at Fake News Sites*, The New York Times, November 14, 2016, https://www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html (last visited Nov 21, 2017); Justin Ling, Google Chief Says Google News Will "Engineer" Russian Propaganda Out of the Feed Motherboard (2017), https://motherboard.vice.com/en_us/article/pa39vv/eric-schmidt-says-google-news-will-delist-rt-sputnik-russia-fake-news (last visited Nov 21, 2017).

facilitating political disinformation. Crowd activity - often performed anonymously - to shock and harass private citizens, public figures, and institutions for pure entertainment purpose has been a longstanding feature of social behavior on the Internet.[23] These activities in the past have leveraged a wide array of tactics, from the manipulation of online polls, to the strategic targeting of journalists and "swatting" - false emergency reports to law enforcement aimed at bringing police officers to a targeted address.[24] In recent years, many of these communities have been radicalized by far-right groups to "spread white supremacist thought, Islamophobia, and misogyny through irony and knowledge of internet culture", as researchers Alice Marwick and Rebecca Lewis have documented.[25]

In the context of the 2016 US presidential election, many of these communities were involved in coordinated campaigns to spread political disinformation. This included promoting conspiracy theories that philanthropist George Soros was engaged in a nationwide campaign to fund protests against Trump, and claims that DNC staffer Seth Rich was assassinated as part of a cover-up connected to the 2016 leak of e-mails from the DNC.[26] Effectively, these campaigns drew on the efforts of volunteers, a loosely coordinated, informal coalition of overlapping "alt-right" groups. This brought together a wide range of actors, including gamer communities, users of the popular online discussion board Reddit, members of the white supremacist community Stormfront, and "alt-light" news outlets echoing some of the messages of the far-right but excluding some of the more controversial views, to name a few.[27] These techniques drew explicitly on these earlier "trolling" efforts. As Mike Cernovich, one prominent alt-right figure involved in both earlier campaigns against

---

[23] *See generally* GABRIELLA COLEMAN, HACKER, HOAXER, WHISTLEBLOWER, SPY: THE MANY FACES OF ANONYMOUS (2015).
[24] *See* Anna North, *Opinion | When a SWAT Team Comes to Your House*, The New York Times, July 6, 2017, https://www.nytimes.com/2017/07/06/opinion/swatting-fbi.html (last visited Nov 21, 2017).
[25] *See* Alice Marwick & Rebecca Lewis, *Media manipulation and disinformation online*, New York: Data & Society Research Institute (2017), *and* Ben Schreckinger, World War Meme POLITICO Magazine, https://www.politico.com/magazine/story/2017/03/memes-4chan-trump-supporters-trolls-internet-214856 (last visited Nov 21, 2017).
[26] *See id.*; Bob Dreyfuss, *Seth Rich, Conspiracy Theorists, and Russiagate "Truthers,"* The Nation, 2017, https://www.thenation.com/article/seth-rich-conspiracy-theorists-and-russiagate-truthers/ (last visited Nov 21, 2017).
[27] *See* Marwick, *supra* note 26.

feminists in the video-game industry and in the 2016 election, put it, "troll tactics" were a means with which to "build [his] brand."[28]

The involvement of these communities in targeted campaigns of political information highlights the important point that these three sources of political information - state-run, financially-driven, and trolling - do not operate independently. Instead, numerous ties link these engines of online disinformation into an ecosystem of overlapping, occasionally cooperating groups. Notably, state-run efforts coordinated by Russia leveraged paid agents who in turn worked to infiltrate and mobilize online communities to spread political disinformation.[29] Similarly, state-run efforts also subsidize and support a variety of financially motivated media channels to spread "fake news" and disinformation through the web.[30] These groups also operate on their own, acting independently for their own reasons to engage in the distribution of disinformation.

The (Ambiguous) Impact of Online Disinformation

While all the activities discussed in this section are well documented, it is important to recognize that, at the time of writing, clear empirical evidence of their actual influence over political outcomes is still unclear. While some researchers have concluded that disinformation efforts did have an impact on the 2016 US presidential election, the issue remains a matter of scholarly debate.[31] Given the limited visibility into the operations of various disinformation activities and the data around overall political

---

[28] Andrew Marantz, *Trolls for Trump*, The New Yorker, 2016, https://www.newyorker.com/magazine/2016/10/31/trolls-for-trump (last visited Dec 17, 2017).

[29] *See, e.g.,* Maya Kosoff, *The Russian Troll Farm That Weaponized Facebook Had American Boots on the Ground,* The Hive, https://www.vanityfair.com/news/2017/10/the-russian-troll-farm-that-weaponized-facebook-had-american-boots-on-the-ground (last visited Nov 21, 2017).

[30] *See, e.g.,* Aubrey Belford Dojčinović Saska Cvetkovska, Biljana Sekulovska and Stevan, *Leaked Documents Show Russian, Serbian Attempts to Meddle in Macedonia*, OCCRP, https://www.occrp.org/en/spooksandspin/leaked-documents-show-russian-serbian-attempts-to-meddle-in-macedonia/ (last visited Nov 21, 2017).

[31] *See, e.g.,* Philip Howard & Bence Kollanyi, *Social media companies must respond to the sinister reality behind fake news*, The Observer, September 30, 2017, http://www.theguardian.com/media/2017/sep/30/social-media-companies-fake-news-us-election (last visited Nov 21, 2017); Bence Kollanyi, Samantha Bradshaw & Lisa-Maria Neudert, *Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States?*, *available at* http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/09/Polarizing-Content-and-Swing-States.pdf.

participation on social media and other platforms, it is likely that this issue will remain ambiguous for some time.

If they are indeed effective, the potential risk to democratic institutions and processes seem clear. The capability of foreign powers to effectively manipulate political discourse within a country raises difficult questions about the representativeness of elected officials and the decisions made by them. To the extent that much disinformation seen during the 2016 US presidential campaign focused on exacerbating political conflict and cementing polarization, such activities might also erode the ability for democracies to effectively act as engines for compromise between segments of society.[32] But, evidence on this front is ambiguous. It is unclear that the Internet is in fact increasing polarization.[33] Moreover, it is unclear whether a more partisan media writ large is in turn making the public more polarized.[34]

However, regardless of whether or not they are indeed effective, these politically-targeted activities - and public knowledge about them - still may raise threats to the health of democratic processes. Disinformation campaigns might accelerate erosion in public trust of institutions seen as critical to the maintenance of democracy. First, skepticism around the veracity of online information generally might also limit the influence of journalistic channels producing and distributing accurate information.[35] This may hinder the ability for democracies to engage in authentic, effective deliberation and arrive at decisions considered "legitimate".[36]

---

[32] *See* Diana Epstein & John David Graham, Polarized politics and policy consequences (2007), 17-18, *available at* https://www.rand.org/content/dam/rand/pubs/occasional_papers/2007/RAND_OP197.pdf (reviewing research on the impact of political polarization).

[33] *See, e.g.,* Levi Boxell, Matthew Gentzkow & Jesse M. Shapiro, *Greater Internet use is not associated with faster growth in political polarization among US demographic groups*, 114 PNAS 10612–10617 (2017) (showing that age cohorts with greater exposure to the Internet do not show higher levels of polarization).

[34] *See, e.g.,* Markus Prior, *Media and Political Polarization*, 16 Annual Review of Political Science 101–127 (2013).

[35] *See generally* Michael Barthel & Amy Mitchell, *Democrats, Republicans now split on support for watchdog role Pew Research Center's Journalism Project* (2017), http://www.journalism.org/2017/05/10/democrats-republicans-now-split-on-support-for-watchdog-role/ (last visited Nov 21, 2017) (reviewing recent levels of trust in media and social media).

[36] For a review of theories of democracy based on the role of deliberation*, see generally* JOHN S. DRYZEK, DELIBERATIVE DEMOCRACY AND BEYOND: LIBERALS, CRITICS, CONTESTATIONS (1 edition ed. 2002). *See also* Hannah Arendt, *Lying in Politics: Reflections on The Pentagon Papers*, The New York Review of Books, 1971,

Second, regardless of actual effectiveness, a broadly held perception that these disinformation campaigns do indeed have an impact may itself create distrust in the legitimacy of elected officials, particularly those supported by foreign governments and interests. This has been the case in the aftermath of the 2016 campaign, with numerous congressional inquiries and an ongoing special investigation attesting to the continued concerns by policymakers and the public as a whole

Though ambiguity still exists, these risks and others have encouraged a live debate as to the set of responses - regulatory or otherwise - needed to combat these activities and limit their potential influence on the media and information ecosystem. These proposals confront the framework of CDA 230.

PART II: HOW DOES CDA 230 SHAPE EFFORTS TO COMBAT ONLINE POLITICAL DISINFORMATION?

