State Power to Regulate Social Media Companies to Prevent Voter Suppression

Spencer Overton[*]

Fake social media accounts and ads did not merely polarize the American electorate in 2016 — these tactics also targeted and suppressed Black votes. While African Americans made up just 12.7% of the United States population, Black audiences accounted for over 38% of U.S.-focused ads purchased by the Russian Internet Research Agency and almost half of the user clicks. The social media accounts generally built a following by posing as being African American-operated and by paying for ads that social media companies distributed largely to Black users. Near Election Day, the accounts urged African Americans to "boycott the election." Federal policymakers have failed to respond immediately to enact strong and clear laws to prevent similar deceptive practices and voter-suppression schemes in the future, and thus States should take the initiative. State lawmakers should not be deterred by arguments that Section 230 of the federal Communications Act of 1934 "immunizes" social media companies from State liability. This Essay explains that Section 230 does not limit the power of States to hold social media companies legally responsible for using data collection and algorithms to target protected classes of voters with

 Professor of Law, The George Washington University Law School and President, the Joint Center for Political and Economic Studies. This Essay is dedicated to the memory of Professor Floyd Feeney, a dear friend and fellow election law scholar who mentored me during my early years of teaching at King Hall. Carol Bruch, Naomi Cahn, Danielle Keats Citron, David Fontana, Jim Gardner, Yosef Getachew, Ed Imwinkelried, Pauline Kim, Justin Levitt, Bill Marshall, Laura Murphy, Larry Norden, Dawn Nunziato, Nate Persily, Aaron Rieke, Olivier Sylvain, Dan Tokaji, Ian Vandewalker, LaShawn Warren, Dan Weiner, and Fane Wolfer provided invaluable comments on an earlier draft of this Essay. Exchanges with David Brody, Joan Donovan, Rick Hasen, Darin Johnson, Orin Kerr, Young Mie Kim, Ann Ravel, Bret Schafer, Dan Solove, and Paul Waters also helped develop the ideas in this Essay. Sheya Jabouin and Mark Peterson provided invaluable research assistance. Thanks also for the patience and support of *UC Davis Law Review* editors Sophia Armstrong and Jessica Gillotte.

suppressive ads. By using such techniques, social media companies contribute materially to discrimination and are thus ineligible for Section 230 immunity.

TABLE OF CONTENTS

INTRODUCTION

On Election Day 2016, the operators of the Williams & Kalvin Facebook page — ostensibly two Black men from Atlanta who ran a popular Facebook page focused on Black media and culture — paid for and posted a Facebook ad. The ad proclaimed: "We don't have any other choice this time but to boycott the election. This time we choose between two racists. No one represents Black people. Don't go to vote."[1]

The Election Day Facebook ad discouraging Black voting targeted the advertising categories of those interested in "Martin Luther King, Jr."; "African American Civil Rights Movement (1954-68)"; and "African American history or Malcolm X."[2] A video with the same message appeared on the Williams & Kalvin YouTube account and was also promoted on the Williams & Kalvin Twitter account.

After the November 2016 election, an investigation revealed that the Williams & Kalvin Facebook, Twitter, and YouTube accounts were fake accounts set up and operated by the Russian Internet Research Agency (the "Russian Agency"). The Williams & Kalvin Facebook page started operating at least as early as January 2016.[3] Many of its posts showcased Black achievements, Black dignity, and other positive affirmations of Black community.[4] Over time, regular posts on police violence, disproportionate levels of incarceration, disparate treatment in news media, and other structural inequalities had allowed Williams & Kalvin to establish a significant following among and credibility with Black users.[5]

---

[1] YOUNG MIE KIM, PROJECT DATA, UNCOVER: STRATEGIES AND TACTICS OF RUSSIAN INTERFERENCE IN US ELECTIONS: RUSSIAN GROUPS INTERFERED IN ELECTIONS WITH SOPHISTICATED DIGITAL CAMPAIGN STRATEGIES 9 (2018), https://journalism.wisc.edu/wp-content/blogs.dir/41/files/2018/09/Uncover.Kim_.v.5.0905181.pdf [https://perma.cc/5ZF2-URT5].

[2] *Id.*

[3] *See* Benjamin Fearnow, *Williams & Kalvin: Pro-Trump Facebook Stars Reportedly Worked for Kremlin, Accounts Removed*, INT'L BUS. TIMES (Oct. 10, 2017, 1:51 PM), https://www.ibtimes.com/williams-kalvin-pro-trump-facebook-stars-reportedly-worked-kremlin-accounts-removed-2599559 [https://perma.cc/8V5X-EWR7] (noting the "personal" account for Kalvin Johnson last posted in 2015); Issie Lapowsky, *House Democrats Release 3,500 Russia-Linked Facebook Ads*, WIRED (May 10, 2018, 10:00 AM), https://www.wired.com/story/house-democrats-release-3500-russia-linked-facebook-ads/ [https://perma.cc/PW8C-YBFU].

[4] *See* Josh Russell (@josh_emerson), TWITTER (Oct. 9, 2017, 7:36 AM), https://twitter.com/josh_emerson/status/917398442661605377 [https://perma.cc/S3LU-NVWN] (initiating a Twitter thread of archived posts from disabled social media accounts of Williams & Kalvin).

[5] *See* PHILIP N. HOWARD ET AL., COMPUTATIONAL PROPAGANDA RESEARCH PROJECT, THE IRA, SOCIAL MEDIA AND POLITICAL POLARIZATION IN THE UNITED STATES, 2012-2018,

Fake social media accounts and targeted digital advertising did not just "polarize" the American electorate in 2016. They did not simply facilitate "foreign interference" with U.S. elections. These tactics also targeted and suppressed Black votes.[6]

While African Americans make up just 12.7% of the U.S. population, 37.04% of the unique Facebook pages believed to be created by the Russian Agency were focused on Black audiences,[7] and these pages attracted 35.72% of the followers of the pages created by the Russian Agency.[8] Of the twenty U.S.-focused audience segments that the Russian Agency targeted on Facebook, just two segments — "African American Politics and Culture" and "Black Identity and Nationalism" — accounted for over 38% of the ads purchased, 46.96% of the user impressions, and 49.84% of the user clicks.[9] The Russian Agency paid Facebook 1,350,489 rubles (about $20,257) for 1,087 different ads for these two Black audience segments. The ad campaign resulted in

at 35 tbl.5 (2018), https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report.pdf [https://perma.cc/7XR5-VG8Y] (listing the top twenty Russian Agency-backed Facebook pages).

[6] RENEE DIRESTA ET AL., THE TACTICS & TROPES OF THE INTERNET RESEARCH AGENCY 12, 87-88 (2019), https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1003&context=senatedocs [https://perma.cc/PN9Y-8B3C] ("While other distinct ethnic and religious groups were the focus of one or two Facebook Pages or Instagram accounts, the Black community was targeted extensively with dozens . . . .").

[7] *See id.* at 21 (calculating a total percentage of Black pages at 37.037%, based on numbers indicating that the "Facebook data provided posts from 81 unique pages" (the denominator) and that "[o]verall, 30 targeted Black audiences" (the numerator)); *ACS 2013-2017 Five Year Estimates*, U.S. CENSUS BUREAU (2017), https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_5YR_DP05&src=pt [https://perma.cc/YZW7-ETB6] (indicating a Black population in the United States of 12.7%); *see also* HOWARD ET AL., *supra* note 5, at 6 (indicating that Facebook provided data on 3,393 individual ads published from 2015-2017 that it believed originated from the Russian Agency to the U.S. Senate Select Committee on Intelligence, and the U.S. House Permanent Select Committee on Intelligence released details on 3,517 of such ads).

[8] *See* DIRESTA ET AL., *supra* note 6, at 21 ("The Facebook data provided included posts from 81 unique Pages . . . . Overall, 30 targeted Black audiences and amassed 1,187,810 followers; 25 targeted the Right and amassed 1,446,588 followers, and 7 targeted the Left and amassed 689,045 followers. The remaining 19 were a sporadic collection of pages with almost no posts and approximately 2000 followers across them.").

[9] *See* HOWARD ET AL., *supra* note 5, at 23 tbl.4 (providing raw numbers of the twenty audience segments on Facebook targeted by the Russian Agency, including the two audience segments of "African American Politics and Culture" and "Black Identity and Nationalism").

15,815,597 user impressions (users seeing the ad) and 1,563,584 user clicks (users engaging with the ad).[10]

Similar trends occurred on other platforms. Of all of the U.S.-focused Russian Agency-generated YouTube content, 96% was related to the Black Lives Matter movement and police brutality.[11] The Russian Agency Instagram account with the most interactions was @blackstagram__, with 303,663 followers, over 27.8 million likes, and over 450,000 comments.[12] The Russian Agency also disproportionately focused on African Americans on its Twitter accounts.[13]

While the Russian Agency also created pages and ads that were targeted at and delivered to conservative groups in the United States, those pages warned of voter fraud and encouraged audiences to vote.[14] In contrast, the messages on Black-oriented pages either ignored the election, discouraged African Americans from voting, or encouraged African Americans to vote for a third-party candidate unlikely to win.[15]

The 2016 presidential election marked the most significant decline in Black voter turnout on record — falling from 66.6% in 2012 to 59.6%

---

[10] *See id.*

[11] DIRESTA ET AL., *supra* note 6, at 16.

[12] *Id.* at 27 (showing that the number one Russian Agency account in terms of interactions was @blackstagram__, with 303,663 followers and over 28 million interactions (over 27.8 million likes and over 450,000 comments).

[13] *See* HOWARD ET AL., *supra* note 5, at 26 ("[T]he IRA focused their political messaging [on Twitter] on two targets above others: conservative voters and African Americans.").

[14] DIRESTA ET AL., *supra* note 6, at 83 ("[T]he strategy for Right-leaning groups appears to have been to generate extreme anger and suspicion, in hopes that it would motivate people to vote; posts darkly hinted at . . . voter fraud."); KIM, *supra* note 1, at 8, 10 (indicating that the Russian Agency "deliberately targeted nonwhite voters, particularly African Americans, by promoting their racial/ethnic identity, then suppressing their votes when closer to the elections . . . . No evidence suggested that the same type of voter suppression strategy was also employed on the other side of the political spectrum, however").

[15] *See* DIRESTA ET AL., *supra* note 6, at 83 ("The Black-targeted content . . . largely ignored the election until the last minute, instead continuing to produce posts on themes about societal alienation and police brutality. As the election became imminent, those themes were then tied into several varieties of voter suppression narratives: don't vote, stay home, this country is not for Black people, these candidates don't care about Black people."); HOWARD ET AL., *supra* note 5, at 18 ("Messaging to African Americans sought to divert their political energy away from established political institutions by preying on anger with structural inequalities faced by African Americans, including police violence, poverty, and disproportionate levels of incarceration. These campaigns pushed a message that the best way to advance the cause of the African American community was to boycott the election and focus on other issues instead . . . . This accounts for the majority of content in the dataset that targeted this group.").

in 2016.[16] Political scientists, however, find it difficult to quantify the precise impact of voter deception through online targeted ads on election outcomes relative to other possible factors, such as the absence of a popular Black major-party candidate in the presidential general election.[17]

Nevertheless, the 2016 presidential election established that the targeting of Black communities with deceptive and suppressive ads by social media companies is a tangible threat, and federal officials should craft strong and clear federal guidelines to prevent future problems. Deceptive social media ads are quick-hitting and anonymous, able to be targeted precisely at their intended audience (known as "microtargeting"), and have great potential to "go viral" online ("virality"). Thus they present unprecedented dangers in facilitating voter suppression.[18]

Anonymity facilitates voter suppression through racial impersonation. Instead of using intimidation or regulatory barriers to suppress votes, the anonymity of social media allows anyone to pose as a community member, build trust, and later undermine community interests — all while avoiding responsibility for misinformation and suppression.[19] The "Blackface" is not used to lampoon and mock but to

[16] Jens Manuel Krogstad & Mark Hugo Lopez, *Black Voter Turnout Fell in 2016, Even as a Record Number of Americans Cast Ballots*, PEW RES. CTR. (May 12, 2017), https://www.pewresearch.org/fact-tank/2017/05/12/black-voter-turnout-fell-in-2016-even-as-a-record-number-of-americans-cast-ballots/ [https://perma.cc/HS77-VBG4].

