

POLICING THE DIGITAL PUBLIC SQUARE: THE DUTY OF NON-MANIPULATION AS AN ALTERNATIVE TO FREE SPEECH RESTRICTIONS ON SOCIAL MEDIA

BEN KABE[†]

INTRODUCTION

James Madison was profoundly concerned with preventing passion from undermining reason in the realm of popular government.¹ Factions—which Madison defined as groups of citizens united by a common impulse or interest—were, in his mind, inevitable.² His addition to our Constitution makes clear that free speech is a vital and inviolable right, but one that it is not without dangers. Though its benefits surely outweigh its costs, its ability to provoke and persuade the people’s passions is an ever-present hazard. These democratic dilemmas are as salient in today’s hyper-communicative world as they were at the Founding. Madison believed that America’s sprawling geography and growing population would help “prevent passionate mobs from mobilizing.”³ He obviously did not predict social media.

At present, YouTube may be one of this century’s “most powerful radicalizing instruments.”⁴ Like its contemporaries—Twitter, Facebook, Instagram, and others—it has been used to rapidly spread misinformation and conspiracy theories to broad, susceptible swaths of people.⁵ The

[†] J.D., Northwestern Pritzker School of Law, 2021. I am grateful to Professor Janice Nadler for her guidance on this Note. I also want to thank the editors of the Connecticut Public Interest Law Journal for all of their great work on this project. Finally, I want to thank my friends and family for their unending support.

¹ See THE FEDERALIST NO. 10 (James Madison), available at https://avalon.law.yale.edu/18th_century/fed10.asp.

² *Id.*

³ Jeffrey Rosen, *America Is Living James Madison’s Nightmare* THE ATLANTIC (Oct. 2018), <https://www.theatlantic.com/magazine/archive/2018/10/james-madison-mob-rule/568351/>.

⁴ Max Fisher & Amanda Taub, *How YouTube Radicalized Brazil*, N.Y. TIMES (Aug. 11, 2019), <https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html> (quoting social media scholar Zeynep Tufekci).

⁵ “Examples include the lunacy of the Comet Pizza story (a.k.a. Pizzagate), the various anti-Obama

algorithms utilized by these sites capture vast amounts of data and continue to get better at ensuring that users stay engaged for longer periods of time. Social media companies are able to compile their users' data into detailed psychological profiles, which can be used to "nudge" users toward ever more extremist and conspiratorial content in pursuit of virality and advertising revenue.⁶ An experiment by *The Atlantic* revealed the distance on YouTube between a pasta cooking tutorial and anti-LGBT junk science—three clicks.⁷

Facebook alone has over two billion users across the planet.⁸ The company's task now is to keep users scrolling the site for an incessantly greater portion of their day.⁹ In this endeavor, Facebook uses engagement algorithms as closely held as YouTube's, as do many other comparable sites.¹⁰ This is startling considering that these algorithms often alter user behavior without identifying profoundly biased content or stories designed to promote fear, mistrust, or outrage.¹¹

Still, social media may also be a powerful tool for good. These platforms allow for instantaneous communication with any other user capable of accessing the internet. They can help spread democratic ideals in otherwise closed regimes.¹² The seemingly infinite sources of

birther conspiracies, and Alex Jones's claim that the Sandy Hook Elementary School shooting that left 20 children dead was a 'complete fake' staged by the government to promote gun control." Thomas B. Edsall, *The Trump Voters Whose 'Need for Chaos' Obliterates Everything Else*, N.Y. TIMES (Sept. 4, 2019), <https://www.nytimes.com/2019/09/04/opinion/trump-voters-chaos.html> (citing Michael Bang Petersen, Mathias Osmundsen, & Kevin Arceneaux, *The "Need for Chaos" and Motivations to Share Hostile Political Rumors*, PSYARXIV (May 2020), <https://psyarxiv.com/6m4ts>)

⁶ See Fisher & Taub, *supra* note 4; see McKay Coppins, *The Billion-Dollar Disinformation Campaign to Reelect the President*, THE ATLANTIC (Feb. 10, 2020), <https://www.theatlantic.com/magazine/archive/2020/03/the-2020-disinformation-war/605530/>

⁷ Derek Thompson, *Why the Internet is So Polarized, Extreme, and Screamy*, THE ATLANTIC (May 23, 2019), <https://www.theatlantic.com/ideas/archive/2019/05/how-did-the-far-right-take-over-the-web/590047/>

⁸ Jack M. Balkin, *The First Amendment in the Second Gilded Age*, 66 BUFF. L. REV. 979, 995 (2018).