Because of its potential threat to democratic processes and institutions, policymakers and scholars have begun to propose a range of legal and regulatory responses to online political disinformation. As is the case in other contexts, attention has turned towards the central role that online platforms play in hosting and facilitating the objectionable activity.[37] One recent examination of online media and sharing behavior during the 2016 election season concluded simply that "[d]isinformation and propaganda are rooted in partisanship and are more prevalent on social media".[38] Specifically, the study found that the set of websites which receive a disproportionate amount of attention on Facebook were also cited by independent sources and media reporting as creators and distributors of "inaccurate if not blatantly false reporting."[39]

---

http://www.nybooks.com/articles/1971/11/18/lying-in-politics-reflections-on-the-pentagon-pape/ (last visited Nov 21, 2017).

[37] For a preliminary review of the role that different platforms have played in distributing disinformation, *see* Russ Feingold et al., Stanford Law School Law and Policy Lab, *Fake News & Misinformation: The role of the nation's digital newsstands Facebook, Google, Twitter, and Reddit, available at* https://www-cdn.law.stanford.edu/wp-content/uploads/2017/10/Fake-News-Misinformation-FINAL-PDF.pdf.

[38] Robert Faris et al., *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 US Presidential Election*, (2017).

[39] *Id.* at 15.

Across a range of issues, these platforms serve as "least cost avoiders" - actors best positioned to manage the risk from certain activities. In the online disinformation context, platforms possess highly granular data about the activity across their services, and have the ability to influence the distribution of content. This might be done algorithmically by modifying systems of recommendation that promote certain content over others, or financially through the barring of ads supporting certain types of content online. In the case of large platforms such as Google and Facebook, the companies are also perceived to possess the resources and technical competence to effectively create effective systems of detection and mitigation.[40]

Without the cooperation of these platforms, proposals to address political disinformation confront a recurring, more general challenge with the enforcement of policy online, namely the challenge of identifying and pursuing particular actors that break rules.[41] Laws broken by perpetrators of online political disinformation, and new laws that might be passed against these types of activities, will likely be limited by the costly requirements of identifying and enforcing rules against a disparate and continually evolving ecosystem of disinformation perpetrators.

To the extent that legislative and regulatory action will seek to shape the incentives that online platforms have to combat online political disinformation, these efforts will take place in the shadow of CDA 230, which provides strong protections shielding platforms from liability for actions taken by their users. This provision does in part inhibit the effectiveness of existing and novel legal levers that would target online political disinformation.

Two routes remain by which legal action might combat these campaigns while leaving CDA 230 untouched. One, which would build on the legal precedent set by the 9th Circuit in the *Roommates.com* case, would leave courts to engage in line drawing around the degree to which platforms might elicit illegal disinformation activity. Second, legislation and regulation targeting the platforms themselves and focus on changing the information environment surrounding online political disinformation, rather

---

[40] This perception is frequently bolstered by projects launched by the companies themselves. *See, e.g.,* YouTube Help, YouTube Content ID, https://www.youtube.com/watch?time_continue=2&v=9g2U12SsRns (automated systems for detecting IP infringement); Perspective, https://www.perspectiveapi.com/#/ (last visited Nov 21, 2017) (automated system for detecting "toxic" comments online).

[41] *See* LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE 38-60 (1999).

than creating liability for the acts themselves, would also avoid having to amend CDA 230.

A Brief History of CDA 230

Passed in 1996, CDA 230 provides that "[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider", subject to a set of exceptions for criminal, intellectual property, state, and communication privacy laws.[42]

This legislation was passed in response to the decision in *Stratton Oakmont v. Prodigy Services*, a 1995 decision which suggested that online service providers could be held liable for the defamatory content posted by users on their platforms to the extent that they exercised editorial control over that content.[43] This decision represented an application of established common law principles around the liability of distributors and publishers. Under that framework, "distributors" exercising limited editorial control over content they distributed - such as bookstores and libraries - only faced liability for defamation if they had knowledge of the content and failed to remove it. In contrast, "publishers" exercising more active editorial control and judgment - such as newspapers and magazines - were deemed to be liable for defamatory content as if they had originally published it regardless of knowledge.[44] *Stratton Oakmont* raised concerns that platforms would be unsustainable if exposed to liability for the acts of any individual user, and be deterred from taking proactive action to filter for offensive content.[45]

To that end, the original impetus for CDA 230, as evidenced by its caption, was to protect platforms from liability for "Good Samaritan" acts to remove offensive content.[46] However, Congress also had a range of other objectives in the passage of CDA 230, including an intent to "promote the continued development of the Internet and other interactive computer services and other interactive media" and "preserve the vibrant and

[42] 47 U.S.C. § 230.
[43] *See* Stratton Oakmont, Inc. v. Prodigy Services Co., 1995 WL 323710 (N.Y. Sup. Ct. 1995).
[44] *Id.*
[45] *See* Joel R. Reidenberg et al., Section 230 of the Communications Decency Act: A Survey of the Legal Literature and Reform Proposals, 5-6 (2012), https://papers.ssrn.com/abstract=2046230 (last visited Sep 6, 2017).
[46] *Id.* at 7.

competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation".[47]

Over the subsequent two decades, courts reviewing CDA 230 interpreted the doctrine to shield online platforms from liability for a broad range of acts taken by their users.[48] In 1997, the Fourth Circuit in *Zeran v. American Online* concluded that CDA 230 worked to shield platforms from both traditional categories of publisher and distributor liability, rejecting an argument by the plaintiff that the provision only worked to block publisher liability.[49] In 2003, the Ninth Circuit in *Batzel v. Smith* concluded that the phrase "interactive computer services" in CDA 230 was not limited to services providing access to the Internet as in *Zeran* and earlier cases, but also included "any information service or other systems" such as a listserv. [50] Later cases also confirmed that users who were independent of an online service provider could invoke the protection of CDA 230.[51] In 2008, the Fifth Circuit in *Doe v. Myspace* found that CDA 230 immunity applied broadly to tort claims, not just those premised on defamation as in the *Stratton Oakmont* decision.[52] This broad view of CDA 230 was followed a year later by the Ninth Circuit in *Barnes v. Yahoo!*[53]. In that case, the Ninth Circuit clarified that the scope of CDA 230 could apply beyond causes of action sounding in tort to include any cause of action which "inherently requires the court to treat the defendant as the 'publisher or speaker' of content provided by another."[54]

Together, these decisions have established CDA 230 as a broad shield for online intermediaries, and influences the scope of available legal options in combatting disinformation.

CDA 230 Shields Platforms from Acts of Online Disinformation Committed by Users

[47] 47 U.S.C. § 230(a).
[48] For a general review of leading cases in this space, *see* Eric Goldman, The Ten Most Important Section 230 Rulings (2017), https://papers.ssrn.com/abstract=3025943 (last visited Nov 13, 2017).
[49] Zeran v. American Online, Inc., 129 F.3d 32 (4th Cir. 1997).
[50] Batzel v. Smith, 333 F.3d 1018 (9th Cir. 2003).
[51] Barrett v. Rosenthal, 9 Cal.Rptr.3d 142 (Cal. Ct. App. 2004).
[52] Doe v. MySpace, Inc. 528 F.3d 413 (5th Cir. 2008).
[53] Barnes v. Yahoo!, 570 F.3d 1096 (9th Cir. 2009).
[54] *Id.* at 1102.

CDA 230 and its impacts have been controversial, and the provision has remained the target of recurring efforts to modify it in various ways.[55] In the context of online political disinformation, CDA 230 conflicts with efforts to use existing causes of action and create new causes of action that would hold platforms liable for the illegal actions of its users.

There are a number of potentially applicable causes of action. Online political disinformation is often false information about an individual, and to that end might give rise to the tort of defamation or libel. Cases might include activities to spread conspiracy theories such as the sex trafficking "Pizzagate" rumor discussed above.[56] Consistent with the cases discussed above, CDA 230 would prevent an online platform which hosted such defamatory content posted by a user from itself being held liable for defamation.[57]

Online political disinformation might also violate a number of other laws which are less prototypical cases for CDA 230. For instance, under federal law foreign nationals are prohibited from "[m]aking any contribution or donation of money or other thing of value, or making any expenditure, independent expenditure, or disbursement in connection with any federal, state or local election in the United States."[58] This would include efforts by foreign actors to interfere in US elections through the purchase of advertising spreading falsehoods about a particular candidate. Courts reviewing the illegality of advertising in other circumstances beyond the election context have generally refused to impose liability on the platforms which host this material, absent some specific cases applying the holding in *Roommates.com* discussed below.[59] Accordingly, actions by agencies like the Federal Election Commission to enforce these laws against the platforms themselves would confront the limitations of CDA 230.[60]

---

[55] *See* Reidenberg, *supra* note 44, at 46-49 (detailing reform efforts seeking to amend CDA 230 to address online pharmacies, data security, child safety, unsolicited email, foreign judgments of defamation, and more). *See also* Stop Enabling Sex Traffickers Act of 2017, S. 1693, 115th Cong. (2017) (currently being debated at the time of writing).

[56] *See* Robb, *supra* note 5.