[17] *See, e.g.*, DIRESTA ET AL., *supra* note 6, at 58 ("When we talk about the 'impact' of the Russian influence operation, most conversations focus on whether the Russian Agency operation swayed voters and swung the Presidential Election in 2016. The answer is, we can't tell from this data.") (emphasis omitted); Scott Shane & Sheera Frenkel, *Russian 2016 Influence Operation Targeted African-Americans on Social Media*, N.Y. TIMES (Dec. 17, 2018), https://www.nytimes.com/2018/12/17/us/politics/russia-2016-influence-campaign.html [https://perma.cc/FL6M-NEA4] ("Black voter turnout declined in 2016 for the first time in 20 years in a presidential election, but it is impossible to determine whether that was the result of the Russian campaign.") (emphasis omitted).

[18] *See* NATHANIEL PERSILY, KOFI ANNAN FOUND., THE INTERNET'S CHALLENGE TO DEMOCRACY: FRAMING THE PROBLEM AND ASSESSING REFORMS 5-6, 21-22 (2019), https://pacscenter.stanford.edu/wp-content/uploads/2019/03/a6112278-190206_kaf_democracy_internet_persily_single_pages_v3.pdf [https://perma.cc/5XA9-9CQ3] (discussing the velocity, virality, and anonymity of online communications, as well as the power of Google and Facebook platforms).

[19] *See id.* at 16 ("For purposes of democratic discourse . . . the pervasiveness of internet anonymity facilitates kinds of speech that are harmful to democracy, hinders audiences' capacity to discount messages by the identity of the speaker . . . . Consequently, the speaker bares no cost for repeating lies and promoting false content."). *Voter suppression* is defined as an effort to support a favored candidate by

infiltrate, deceive, and dilute political influence. While engaging in deceptive practices is not a new tactic of those seeking to suppress votes, targeted ads on social media platforms super-charge the tactic.

As discussed below, microtargeting gives social media platforms an unprecedented opportunity to gather information about the preferences and interests of individuals and to build audiences of color to target with ads designed by third parties to appeal to them, deceive them, and suppress votes.[20] The velocity and virality of deceptive, suppressive ads allow them to be quickly disseminated and shared among networks of friends before the deceptive ads can be rebutted — sometimes just before Election Day.[21] Preventing deceptive, suppressive targeted ads becomes increasingly important as the electorate grows more racially diverse and shifts more of its political engagement to social media platforms.

Recognizing that Congress and federal agencies have failed to act immediately to prevent this problem, States should take the initiative.

impeding participation by voters believed to oppose the favored candidate through deception, intimidation, or other means. *See* CRAIG C. DONSANTO, FEDERAL PROSECUTION OF ELECTION OFFENSES 56 (Richard C. Pilger ed., 8th ed. 2017), https://www.justice.gov/criminal/file/1029066/download [https://perma.cc/35ZP-XDZR] (defining voter suppression, and citing examples as providing false information about the time, place, and manner of elections or voter qualifications, and jamming the phone lines of voter mobilization organizations); Gilda R. Daniels, *Voter Deception*, 43 IND. L. REV. 343, 358-59 (2010) (explaining that voter suppression "seeks to decrease the number of eligible voters and, generally, take the electoral power away from individuals or groups; it also often uses deception or threats to accomplish this goal"). While some limit the definition of *deceptive practices* to "knowingly deceiving voters regarding the time, place, or manner of conducting elections" or voter qualifications, many states also have laws prohibiting the deception of voters regarding candidates or issues. *Id.* 354, 369 (describing various types of state voter deceptive practices laws, including those that prohibit dissemination of "false information on candidates or issues, such as making a false statement about a candidate or a proposition"). As discussed briefly below, regulating the broader definition of deceptive practices raises more extensive First Amendment issues.

[20] *See* PERSILY, *supra* note 18, at 21-23 ("While targeted advertising is as old as advertising, microtargeting in the digital age represents an extreme difference in degree if not in kind . . . . [T]he internet enables unprecedented gathering of information on individuals (including search histories, friendship networks, and buying habits) and therefore the crafting of messages designed to appeal to their particular preferences and prejudices.").

[21] *See id.* at 11 ("As bad as the rapid dissemination of falsehoods may be, it is compounded by the inability to timely correct or combat disinformation . . . . A correction is unlikely to reach either the same audience . . . . The speed of information transfer poses particular challenges for democracy, because elections occur at a certain period in time.").

States routinely enact civil rights antidiscrimination measures.[22] Indeed, States can go even further than the federal government in the regulation of private behavior and protecting civil rights because they are governments of general, not limited, powers.[23] While no federal law directly criminalizes deceptive practices that result in voter suppression,[24] several States have such laws.[25]

[22] *See Discrimination and Harassment in the Workplace*, NAT'L CONF. ST. LEGISLATURES, https://www.ncsl.org/research/labor-and-employment/employment-discrimination.aspx (last visited Feb. 10, 2020) [https://perma.cc/M5H8-ESDC]; *State Public Accommodation Laws*, NAT'L CONF. ST. LEGISLATURES, https://www.ncsl.org/research/civil-and-criminal-justice/state-public-accommodation-laws.aspx (last visited Feb. 10, 2020) [https://perma.cc/8PUN-587Z].

[23] *See generally* U.S. CONST. amend. X ("The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people."); Ry. Mail Ass'n v. Corsi, 326 U.S. 88, 94 (1945) ("We see no constitutional basis for the contention that a state cannot protect workers from exclusion solely on the basis of race, color or creed by an organization, functioning under the protection of the state, which holds itself out to represent the general business needs of employees."); Commonwealth v. Alger, 61 Mass. 53, 85 (1851) ("It is much easier to perceive and realize the existence and sources of [state police] power, than to mark its boundaries, or prescribe limits to its exercise."). Also, the U.S. Constitution explicitly gives each State the authority to regulate the manner of federal elections unless Congress legislates otherwise, and States possess the inherent authority to regulate State and local elections. U.S. CONST. art. I, § 4 ("The Times, Places and Manner of holding Elections for Senators and Representatives, shall be prescribed in each State by the Legislature thereof; but the Congress may at any time by Law make or alter such Regulations . . . ."). The U.S. Supreme Court has declared that a State has "a compelling interest in preserving the integrity of its election process." Burson v. Freeman, 504 U.S. 191, 199 (1992) (quoting Eu v. San Francisco Cty. Democratic Cent. Comm., 489 U.S. 214, 231 (1989)). In regulating State and local elections, States must comply with federal constitutional provisions prohibiting discrimination on the basis of race and gender, prohibiting poll taxes, and prohibiting age restrictions on those eighteen and older. U.S. CONST. amends. XV, XIX, XXIV, XXVI.

[24] Daniels, *supra* note 19, at 359 ("Although voter intimidation and deception are similar and statutes exist specifically for intimidation and fraud, no federal legislation directly addresses deception.").

[25] *See, e.g.*, ALA. CODE § 17-17-38 (1975) (considering it a misdemeanor for a person to use corrupt means to attempt to influence or deter a voter from giving his or her vote, or to hinder "the elector in the free exercise of the right of suffrage . . . ."); ALASKA STAT. § 15.56.14 (2010) (prohibiting knowingly making false statements about a candidate that would "provoke a reasonable person to . . . construe as damaging the candidate's reputation . . . ."); COLO. REV. STAT. § 1-13-109 (2005) (considering it a misdemeanor to knowingly disseminate false statements designed to affect the vote related to any issue or candidate running for public office); FLA. STAT. § 104.0615 (2008) (including in the purview of the statute false information to induce or compel an individual to vote or refrain from voting); IDAHO CODE § 18-2305 (1972) (determining that "[a] person who . . . defrauds any elector . . . is guilty of a misdemeanor"); 10 ILL. COMP. STAT. 5/29-4 (2003) (penalizing "[a]ny person who, by . . . deception . . . knowingly prevents" another from voting or registering to vote); KAN. STAT. ANN. § 25-2415 (1974)

States should adopt explicit standards of liability for social media companies that develop algorithms and collect data to build audiences of color and then target them with deceptive ads that suppress voting.[26]

(including the mailing or publishing of false information as proscribed voter intimidation); LA. STAT. ANN. § 18:1463 (2004) (precluding the dissemination of any "oral, visual, or written material containing . . . a false statement about a candidate . . . or about a proposition."); MD. CODE ANN., ELEC. LAW § 16-201 (2009) (maintaining that "[a] person may not willfully and knowingly influence or attempt to influence a voter's decision through . . . fraud"); MINN. STAT. § 204C.035 (2006) (prohibiting a person from "knowingly deceiv[ing] another person" about election information); N.M. STAT. ANN. § 1-20-9 (2009) (prohibits "printing, causing to be printed, distributing or displaying false or misleading" information relating to the voting or election process); S.C. CODE ANN. § 7-25-190 (1994) (considering it a felony for a person to use force, intimidation, deception, undue influence or fraud to control the vote of any voter); VA. CODE ANN. § 24.2-1005.1 (2007) (considering it a misdemeanor to "[knowingly] communicate . . . false [election] information [to a registered voter] . . . about the time, date and place of [voting] or the voter's precinct, polling place or registration status"); W. VA. CODE § 3-9-10 (2002) (declaring that "[a]ny person who shall, by . . . fraud . . . prevent or attempt to prevent any voter . . . from freely exercising his right of suffrage at any election" is guilty of a misdemeanor); WIS. STAT. § 12.05 (2004) (prohibiting knowingly making false representations about a candidate or referendum with the intent of affecting voting at an election). For references to many of these statutes, see Daniels, *supra* note 19, at 369-71.

[26] *See* WENDY WEISER & VISHAL AGRAHARKAR, BRENNAN CTR. FOR JUSTICE, BALLOT SECURITY AND VOTER SUPPRESSION: WHAT IT IS AND WHAT THE LAW SAYS 9 (2012), https://www.brennancenter.org/sites/default/files/2019-08/Report_Ballot_Security_Voter_Suppression.pdf [https://perma.cc/3FM3-BM3G]. *See generally* CTR. FOR THE ADVANCEMENT OF PUB. INTEGRITY, PROSECUTING VOTE SUPPRESSION BY MISINFORMATION (2019), https://www.law.columbia.edu/sites/default/files/microsites/public-integrity/files/voter_suppression_final.pdf [https://perma.cc/6MKP-H4GS] [hereinafter PROSECUTING VOTE SUPPRESSION] (discussing various forms of voter suppression, the historical development of voting rights in the United States, current voting protections, and potential policy options for reducing contemporary voter suppression); COMMON CAUSE, DECEPTIVE ELECTION PRACTICES AND VOTER INTIMIDATION: THE NEED FOR VOTER PROTECTION (2012), https://lawyerscommittee.org/wp-content/uploads/2015/07/DeceptivePracticesReportJuly2012FINALpdf.pdf [https://perma.cc/UZM3-UEA8] (providing an updated analysis of the various ways deceptive practices have been used to manipulate voters and providing policy options to mitigate their use and effects); COMMON CAUSE ET AL., DECEPTIVE PRACTICES 2.0: LEGAL AND POLICY RESPONSES (2008), https://www.commoncause.org/wp-content/uploads/2018/03/0064.pdf [https://perma.cc/TB6V-WW49] (discussing the various ways deceptive practices have been used to manipulate voters and providing policy options to mitigate their use and effects); Daniels, *supra* note 19, at 348-49 ("Voter deception involves, inter alia the distribution of misinformation regarding the time, place, and manner of elections as well as voter eligibility. These deceptive practices regularly have as their main objective to misinform unwanted minority, elderly, disabled, and language-minority voters in an effort to suppress votes.") (footnote omitted); Nichole Rustin-Paschal, *Online Behavioral Advertising and Deceptive Campaign Tactics: Policy Issues*, 19 WM. & MARY BILL RTS. J. 907 (2011).