⁹ *Id.* at 995–96.

¹⁰ Tobias Rose-Stockwell, *This is How Your Fear and Outrage Are Being Sold for Profit*, QUARTZ (July 28, 2017), <https://qz.com/1039910/how-facebooks-news-feed-algorithm-sells-our-fear-and-outrage-for-profit/>.

¹¹ *Id.* Notably, in an October 2019 hearing before Congress, Facebook's founder and CEO, Mark Zuckerberg, said that Facebook would not remove objectively false political advertising. *Hearing Before the House Committee on Financial Services*, 116th Cong. 63 (Oct. 23, 2019) (testimony of Mark Zuckerberg), available at <https://www.c-span.org/video/?465293-1/facebook-ceo-testimony-house-financial-services-committee>. This is particularly distressing in light of the advancing viability of the "deepfake." This practice makes it possible to alter videos using complex machine learning in a manner which makes manipulation quite difficult to detect.

¹² Philip N. Howard et al., *Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?* (Project on Info. Tech. & Pol. Islam, Working Paper No. 2011.1, 2011), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2595096. A good example of this is social

information online, coupled with the user engagement cultivated by these sites, may be helping to facilitate a more participatory democracy.¹³ Social media is neither devil nor angel. It is simply an instrument which may be employed to complement human nature, for better or worse.

Courts have been cautious in attempting to balance individual interests in free association and communication with the dangers of our increasingly algorithmic and anonymous world.¹⁴ In *Packingham v. North Carolina*, the Supreme Court struck down a law that prohibited sex offenders from accessing social networking sites.¹⁵ Justice Kennedy, holding that the law violated the First Amendment, noted that this was one of the first cases taken by the Court to address the relationship between free speech and the modern internet.¹⁶ When addressing this relationship, he said, the Court “must exercise extreme caution.”¹⁷

Congress also finds itself in a difficult position legislatively. Section 230 of the Communications Decency Act provides social media companies with immunity from liability based on speech shared on their platforms by their users.¹⁸ Although this law is increasingly unpopular, experts warn against its hasty alteration.¹⁹ Only one real change has been made to the provision since its enactment.²⁰ As it stands, social media companies enjoy broad immunities from civil or criminal liability based on content that they merely host rather than create.²¹

Neither courts nor lawmakers have sufficiently addressed potential alternatives to this challenge. Professor Jack Balkin, however, has argued for a new regime to deal with some of the questions presented by our new

media’s role in the Egyptian uprising during the Arab Spring, which is discussed further in Section I(a) below.

¹³ Homero Gil de Zúñiga, Nakwon Jung, & Sebastián Valenzuela, *Social Media Use for News and Individuals’ Social Capital, Civic Engagement and Political Participation*, 17 J. COMPUT.-MEDIATED COMM’N 319 (2012),

<https://academic.oup.com/jcmc/article/17/3/319/4067682#94905782>.

¹⁴ See *Packingham v. North Carolina*, 137 S. Ct. 1730 (2017) (addressing First Amendment protections on social media); *Liverman v. City of Petersburg*, 844 F.3d 400 (4th Cir. 2016) (finding that the social networking policy, which operated as a prior restraint on speech, violated the First Amendment’s Free Speech Clause).

¹⁵ 137 S. Ct. 1730.

¹⁶ *Id.* at 1736.

¹⁷ *Id.*

¹⁸ 47 U.S.C. § 230 (1996).