[57] *See, e.g.,* Zeran v. American Online, Inc., 129 F.3d 32 (4th Cir. 1997).

[58] 52 U.S.C. § 30121. *See also,* 11 CFR 110.20.

[59] *See* Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12 (1st Cir. 2016) (ads for prostitution); Chicago Lawyers' Committee for Civil Rights under the Law v. Craigslist, 519 F.3d 666 (housing ads violating the Fair Housing Act). *But see infra* text accompanying notes 62-76.

[60] *Cf.* Federal Trade Commission v. Accusearch, Inc., 570 F.3d 1187 (10th Cir. 2009) (suits by an agency to enforce federal law still subject to CDA 230 analysis).

New causes of action would likely face similar barriers, too. In the wake of the 2016 presidential election, California state legislators have proposed a bill which would make it illegal "for a person to knowingly and willingly make, publish or circulate on an Internet Web site…a false or deceptive statement designed to influence the vote on…(a) Any issue submitted to voters at an election, (b) Any candidate for election to public office."[61] Beyond the range of potential First Amendment challenges to laws attempting to make creating or spreading political disinformation illegal, CDA 230 would still work to inhibit the enforceability of those rules on platforms.[62]

The impact of these limitations parallels long-standing critiques of CDA 230. Critics have argued since its passage that the shield provided by CDA 230 makes online platforms less responsive and proactive than they otherwise would in dealing with defamatory content.[63] Where information about the perpetrator of the defamation is scant, victims of defamation may be left without adequate routes for recovery.[64] Similar discussions have played out in the context of enforcing laws against harassment in cyberspace.[65] This challenge may be compounded where the perpetrators of this activity are operating outside of the United States, as it was during the 2016 US presidential election.

It should be observed that CDA 230 only operates to preclude the bringing of a suit against the platform seeking to find it liable as if it was the publisher of the defamatory content. Even if it did not, the fact that many acts of political disinformation will target public figures of various kinds may mean that claims like defamation and libel may as yet be relatively weak legal tools to bring to bear. For instance, even without CDA 230, a successful suit by a public figure would need to meet the standard set under *New York Times Co. v. Sullivan* which requires proof of "actual malice". This is a challenging burden which requires plaintiffs to show that the act was committed with "sufficient evidence to permit the conclusion

---

[61] A.B. 1104, 2017-18 Cal. State Leg. (Cal. 2017).
[62] *See* California A.B. 1104 Opposition Letter, Electronic Frontier Foundation (2017), https://www.eff.org/document/california-ab-1104-opposition-letter (last visited Nov 21, 2017) (noting a number of First Amendment challenges to this type of legislation).
[63] *See* Reidenberg, *supra* note 44, at 26. For a more recent treatment, *see Vanessa S. Brown Barbour, Losing Their License to Libel: Revisiting § 230 Immunity,* 30 BERKELEY TECH. L.J. 1505, 1559 (2015).
[64] For a review of the large literature on this topic. *See id.*, at 29-31.
[65] *See id.*, at 26-27; SARAH JEONG, THE INTERNET OF GARBAGE (2015).

that the defendant in fact entertained serious doubts as to the truth of his publication."[66]

Even in light of the limitations imposed by CDA 230, it is important to underscore that the law does not in effect bar all legal or regulatory interventions that would incentivize platforms to combat online political disinformation. Two routes provide a potential basis for changing the state of play around this issue.

Option One: Court-Driven Regulation via CDA 230

Since *Zeran*, courts have consistently found that CDA 230 provides broad protections against online platforms being held liable for the activities of their users. However, a set of cases suggest that, under certain circumstances, courts may be willing to narrow the scope of the shield provided by CDA 230.

Fair Housing Council of San Fernando Valley v. Roommates.com concerned a claim against a website which provided a service connecting prospective renters with open apartments and rooms.[67] The plaintiffs in the case alleged that the platform violated the federal Fair Housing Act (FHA) by eliciting information about the renters preferences around gender, sexual orientation, and family status. By publishing these preferences and allowing users to choose renters based on this criteria, the suit alleged a violation of FHA provisions which prohibited discrimination by landlords and tenants on this basis. Roommates.com invoked CDA 230, arguing that this would treat the platform as the one engaging in the discriminatory activity.

The Ninth Circuit - adopting a rationale parallel to that of the Seventh Circuit - held that Roommates.com did not receive immunity from CDA 230 since it played the role of a "information content provider."[68] By designing a website registration process which included questions around categories like gender and sexual orientation, and providing a service which

---

[66] New York Times Co. v. Sullivan, 376 U.S. 254 (1964).
[67] Fair Housing Council of San Fernando Valley v. Roommates.com, 521 F. 3d 1157.
[68] *See* Mark D. Quist, *" Plumbing the Depths" of the CDA: Weighing the Competing Fourth and Seventh Circuit Standards of ISP Immunity Under Section 230 of the Communications Decency Act*, 20 Geo. Mason L. Rev. 275 (2012) (reviewing the details of this circuit split).

filtered based on these preferences, the court held that Roommates.com contributed "materially to the alleged illegality of the conduct."[69]

The end result of the holding in *Roommates.com* is that the specific design decisions made around a website can contribute to the determination of whether or not it can claim immunity under CDA 230. Notably, the Ninth Circuit rejected a claim by the plaintiffs that the platform should be liable for discriminatory posts made by users in an open-ended, optional "Additional Comments" text field on user profiles[70]. Since Roommates.com did not solicit a specific type of content in this text field, and published them as written, it was not a co-developer of the content and therefore received CDA 230 immunity for those activities.[71]

The holding in *Roommates.com* has been inconsistently applied across jurisdictions in the years following the decision, leading some scholars to suggest that the case is not settled law and in fact has a "checkered legacy." [72] The Tenth Circuit in *FTC v. Accusearch* adopted the reasoning in *Roommates.com* in 2009, finding that a site that sold illegally acquired phone records was not entitled to CDA 230 immunity because it "specifically [encouraged] development of what [was] offensive about the content."[73] In contrast, the application of *Roommates.com* by the First Circuit in *Doe v. Backpage* in 2016 allowed a site publishing ads for prostitution to obtain CDA 230 immunity, noting that an online platform's decisions in "structur[ing] its website and posting requirements are publisher functions entitled to section 230(c)(1) protection."[74] Such a line of reasoning would seem to significantly limit the holding in *Roommates.com*. [75] These differing applications of the rule raise doubts as to whether existing legal precedent will be a stable basis on which to combat online political disinformation.[76]

---

[69] Roommates.com*,* 521 F.3d at 1168.
[70] *Id.* at 1173-75.
[71] *Id.* at 1175.
[72] Goldman, *supra* note 47 at 2.
[73] Federal Trade Commission v. Accusearch, Inc., 570 F.3d 1187, 1199 (10th Cir. 2009).
[74] Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 22 (1st Cir. 2016).
[75] It should be noted that at the time of writing the holding in *Backpage* is the subject of congressional scrutiny. Two proposed laws, SESTA and FOSTA, would create exceptions to CDA 230 targeted at reversing the outcome in that case. *See* Stop Enabling Sex Traffickers Act of 2017, S. 1693, 115th Cong. (2017); Allow States and Victims to Fight Online Sex Trafficking Act of 2017, H.R. 1865, 115th Cong. (2017).
[76] *See also* Matthew Feuerman, *Court-Side Seats? The Communications Decency Act and the Potential Threat to StubHub and Peer-to Peer Marketplaces*, 57 Boston College Law Review 227 (2016) (discussing the more permissive standard looking at platform

But, if it is applied, the *Roommates.com* holding suggests that - even absent a legislative modification - the immunities provided by CDA 230 might be effectively thinned by courts assessing whether or not activities associated with campaigns of political disinformation should create liability for online platforms. At issue will be the question of whether or not the specific design of the platform makes it "responsible, in whole or in part, for the creation or development of" the offending content.[77] While this is an inquiry that will turn on the particular claim, the service under question, and the type of disinformation activity, the case provides some rough guidelines around what might not receive the benefit of CDA 230 immunity.

At the most basic level, under the *Roommates.com* holding it is unlikely that simply providing a space through which to engage in illegal acts around political disinformation will expose the platforms themselves to liability. As mentioned above, creating an open-ended box for posting content did not constitute co-development, enabling Roomates.com to obtain the benefit of CDA 230 at least for those elements of its platform. On this count, it is likely that simply making available the means of posting - even if leveraged by trolls, bots, and foreign agents to spread disinformation - will be a point on which platforms will be able to claim immunity.

However, the outcome is more ambiguous when considering other features common to web services. To the extent that disinformation campaigns operate through advertising channels provided by the platforms, a claim might be made that companies like Facebook and Google materially contribute to the illegality. Cases applying the *Roommates.com* holding have at times rejected the application of CDA 230 immunity in the advertising context. Merely profiting from advertisements which promote illegal services or are themselves illegal is insufficient by itself to make the platform liable.[78] However, a pricing arrangement which encourages the activity at issue may qualify as a material contribution. Courts have articulated a few possible scenarios here, including offering discounts to the problematic advertising and variable pricing structures that increase the profit of the platform in proportion to the value and volume of the illegal

---

"encouragement", and the higher standard which looks to whether platforms "require the information at issue").