Social media companies are best positioned to monitor their platforms and prevent these problems. Governments lack the technical expertise and resources to consistently and effectively detect fake accounts and police voter-suppression schemes, especially in an ever-evolving social media landscape. The third-party users that buy ads to promote their posts may be hard-to-detect, unaccountable foreign or domestic special interests such as the Russian Agency, a domestic hate group, a dark-money 501(c)(4) group, an underfunded shell organization, or an unknown individual with modest resources. In contrast, social media companies — especially those with an audience large enough to swing election outcomes, such as Facebook or Twitter — are generally easier to hold accountable because they are often more identifiable and better resourced. Social media companies also are best positioned to detect fake accounts and reject ads designed to suppress votes.

Social media companies also make material contributions to the discriminatory distribution itself: They design platforms that collect data from users, build defined demographic audiences for advertisers, decide which users will see suppressive ads based on algorithmic predictions about which users will click on and engage with an ad, and profit from ads designed to suppress votes of underrepresented communities. Social media companies are also best situated to use Data Protection Impact Assessment tools to identify the risks of discrimination in ad distribution and make revisions to algorithms before they are deployed.[27]

Several other scholars have grappled with unresolved First Amendment issues confronting regulation of deceptive political speech.[28] The United States Supreme Court has suggested that a State

[27] *See* Pauline T. Kim, *Manipulating Opportunity*, 106 VA. L. REV. (forthcoming 2020) (manuscript at 48), https://ssrn.com/abstract=3466933 [https://perma.cc/84BJ-7P9M] [hereinafter *Manipulating Opportunity*] ("[Impact assessments] can force entities to 'think hard' about . . . 'collateral effects' . . . [and] permit interventions in the design and model-building stages, thereby avoiding sources of unfairness or bias before they are baked in."). *See generally* Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L.J. 143 (2019) (discussing the development of the GDPR and its interaction with data protection and algorithmic learning); Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529 (2019) (discussing the development of algorithm accountability in the EU under the GDPR); Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109 (2017) (discussing generally different application of data driven systems and how and in what ways they are susceptible to unconscious biases).

[28] Daniels, *supra* note 19, at 372-380 (reviewing various First Amendment concerns with voter deception); Richard L. Hasen, *Deep Fakes, Bots, and Siloed Justices: American*

could "prohibit messages intended to mislead voters about voting requirements and procedures."[29] Less clear is the constitutional status of regulations of other deceptive and suppressive activity, such as a social media ad from someone falsely posing as African American that states, "Democrats don't care about us. . . . Let's protest and not vote."[30]

This Essay addresses a different critical issue and explains why federal statutory law does not limit the power of States to hold social media companies legally responsible for using data collection and algorithms to target protected classes of voters and deliver suppressive ads to them. Section 230 of the federal Communications Act of 1934 (commonly known as the Section 230 of the Communications Decency Act) immunizes website operators such as Facebook, Twitter, and YouTube from liability for claims based on content created by third-party users — such as a post by a person with a Facebook page or a tweet by a person with a Twitter account. Some social media companies may argue

Election Law in a "Post-Truth" World, ST. LOUIS U. L.J. (forthcoming 2020) (manuscript at 11) (asserting that regulation of political communications would be "subject to heightened First Amendment scrutiny," and proposing disclosure of deceptive practices); William Marshall, *Internet Service Provider Liability for Disseminating False Information About Voting Requirements and Procedures*, OHIO ST. TECH. L.J. (forthcoming 2020) (examining "the constitutionality of deceptive campaign practices acts under" the First Amendment including various cases invalidating and upholding regulation of false statements, and also analyzing whether internet service providers may be held liable for deceptive campaign messages); Daniel P. Tokaji, *Bullshit, Democracy, and the Limits of Law*, ST. LOUIS U. L.J. (forthcoming 2020) (asserting that "under *New York Times v. Sullivan* and its progeny, the government can prohibit defamation of public officials and public figures, if it satisfies the actual malice standard" and "[o]utside the realms of defamation and the polling place, bans on false campaign speech would have to meet strict scrutiny"); *see also* Lori A. Ringhand, *First Amendment (Un)Exceptionalism: A Comparative Taxonomy of Campaign Finance Reform Proposals in the US and UK*, OHIO ST. L.J. (forthcoming 2020) (asserting that some false and misleading campaign statements may constitutionally be prohibited).

[29] Minn. Voters All. v. Mansky, 138 S. Ct. 1876, 1889 n.4 (2018) ("We do not doubt that the State may prohibit messages intended to mislead voters about voting requirements and procedures.").

[30] *See* Care Comm. v. Arneson, 766 F.3d 774, 785 (2014) (invalidating Minnesota statute prohibiting false campaign speech about a ballot initiative because the law was not narrowly tailored); Susan B. Anthony List v. Driehaus, 573 U.S. 149 (2014), *remanded to* 814 F.3d 466 (6th Cir. 2014) (invalidating an Ohio false campaign speech statute); United States v. Alvarez, 567 U.S. 709, 719 (2012) (holding that false statements may receive First Amendment protection even when made with knowledge of falsity or reckless disregard to their truth). *But see* Linert v. MacDonald, 901 N.W.2d 664, 670 (Minn. Ct. App. 2017) (holding that Minnesota's law prohibiting false statements about endorsements was not unconstitutionally overbroad because counterspeech was not an effective means to combat false information, particularly just before an election).

that Section 230 immunizes them from liability for violations of State laws regulating discriminatory targeting by suppressive ads.

This Essay explains why this argument is flawed. Section 230 does not limit the power of States to hold social media companies responsible for using data collection and algorithms to engage in discriminatory dissemination of deceptive ads that suppress voting. By engaging in this activity, social media companies are not simply acting as neutral platforms that passively post the information of third parties — like a Facebook post or a tweet. Instead, through data collection and algorithms that identify which users see the suppressive ads, social media companies make a "material contribution" to the illegal racial targeting. The discrimination is facilitated not simply by the third party's ad content (e.g., a post impersonating a Black activist that says, "Don't go to vote"), but also by the social media company's microtargeting conduct that steers the ad to the feeds of Black users and away from the accounts of others.

Part I of this Essay details the scope of Section 230 and explains the boundaries of Section 230's protection from liability. Part II explains that Section 230 does not immunize social media companies from liability for using data collection and algorithms to identify and target protected classes of voters and deliver suppressive ads to them, and thus States possess the power to establish and enforce election laws to hold social media companies accountable. Part III reviews next steps, which include tailoring State deceptive-practices laws to be both effective and constitutional, as well as ultimately enacting strong and clear federal rules that protect civil rights in voting, employment, housing, and financial services in the online landscape.

I. THE LIMITS OF SECTION 230

Section 230 of the Communications Act of 1934 immunizes "interactive computer services" such as Facebook, Twitter, and YouTube from liability for claims based on content created entirely by third-party users, such as a person with a Facebook page or a Twitter or YouTube account.[31]

Section 230(c) provides:

> Protection for "Good Samaritan" blocking and screening of offensive material

[31] *See* 47 U.S.C. § 230(f)(2) (2019) (indicating that an "interactive computer service" is "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server").

> (1) Treatment of publisher or speaker. No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.
>
> (2) Civil liability. No provider or user of an interactive computer service shall be held liable on account of — (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).[32]

Section 230(c)(1) attempts to protect online platforms from being liable as publishers or speakers due to the content of a third party by stating: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."[33] In addition to facilitating the growth of the internet by limiting platform liability for third-party content, Section 230 ensures platforms can freely remove unsavory third-party content without fear of becoming "publishers" who are suddenly liable for all third-party content.[34] Section 230 also allows for the development of movements like Black Lives Matter whose members make controversial allegations against powerful interests, because social media platforms can display this content without fear of being sued.[35]

---

[32] *Id.* § 230(c).

[33] *Id.* § 230(c)(1). An "interactive computer service" is "any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server . . . ." *Id.* § 230(f)(2).

[34] The "good Samaritan" provision of Section 230 proclaims platforms will not "be held liable on account of . . . any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene . . . excessively violent, harassing, or otherwise objectionable . . . ." *Id.* § 230(c)(2)(A); *see also* Doe v. Backpage.com, 817 F.3d 12, 18-19 (1st Cir. 2016) (explaining that "Congress sought to encourage websites to make efforts to screen content without fear of liability").

[35] *See, e.g.*, Ron Wyden, *Corporations Are Working with the Trump Administration to Control Online Speech*, WASH. POST (Feb. 17, 2020, 6:30 AM), https://www.washingtonpost.com/opinions/corporations-are-working-with-the-trump-administration-to-control-online-speech/2020/02/14/4d3078c8-4e9d-11ea-bf44-f5043eb3918a_story.html [https://perma.cc/T9Z4-R84E] (arguing that "[w]ithout 230, social media couldn't exist . . . . Movements such as Black Lives Matter or #MeToo, whose advocates post controversial

Section 230 provides immunity[36] only if the platform is not "responsible, in whole or in part, for the creation or development" of the information.[37] When a platform provides "'neutral tools' that others use to create discriminatory content"[38] and "passively displays content that is created entirely by third parties," Section 230 immunizes the platform from liability for claims such as negligence,[39] assault,[40] and defamation.[41] However, Section 230 does not extend immunity to an "information content provider," which Section 230(f) defines as "any person or entity that is responsible, in whole or in part, for the creation or **development** of information provided through the Internet or any other interactive computer service."[42] Circuits considering the issue

accusations against powerful figures on social media, would have remained whispers, not megaphones for oppressed communities," and asserting that repealing Section 230 would harm start-up companies more than big tech companies that can afford extensive legal representation).

[36] Section 230(c)(1) does not mention the term immunity, and it is technically a defense to a claim. *See* § 230(c)(1). Nevertheless, many scholars and judges have characterized Section 230's protection from liability as "immunity," and thus this Essay uses that common term.

[37] Fair Hous. Council v. Roommates.com, 521 F.3d 1157, 1166 (9th Cir. 2008) ("[S]ection 230 provides immunity only if the interactive computer service does not 'creat[e] or develop[ ]' the information 'in whole or in part.'").

[38] Amit Datta et al., *Discrimination in Online Advertising: A Multidisciplinary Inquiry*, 81 PROC. MACHINE LEARNING RES. 1, 11 (2018) ("Generally, entities are treated as an interactive computer service (ICS) if they provide 'neutral tools' that others use to create discriminatory content."); *see, e.g.*, Saponaro v. Grindr, LLC, 93 F. Supp. 3d 319, 324 (D.N.J. 2015) ("Similarly, in this case, Defendant merely provid[ed] neutral tools to carry out what may be unlawful or illicit [conduct].") (internal quotations omitted); Dart v. Craigslist, Inc., 665 F. Supp. 2d 961, 969 (N.D. Ill. 2009) ("The word-search function is a 'neutral tool' that permits users to search for terms that they select in ads created by other users."); Xcentric Ventures, L.L.C. v. Smith, 2015 WL 4940812, at *17 (N.D. Iowa Aug. 19, 2015), *report and recommendation adopted*, 2015 WL 5184114 (N.D. Iowa Sept. 4, 2015) ("Based on the current record, I predict it is more likely than not that at trial, the plaintiffs will be found to have been 'more than a neutral conduit' for the allegedly-harassing content . . . .").