¹⁹ Matt Laslo, *The Fight Over Section 230—and the Internet as We Know It*, WIRED (Aug. 13, 2019), <https://www.wired.com/story/fight-over-section-230-internet-as-we-know-it/>

²⁰ Platforms are no longer protected from liability related to prostitution and sex trafficking. *Id.*

²¹ *Id.*

digital age.²² Professor Balkin put forward the idea of treating large social media companies as “information fiduciar[ies]” based on end-users’ limited information and vulnerability and the companies’ expertise and position of trust.²³ As a result of this power imbalance, he argues, social media companies owe their end-users fiduciary duties, including a duty of non-manipulation.²⁴

Through relevant common law doctrine, this Note will expand on the duty of non-manipulation as an alternative to abridging First Amendment rights related to dangerous behavior online. Social media companies use algorithms that incentivize extremist and sensationalist content.²⁵ By imposing a duty of non-manipulation, these companies would be forced to halt the mechanisms that actively encourage radicalization. As a result, much of the speech that courts are hesitant to proscribe would be less likely to materialize or spread.

This Note will begin (in Part I) by discussing social science that highlights aspects of social media as a social good as well as research that describes many of the platforms’ unintended societal harms. Part II will review the relationship between social media and the First Amendment in modern case law and legal scholarship. It will also assess online platforms’ current statutory liability for their users’ speech and content. Finally, in Part III, this Note will explain Professor Jack Balkin’s theory of social media companies as “information fiduciaries” and expand on the duty of non-manipulation as an alternative to some restrictions of First Amendment rights.

I. SOCIAL MEDIA AND BEHAVIOR

Social media is comprised of “forms of electronic communication . . . through which users create online communities to share information, ideas, personal messages, and other content”²⁶ This definition is illuminating because, at face value, these platforms serve only to facilitate the creation and diffusion of ideas. Social networking sites often open up lines of communication that were previously beyond reach and, in doing so, benefit their users and society generally. What this definition does not capture, however, is the way in which communication is altered on these sites. The online public sphere, in some cases, can transform our behavior and

²² Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2016).

²³ *Id.* at 1186, 1222.

²⁴ *Id.* at 1233.

²⁵ See Fisher & Taub, *supra* note 4.

²⁶ *Social Media*, MERRIAM-WEBSTER.COM,

<https://www.merriam-webster.com/dictionary/social%20media> (last visited Apr. 9, 2021).

dialogue in dangerous ways—based in part on human nature, but also on the algorithmic exploitation of our innate tendencies. In order to understand and properly prevent potential social harms, it is necessary to review some of the positive and negative aspects of our modern forms of connection.

A. *The Promise of Connectivity*

The Arab Spring shows just how quickly social media can accelerate social change. In December 2010, a Tunisian street vendor named Mohammed Bouazizi set himself on fire in protest of his government.²⁷ About one month later, in response to protests sparked by Bouazizi's act of defiance, then-Tunisian president Zine al-Abidine Ben Ali abandoned his more than twenty-year rule and fled the country.²⁸ Tunisia, one of the Arab world's "most repressive regimes," held its first democratic parliamentary elections in October of that same year.²⁹ This astonishingly swift change in power set in motion similar protests in other countries across the region.³⁰ Of course, as with any large-scale social change, there was not one single causal factor. Even so, it appears now that social media was pivotal in the acceleration of these events.³¹

According to research from the Project on Information Technology & Political Islam, social media played a central role in the Arab Spring by facilitating and shaping political discussion, allowing for the conversations that preceded important protests on the ground, and spreading democratic ideas internationally.³² This is not to say that these revolutions could not have occurred without social media, nor that the outcomes of the Arab Spring were universally successful. Instead, this example simply shows that social networking sites provide a new avenue for polities to effectively organize for some form of social change.

People who have formerly been blocked from political conversation are now widely provided some access through social media. Media Professor Clay Shirky argues that political freedom requires a society that is "densely connected enough to discuss the issues presented to the

²⁷ *Arab Spring*, HISTORY.COM (Jan. 10, 2018), <https://www.history.com/topics/middle-east/arab-spring>

²⁸ Angeliqe Chrisafis, *Zine al-Abidine Ben Ali Forced to Flee Tunisia as Protesters Claim Victory*, THE GUARDIAN (Jan. 14, 2011), <https://www.theguardian.com/world/2011/jan/14/tunisian-president-flees-country-protests>

²⁹ HISTORY.COM, *supra* note 27; Chrisafis, *supra* note 28.