[77] Roommates.com, 521 F.3d at 1174.

[78] *See, e.g.,* Jane Doe No. 1 v. Backpage.com, LLC, 817 F.3d 12, 17 (1st Cir. 2016) (CDA 230 immunity exists even when platform specifically charges for advertisements promoting prostitution).

activity.[79] Variable pricing is the reality online: most large platforms rely on auction based systems for delivering advertising, with many buyers competing to deliver their content to a given user.[80] To the extent that it can be shown that platforms systematically offer advertising which violates the law at a lower rate, there is potentially a claim of contribution which would block the application of CDA 230 immunity.

Also, content is typically curated, shaped, and personalized to users through an algorithmically generated "feed".[81] This might be grounds to make an argument of co-development, particularly when one takes into account the fact that - in response to interest in a single piece of defamatory content - the platform may recommend further defamatory content. In this sense it matches the search functionality in *Roommates.com*, in which the provision of a mechanism which highlighted content based on discriminatory criteria was seen to facilitate the illegal activity in a way that shed immunity under CDA 230.[82] Similar claims might also be made based on issues that have emerged in the advertising targeting context as well, in which the algorithmic generation of targeting criteria facilitates potentially illegal activity.[83]

In spite of this, *Roommates.com* also provides ample opportunity to reframe the curation and co-creation of content in a feed in a more charitable light. Under the holding a "website operator who edits user-created content…retains his immunity for any illegality in the user-created content, provided that the edits are unrelated to the illegality" - examples

---

[79] *See Chicago Lawyers' Committee for Civil Rights under the Law v. Craigslist, 519 F.3d* 666, 672 (applying CDA 230 in part because platform did "not offer a lower price to people who include discriminatory statements in their postings")*;* NPS LLC v. StubHub, Inc., 25 Mass. L. Rptr. 478 (Super. Ct. 2009) (rejected CDA 230 immunity in part because the platform's "revenue increased in direct proportion to the price of the ticket sold", in contrast to a newspaper which "generally charges a fixed price").

[80] *See, e.g.,* About the ad auction - AdSense Help, https://support.google.com/adsense/answer/160525?hl=en (last visited Dec 17, 2017) (explaining the Google auction system); Help Center, https://www.facebook.com/business/help/430291176997542 (last visited Dec 17, 2017) (explaining the Facebook auction system).

[81] *See* Facebook, Welcome to News Feed, https://newsfeed.fb.com/?lang=en (reviewing how the system curates content for users).

[82] *See also* Catherine Tremble, *Wild Westworld: Section 230 of the CDA and Social Networks' Use of Machine-Learning Algorithms*, 86 Fordham Law Review 825 (2017).

[83] *See, e.g.,* Madeleine Varner and Julia Angwin, Facebook Enabled Advertisers to Reach "Jew Haters" ProPublica (2017), https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters (last visited Nov 13, 2017).

include editing for length, correcting for spelling, or removing obscenity.[84] Arguably platforms like Facebook do the same here, simply collaging and presenting the content as posted, as opposed to actually changing its meaning or otherwise "[contributing] to the alleged illegality."[85]

It is also not so clear that simply recommending additional defamatory content based on the interest of a user in defamatory content rises to the level of "co-development." In *Roommates.com,* the court focused on the fact that the platform used impermissibly discriminatory criteria in filtering and delivering apartment listings to users. However, in an effort to distinguish this case from other platforms that might arguably perform the same function, the court observed that the use of a "*neutral* [tool] to carry out what may be unlawful or illicit searches" on an "ordinary search engine" does not expose that service itself to liability.[86] This is arguably the case here: a platform like Facebook does not explicitly solicit and then filter based on defamatory content. This sets the case apart from the design in *Roommates.com*, in which the platform presented a set of pre-defined categories which themselves were characteristics it was illegal to discriminate against. Instead, on a platform like Facebook, the user effectively "searches" on a neutral tool through her browsing behavior, and the feed returns more content responsive to that behavior, regardless of the specific topic.

As this brief analysis highlights, the end result of the *Roommates.com* holding is ambiguous and depends a great deal on the platform in question, a critique that was voiced by the dissent in that case.[87] However, it seems clear that under certain circumstances platforms might not enjoy the benefits of CDA 230 immunity, and there appears to be at least a colorable claim that this would allow liability to be applied to the platforms for at least some of the tactics used by those driving political disinformation campaigns.

Option Two: Legislative Actions Beyond Amending CDA 230

CDA 230 is simultaneously a broad and narrow provision. It is broad in the sense that platforms are shielded from liability against a wide range of illegal acts that their users might perpetrate. At the same time, CDA 230

---

[84] Fair Housing Council of San Fernando Valley v. Roommates.com, 521 F. 3d 1157, 1169.
[85] *Id.*
[86] *Id.*
[87] *Id.* at 1176-89.

is narrow in the sense that it does not preclude a wide range of actions that might address campaigns of political disinformation outside the lever of applying user level liability to the platform. Three recent proposals made by researchers and policymakers focusing on this issue provide examples of the types of activities not inhibited by the framework laid down by CDA 230 and the caselaw interpreting the provision.

First, CDA 230 does not preclude the imposition of transparency requirements which would mandate that platforms disclose information relevant to evaluating the credibility of information. These might be interventions at the level of the user - helping to inform consumers about the provenance and verification of content. This might manifest as rules locking in a set of standards around "dispute flags" that would appear alongside posts online to signal that a story has been contested by an approved fact-checking organization, like those being experimented with at the time of writing.[88] It might also include more extensive disclosures to particular regulators or expert research groups who might then work to enforce rules and inform the public at large.[89] This would formalize the more ad hoc data provided by Facebook and other companies about advertising activity during the 2017 congressional hearings on this issue.[90]

Second, beyond greater transparency about the data platforms have "on hand," CDA 230 does not preclude measures which would mandate that platforms require greater disclosure from their users, as well. Proposals on this front have focused on the processes around online advertising, seen to be one channel for political disinformation in the 2016 US presidential election. Researchers have proposed "know your customer" requirements for online advertisers paralleling similar rules imposed in the financial sector, as well as more stringent rules around the labeling of anonymous

---

[88] *See* Peter Kafka, Facebook has started to flag fake news stories Recode (2017), https://www.recode.net/2017/3/4/14816254/facebook-fake-news-disputed-trump-snopes-politifact-seattle-tribune (last visited Nov 21, 2017).

[89] *See* Renee Diresta & Tristan Harris, Why Facebook and Twitter Can't Be Trusted to Police Themselves POLITICO Magazine, http://politi.co/2zppJMA (last visited Nov 13, 2017) (proposing an agency modeled on financial regulation).

[90] *See* Deepa Seetharaman & Georgia Wells, *Tech Giants Disclose Russian Activity on Eve of Congressional Appearance*, Wall Street Journal, October 30, 2017, https://www.wsj.com/articles/facebook-estimates-126-million-people-saw-russian-backed-content-1509401546 (last visited Nov 21, 2017); Tim Hwang and Samuel Woolley, *The Most Important Lesson From the Dust-Up Over Trump's Fake Twitter Followers*, Slate, 2017, http://www.slate.com/articles/technology/future_tense/2017/06/the_lesson_of_the_dust_up_over_trump_s_fake_twitter_followers.html (arguing for stronger transparency requirements).

and automated accounts.[91] The Honest Ads Act - bipartisan legislation proposed in October 2017 - would require that large online platforms maintain a public file of all electioneering communications beyond a certain monetary threshold[92]. This file would include a copy of the advertisement, targeting data, as well as information about the purchaser of the advertisement.[93] Similar approaches outside the advertising context might attempt to prevent bots and astroturfing by mandating more stringent requirements on the creation of new user accounts and profiles on a service.