[39] Doe v. MySpace, Inc., 528 F.3d 413, 422 (5th Cir. 2008) (affirming dismissal of claims for negligence and gross negligence against MySpace.com for failing to prevent 13-year-old from lying about her age that led to her meeting and being sexually assaulted by an alleged predator).

[40] Klayman v. Zuckerberg, 753 F.3d 1354, 1358 (D.C. Cir. 2014) (dismissing claim of intentional assault and negligence against Facebook for page that called for "Muslims to rise up and kill the Jewish people").

[41] Jones v. Dirty World, 755 F.3d 398, 406 (6th Cir. 2014) (finding user-generated tabloid site Dirty World immune from state-law defamation claims for posts anonymously uploaded to website).

[42] 47 U.S.C. § 230(f)(3) (2019) (emphasis added).

have generally adopted the "material contribution test"[43] — a website is responsible for helping to develop content "if it contributes materially to the alleged illegality of the conduct"[44] and does not enjoy Section 230 immunity. "Development" is not just content creation — it includes making information "usable," "available," or "visible."[45]

Section 230 has several exceptions written into the statute. The provision does not give platforms a defense to violations of federal criminal law, intellectual property law (e.g., copyright violations), the federal Electronic Communications Privacy Act of 1986 and similar State laws, and federal sex trafficking law.[46] Otherwise, States may enforce State laws that are consistent with Section 230, but not those that are inconsistent with the section.[47]

---

[43] *See, e.g.*, Kimzey v. Yelp! Inc., 836 F.3d 1263, 1269 n.4 (2016) ("Our sister circuits have generally adopted *Roommates.com*'s 'material contribution' to activity test . . . ."); Fed. Trade Comm'n v. LeadClick Media, LLC, 838 F.3d 158, 176 (2d Cir. 2016) ("It participated in the development of its affiliates' deceptive websites, 'materially contributing to [the content's] alleged unlawfulness.' . . . Accordingly, LeadClick is an information content provider . . . ."); Nemet Chevrolet, Ltd. v. Consumeraffairs.com, Inc., 591 F.3d 250, 257 (4th Cir. 2009); Jones v. Dirty World Entm't Recordings LLC, 755 F.3d 398, 413 (6th Cir. 2014) (adopting the material contribution test); Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1167-68 (9th Cir. 2008) ("We interpret the term 'development' as referring not merely to augmenting the content generally, but to materially contributing to its alleged unlawfulness."); Baldino's Lock & Key Serv., Inc. v. Google LLC, 285 F. Supp. 3d 276, 282-83 (D.D.C. 2018), *aff'd sub nom.* Marshall's Locksmith Serv. Inc. v. Google, LLC, 925 F.3d 1263 (D.C. Cir. 2019) ("The Providers' mapping information does not materially contribute to the alleged unlawfulness of the underlying information.").

[44] *Roommates.com*, 521 F.3d at 1167-68 (interpreting the term "development" as not merely contributing to the content itself — "but to materially contributing to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct"); Datta et al., *supra* note 38, at 11 ("However, if an interactive computer service materially contributes to the development of discriminatory content they are treated like an 'information content provider,' and lose the protection §230 offers.").

[45] FTC v. Accusearch, Inc., 570 F.3d 1187, 1198 (10th Cir. 2009) ("The dictionary definitions for *develop* correspondingly revolve around the act of drawing something out, making it 'visible,' 'active,' or 'usable.'"); *Roommates.com*, 521 F.3d at 1167 (indicating "development" is not merely content creation — but includes another dictionary definition "that is far more suitable to the context in which we operate: 'making usable or available'") (quoting WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 618 (2002)).

[46] 47 U.S.C. § 230(e)(1)-(5) (2019).

[47] *Id.* § 230 (e)(3) ("Nothing in this section shall be construed to prevent any State from enforcing any State law that is consistent with this section. No cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section.").

While the U.S. Supreme Court has not yet interpreted Section 230, many of federal and State courts have construed the provision broadly.[48]

For example, in *Chicago Lawyers' Committee for Civil Rights Under Law v. Craigslist, Inc.*,[49] the Seventh Circuit held that Section 230 protection applied to prevent a claim against Craigslist for a violation of the federal Fair Housing Act's prohibition on publishing advertisements for the sale or rental of housing that indicate a preference based on race. In that case, third-party users posted on Craigslist (an electronic meeting place) notices advertising housing that proclaimed "NO MINORITIES."[50] Craigslist did not actively participate in the discrimination by steering housing ads away from Black users or toward White users, and it provided unlimited access to the ads on the site.[51] Craigslist was a neutral forum. The court emphasized that Craigslist *caused* "postings only in the sense of providing a place where people can post," much like Microsoft may provide software and Dell may provide a computer that owners use to create discriminatory notices.[52]

Outside of the discrimination context, the Second Circuit held that the use of algorithms to match users does not render Section 230 protections inapplicable. In *Force v. Facebook, Inc.*,[53] the relatives of victims of terrorism claimed Facebook violated civil provisions of the federal Anti-Terrorism Act.[54] The relatives alleged that Hamas used Facebook to post content that encouraged terrorism, that the attackers

[48] Danielle Keats Citron & Benjamin Wittes, *The Problem Isn't Just Backpage: Revising Section 230 Immunity*, 2 GEO. L. TECH. REV. 453, 458 (2018) [hereinafter *The Problem Isn't Just Backpage*] ("The Supreme Court has declined to weigh in on the meaning of Section 230, but state and lower federal courts have reached a 'near-universal agreement' that it should be construed broadly."); *id.* at 460 ("Platforms have been protected from liability even though they republished content knowing it might violate the law, encouraged users to post illegal content, changed their design and policies to enable illegal activity, or sold dangerous products.").

[49] 519 F.3d 666 (2008).

[50] *Chi. Lawyers' Comm. for Civil Rights Under Law*, 519 F.3d at 668.

[51] *See* Statement of Interest of the United States of America at 14, Nat'l Fair Hous. All. v. Facebook, Inc., No. 18-cv-02689-JGK (S.D.N.Y. Aug. 17, 2018) ("[T]he owners and operators of Craigslist.com engaged in *no* efforts to categorize or classify its users, or to review or promote the postings, and provided unlimited access to material on the site . . . . In this way, Craigslist.com did not even partially "develop" or "create" the content at issue.").

[52] 519 F.3d at 672 (2008) ("If craigslist "causes" the discriminatory notices, then so do phone companies and courier services (and, for that matter, the firms that make the computers and software that owners use to post their notices online), yet no one could think that Microsoft and Dell are liable for "causing" discriminatory advertisements.").

[53] 934 F.3d 53 (2d Cir. 2019).

[54] *Id.* at 61.

viewed the content and subsequently murdered the victims, and that Hamas used Facebook to celebrate the attacks, transmit political messages, and promote additional violence.[55] The relatives of the victims argued that Facebook was not a publisher protected from liability under the terms of Section 230 because Facebook's algorithms engaged in "matchmaking" that suggests friends and connects users with content that is most likely to interest them.[56]

In deciding that Section 230 barred plaintiffs' claims, the court reasoned that Facebook's use of algorithms to match the interests of users was not a sufficiently material contribution to hold Facebook responsible in part as the developer or creator of the content.[57] The plaintiffs did not allege Hamas used Facebook's ad-targeting functions,[58] and the court did not consider whether Facebook "developed content" through its advertising functions.[59] The court distinguished the facts in *Force* from the unique context of a discrimination legal claim,[60] and there was no suggestion that Facebook's use of algorithms to direct Hamas's posts to the newsfeeds of the perpetrators of violence was a material part of the underlying federal Anti-Terrorism Act violation.

Despite these holdings, Section 230 protection from liability is not absolute. One empirical study found that a third of claims survived a Section 230 defense.[61]

---

[55] *Id.* at 59.

[56] *Id.* at 65.

[57] *Id.* at 70 ("Merely arranging and displaying others' content to users of Facebook through such algorithms — even if the content is not actively sought by those users — is not enough to hold Facebook responsible as the "develop[er]" or "creat[or]" of that content.").

[58] Complaint, Force v. Facebook, Inc., No. 1:16CV05158 (E.D.N.Y Jul. 10, 2016) (describing HAMAS's use of Facebook, and not alleging use of advertising services).

[59] *See* Statement of Interest of the United States of America at 17, Nat'l Fair Hous. All. v. Facebook, Inc., No. 18-cv-02689-JGK (S.D.N.Y. Aug. 17, 2018) ("In none of these cases [including Force v. Facebook, Inc.] did the court consider whether Facebook developed, in whole or in part, content through its advertising functions, let alone its advertising targeting functions, thereby making it a "content provider.").

[60] Force v. Facebook, Inc., 934 F.3d 53, 70 (2d Cir. 2019) (finding that unlike the instant case, the website in the *Roommates.com* case (discussed below) required "users to select from 'a limited set of pre-populated answers' to respond to particular 'discriminatory questions'" and that this "had a content-development effect that was actionable in the context of the Fair Housing Act").

[61] David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 493 (2010) ("While section 230 has largely protected intermediaries from liability for third-party speech, it has not been the free pass many of its proponents

For example, in *Fair Housing Council of San Fernando Valley v. Roommates.com*,[62] the Ninth Circuit held that Section 230 protection did not apply to prevent a discrimination claim against a website designed to match people with spare rooms with those looking for a place to live. The site required users looking for roommates to answer questions about their gender, sexual orientation, and whether they would bring children into the household.[63] Based on these criteria, the site then created searchable user profiles and sent emails to users informing them of other profiles that matched their preferences.[64] Plaintiffs argued that Roommates.com violated the federal Fair Housing Act and California State antidiscrimination laws.[65]

The Ninth Circuit held that Section 230 immunity was not available to Roommates.com because the site became "much more than a passive transmitter of information provided by others; it [became] the developer, at least in part, of that information."[66] The court wrote:

> We believe that both the immunity for passive conduits and the exception for co-developers must be given their proper scope and, to that end, we interpret the term "development" as referring not merely to augmenting the content generally, but to *materially contributing* to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus falls within the exception to section 230, if it *contributes materially* to the alleged illegality of the conduct.[67]

Roommates.com extracted information from potential customers, and its "connection to the discriminatory filtering process [was] direct and palpable: Roommates.com designed its search and email systems to limit the listings available to subscribers based on sex, sexual

claim . . . . [M]ore than a third of the claims at issue in the cases survived section 230's broad protection.").

62 521 F.3d 1157 (9th Cir. 2008).

63 *Id.* at 1161 ("In addition to requesting basic information — such as name, location and email address — Roommate requires each subscriber to disclose his sex, sexual orientation and whether he would bring children to a household.").

64 *Id.* at 1161-62 ("After a new subscriber completes the application, Roommate assembles his answers into a 'profile page.' . . . Those using the site's free service level can create their own personal profile page, search the profiles of others and send personal email messages. They can also receive periodic emails from Roommate, informing them of available housing opportunities matching their preferences.").

65 *Id.* at 1162.

66 *Id.* at 1166.