³⁰ HISTORY.COM, *supra* note 27.

³¹ Philip N. Howard et al., *supra* note 12.

³² *Id.*

public.”³³ With low barriers to entry, people from every economic and social class can use social media’s dense web of connection to ingest—and express views on—the important issues of the day. Indeed, there is some evidence that suggests social media’s ubiquity may lead to a more participatory democracy.³⁴

The Arab Spring is the most often used example of social media’s informing and organizing power, and for good reason. Beyond the Arab Spring, however, social media has been implicated in a range of pro-democratic outcomes—from increasing government accountability in the small town of Jun, Spain,³⁵ to providing the means to organize and overturn a fraudulent election in Moldova.³⁶ People are empowered to resist oppression when supplied with a free flow of information and a means to coordinate action. Those same hopeful characteristics, however, can be manipulated to cause harm to society.

B. Unintended Social Harms

Social media companies’ most valuable asset is their end-users’ data.³⁷ In 2017, Google, Facebook, Amazon, Apple, and Microsoft made over \$25 billion in profit.³⁸ The rapid growth of personal data collection, and the power that it grants these tech behemoths, prompted the Economist to claim that data has surpassed oil as the world’s most valuable resource.³⁹ In response to this power, many authorities have initiated antitrust

³³ Sarah Joseph, *Social Media, Political Change, and Human Rights*, 35 B.C. INT’L & COMP. L. REV. 145, 152 (2012) (quoting Clay Shirky, *The Political Power of Social Media*, 90 FOREIGN AFF. 28, 34 (2011)).

³⁴ Homero Gil de Zúñiga, et al., *supra* note 13, at 329.

³⁵ Jun is a small town in southern Spain. For many of its 3,500 residents, Twitter is the main way that they can communicate with their local government. Communicating in this way not only shaved 13% off the local budget, but also increased the interaction between residents and local officials. Mark Scott, *The Spanish Town That Runs on Twitter*, N.Y. TIMES (Jun. 7, 2016), <https://www.nytimes.com/2016/06/09/technology/the-spanish-town-that-runs-on-twitter.html>.

³⁶ In 2009, the Moldovan Communist Party tried to maintain control of Parliament through a fraudulent election, but lost power because of massive protests coordinated by text, Facebook, and Twitter. Joseph, *supra* note 33.

³⁷ See Matthew Johnson, *How Facebook Makes Money: Advertising, Payments, and Other Fees*, INVESTOPEDIA, <https://www.investopedia.com/ask/answers/120114/how-does-facebook-fb-make-money.asp> (last updated Jan. 30, 2021).

³⁸ *The World’s Most Valuable Resource is no Longer Oil, but Data*, THE ECONOMIST (May 6, 2017), <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>.

³⁹ *Id.*

investigations into some of these large companies.⁴⁰

There are a couple of major concerns over social media companies' capture and use of data. The first, and most natural, is privacy. For instance, Google recently partnered with the second-largest health system in the U.S.⁴¹ Through their secret "Project Nightingale," they began to collect and utilize massive amounts of health data, including lab results, patient names, and complete health histories.⁴² This was done without notifying doctors or patients.⁴³ Beyond its Orwellian nature, this project is concerning in light of Google's past data protection failures, such as an undisclosed flaw that exposed hundreds of thousands of users' personal data.⁴⁴ Data privacy is an urgent concern. Yet another urgent concern, and the focus of this Note, is data collecting companies' algorithmic manipulation of users.