Third, CDA 230 does not preclude more dramatic interventions which would change the actual flow of information through platforms. As a means of limiting the influence of online platforms in shaping public discourse, policymakers have called for a form of "net neutrality" to apply to the content layer of the web, such that platforms like "Facebook, Google, and Amazon – like ISPs – should be 'neutral' in their treatment of the flow of lawful information and commerce on their platforms."[94] Other approaches might require that algorithms take into account certain machine-readable indicators of "credibility" in promoting and ranking information.[95]

All three of these approaches operate within the structure of CDA 230, enabling policymakers to address the tactics used by political disinformation campaigns without necessarily applying liability for individual acts to the platforms. This does not mean that they will not be otherwise rendered invalid. Courts have affirmed in a number of cases that algorithmic outputs are an exercise of the First Amendment rights of the platforms themselves.[96]

---

[91] *See* Diresta, *supra* note 82.
[92] The Honest Ads Act, S. 1989, 115th Cong. (2017).
[93] *Id.* at §8.
[94] Al Franken, *We must not let Big Tech threaten our security, freedoms and democracy | Al Franken*, The Guardian, November 8, 2017, http://www.theguardian.com/commentisfree/2017/nov/08/big-tech-security-freedoms-democracy-al-franken (last visited Nov 13, 2017).
[95] *See e.g.,* An Xiao Mina, Knight Prototype Fund Supports the Credibility Working Group MisinfoCon (2017), https://misinfocon.com/knight-prototype-fund-supports-the-credibility-working-group-c3dcc6667569 (last visited Nov 21, 2017) (discussing one initiative to develop "credibility indicators").
[96] *See, e.g., Search King, Inc. v. Google Tech., Inc., No. 02-1457, 2003 WL 21464568, at *4 (W.D. Okla. May 27, 2003). See also,* Eugene Volokh & Donald M. Falk, *First Amendment Protection for Search Engine Search Results–White Paper Commissioned by Google* (2012), *available at* http://www.volokh.com/wp-content/uploads/2012/05/SearchEngineFirstAmendment.pdf (last visited Dec 17, 2017) (reviewing these cases in the context of search results).

Regulation which would shape these outputs will thereby confront these constitutional protections. Interestingly, as Tim Wu has argued, First Amendment protections will not cover the algorithmic outputs of "functional" platforms whose "involvement with information is too distant or mechanical to be speech."[97] These are circumstances in which CDA 230 immunity will be most likely to apply under *Roommates.com* because these platforms do not "materially contribute" to the offending content.[98] To that end, the doctrines are somewhat complimentary since regulations to directly shape algorithmic output will be most likely to survive First Amendment challenge in circumstances where the *Roommates.com* doctrine is likely to block attempts to impose liability on the platform.

Assuming an intervention met these constitutional requirements, these three approaches might enable action to be taken around these threats without a modification of the underlying law alongside the exception articulated in the *Roommates.com* case.

Conclusion: Some Routes Closed, Others Remain Open

Consistent with long-standing critiques of the provision on issues such as defamation and harassment, CDA 230 may provide perverse incentives for platforms to be less proactive on combatting disinformation than would be preferable. It may also make online platforms less active on gathering and sharing information about perpetrators in a way that may hinder efforts to pursue these actors directly.[99]

At the same time, CDA 230 does not function as an absolute bar to action in the space. Under the holding in *Roommates.com*, judicial action might serve to impose liability on platforms to the extent that their specific design rises to the level of "co-development" which would make them complicit in the commission of illegal acts. Furthermore, since CDA 230 narrowly applies to claims which would treat the platform as a "publisher or speaker" of content, it does not conflict with legislative interventions which would place obligations on the platforms directly. Many of the proposals which would require greater transparency, user disclosure, and modifications to underlying content algorithms continue to be open options under the structure of CDA 230.

---

[97] Tim Wu, *Machine Speech*, 161 University of Pennsylvania Law Review 1495, 1521 (2013).
[98] *See supra* text accompanying note 64.
[99] *See* Reidenberg, *supra* note 60.

The question of whether or not to modify CDA 230 to contend with disinformation threats therefore depends on a careful weighing. At issue is whether or not these remaining options are sufficient to meet the threat posed by political disinformation. And, relatedly, the potential practicality, benefit, and cost of modifying CDA 230 to impose individual liability more directly on the platforms themselves.

PART III: SHOULD CDA 230 BE MODIFIED TO ADDRESS POLITICAL DISINFORMATION?

As public discussion around the challenge posed by online disinformation continues, there have been an increasing number of voices advocating for modification or removal of CDA 230. One recent op-ed in the *Financial Times* characterized the provision as a "loophole", arguing simply that platform operators "no longer deserve the sort of blanket exemptions from liabilities that companies in every other industry incur as a cost of doing business."[100] *The Economist* characterized the provision as an "implicit subsidy" for online platforms, and arguing that "giving platforms a free pass is increasingly difficult for regulators and courts: they simply have become too important for the economy and society more generally."[101]

The landscape CDA 230 gives rise to in the disinformation context is complex. This section seeks to assess the argument for modification or elimination of CDA 230 by answering the following questions. First, given the status quo, are the range of possible interventions sufficient to address the threat posed by campaigns of political disinformation? Second, what would be the potential positive and negative impacts produced by such a modification? Third, practically speaking, if one were to modify CDA 230 - what modification would be appropriate to address the challenge posed by political disinformation?

Are Interventions Within the CDA 230 Framework Sufficient?

As discussed in Part II, CDA 230 does not function as a categorical block to potential legal interventions to address the challenges of political

[100] Rana Foroohar, Facebook's self-policing needs an update Financial Times (2017), https://www.ft.com/content/f5d04d7e-9481-11e7-a9e6-11d2f0ebb7f0 (last visited Nov 21, 2017).
[101] Internet firms' legal immunity is under threat, The Economist, 2017, https://www.economist.com/news/business/21716661-platforms-have-benefited-greatly-special-legal-and-regulatory-treatment-internet-firms.

disinformation. Its impact is considerably more specific: it limits interventions which would serve to treat the platform as the publisher or speaker of an act, applying the liability of a given user to the platform as a whole. It does not hinder a range of potential legislative actions to mandate greater transparency from the platforms, enforce more robust disclosure on the part of users, or even modify the mechanics of how information is distributed by services like Facebook or Google. Nor does CDA 230 serve to block potential actions by courts using the precedent set in *Roommates.com* to selectively eliminate immunity. An immediate question is whether or not CDA 230 is merely a distraction. Are the potential tools that are available without modifying CDA 230 sufficient by themselves to contend with modern campaigns of political disinformation?

One significant challenge to regulatory or court-driven action in this space is speed at which online disinformation campaigns are evolving. Russian political disinformation tactics have continuously incorporated new techniques, and their intervention in 2016 represents the culmination of years of previous experimentation in the space.[102] To that end, even if a given fix is successful in limiting the influence of these tactics in the short-term, it may become rapidly obsolete as perpetrators change their approach. A robust solution will be able to adapt quickly as the landscape of strategies evolves, and in that respect it is unclear if judicial decision making or the regulatory proposals discussed above will provide a sufficiently nimble response.

For example, laws that would mandate that platforms require greater disclosure of information from users and advertisers might be quickly rendered a dead letter as perpetrators of these campaigns find new ways to mask their identity. Particularly in the case of sophisticated disinformation campaigns of the most concern such as those seen in the 2016 presidential election, perpetrators of these efforts may have the means by which to mask their involvement through corporate shells and other aliases.[103] We might expect in this case that defining a fixed set of reporting requirements would be easily evaded.

---

[102] *See* Chivvis, Christopher S.. Understanding Russian "Hybrid Warfare": And What Can Be Done About It. RAND Corporation, 2017, *available at* https://www.rand.org/pubs/testimonies/CT468.html (detailing the historical precedent for Russian tactics in 2016).

[103] Issie Lapowsky, Russia Wouldn't Need Trump's Digital Team to Spread Fake News, WIRED, https://www.wired.com/story/russia-trump-targeting-fake-news/ (last visited Nov 21, 2017) (describing such a scenario).

The current focus on advertising may also be too narrow. While tools to block perpetrators of political disinformation from advertising platforms may limit easy access to powerful tools for targeting a given message, it is important to note that these campaigns can proceed even without access to paid promotion. Well-resourced campaigns may have access to the distribution capabilities of state-run media infrastructure, informal promotion relationships existing outside a platform's advertising tools, and "grassroots" supporters willing to spread a given message.[104] Even narrower regimes focusing on limitations around *political* advertising in particular would miss efforts that seek to target and produce conflict outside the context of an election or campaign. The Facebook events staged by Russia to stoke conflict between "Black Lives Matter" and "Blue Lives Matter" groups, for instance, may not be activities hindered by "know your customer"-style laws.[105]

Intervention by the courts to selectively apply CDA 230 immunity without amendment of the underlying language seems similarly fraught. *Roommates.com* and its progeny depend on a highly fact-specific inquiry which turn on the precise design of an online platform. This is likely to leave untouched a number of channels through which campaigns of political disinformation might flow and remain effective. Recall in that decision that the provision of an open-ended text box for posting content was sufficiently "hands off" such that the Roommates.com service was still able to obtain CDA 230 immunity for discriminatory content posted on that portion of the site. Extending such a rule to the political context would mean that platforms might, for instance, be granted immunity for activities occurring on its free-form posting features but not for algorithmic feeds where the platform plays a more active "co-development" role. The outcome might be that platforms take more proactive action to combat disinformation on some portions of their services more than others, leaving the overall problem unchecked.

Platforms may be left with considerable legal ambiguity as to the bounds of what may or may not incur intermediary liability, producing inconsistent implementation of policy across platforms. Beyond the fact-specific nature of decisions based on *Roommates.com*, courts have also applied the rule inconsistently across jurisdictions and in some cases articulated reasoning which would seem to reject the reasoning in that case.