67 *Id.* at 1167-68 (emphasis added).

orientation, and presence of children."[68] The court held that Roommates.com enjoyed no immunity for filtering listings and disseminating "emails to subscribers according to discriminatory criteria,"[69] or for "using the answers to the unlawful questions to limit who has access to housing."[70] Despite Section 230's protections, Roommates.com did not enjoy unfettered freedom to choose an audience for particular housing notices along discriminatory lines.

While courts have not squarely addressed the question of whether social media companies are entitled to Section 230 immunity when they collect data and use algorithms to target protected classes of voters and deliver suppressive ads to them, the Department of Justice has taken the position that the use of algorithms to target and deliver housing advertisements along racial lines may not warrant Section 230 immunity from a federal Fair Housing Act violation. Asserting that the court should not grant immunity to Facebook at the motion-to-dismiss stage, the Justice Department argued:

> [T]he CDA does not immunize Facebook for materially contributing to the alleged illegality, namely excluding users from seeing ads based on protected characteristics. First, like Roommates.com, Facebook allegedly solicits discriminatory preferences through drop-down menus and other tools that offer advertisers discriminatory options. Second, Facebook allegedly mines its users' information and activity for data about their race, and national origin, and other protected characteristics, which is the touchstone in making discriminatory targeting possible. And third, Facebook is not entitled to the protection of the CDA because it is Facebook that ultimately decides for each ad which users will see it and which users will not. If Facebook engaged in this conduct, it was not simply providing its advertisers with a neutral tool or a blank slate to express their own content; it was materially contributing to an alleged violation of the FHA.[71]

---

[68] *Id.* at 1169.

[69] *Id.* at 1167.

[70] *Id.*

[71] United States' Statement of Interest at 7, Onuoha v. Facebook, Inc., No. C 16-06440-EJD (N.D. Cal. Nov. 16, 2018); *see also* Statement of Interest of the United States of America, Nat'l Fair Hous. All. v. Facebook, Inc., No. 18-cv-02689-JGK (S.D.N.Y. Aug. 17, 2018) (asserting that Facebook is not entitled to immunity for violations of the federal Fair Housing Act because it uses its algorithms to deliver discriminatorily targeted ads).

In March 2019, Facebook settled several discrimination lawsuits and agreed to make significant changes to prevent advertisers for housing, employment, or credit from discriminating based on race, national origin, ethnicity, age, sex, sexual orientation, disability, or family status.[72]

As in *Roommates.com*, social media companies' collection of data and use of algorithms to target protected classes of voters and deliver suppressive ads to them makes the social media companies more than "neutral tools" or passive transmitters of information. Third-party advertisers generally exercise minimal control in the targeting of suppressive ads and no control over which users actually see suppressive ads — social media companies often exercise the bulk of the control in the selective targeting and all of the control over which users actually see the suppressive ads. Such selective targeting and delivery by social media companies contributes materially to the discrimination against protected classes of voters, and as explained below, social media companies do not enjoy Section 230 immunity for this activity.

II. STATES CAN HOLD SOCIAL MEDIA COMPANIES ACCOUNTABLE FOR DISCRIMINATORY DISTRIBUTION OF ADS

In combatting deceptive practices and voter suppression, States can hold accountable social media companies that collect data from users, build demographic audiences of protected classes, and use algorithmic predictions to target protected classes of people with suppressive ads. Section 230 immunity is unavailable to platforms that are "responsible, in whole or in part, for the creation or development" of content,[73] and a platform is responsible for helping to develop content "if it contributes materially to the alleged illegality of the conduct."[74] "Development"

[72] *Summary of Settlements Between Civil Rights Advocates and Facebook, Housing, Employment and Credit Advertising Reforms*, ACLU (Mar. 19, 2019), https://www.aclu.org/other/summary-settlements-between-civil-rights-advocates-and-facebook [https://perma.cc/X52Q-QX6P].

[73] *Roommates.com*, 521 F.3d at 1166 ("[S]ection 230 provides immunity only if the interactive computer service does not 'creat[e] or develop[ ]' the information 'in whole or in part.'").

[74] *Id.* at 1167-68 (interpreting the term "development" as not merely contributing to the content itself — "but to materially contributing to its alleged unlawfulness. In other words, a website helps to develop unlawful content, and thus falls within the exception to section 230, if it contributes materially to the alleged illegality of the conduct"); Datta et al., *supra* note 38, at 11 ("However, if an interactive computer service materially contributes to the development of discriminatory content they are treated like an 'information content provider,' and lose the protection §230 offers.").

includes making information "usable," "available," or "visible."[75] By deciding that a protected group like African Americans will be shown a voter-suppression ad and by steering the ad away from other groups, social media companies contribute materially to voter suppression. Such efforts are not simply a "passive" posting of third-party content on a website like Craigslist for all to see, which is the activity Section 230 was intended to immunize. Platforms do not enjoy Section 230 immunity for targeting and delivering suppressive ads to protected classes.

A. Platforms Exercise Significant Control Over Ad Targeting and Delivery

Many social media companies would argue that third-party advertisers — not social media companies — control the discriminatory distribution of ads by selecting targeting options. A platform, the argument goes, should not be held liable if a few unsavory advertisers choose to use a platform's "neutral tool" for an unlawful purpose like voter suppression. Algorithms for targeted ads, the argument goes, are similar to the algorithms used to match Hamas to its potential supporters that the court in *Force* held were neutral tools that did not render Facebook liable under the Anti-Terrorism Act.[76]

This argument ignores, however, that platforms materially contribute to the underlying illegality of discriminatory distribution of suppressive

---

[75] FTC v. Accusearch, Inc., 570 F.3d 1187, 1198 (10th Cir. 2009) ("The dictionary definitions for *develop* correspondingly revolve around the act of drawing something out, making it 'visible,' 'active,' or 'usable.'"); *Roommates.com*, 521 F.3d at 1167 (indicating "development" is not merely content creation — but includes another dictionary definition "that is far more suitable to the context in which we operate: 'making usable or available'") (quoting WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 618 (2002)).

[76] Force v. Facebook, Inc., 934 F.3d 53 (2d Cir. 2019). Facebook has made these arguments in responding to claims of housing discrimination stemming from ads. *See* Notice of Motion & Motion to Dismiss First Amended Complaint for Defendant at 2, Onuoha v. Facebook, Inc., No. 16-cv-06440-EJD (N.D. Cal. Apr. 3, 2017) ("Advertisers, not Facebook, are responsible for both the content of their ads and what targeting criteria to use, if any. Facebook's provision of these neutral tools to advertisers falls squarely within the scope of CDA immunity."). In 2019, Facebook settled several legal actions and agreed to make significant changes to prevent advertisers for housing, employment, or credit, from discriminating based on race, national origin, ethnicity, age, sex, sexual orientation, disability, or family status. *Summary of Settlements Between Civil Rights Advocates and Facebook, Housing, Employment and Credit Advertising Reforms*, *supra* note 72.

messages in the *advertising* context, which was not at issue in *Force*.[77] Platforms accept significant funds to create particular audiences in the advertising context (Google and Facebook together accounted for nearly 60% of the $107.5 billion in internet advertising revenues in the United States in 2018)[78] and exercise significant control over which users actually see a suppressive ad.[79]

For example, Facebook ads require two phases to determine which users see an ad — the targeting phase and the delivery phase.[80]

[77] Facebook, for example, could simultaneously enjoy Section 230 immunity for pages and posts erected by HAMAS analyzed in *Force*, while not enjoying Section 230 immunity for suppressive ads that Facebook disseminates along racial lines because of the significant control that Facebook exercises in determining *which users* will see the suppressive ad and *when* they will see it. *See Roommates.com*, 521 F.3d at 1162-63 ("A website operator can be both a service provider and a content provider: If it passively displays content that is created entirely by third parties, then it is only a service provider with respect to that content. But as to content that it creates itself, or is 'responsible, in whole or in part' for creating or developing, the website is also a content provider. Thus, a website may be immune from liability for some of the content it displays to the public but be subject to liability for other content.").

[78] PWC, IAB INTERNET ADVERTISING REVENUE REPORT: 2018 FULL YEAR RESULTS 3 (2019), https://www.iab.com/wp-content/uploads/2019/05/Full-Year-2018-IAB-Internet-Advertising-Revenue-Report.pdf [https://perma.cc/H7YM-SG5P] (showing total 2018 U.S. digital revenues of $107.5 billion); *US Digital Ad Spending Will Surpass Traditional in 2019*, EMARKETER (Feb. 20, 2019), https://www.emarketer.com/newsroom/index.php/us-digital-ad-spending-will-surpass-traditional-in-2019/ [https://perma.cc/FHK4-XMLW] (reporting that in 2018, Google was responsible for 38.2% of the digital ad revenue in the United States and that Facebook was responsible for 21.8%).

[79] *See generally* Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203, 218 (2018) ("Today, online services do so much more than relay or store user-generated content in the way that the early proponents of immunity and nongovernmental interference presumed. They actively shape every aspect of the user experience. Many of the most successful internet companies . . . design their applications to collect, analyze, sort, reconfigure, and repurpose user data for their own commercial reasons . . . . These developments belie any suggestion that online intermediaries are merely conduits of user information anymore.") [hereinafter *Intermediary Design Duties*]; Danielle Keats Citron, *Section 230's Challenge to Civil Rights and Civil Liberties: Response to Olivier Sylvain's Essay "Discriminatory Designs on User Data*," KNIGHT FIRST AMEND. INST. COLUM. U. (Apr. 6, 2018), https://knightcolumbia.org/content/section-230s-challenge-civil-rights-and-civil-liberties [https://perma.cc/2XPG-79LW] (asserting that the wrongful activity of platforms is not the republication of third party activity but the platform design that induces and enables illegal discrimination).

[80] Brief of Amicus Curiae Upturn in Support of Plaintiffs' Opposition to Facebook's Motion to Dismiss First Amended Complaint at 1-2, Onuoha v. Facebook, Inc., No. 16-cv-06440-EJD (N.D. Cal. Nov. 16, 2018) ("On Facebook's Ad Platform, any ad that is seen by a user must first go through two phases: targeting and delivery. During the ad targeting phase, Facebook helps the advertiser create a "target audience": a list of users who are eligible, but not guaranteed, to see a given ad . . . . During the ad delivery phase, Facebook itself makes decisions, independently of the advertiser, about which users

Facebook exerts varying levels of control in helping the advertiser target an audience (which depends on the method of targeting an advertiser selects) and exercises almost complete control over the delivery phase (who actually sees the ad).

With regard to *targeting*, for example, Facebook's "Ad Manager" allows an advertiser to select, from a series of dropdowns, 52,000 targeting attributes, including demographics/ethnic affinity (e.g., African American), issue interests (e.g., "Malcolm X" or the "Civil Rights Movement"), and Facebook engagement (e.g., liked a particular post).[81] About 73% of the Russian Agency ads used interest-based targeting,[82] and most of the "interest-based targeting focused on African American communities and interests"[83] like "Martin Luther King, Jr., Nelson Mandela, Malcolm X and Muhammad Ali."[84] Facebook develops these profiles by collecting vast amounts of data on its two billion users — including zip codes, posts, comments, likes, clicks, and other information — and by utilizing predictive modeling techniques to make inferences.[85] This microtargeting "is also enhanced by real-time re-

within an ad's target audience will actually see the ad. . . . based on its own predictions about what kinds of users are most likely to engage with that ad. Advertisers have no meaningful control over Facebook's delivery decisions.").