Social media companies depend on capturing ever-more of their users' attention to continue their increasing profitability.⁴⁵ They do this by algorithmically maximizing user engagement.⁴⁶ For example, Facebook, like many of these companies, uses an algorithmic news-feed editor that follows every move users make on the site, allowing it to constantly get better at predicting what each user will click and what will capture their attention.⁴⁷ It then curates users' news-feeds to include the most engaging material while filtering out the material less likely to keep the user on their site.⁴⁸ It does not do so, however, with any consideration for the material it is propagating; meaning that it "doesn't identify content that is profoundly biased, or stories that are designed to propagate fear, mistrust, or outrage."⁴⁹

This method of engagement has been enormously influential across the globe. In Brazil, YouTube's recommendation algorithm "systematically

⁴⁰ See, e.g., Steve Lohr, *New Google and Facebook Inquiries Show Big Tech Scrutiny Is Rare Bipartisan Act*, N.Y. TIMES (Sep. 6, 2019), <https://www.nytimes.com/2019/09/06/technology/attorney-generals-tech-antitrust-investigation.html>

⁴¹ Rob Copeland, *Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans*, WALL. ST. J., <https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790> (last updated Nov. 11, 2019).

⁴² *Id.*

⁴³ *Id.*

⁴⁴ In 2018, a Wall Street Journal investigation revealed that Google decided not to disclose this flaw, in part, out of fear of a regulatory backlash. *Id.*

⁴⁵ See generally, TIM WU, ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS (Alfred A. Knopf 2016).

⁴⁶ Rose-Stockwell, *supra* note 10.

⁴⁷ *Id.*

⁴⁸ *Id.*

⁴⁹ *Id.*

diverted users to far-right and conspiracy channels.”⁵⁰ This aided in the election of a formerly fringe, far-right politician to the presidency.⁵¹ In one instance, YouTube led a young man from amateur guitar lesson videos deep into far-right conspiracies.⁵² Many parents were given disinformation while in search of medical advice, which impinged authorities’ efforts to fight the Zika outbreak.⁵³ The false information in these videos was so compelling that it incited death threats against public health advocates.⁵⁴ Brazil is just one example of the effects of manipulation on a mass scale.⁵⁵

This manipulation can compel individual users to carry out extremist acts. Professor Taina Bucher argues that social media algorithms like Facebook’s create a “threat of invisibility” which manifests in a “constant possibility of disappearing and becoming obsolete.”⁵⁶ Just as these algorithms drive consumers toward more radical content in order to keep them engaged, they incentivize creators to produce more radical content if they hope to remain relevant. One example of this phenomenon is Cesar Sayoc, who began his social media career posting about food, workouts, and sports.⁵⁷ He soon, however, resembled a modern extremist, radicalized online by misinformation and right-wing conspiracies.⁵⁸ In 2018 he was arrested in Florida and charged with sending at least twelve pipe bombs to President Trump’s critics.⁵⁹ In March 2019, an Australian extremist shot and killed fifty-one Muslim worshipers at their mosques in New Zealand

⁵⁰Fisher & Taub, *supra* note 4, (citing Virgilio Almeida et al., *Understanding Video Interactions in YouTube*, 16TH ACM INT’L CONFERENCE ON MULTIMEDIA, MM ’08 (2008), available at <http://www.decom.ufop.br/fabricio/download/multimedia08>).

⁵¹*Id.*

⁵²This is the story of then-sixteen year old Brazilian Matheus Dominguez. When he began using YouTube, he was only interested in learning to improve his skills on the guitar. YouTube’s artificial intelligence system, which learns from users’ behavior, recommended videos for Matheus. It eventually ended up recommending videos of “paranoid far-right rants” and conspiracies. Matheus bought in. Members of Brazil’s far-right, who now control the Brazilian presidency, said their ascendancy would not have been possible without YouTube’s recommendation engine. *Id.*

⁵³*Id.*

⁵⁴*Id.*

⁵⁵Another often-discussed instance of data manipulation is Cambridge Analytica’s role in the 2016 U.S. presidential election. The company collected data from millions of users and utilized psychographics to target political advertising. See Carole Cadwalladr et al., *How Trump Consultants Exploited the Facebook Data of Millions*, N.Y. TIMES (Mar. 17, 2018), <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>.