---

[104] *See supra* Part I.
[105] *See supra,* note 6.

A second difficulty is that currently much remains unknown about the societal impacts of political disinformation campaigns, which makes crafting an effective response in the near-term a challenge. For instance, proposals have proliferated that would require better signaling to consumers about the quality and provenance of content they encounter on online platforms.[106] But, it remains unclear whether or not labels indicating when a piece of content has been challenged by a fact-checking organization are indeed effective. One recent study suggests that simple repetition of "fake news" headlines are sufficient to increase user perceptions of accuracy, even when labeled as false or disputed.[107] There is also evidence to suggest an "implied truth" effect, in which labeling "fake news" as such modestly reduces its perceived accuracy while perversely raising the perceived accuracy of disinformation which goes untagged.[108] In the political context, recent experimental results suggest that, even when a political leader's own statements are exposed as disinformation, the impact on actual voting intention may be limited.[109] Further study will be needed to craft a meaningful response to these threats.

Finally, it is unclear if legislative and judicial bodies will have the technical competence to effectively administrate these interventions. Adopting a *Roommates.com* based approach requires courts to play the primary role in interrogating specific design decisions and evaluating the extent to which they contribute to illegal conduct. Given that even specialists in the industry admit that the risks and complexity of managing these systems drive them to be conservative in their attempts to solve disinformation challenges, generalist courts may not do much better.[110]

Crafting a legislative intervention encounters similar hurdles. Many of the proposals discussed above suffer from a lack of sufficient coverage,

---

[106] *See, e.g.,* Santa Clara University, Trust Project Launches Indicators, https://www.scu.edu/ethics/focus-areas/journalism-ethics/programs/the-trust-project/trust-project-launches-indicators/ (last visited Nov 21, 2017).

[107] *See* Gordon Pennycook, Tyrone D. Cannon & David G. Rand, Prior Exposure Increases Perceived Accuracy of Fake News (2017), https://papers.ssrn.com/abstract=2958246 (last visited Oct 12, 2017).

[108] *See* Gordon Pennycook & David Rand, Assessing the Effect of "Disputed" Warnings and Source Salience on Perceptions of Fake News Accuracy (2017), https://papers.ssrn.com/abstract=3035384 (last visited Nov 15, 2017).

[109] *See* Briony Swire et al., *Processing political misinformation: comprehending the Trump phenomenon*, 4 Royal Society Open Science 160802 (2017).

[110] *See, e.g.,* Josh Constine, Facebook security chief rants about misguided "algorithm" backlash TechCrunch, http://social.techcrunch.com/2017/10/07/alex-stamos/ (last visited Nov 21, 2017) (lead security staff at Facebook highlighting that "an understanding of the risks of machine learning (ML) drives small-c conservatism in solving some issues").

making some techniques of disinformation more difficult while continuing to leave others open for exploitation. One option for overcoming this limitation is to expand the scope of legislation to more comprehensively deal with disinformation by, for instance, prescribing certain algorithms which would take defined elements of information quality into account, or mandating the regular audit of algorithmic behavior. But increasing the scope simultaneously expands the level of complexity, requiring legislators to engage more fully with the technical design of platforms at a detailed level. As in the judicial case, it is unclear if the legislative process will be able to do so effectively and in a timely manner.[111]

What Are the Benefits and Potential Costs of Modification?

Modification of CDA 230 overcomes many of the deficiencies that are likely to hinder a regulatory or court-driven approach to the challenges posed by political disinformation. Importantly, exposing an "interactive computer service" to liability for illegal acts taken by users modifies the overall structure of incentives by shifting the burden to the platform. In essence, rather than specifying a detailed set of actions that should be taken to address disinformation, the government would set a priority about the activity to be minimized, and delegate the decisions about how to achieve that end to the platform.

Such an arrangement avoids the challenges faced by the regulatory or court-driven approaches. Platforms are able to rapidly develop and implement measures to mitigate the impact of political disinformation. Importantly, they will be able to change tactics nimbly as the landscape of these campaigns continues to evolve, ensuring a more robust bulwark against these threats moving forwards. Moreover, platforms are also best situated to assess the efficacy of certain proposed solutions and shed light on the current ambiguities about the impact and behavioral mechanisms underlying disinformation. The association of financial risk with failure to address the challenge would promote investment in this research work, and empirically supported interventions based on it. Finally, modification of CDA 230 would shift responsibility to the actors with the technical expertise and deep understanding of the products necessary to develop a nuanced response to political disinformation.

---

[111] Cf. Cade Metz, Google Is 2 Billion Lines of Code—And It's All in One Place WIRED, https://www.wired.com/2015/09/google-2-billion-lines-codeand-one-place/ (last visited Nov 21, 2017) (describing the complex infrastructure for managing Google's code).

That being said, the broad scope of CDA 230 means that modifications will create a range of downstream effects. While amendment or wholesale removal might limit the influence of political disinformation, it may produce a range of harms that will on net make such a change unwise.

First, applying user level liability to platforms may render these services either insolvent or overly reactive in ways that harm freedom of expression. These are in some sense the "classic" arguments against creating exceptions within the CDA 230 framework.[112] On one hand, liability for the acts of any one of a large pool of users may threaten the financial viability of certain platforms.[113] This is particularly the case given the broad scope of "interactive computer service" under legal precedent, covering everything from a small blog or listserv to the biggest platforms operated by companies like Google and Facebook.[114] Larger platforms will have the financial resources and legal expertise to absorb this risk, while smaller businesses and services run by volunteers may not be able to manage litigation based on the acts of their users. One effect may be to further accelerate and reinforce consolidation to the set of largest companies as less well-resourced platforms exit the market or merge with better positioned competitors.

On the other hand, platforms may become overly reactive to even the minute threat of legal liability, favoring removal of potentially offending content by default rather than making a considered evaluation of the risk.[115] Insofar as modification to CDA 230 attempts to confront the challenge of disinformation, the incentive to minimize legal risk might prompt platforms to preemptively remove content which is true and valuable but likely to be a source of controversy. These takedowns may also disproportionately reflect the interests of those most willing and able to pursue legal claims against the platforms. Ironically, the creation of platform liability for disinformation may create another route by which to suppress true information.

Second, modification of CDA 230 may render platforms substantially less transparent and participatory than they would otherwise be. One concern motivating passage of CDA 230 in the 1990s was the notion that there was no practical means by which companies could effectively monitor

---

[112] *See* Reidenberg, *supra* note 44, at 35-37.
[113] *Id.* at 36 (reviewing literature arguing that "Section 230 immunity has allowed companies to explore and develop new services and is necessary to maintain continued innovation").
[114] *See* Batzel v. Smith, 333 F.3d 1018 (9th Cir. 2003).
[115] *See* Reidenberg, *supra* note 44, at 36 (reviewing literature on the freedom of expression impact of CDA 230).

the massive flows of user-generated content published through their services each day.[116] However, as others have pointed out, this limitation has become less daunting with time, as advances in machine learning and processing power have made it more possible to monitor and moderate content at massive scale. Indeed, many such systems are today used to administrate monitoring of child pornography and intellectual property violations, laws which were exempted from the purview of CDA 230.[117]

To the extent that modification of CDA 230 exposes platforms to liability around the disinformation activities perpetrated by their users, it is likely that the same automated, algorithmic approach will be deployed to maximize and accelerate identification and removal of offending content.[118] Adoption of these automated methods to deal with questions of truth, falsity, and information quality may hinder other, alternative models that leverage user and community participation to filter for these criteria.[119] Wikipedia, the collaboratively edited encyclopedia, has been relatively successful in resisting the influence of "fake news" through human moderation.[120] Research indicates that users are driven by a sense of ownership and community identity to take an active role in disputing facts and eliminating falsehoods.[121] The presence of automated algorithms which moderate content can erode these motivations and inhibit the contributions of volunteers.[122] In that respect, automated systems are in tension with community-driven approaches to the problem of disinformation.

---

[116] *See* Zeran v. America Online, Inc., 129 F.3d 327, 331 (4th Cir. 1997) ("The amount of information communicated via interactive computer services is therefore staggering…It would be impossible for service providers to screen each of their millions of postings for possible problems.")

[117] *See, e.g., supra* note 40; Microsoft's PhotoDNA: Protecting children and businesses in the cloud, News Center, https://news.microsoft.com/features/microsofts-photodna-protecting-children-and-businesses-in-the-cloud/ (last visited Nov 21, 2017) (describing one automated system in the combatting child sex trafficking).

[118] *See, e.g.* Fake News Challenge, , http://www.fakenewschallenge.org/ (last visited Nov 21, 2017) (competition to build machine learning models to assist in the detection of "fake news").