[81] *See* ANTHONY NADLER ET AL., DATA & SOC'Y RESEARCH INST., WEAPONIZING THE DIGITAL INFLUENCE MACHINE: THE POLITICAL PERILS OF ONLINE AD TECH 11-12 (2018), https://datasociety.net/wp-content/uploads/2018/10/DS_Digital_Influence_Machine.pdf [https://perma.cc/63Y8-RP7C]; Young Mie Kim et al., *The Stealth Media? Groups and Targets Behind Divisive Issue Campaigns on Facebook*, 35 POL. COMM. 515, 520 (2018) [hereinafter *The Stealth Media?*].

[82] KIM, *supra* note 1, at 6 (2018) ("Among the various targeting methods Facebook/Instagram offers, the Russian Agency predominantly utilized Interest-based targeting. About 73% of the Russian Agency ads used interest-based targeting.").

[83] DIRESTA ET AL., *supra* note 6, at 34 ("Most of the interest-based targeting focused on African American communities and interests.").

[84] KIM, *supra* note 1, at 6 (indicating "the IRA group BM targeted Facebook users who selected Martin Luther King, Jr., Nelson Mandela, Malcolm X and Muhammad Ali as 'interests' and presented them with an ad highlighting the issue of police brutality").

[85] NADLER ET AL., *supra* note 81, at 11-12 ("Social media platforms are among the most prodigious hoarders of consumer information. Facebook reportedly employs a classification scheme of some 52,000 attributes to categorize its 2 billion monthly active users. Among the information that Facebook routinely captures are data submitted directly by users such as posts, likes, profile information and social connections, data extracted from photographs and video (including facial recognition data), and many types of behavioral data, such as when and where users log in, what devices they use, and even, for a time at least, so-called 'self-censored' posts that users composed but did not actually publish."); Kim et al., *The Stealth Media?* , *supra* note 81, at 520 ("By gathering a vast amount of data, including digital trace data, and by utilizing predictive modeling techniques, campaigns create enhanced profiles that identify and target specific types of individuals, and then customize their messages."); Till Speicher et al.,

targeting algorithms, a constant loop between users' voluntary choices (e.g., liking) and the machine's feedback on their choices."[86]

Another targeting tool, Facebook's Lookalike Audience, allows advertisers to ask Facebook to create target audiences that are demographically similar to (i.e., "look like") another clearly defined audience — and by doing so, Facebook "clones" audiences.[87] Research has found that the tool accurately replicates biases along racial and other demographic lines.[88] While the advertiser may identify the model of the type of audience he or she wants (e.g., a list of voters who happen to be Black), the lookalike audience is created entirely by Facebook. Indeed, an advertiser placing a suppressive ad has no real understanding of how the lookalike audience is created and has no opportunity to decide which users will receive the ad — and thus requires that Facebook materially contribute to the discriminatory distribution of the suppressive ad.[89]

By comparison, Facebook exercises less control in determining which users are targeted when advertisers choose their own "custom audience." The Custom Audience function requires that an advertiser give Facebook personally identifiable information (e.g., voter records, email addresses) of the precise people the advertiser wants to target,[90] and Facebook uses that data to identify corresponding social media accounts.

Potential for Discrimination in Online Targeted Advertising, 81 PROC. MACHINE LEARNING RES. 1, 5, 7 (2018) ("For each user in the US, Facebook tracks a list of over 1,100 binary attributes spanning demographic, behavioral and interest categories that we refer to as curated attributes. Additionally, Facebook tracks users' interests in entities such as websites, apps, and services as well as topics ranging from food preferences (e.g., pizza) to niche interests (e.g., space exploration).").

86 Kim et al., *The Stealth Media?* , *supra* note 81, at 520.

87 *About Lookalike Audiences*, FACEBOOK, https://www.facebook.com/business/help/164749007013531?helpref=page_content (last visited Jan. 4, 2020) [https://perma.cc/UMW7-MPQG].

88 Speicher et al., *supra* note 85, at 13-14 (examining gender, age, ethnicity and political affiliation, and finding "the look-alike audience feature in Facebook is able to both capture the biases in a source audience and propagate the biases to the larger audiences it helps construct" and concluding that "look-alike audiences selected using highly biased source audiences can be highly discriminatory").

89 *See id.* at 14 ("As Facebook is actively involved in the selection of the look-alike audience, one might argue that Facebook needs to be more accountable for the selection of such a discriminatory audience.").

90 *About Website Custom Audiences*, FACEBOOK, https://www.facebook.com/business/help/610516375684216?helpref=page_content (last visited Jan. 4, 2020) [https://perma.cc/Z9EE-XPQR].

After shifting to the *delivery phase*, the social media company generally controls which users will actually see the suppressive ad, and when and where they see it. Third-party advertisers like the Russian Agency have no meaningful control over who will see the ad. For example, Facebook and other social media platforms may determine which users see which ads based on which advertisers have higher budgets and are willing to pay more (an auction),[91] and whether the user is more likely than others to engage with the ad (e.g., like, share, comment, retweet). As Professor Pauline Kim writes:

> Because platforms seek to optimize revenue, their algorithms try to predict which ads will be most relevant to which users. These predictions are based on not only the known interests and behaviors of that particular individual, but also what is inferred about her from the behavior of similar users. Precisely which ads an individual will see is ultimately determined through an algorithmic process controlled by the platform. [92]

Social media companies do not simply sell advertisers the ability to blast out information to a list of individuals predetermined by the advertiser. Instead, a social media company is using data it has collected on the likes and preferences of its users to select who will receive the ad.[93] Just as the site in *Roommates.com* was responsible for developing content because it forced users to provide information as a condition to use the site, many platforms today compel users to provide information by mandating that users consent to the platform collecting and using the user's data as a condition to subscribing to the platform.[94] Platforms curate information from users, and their algorithms affirmatively

---

[91] *See, e.g.*, Datta et al., *supra* note 38, at 12-14 (detailing employment ads on Google AdWords platform which the company's machine learning steered away from women and toward men, and asserting that Google is making a "material contribution to employment ads that express a preference for men").

[92] Kim, *Manipulating Opportunity*, *supra* note 27 (manuscript at 15).

[93] *See* Muhammad Ali et al., Discrimination Through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes 12 (Apr. 4, 2019) (unpublished manuscript), https://arxiv.org/abs/1904.02095v2 [https://perma.cc/FQ79-454Z] ("While ad targeting is facilitated by an advertising platform — but nominally controlled by advertisers — ad delivery is conducted and controlled by the advertising platform itself. We demonstrate that, during the ad delivery phase, advertising platforms can play an independent, central role in creating skewed, and potentially discriminatory, outcomes.").

[94] Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1166 (9th Cir. 2008) ("When a business enterprise extracts such information from potential customers as a condition of accepting them as clients, it is no stretch to say that the enterprise is responsible, at least in part, for developing that information.").

recommend highlighting particular posts to particular users. The delivery of this information is not just created by "another information content provider" — it is created by the platform itself.[95] Often, advertisers and other outsiders do not determine *who* will receive an ad, nor do they generally understand how the particular algorithms work.[96] Indeed, research shows that Facebook ad-delivery algorithms inadvertently inhibit political campaigns' ads from reaching voters with diverse political views — which may be "invisible to political campaigns."[97]

The prediction by the platform about which users have a particular issue interest (e.g., Martin Luther King, Jr., the Civil Rights Movement, African American history), the platform's prediction about which users are most likely to click on the ad,[98] and the decision to show those users the ad (and not show it to other users) are all "material contributions" that make the platform responsible, in part, for the voter suppression.

While the ad targeting and delivery examples above use Facebook to illustrate various levels of platform control in the targeting and delivery process, a similar analysis could be done with other platforms — such as Google's gender segmentation, machine learning, keywords, and auction functions.[99]

---

[95] Kim, *Manipulating Opportunity*, *supra* note 27 (manuscript at 44 n.206) ("In contrast, in challenging discriminatory ad targeting, it is not necessary to consider the content of the ads at all. The charge of discriminatory targeting would stand regardless of which jobs were advertised or how the ads were formatted.").

[96] *See* Datta et al., *supra* note 38, at 14 (observing that while researchers conduct outside experiments, the fact that so many online platforms' algorithms are deemed proprietary and go unexamined (essentially a Black Box) is a significant part of the problem).

[97] Muhammad Ali et al., Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging (Dec. 17, 2019) (unpublished manuscript), https://arxiv.org/pdf/1912.04255.pdf [https://perma.cc/33W4-MA5C] (finding that "Facebook's ad delivery algorithms effectively differentiate the price of reaching a user based on their inferred political alignment with the advertised content, inhibiting political campaigns' ability to reach voters with diverse political views," and that this phenomenon may be "invisible to political campaigns" and increase political polarization).

[98] Datta et al., *supra* note 38, at 12-13 ("Google, using programs that are part of its AdWords platform, decides who sees an ad based on Google's opinion of who is most likely to click on it. Advertisers are not part of the decision, and in fact they may be unaware that such a decision is being made . . . . As a result, Google is making a material contribution to the publishing enterprise.").

[99] *See, e.g.*, Datta et al., *supra* note 38, at 11-14 (examining various ad targeting and delivery functions on Google ad serving platform and the varying levels of control exercised by Google to assess whether the function is a "neutral tool" that warrants 230 immunity, or result in Google making a material contribution to discrimination removing Section 230 protection).

B. Platforms' Discriminatory Ad Targeting and Delivery Make a Material Contribution to Voter Suppression

By selectively steering deceptive ads toward Black voters and not toward others, social media companies' ad targeting and delivery mechanisms make material contributions to discriminatory voter suppression.

The harm of discriminatory distribution of deceptive voting ads is unique. In the housing and employment contexts, the harm stems not from deception of the advertiser or the content of the ad, but from discriminatory distribution that steers housing and employment opportunities away from protected classes such as African Americans and Latinos.[100]

With regard to the voting context, the harm of targeted ads stems from a combination of three factors: (1) the deception (e.g., the author is allegedly Black and committed to racial justice issues or is another trusted source of information, the post provides false information such as "you need three different forms of photo identification to vote" or "you can't vote if anybody in your family has been in prison"); (2) the content of the ad discouraging participation (e.g., "let's protest that neither Democrats nor Republicans care about us by not voting," or "if you vote you'll be arrested and convicted"); and (3) the discriminatory dissemination of the deceptive ad to the protected class.

Targeting the deceptive vote-suppression message almost exclusively to a protected class produces different harms than those that stem from discouraging all voters from casting ballots (as Craigslist made discriminatory housing ads visible to all users). Voting is relative — it is not exercised in isolation. Lower turnout among just African Americans, for example, dilutes the voting strength of Black communities relative to other communities and prevents Black communities from electing their preferred candidates. Also, by targeting suppressive ads at Black communities, vote suppressors can concentrate their limited resources on more effectively deceiving Black voters.

Granted, individual African Americans may scrutinize deceptive messages and still decide to vote. But even those who choose to vote are harmed if they are politically cohesive (e.g., they share preferred candidates) with other African Americans who are discouraged from voting by the targeted ads, and if other communities with different

---

[100] Kim, *Manipulating Opportunity*, *supra* note 27 (manuscript at 45 n.217) ("Where the content itself is not harmful, I believe the more straightforward argument is that 230 does not apply at all, because the platform is not held liable as a speaker or publisher, but because of the entirely separate function of distributing the content.")

political interests are not targeted and thus are not discouraged from voting. Targeting Black voters prevents African Americans from identifying with one another and their allies and collectively using their votes to enact change through the democratic process.