⁵⁶Taina Bucher, *Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook*, NEW MEDIA & SOC’Y 1164, 1171–75 (2012), <https://journals.sagepub.com/doi/10.1177/1461444812440159>

⁵⁷Kevin Roose, *Cesar Sayoc’s Path on Social Media: From Food Photos to Partisan Fury*, N.Y. TIMES (Oct. 27, 2018), <https://www.nytimes.com/2018/10/27/technology/cesar-sayoc-facebook-twitter.html>

⁵⁸*Id.*

⁵⁹*Id.*

and livestreamed the entire massacre on Facebook.⁶⁰ Social media algorithms fostered, validated, and encouraged these hateful views.⁶¹

As machine learning continues to improve, another worrying development has entered the public discourse—the “deepfake.”⁶² In a deepfake, artificial intelligence is used to alter faces, voices, or both to create a believable fake video.⁶³ It is incredibly difficult to detect manipulation in a deepfake.⁶⁴ Recently, a video of a speech by Democratic House Speaker Nancy Pelosi was slowed to make her appear drunk.⁶⁵ This was believed and spread by millions of people online.⁶⁶ In fact, the President of the United States shared a doctored video of the Speaker, and his personal attorney, Rudolph Giuliani, tweeted a link to the video with the message, “What is wrong with Nancy Pelosi? Her speech pattern is bizarre.”⁶⁷ This video was merely slowed.

Deepfake technology, however, could have believably changed the face of the speaker and altered the content of her speech.⁶⁸ Facebook’s policy against policing political speech—even blatant lies—would allow these videos to stand.⁶⁹ This presents a catch-22; either allow disinformation to spread online, or allow a private company to decide what speech needs to be censored. Neither is an ideal outcome for democracy.

Big technology companies are understandably protective of their algorithms.⁷⁰ But these tools dramatically affect users’ online experiences. Some algorithms utilize around 100,000 different variables to “optimize”

⁶⁰ Allyson Haynes Stuart, *Social Media, Manipulation, and Violence*, 15 S.C. J. INT’L. L. & BUS. 100, 117–118 (2019) (citing Daniel Victor, *In Christchurch, Signs Point to a Gunman Steeped in Internet Trolling*, N.Y. TIMES (Mar. 15, 2019), <https://www.nytimes.com/2019/03/15/world/asia/new-zealand-gunman-christchurch.html>

⁶¹ *Id.*

⁶² Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J. L. & TECH. 1 (2020).

⁶³ David Güera & Edward J. Delp, *Deepfake Video Detection Using Recurrent Neural Networks*, VIDEO AND IMAGE PROCESSING LAB. (VIPER), PURDUE UNIV. (Feb. 14, 2019), <https://ieeexplore.ieee.org/abstract/document/8639163/citations#citations>

⁶⁴ *Id.*

⁶⁵ Drew Harwell, *Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread Across Social Media*, WASH. POST (May 24, 2019), <https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/>

⁶⁶ *Id.*

⁶⁷ *Id.*

⁶⁸ *Id.*

⁶⁹ See *Hearing before the House Committee on Financial Services*, *supra* note 11.

⁷⁰ See Ryan Holmes, *Do Social Networks have an “Algorithmic Responsibility” to Users?*, MEDIUM (Jan 25, 2017), <https://medium.com/@invoker/do-social-networks-have-an-algorithmic-responsibility-to-users-44e481332420> (citing Josh Constine, *How Facebook News Feed Works*, TECH CRUNCH (Sept. 6, 2016), <https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/>).

users' news feeds, and all of them are constantly changing.⁷¹ Of course, these algorithms are only exploiting what is already innate in human behavior.⁷² The challenge for the law is to prevent them from indulging humanity's worst impulses while protecting these companies' rights and the rights of the individuals who use their platforms.

II. FIRST AMENDMENT CONCERNS

Many of the previously discussed benefits of online social networks stem from users' ability to speak and associate without restriction. Concerned parties have tried to alter the free speech landscape on social media platforms in a few ways. Proponents of First Amendment protections on social media have argued that the platforms are public fora that should be governed by First Amendment doctrine.⁷³ Legislators, experts, and activists have argued over a section of the U.S. Code that frees platforms from liability for their users' speech.⁷⁴ Lawmakers have also attempted to restrict speech on social media in certain circumstances, but courts have been resistant of their efforts.⁷⁵ In most cases, when authorities have attempted to prevent speech online, they have been unable to clear the significant hurdles created by constitutional and statutory protection of free speech.