[119] *See* Jacob Rogers, *Wikipedia and Intermediary Immunity: Supporting Sturdy Crowd Systems for Producing Reliable Information*, 127 Yale LJF 358 (2017); James Grimmelmann, *The virtues of moderation*, 17 Yale JL & Tech. 42, 104 (2015) (arguing that CDA 230 protects against a "judicially enforced standard of conduct [that] risks…stomping on valuable experiments in self-governance").

[120] *Id.* at 359-62.

[121] *Id.* at 363 (reviewing literature on this topic).

[122] *See* Aaron Halfaker et al., *The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline*, 57 American Behavioral Scientist 664–688 (2013) (describing how automation can create perverse effects that reduce volunteer contributions over time in the context of Wikipedia).

There are qualities of community-driven filtration of disinformation which make it preferable to automated, algorithmic approaches. Algorithms are opaque, possessing hidden biases that can be difficult to ascertain as a user. Similarly, algorithms are designed and maintained by the platform, effectively delegating decisions over truth and falsity to the company operating the service. In contrast, a filtration approach that leverages community debate and moderation can take place in a more transparent manner, and leaves decisions about disinformation to the users. Amending CDA 230 may incentivize platforms to minimize risk by adopting the algorithmic approach, thereby squeezing out better, more participatory options.[123]

Third, as Nicholas Bramble has written, the immunity provided under CDA 230 represents a regulatory strategy to avoid data enclosure and regulatory capture in information infrastructure.[124] Specifically, CDA 230 - and the immunity provided to platforms under Section 512 of the Digital Millennium Copyright Act - positions online intermediaries as a balancing force against the influence of network providers such as Comcast and AT&T on one hand, and the influence of content providers like Disney, Viacom, and the *New York Times* on the other.[125] By limiting platform exposure to user-level liability, "network providers and content owners are no longer the sole entities to determine under what conditions user access, participation, and innovation shall take place within these [online] spaces."[126] These exceptions also arrange financial incentives in a manner which, at least theoretically, positions online intermediaries as advocates for the communicative interests of its users against these other actors.[127] The modification of CDA 230 would rewrite the balance of power between different interests with the ability to shape the broader information infrastructure in an undesirable way. This may outweigh the potential

---

[123] This is not to dispute that there are circumstances in which automation and bots can work productively with community-driven models. *See* R Stuart Geiger, *Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture*, 4 Big Data & Society 205395171773073 (2017).
[124] Nicholas W. Bramble, *Safe Harbors and the National Information Infrastructure*, 64 Hastings LJ 325 (2012).
[125] *Id.* at 364.
[126] *Id.*
[127] *See id.* at 359-61. *But see* Frank Pasquale & Oren Bracha, Federal Search Commission? Access, Fairness and Accountability in the Law of Search (2007), https://papers.ssrn.com/abstract=1002453 (last visited Nov 21, 2017) (contesting the notion that a "balance of power" exists).

benefit gained by addressing the more narrow threats posed by political disinformation.

Meeting the specific challenge of political disinformation with a wholesale repeal of CDA 230 is unwarranted given the broader negative impacts that may result. To that end, the central question is one of tailoring: precisely *what kind* of acts should be targeted by the crafting of an exception to CDA 230? Will the attribution of existing causes of action to the platforms be sufficient, or will new causes of action be needed?

What Modifications are Practicable?

Particularly in the context of regulating information falsehood and quality, crafting an appropriate exception to CDA 230 is challenging. For one, as legal scholar Cass Sunstein has pointed out, "we do not know what a well-functioning marketplace of ideas would look like."[128] Since it is difficult to specify an ideal end state with precision, it becomes similarly difficult to identify the set of incentives that society should impose on the platforms in addressing online disinformation.

For another, there is the difficult and practical choice of ascertaining precisely what causes of action, when taken by an individual user, should expose the platform to liability. There are not many relevant laws which make the spreading of falsehoods illegal. Defamation is one obvious tort that could be given an exception under CDA 230 given that political disinformation often concerns a specific individual's reputation or character. But as discussed above, disinformation campaigns may seek to spread falsehood about a far broader set of topics than those concerning an individual's reputation, and the standard for proving defamation against public figures is particularly high.[129]

Other activities that have been associated with political disinformation campaigns in the past would potentially run afoul of a host of laws, including statutes against cyberbullying and the tort of intentional infliction

[128] Cass R. Sunstein, *Free speech now*, 59 The University of Chicago Law Review 255, 296 (1992).
[129] *See supra* text accompanying note 63.

of emotional distress.[130] Insofar as a disinformation campaign made an effort to acquire and leak information, it might also commit invasion of privacy and violations of a state right to publicity.[131] These are all claims which could create liability for the platform if excepted from CDA 230, and in doing so encourage those platforms to combat perpetrators of these campaigns. While creating exceptions around these collateral acts might indirectly hinder the efficacy of a disinformation effort, they might still fail in addressing the core challenge: media manipulation and the spread of propaganda.

This scarcity of causes of action against individuals who perpetuate disinformation should come as no surprise. The First Amendment heavily limits regulations on the truth, falsity, or quality of information as *content-based* restrictions. The Constitution "demands that content-based restrictions on speech be presumed invalid."[132] In 2012, the Supreme Court examined the specific question on laws against false statements in *United States v. Alvarez.*[133] At issue in that case was the Stolen Valor Act of 2005, which made false statements about decorations awarded by the armed forced punishable by a fine or imprisonment up to six months.[134]

In evaluating the constitutionality of that law, a plurality of the Court noted that "falsity alone may not suffice to bring the speech outside the First Amendment" and rejected the argument that a government "interest in truthful discourse alone [was] sufficient to sustain a ban on speech."[135] The Court argued for battling disinformation with counterspeech stating "[t]he remedy for speech that is false is speech that is true. This is the ordinary course in a free society…[Society is] not well served when the government seeks to orchestrate public discussion through content-based mandates"[136]. Facing "the most exacting scrutiny", laws which would punish the

---

[130] *See generally* David O. Klein & Joshua R. Wueller, *Fake News: A Legal Perspective*, 7-9 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958790 (last visited Oct 12, 2017).

[131] State law rights to publicity have occasionally been cast as an intellectual property claim, attempting plaintiffs to use the exception under 47 U.S.C. § 230(e)(2). *See, e.g,* Cross v. Facebook, CIV 537384, 2016 WL 7785723 (Cal. Super. Ct. May 31, 2016), aff 'd in part and rev'd in part, No. A148623, 2017 WL 3404767 (Cal. Ct. App. Aug. 9, 2017).

[132] Ashcroft v. American Civil Liberties Union, 542 U.S. 656, 660.

[133] United States v. Alvarez, 567 U.S. 709.

[134] *See id.,* at 715.

[135] *Id.,* at 719, 723.

[136] *Id.*, at 727.

distribution of information on basis of its falsity must pass a very high constitutional bar to be permissible.[137]

One approach may be to avoid the question of attempting to regulate against falsehood, or even political falsehood, *per se*. As in *Alvarez*, "some false statements are inevitable if there is to be an open and vigorous expression of views in public and private conversation, expression the First Amendment seeks to guarantee."[138] The threat posed by coordinated campaigns of disinformation like those seen in the 2016 US presidential election and elsewhere is not simply that inaccurate information is being spread in the political realm. The distribution of political falsehoods is a long-standing feature of the history of US democratic institutions.[139]

What is unique is that disinformation is being spread through means which by themselves erode trust in the outcomes of democratic processes and hinders effective discourse around policy. Part of this is the perceived - and potentially actual - advantage that tools such as bots, advertisement microtargeting, the financial resources of a foreign government, and other tactics confer to actors spreading a message and influencing the public. This advantage is independent of whether or not the information being spread is true or false, though in the immediate context it gives rise to the dramatic sense that the current state of affairs is one in which society is "counter[ing] a firehose of falsehood with a squirt gun of truth."[140] Rectifying that balance of power focuses on better equalizing the instrumentalities of discourse. Such a goal may be more tractable politically, legally, and intellectually than defining what the threshold of "truthiness" should be and delegating that to private actors and the government to interpret and enforce.

In short, while it may be difficult to specify concretely what an "ideal" marketplace of ideas looks like, it is more straightforward to articulate and halt what might be considered methods of unfair competition in the marketplace of ideas. Amendments to CDA 230 should be directed towards this aim, rather than the broader objective of encouraging platforms to eliminate political falsehoods from the web writ large. To adopt the latter as a goal raises the risk of exceptions to CDA 230 that are entirely too

---

[137] *Id.*, at 724 (citing Turner Broadcasting System, Inc. v. FCC, 512 U.S. 622, 642).
[138] *Id.*, at 718.
[139] *See, e.g,* Jon Meacham, Thomas Jefferson: The Art of Power, Chapter 30 (Reprint edition ed. 2013) (detailing political lies during the 1800 US presidential campaign).
[140] Christopher Paul & William Courtney, Russian Propaganda Is Pervasive, and America Is Behind the Power Curve in Countering It, https://www.rand.org/blog/2016/09/russian-propaganda-is-pervasive-and-america-is-behind.html (last visited Nov 18, 2017).

broad, and in the very least might require the creation of new causes of action that are of dubious legality under the First Amendment.