Under these circumstances, platforms can make material contributions to voting discrimination. Imagine a State law that provides civil liability for individuals, entities, and platforms that target deceptive and suppressive ads at particular racial or ethnic groups. Specifically, the State statute could provide civil penalties for individuals or entities providing funding *and* for platforms accepting such funding for directing deceptive or misleading advertisements at a particular racial or ethnic group. The law would prohibit false or misleading ads about the qualifications for voters to register or to vote, as well as about the time, place, or manner of an election — with knowledge of falsity or reckless disregard for the truth. The law would require that offending individuals, entities, or platforms know or have reason to know that such ads were being directed toward a particular racial group with an intent to discourage voting or with reckless disregard for whether the ads discourage voting by members of the particular racial or ethnic group.[101] The State law would not impose

[101] As discussed above, this Essay focuses on Section 230, and does not grapple with First Amendment challenges presented by deceptive practices regulations. Thus, the hypothetical statute is not the only option or necessarily best option — but is provided simply to illustrate that platform targeting and delivery of deceptive and suppressive ads along racial lines makes a material contribution to the underlying unlawful behavior of discrimination in the voting context. A "knowledge of falsity or reckless disregard for the truth" standard, however, may be a valuable provision in any deceptive practices law regulating platforms. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 279-80 (1964) (requiring that a publisher act with actual malice — meaning with knowledge of falsity or reckless disregard for the truth — to be held liable for defamation actions brought by public figures regarding a matter of public concern). A provision restricting targeted ads that mislead voters about voting requirements and procedures likely poses few constitutional problems. Minn. Voters All. v. Mansky, 138 S. Ct. 1876, 1889 n.4 (2018) ("We do not doubt that the State may prohibit messages intended to mislead voters about voting requirements and procedures."). This hypothetical statute that regulates voting requirements and procedures, however, leaves unregulated significant deceptive activities that could suppress votes, such as ads by those who assume a false identity as African American and state "we should protest and not vote because neither Democrats nor Republicans care about us." Extending the restrictions to prohibit false or misleading statements about the speaker's identity or to restrict other false political speech may raise constitutional problems. *See* Susan B. Anthony List v. Driehaus, 573 U.S. 149, 164 (2014) (expressing significant reservations about the validity of false campaign speech regulations); United States v. Alvarez, 567 U.S. 709, 719 (2012) (holding that false statements may receive First Amendment protection even when made with knowledge of falsity or reckless disregard to their truth). A court interpreting the prohibition on false speaker identity as a simple disclosure requirement of the

liability on individuals, entities, or platforms that take reasonable steps to prevent or address such suppressive ads once warned about them.[102]

Under such a law, many social media platforms would engage in behavior that would fall outside of Section 230's legal shield. They would materially contribute to illegal discrimination for various reasons and thus would be understood as content developers — and thus exempt from the protection of Section 230(c)(1).

A platform would materially contribute when its "terms of service" agreement mandates that users consent to the platform collecting and using the users' personal data as a condition of using the platform, which effectively prevents users from opting out of being targeted by a platform with suppressive ads along racial lines. Platforms would market their ability to classify and target users to potential advertisers, and they would make significant profits from potential advertisers. Also, platforms would make material contributions because they would decide which users will and will not see a deceptive and suppressive ad.

While some platforms might claim they do not explicitly use race or ethnicity to deploy ads, the platforms are aware of the discriminatory potential of such tools, even when race is not explicitly used as a classification.[103] Facebook, for example, announced a rule to prohibit housing, employment, and credit advertisers from using ethnic-affinity

speaker's identity, however, may uphold the provision. *See* Buckley v. Valeo, 424 U.S. 1, 68 (1976) (upholding disclosure requirements in the campaign finance context). *See generally* Marshall, *supra* note 28 (discussing the constitutional status of regulations on false speech in the political context in Section IV).

[102] This hypothetical state statute is inspired by Virginia's voter deception law, the Voting Rights Act, and a proposed amendment to Section 230. *See* 42 U.S.C. § 1973(a) (2000) (providing that no voting procedure shall be imposed that "results in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color"); VA. CODE ANN. § 24.2-1005.1 (2007) (considering it a misdemeanor to "[knowingly] communicate . . . false [election] information [to a registered voter] . . . about the time, date and place of [voting] or the voter's precinct, polling place or registration status"); Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans Section 230 Immunity*, 86 FORDHAM L. REV. 401, 419 (2017) ("No provider or user of an interactive computer service that takes reasonable steps to prevent or address unlawful uses of its services shall be treated as the publisher or speaker of any information provided by another information content provider . . ."); Citron & Wittes, *The Problem Isn't Just Backpage*, *supra* note 48, at 471 (proposing that Section 230 be amended to exempt from liability only platforms that take "reasonable steps to prevent or address unlawful uses of its services once warned about such uses").

[103] Statement of Interest of the United States of America at 18, Nat'l Fair Hous. All. v. Facebook, Inc., No. 18-cv-02689-JGK (S.D.N.Y. Aug. 17, 2018) ("Facebook markets the availability, ease of use, and effect of these classifications to potential advertisers without regard to the possible illegality of these classifications under federal fair housing laws.").

targeting.[104] There are many proxies for race, however, such as users who have indicated an interest in Malcolm X, the U.S. Civil Rights Movement, and BlackNews.com.[105] Even absent clear proxies, the algorithm that shapes the audience may produce a racial skew.[106]

Platforms' data about users are so extensive that the platforms can create audiences with particular racial traits even in the absence of explicitly considering race.[107] One recent study found that the Facebook

[104] Erin Egan, *Improving Enforcement and Promoting Diversity: Updates to Ethnic Affinity Marketing*, FACEBOOK (Nov. 11, 2016), https://about.fb.com/news/2016/11/updates-to-ethnic-affinity-marketing/ [https://perma.cc/GZ6A-ARUE] (announcing the disabling of the use of ethnic affinity targeting for housing, employment, or credit ads and a required affirmation by advertisers that they will not engage in discriminatory advertising); Sheryl Sandberg, *Doing More to Protect Against Discrimination in Housing, Employment and Credit Advertising*, FACEBOOK (Mar. 19, 2019), https://about.fb.com/news/2019/03/protecting-against-discrimination-in-ads/ [https://perma.cc/4A52-VQAG] (placing additional limits on targeting categories for housing, employment, and credit advertisers, including prohibiting targeting by age, gender, or zip code).

[105] Speicher et al., *supra* note 85, at 9 ("For example, 'BlackNews.com' has an audience with 89% of the users with African American affinity . . . the audience of 'Hoa hoc Tro Magazine' is composed of 95% users with Asian American affinity . . . . 'Nuestro Diario' has an audience with 98% of Hispanic affinity . . . . These results suggest that a malicious advertiser could easily find free-form attributes to launch discriminatory ads . . . ."); *id.* at 14 (conducting an empirical study using public voter record data "[d]emonstrating that several user attributes in Facebook, beyond the much-criticized 'ethnic affinity,' show strong positive and negative correlations with users belonging to different races. Worse, Facebook's related attribute suggestions can be exploited by advertisers to discover facially-neutral attributes that can be used for highly discriminatory audience targeting. Thus, simply banning certain attributes is insufficient to solve the problem").

[106] Ali et al., *supra* note 93 (manuscript at 1) ("[W]e observe significant skew in delivery along gender and racial lines for 'real' ads for employment and housing opportunities despite neutral targeting parameters. Our results demonstrate previously unknown mechanisms that can lead to potentially discriminatory ad delivery, even when advertisers set their targeting parameters to be highly inclusive."); Kim, *Manipulating Opportunity*, *supra* note 27 (manuscript at 16) ("[E]ven if [an advertiser] has chosen neutral targeting criteria, the actual audience receiving the ad may be skewed along the lines of race, sex or other protected characteristics due to the platform's targeting algorithm."); Speicher et al., *supra* note 85, at 14 ("[S]everal user attributes in Facebook, beyond the much-criticized 'ethnic affinity,' show strong positive and negative correlations with users belonging to different races. Worse, Facebook's related attribute suggestions can be exploited by advertisers to discover facially-neutral attributes that can be used for highly discriminatory audience targeting.").

[107] Sylvain, *Intermediary Design Duties*, *supra* note 79, at 274-75 ("That the company collects and synthesizes non-racial or non-ethnic user data to create 'ethnic' or 'multicultural affinity' classifications does not necessarily justify the immunity. To the contrary, Facebook's use of big data algorithmic analysis of ostensibly non-racial data is precisely the sort of thing on which we would expect bigots to rely to mask their true

Special Audiences tool — which was built pursuant to a civil rights settlement and was intentionally not provided with demographic features — creates audiences that have nearly the same level of racial bias as the standard Facebook Lookalike audience.[108] As mentioned above, on many platforms users cannot even opt out of this data collection that assembles audiences with particular racial traits because platforms often mandate that users consent to the platform collecting and using their data as a condition of using the platform.

Just as in *Roommates.com*, where Section 230 immunity was inapplicable to protect the platform from allegations of a violation of the federal Fair Housing Act and California State prohibitions on housing discrimination, in the voting context, platforms' "connection to the discriminatory filtering process is direct and palpable."[109] Section 230 does not license social media companies to freely distribute deceptive and suppressive ads along racial lines using ethnic affinity targeting, proxies for race, or simple data collection and algorithms that produce a racially discriminatory effect in suppressive ad delivery.

While targeted advertising is used for many legitimate purposes (e.g., mobilizing voters, treating sickle cell anemia and other diseases, promoting an urban radio station and Tyler Perry movies), in the unique context of voter suppression,[110] algorithms that facilitate racial targeting of suppressive ads make a material contribution to the underlying illegality. These platforms are not neutral tools. When they

intentions."); Brief of Amicus Curiae Upturn in Support of Plaintiffs' Opposition to Facebook's Motion to Dismiss First Amended Complaint at 3, Onuoha v. Facebook, Inc., No. 16-cv-06440-EJD (N.D. Cal. Nov. 16, 2018) ("Facebook's extensive data about its users includes strong proxies for protected class membership, and these proxies can lead to a Lookalike Audience whose protected status traits match those of the source audience.").

[108] Piotr Sapiezynski et al. Algorithms that "Don't See Color": Comparing Biases in Lookalike and Special Ad Audiences (Dec. 17, 2019) (unpublished manuscript), https://arxiv.org/pdf/1912.07579.pdf [https://perma.cc/XRB7-UU7F] (finding that the Facebook Special Audiences tool, which does not consider race, creates audiences that have nearly the same level of racial bias as the standard Lookalike audience); Ava Kofman & Ariana Tobin, *Facebook Ads Can Still Discriminate Against Women and Older Workers, Despite a Civil Rights Settlement*, PROPUBLICA (Dec. 13, 2019, 5:00 AM), https://www.propublica.org/article/facebook-ads-can-still-discriminate-against-women-and-older-workers-despite-a-civil-rights-settlement [https://perma.cc/487B-GB7G].

[109] Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1169 (9th Cir. 2008).

[110] Statement of Interest of the United States of America at 16, Nat'l Fair Hous. All. v. Facebook, Inc., No. 18-cv-02689-JGK (S.D.N.Y. Aug. 17, 2018) (asserting that 230 "these cases must be read in the context of the particular Facebook function and content at issue in each case").

suppress votes, they are not merely "publishing content of others" — the activity that Section 230 was intended to protect.

Targeting has historically made a material contribution to voting discrimination. In Louisiana, for example, there were 127,923 Black voters on the registration rolls in 1888, compared to 730 by 1910 due to targeted efforts at disenfranchisement.[111] In 1957, the Alabama legislature redrew the boundaries of the City of Tuskegee to "remove from the city all save four or five of its 400 Negro voters while not removing a single white voter or resident."[112] In 2013, the North Carolina legislature enacted a series of voting restrictions that the Fourth Circuit later invalidated, finding that they "target African Americans with almost surgical precision . . . ."[113] While the deception and the content communicated are important parts of voter suppression, the targeting of the suppressive ad toward protected classes by the platform makes the ads discriminatory and actionable.

A social media company's *discriminatory distribution* of suppressive ads differs from Craigslist's passive posting of content by third parties, as well as Hamas's creation of free Facebook pages and posts in *Force*. When Facebook, Twitter, or YouTube develops and utilizes an algorithm to determine which users will receive a particular suppressive ad,[114] a State should be able to hold the platform accountable when it chooses to take money to disseminate that ad along racial lines. State regulation of discriminatory ad delivery is not within the immunity contemplated by Section 230, and it is within the scope of State power to enact antidiscrimination civil rights measures.[115]

Section 230 recognizes that requiring online platforms to police the billions of posts, tweets, and comments that appear on such platforms

---

[111] SAMUEL ISSACHAROFF, PAMELA S. KARLAN & RICHARD H. PILDES, THE LAW OF DEMOCRACY: LEGAL STRUCTURE OF THE POLITICAL PROCESS 90 (2d ed. 2002).

[112] Gomillion v. Lightfoot, 364 U.S. 339, 341 (1960).

[113] North Carolina State Conference of the NAACP v. McCrory, 831 F.3d 204, 214 (2016).

[114] Kim, *Manipulating Opportunity*, *supra* note 27 (manuscript at 43-44) ("When an online platform like Google or Facebook designs a targeting algorithm to determine which ads are delivered to which users, it is clearly not acting as a speaker of the ad content. Nor is the platform acting as a 'publisher,' . . . . [H]olding [online platforms] responsible for discriminatory patterns of information delivery would not impinge on any of those functions.").

[115] Ry. Mail Ass'n v. Corsi, 326 U.S. 88, 94 (1945) ("We see no constitutional basis for the contention that a state cannot protect workers from exclusion solely on the basis of race, color or creed by an organization, functioning under the protection of the state, which holds itself out to represent the general business needs of employees."); *Discrimination and Harassment in the Workplace*, *supra* note 22; *State Public Accommodation Laws*, *supra* note 22.

would be overwhelmingly burdensome, and thus shields them from liability for passive display of third-party content.[116] A State requirement that online platforms monitor and prevent the distributive effects of ads discouraging protected liberties such as the right to vote, however, is reasonable. The world's most valuable companies should not have the right to externalize the costs of discriminatory ad distribution onto many of the nation's most economically and politically marginalized communities.[117]

III. NEXT STEPS

Recognizing that States have the power to regulate social media companies for discriminatory dissemination of suppressive ads is just the beginning. Congress and federal agencies have not responded to adequately address this issue, and thus the next step involves States tailoring deceptive practices laws to address online activity and grapple with various issues.[118] For example:

- How do States prohibit as many forms of voter suppression as possible, recognizing that laws need to be narrowly tailored

[116] *See* Kim, *Manipulating Opportunity*, *supra* note 27 (manuscript at 45-46) ("Congress recognized that if websites were held liable for content posted by others, it would impose an enormous burden on them . . . . However, holding platforms responsible when they act in another capacity — not as publisher or speaker — does not create the same sort of existential threat. It does not require them to review all user posts, or to make editorial decisions about which posts to permit and which to remove. Instead, it requires them to be attentive to the distributive effects of their choices regarding who sees what information, and holds them responsible if their choices produce discriminatory effects.").

[117] 2019 Fortune 500, FORTUNE, https://fortune.com/fortune500/2019/search (last visited Mar. 24, 2020) [https://perma.cc/XRC7-YUV5] (showing that Alphabet (the parent company of Google), Amazon.com, Apple, Facebook, and Microsoft are all among the top 6 U.S. companies in market value); *see also* Jack M. Balkin, *2016 Sidley Austin Distinguished Lecture on Big Data Law and Policy: The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217 (2017) (analogizing the harms caused by algorithms to nuisance in analyzing the "socially unjustified use of computational capacities that externalizes costs onto innocent others"); Sylvain, *Intermediary Design Duties*, *supra* note 79, at 207-08 ("Profits, of course, are not unlawful . . . . But profits in this context also are the spoils of a legal regime that effectively absolves online intermediaries from minding the harmful third-party user content that they host and repurpose for commercial gain. They are the benefits of a legal protection that almost no other entity in other legislative fields enjoys.").

[118] *See* Daniels, *supra* note 19, at 372-80 (grappling with various First Amendment issues associated with deceptive practices laws).

to avoid impinging on constitutionally protected political speech?[119]

- If the State requires evidence of an "*intent* to impede or prevent another person from exercising the right to vote,"[120] will liability of social media companies be too difficult to establish in court? Without an intent requirement, however, how does the State ensure that liability is not incurred for a typographical error or other honest mistake?[121]

- How do States craft laws that provide sufficient incentive to social media companies to police their platforms and avoid discriminatory dissemination of suppressive ads without prompting risk-averse social media companies to ban all targeted political ads?[122] Such a ban could harm less wealthy candidates and non-profit voter mobilization groups that rely on targeted social media ads and lack resources to invest in expensive traditional television and radio ads.

While these questions are challenging, they are not insurmountable. They require real thought and an acknowledgment of competing values. The difficulties of these questions do not warrant a conclusion that

---

[119] Facebook now has policies on all of these issues. Guy Rosen et al., *Helping to Protect the 2020 US Elections*, FACEBOOK (Oct. 21, 2019), https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/ [https://perma.cc/6JXM-VVXL] (announcing policies prohibiting "paid advertising that suggests voting is useless or meaningless, or advises people not to vote," "[m]isrepresentation of the dates, locations, times and methods for voting or voter registration," and "[m]isrepresentation of who can vote, qualifications for voting, whether a vote will be counted and what information and/or materials must be provided in order to vote"); *Community Standard #17. Misrepresentation*, FACEBOOK, https://www.facebook.com/communitystandards/misrepresentation (last visited Jan. 4, 2020) [https://perma.cc/KK9C-Z28W] ("We believe that people are more accountable for their statements and actions when they use their authentic identities. That's why we require people to connect on Facebook using the name they go by in everyday life."); *Community Standard #20. Inauthentic Behavior*, FACEBOOK, https://www.facebook.com/communitystandards/inauthentic_behavior (last visited Jan. 4, 2020) [https://perma.cc/SP5W-RT6P] ("In line with our commitment to authenticity, we don't allow people to misrepresent themselves on Facebook [or] use fake accounts.").

[120] For the People Act, H.R. 1, 116th Cong. (2019).

[121] *See* CTR. FOR THE ADVANCEMENT OF PUB. INTEGRITY, PROSECUTING VOTE SUPPRESSION, *supra* note 26, at 4 (acknowledging the challenges with identifying intent, but explaining that the lack of an intent requirement chills protected speech).

[122] *See, e.g.*, Kate Conger, *Twitter Will Ban All Political Ads, C.E.O. Jack Dorsey Says*, N.Y. TIMES (Oct. 30, 2019), https://www.nytimes.com/2019/10/30/technology/twitter-political-ads-ban.html [https://perma.cc/5N69-6KEM].

social media companies should be free to disseminate suppressive ads targeted along racial lines.

This Essay has focused on State power, but the ultimate goal should be federal rules that provide strong civil rights protections and clear standards. As soon as possible, Congress should enact voter deception and suppression legislation that includes regulation of social media companies.[123]

Also, Congress should explicitly acknowledge that Section 230 does not provide a defense to federal and state civil rights claims arising from online ad targeting. While this Essay has established this in the context of targeted voter-suppression ads, Congress should explicitly articulate this carve-out as applied to all types of civil rights claims arising from online ad targeting (e.g., discriminatory dissemination of ads in voting, employment, lending, housing).[124] As discussed above, carve-outs

[123] Several deceptive practices bills have been introduced, but they generally need to do more to address the unique responsibilities of social media platforms in preventing deceptive practices and the discriminatory dissemination of suppressive ads. *See* Voter Empowerment Act, H.R. 1275, 116th Cong. (2019); Voter Empowerment Act, S. 549, 116th Cong. (2019); For the People Act, H.R. 1, 116th Cong. (2019); For the People Act, S. 949, 116th Cong. (2019); Deceptive Practices and Voter Intimidation Prevention Act, H.R. 6607, 115th Cong. (2018); Deceptive Practices and Voter Intimidation Prevention Act, S. 3279, 115th Cong. (2018); Deceptive Practices and Voter Intimidation Prevention Act, H.R. 5815, 112th Cong. (2012); Deceptive Practices and Voter Intimidation Prevention Act, S. 1994, 112th Cong. (2011); Deceptive Practices and Voter Intimidation Prevention Act, H.R. 97, 111th Cong. (2009); Deceptive Practices and Voter Intimidation Prevention Act, S. 453, 110th Cong. (2007); Deceptive Practices and Voter Intimidation Prevention Act, H.R. 1281, 110th Cong. (2007); Deceptive Practices and Voter Intimidation Prevention Act, S. 4069, 109th Cong. (2006); Deceptive Practices and Voter Intimidation Prevention Act, S. 1975, 109th Cong. (2005); *see also* CTR. FOR THE ADVANCEMENT OF PUB. INTEGRITY PROSECUTING VOTE SUPPRESSION, *supra* note 26, at 9 nn.81-83 (2019) (citing and describing various federal deceptive practices bills).

[124] *See* Olivier Sylvain, *Discriminatory Designs on User Data; Exploring How Section 230's Immunity Protections May Enable or Elicit Discriminatory Behaviors Online*, KNIGHT FIRST AMEND. INST. COLUM. U. (Apr. 1, 2018), https://knightcolumbia.org/content/discriminatory-designs-user-data (last visited Jan. 23, 2020) [https://perma.cc/CB3L-NCSJ] ("There is no reason why Congress couldn't also write in an explicit exception to Section 230 immunity for violations of civil rights laws."). *But see* Citron, *supra* note 79 (asserting that Congress should avoid carving out exceptions to 230 like civil rights violations because such a piecemeal approach risks continuing to provide immunity to other areas that do not deserve it, proposing instead immunity for a platform that "takes reasonable steps to prevent or address unlawful uses of its services," and arguing that this reasonable duty standard is sufficiently flexible to adapt to different platforms and evolving technologies); Olivier Sylvain, *Recovering Tech's Humanity*, 119 COLUM. L. REV. F. 252, 275 n.150 (2019) (describing the proposal by Citron and Wittes to reform Section 230 through conditioning immunity on intermediaries' "exercise of a reasonable standard of care" as a "sensible way of reforming the law").

already exist for violations in various areas of the law (e.g., federal criminal law, intellectual property law, the Electronic Communications Privacy Act of 1986 and similar State laws, federal sex trafficking law).[125] Companies should not be able to assert that the Civil Rights Act of 1964, the Fair Housing Act, and other landmark civil rights laws are inapplicable simply because a company discriminates online rather than at a brick-and-mortar storefront.

CONCLUSION

Discriminatory dissemination of deceptive ads by social media companies presents unprecedented dangers in facilitating voter suppression, and Congress should immediately enact strong and clear laws to address these threats. Recognizing that Congress has failed to act, however, States should take the initiative. States routinely enact civil rights antidiscrimination measures. Section 230 does not limit the power of States to hold social media companies legally responsible for using data collection and algorithms to further discriminatory delivery of suppressive ads to protected classes of voters. Social media companies using such techniques are ineligible for Section 230 immunity because they "contribute materially" to discrimination.

[125] *See, e.g.*, 47 U.S.C. § 230(e)(1)-(5) (2019).