This approach would create exceptions to CDA 230 for a number of existing laws and potential new regulations. Portions of the Federal Election Campaign Act (FECA) would be excepted from CDA 230 to block foreign interference into political discourse.[141] FECA prohibits foreign interests from engaging in activities to shape elections through advertising and other electioneering communications as they did during the 2016 US presidential election. By excepting these rules from the immunity provided by CDA 230, platforms would be liable for these acts, and face incentives to minimize or eliminate this activity from their systems.

A range of rules have been proposed which would limit "microtargeting", the use of highly granular data to target messages to users in the elections context and beyond.[142] Some ideas include regulation that would require that data brokers specializing in the collection and distribution of user data provide citizens with the ability to access the dossier compiled about them and opt-out of certain uses.[143] Another proposal would require provision of "comprehensive notices…[of] data processing practices" from those engaging in the collection of voter data.[144] If implemented, these rules might be subsequently reinforced by crafting an exception to CDA 230 which would align the incentives of the platforms with the enforcement of the law. This would increase the level of transparency available to the public around what kinds of targeting are in use, and would block actors unwilling to provide that transparency from the platforms.

Limited exceptions for fraud to be built into CDA 230 might also be justified by the prevalence of unlabeled bots or paid agents purporting to be genuine users for the purposes of persuasion and mobilization. Such an exception would also work to align platforms with the objective of reducing

---

[141] *See* 52 U.S.C. § 30121 and 11 C.F.R. § 110.20

[142] For an overview of some of the proposals around political advertisement microtargeting, *see* Ira Rubinstein, *Voter Privacy in the Age of Big Data* (2014), https://papers.ssrn.com/abstract=2447956 (last visited Nov 18, 2017).

[143] *See id.,* at 41; Executive Office of the President, Big Data: Seizing Opportunities, Preserving Values (2014), *available at* https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf (last visited Dec 17, 2017) (advocating for approaches that give individuals the ability to "participate in the use and distribution of his or her information after it is collected").

[144] *See id.,* at 37.

or eliminating the creation of phony websites imitating or purporting to be local news outlets, as was seen in the 2016 election cycle.[145]

What techniques are "unfair competition in the marketplace of ideas" is rightly a matter of public debate, and such an open-ended inquiry may raise may concerns about the wide range of liability platforms might face. In the very least, the elimination of immunity from platforms that actively support the use of these techniques in influencing public discourse seems warranted and less controversial. Scholars Danielle Keats Citron and Benjamin Wittes have proposed such an approach, proposing an amendment that would explicitly prevent the limitation of liability for "Bad Samaritan" websites and other content hosts that "purposefully encourage" a defined set of illegal acts.[146] While their focus is on issues of sex-trafficking and nonconsensual pornography, a similar approach might be taken to contend with the challenges of political disinformation.

This will be a piecemeal fine-tuning. The incentive to engage in campaigns of political disinformation is not eliminated by simply making these efforts more challenging, and they will continue evolving as they have been in the past. While exceptions to CDA 230 will likely make responses more agile than more rigid regulatory or judicial prescriptions, these campaigns are likely to find new channels through which to operate. We might also expect that the impact of these campaigns may change over time. For example, a public increasingly on guard to the possibility of online disinformation campaigns might result in an overall reduction in the persuasive impact of those campaigns in the future. At the same time, ever improving techniques for fabricating believable fakes in video and other media may increase the persuasive capability of these tactics over time.[147] Enabling an open-ended evolution of these exceptions permits CDA 230 to adapt as our understanding of these techniques and the risk they pose changes over time.

CONCLUSION: THE TWILIGHT OF THE CROWD?

---

[145] *See* Coler, *supra* note 17.

[146] Danielle Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans Section 230 Immunity* 17 (2017), https://papers.ssrn.com/abstract=3007720 (last visited Nov 13, 2017).

[147] *See, e.g.,* Matthias Niessner, Face2Face: Real-time Face Capture and Reenactment of RGB Videos (CVPR 2016 Oral), https://www.youtube.com/watch?v=ohmajJTcpNk (demonstrating the use of machine learning to create believable simulations of political leaders speaking).

The rise of political disinformation, and the pervasiveness of disinformation more generally, represent an unexpected market failure in the figurative online marketplace of ideas. Much of the rhetoric in the early era of social media highlighted the extent to which the organic discourse of many participants - the much vaunted "wisdom of the crowds" - would help to weed out false information and produce a multifaceted representation of "the truth."[148] This was also implicit in the ideology of those designing the platforms now seen to be some of the greatest sources of disinformation. As Ev Williams, co-founder of Twitter, reflected recently, "I thought once everyone could speak freely and exchange information and ideas, the world is automatically going to be a better place…I was wrong about that."[149]

Early successes like Wikipedia did not generalize into a broader principle that crowds could effectively and reliably filter for truth and against falsity.[150] Regardless of its causal impact on voting behavior and political perceptions, the 2016 US presidential election cycle demonstrated in the very least that concerted efforts to spread disinformation can be wildly successful in being shared online, rather than quickly weeded out. Organic filtration by the wisdom of the crowds was less robust against deliberate manipulation than originally expected.[151]

Efforts to amend or eliminate CDA 230 should be seen in the broader context of a retreat from open, participatory approaches to the problem of disinformation. In light of the perceived absence or weakness of a robust crowd, interventions have turned towards managing disinformation through legislative or judicial action or the judgments of private platform intermediaries. Making this shift can and should raise long-standing concerns about the influence and interests of platforms in regulating

---

[148] *See, e.g.,* James Surowiecki, The Wisdom of Crowds (Reprint edition ed. 2005); Seven years after Nature, pilot study compares Wikipedia favorably to other encyclopedias in three languages – Wikimedia Blog, https://blog.wikimedia.org/2012/08/02/seven-years-after-nature-pilot-study-compares-wikipedia-favorably-to-other-encyclopedias-in-three-languages/ (last visited Nov 21, 2017).

[149] David Streitfeld, *"The Internet Is Broken": @ev Is Trying to Salvage It*, The New York Times, May 20, 2017, https://www.nytimes.com/2017/05/20/technology/evan-williams-medium-twitter-internet.html (last visited Nov 18, 2017).

[150] *See, e.g.,* DON TAPSCOTT & ANTHONY D. WILLIAMS, WIKINOMICS: HOW MASS COLLABORATION CHANGES EVERYTHING (2006) (predicting the broader application of the collaborative model used by Wikipedia).

[151] *See* Tim Hwang, *The Madness of the Crowd*, Logic Magazine (2017), https://logicmag.io/01-the-madness-of-the-crowd/ (last visited Nov 21, 2017) (discussing this challenge).

expression.[152] It also raises even longer-standing concerns about the role of government in regulating freedom of expression.[153]

However, the threat from political disinformation, particularly state-supported campaigns, continues to expand worldwide. Russian efforts leveraging these techniques continue to advance, and recent developments suggest that other nations like China are experimenting with the same playbook to advance their interests.[154] In parallel, manipulation by far-right domestic actors continues to advance in the US.[155]

In light of this, well-calibrated modification of CDA 230 may go a long way in helping to give the public and civil society a fighting chance by encouraging platforms to stabilize and balance the marketplaces of ideas they own and operate. Of particular importance is the reduction or elimination of techniques of distribution that - regardless of the truth or falsity of the messages channeled through them - erode trust in public discourse and democratic processes.

Ultimately, the end goal should not be to fully delegate responsibility around the truth value of information to the government or to the platforms. Instead, the primary objective should be the encouragement of publics which are themselves robust against the ever evolving nature of disinformation. If the wisdom of the crowds has been less robust than was expected a decade ago it is in part because the online spaces in which they operate have failed to create the proper circumstances under which they could succeed. Fine-tuning the bounds of CDA 230 represents one step in realizing and revitalizing this original vision.

---

[152] *See* Frank Pasquale, The Automated Public Sphere (2017), https://papers.ssrn.com/abstract=3067552 (last visited Nov 21, 2017).
[153] *Id.*, at 14-15 (responding to these concerns in the context of intermediary liability).
[154] *See* Paul Mozur, *China Spreads Propaganda to U.S. on Facebook, a Platform It Bans at Home*, The New York Times, November 8, 2017, https://www.nytimes.com/2017/11/08/technology/china-facebook.html (last visited Nov 21, 2017); Russian Bots Tweeting Calls To Fire McMaster, Former FBI Agent Says, NPR.org, https://www.npr.org/2017/08/20/544817844/russian-bots-tweeting-calls-to-fire-mcmaster-former-fbi-agent-says (last visited Nov 21, 2017).
[155] *See* Lee Fang & Leighton Akio Woodhouse, Video: How White Nationalism Became Normal Online The Intercept (2017), https://theintercept.com/2017/08/25/video-how-white-nationalism-became-normal-online/ (noting activities following the 2016 election).