A. *Constitutional Restrictions—or Lack Thereof—on Social Media Companies*

Twitter and other social media companies can, and do, ban users from their site.⁷⁶ They can also delete posts that conflict with their content policies.⁷⁷ Given their immense public importance, does this violate users' Constitutional rights? The First Amendment applies to state, rather than private, actors.⁷⁸ Still, advocates have argued that social media sites serve

⁷¹ *Id.*

⁷² Algorithmic racial or gender bias, for example, is based in part on the human bias in the data from which the artificial intelligence was trained. See John Naughton, *To Err is Human – Is That Why we Fear Machines that can be Made to Err Less?*, THE GUARDIAN (Dec. 14, 2019), <https://www.theguardian.com/commentisfree/2019/dec/14/err-is-human-why-fear-machines-made-to-err-less-algorithmic-bias>

⁷³ Mason C. Shefa, *First Amendment 2.0: Revisiting Marsh and the Quasi-Public Forum in the Age of Social Media*, 41 U. HAW. L. REV. 159, 184–86 (2018).

⁷⁴ 47 U.S.C. § 230(c)(2) (2018).

⁷⁵ See, e.g., *Packingham v. North Carolina*, 137 S. Ct. 1730, 1738 (2017) (striking down a North Carolina law that prevented sex offenders from accessing certain social media pages as a violation of the First Amendment).

⁷⁶ See *Twitter Terms of Service*, TWITTER, <https://twitter.com/en/tos> (last visited Mar. 21, 2021).

⁷⁷ *Id.*

⁷⁸ U.S. CONST. amend. I.

as public or quasi-public fora, and as such, are governed by relevant First Amendment law.⁷⁹

This argument often stems from the Supreme Court case *Marsh v. Alabama*, decided in 1946.⁸⁰ In *Marsh*, a Jehovah's Witness, Grace Marsh, was arrested for handing out religious pamphlets in a town wholly owned by a private corporation.⁸¹ The corporation had posted a notice banning solicitation without its permission.⁸² The Court held that the state could not infringe on Marsh's First Amendment right to distribute her pamphlets simply because she was on privately owned property.⁸³ It noted that, other than the community's ownership, it had "all the characteristics of an American town."⁸⁴ Due to this, the Court said, the freedom afforded by the First Amendment outweighs the Constitutional property rights of the private corporation in a balancing test.⁸⁵

In the same vein, the Court discussed the public forum doctrine in *Amalgamated Food Employee Union Local 590 v. Logan Valley Plaza, Inc.*⁸⁶ Here, the Court answered whether union picketers outside of a shopping mall could be enjoined by the mall's owner—again balancing property interests with First Amendment rights.⁸⁷ Justice Marshall reiterated the Court's holding in *Marsh* and found that the shopping center was the equivalent of a "business block" for the purposes of the First Amendment.⁸⁸ As such, the business owners were not allowed to enjoin the peaceful picketers solely by asserting their property rights.⁸⁹

Over the subsequent decades, however, the Supreme Court overruled *Logan Valley*⁹⁰ and, along with U.S. Circuit Courts, limited *Marsh* significantly.⁹¹ In *Lloyd Corp. v. Tanner*, a shopping mall that was open to

⁷⁹ See Shefa, *supra* note 73.

⁸⁰ *Marsh v. Alabama*, 326 U.S. 501 (1946).

⁸¹ *Id.* at 502–03.

⁸² *Id.* at 503.

⁸³ *Id.* at 509.

⁸⁴ *Id.* at 502.

⁸⁵ *Id.* at 509.

⁸⁶ *Amalgamated Food Emp. Union Local 590 v. Logan Valley Plaza, Inc.*, 391 U.S. 308 (1968), abrogated by *Hudgens v. N. L. R. B.*, 424 U.S. 507 (1976).

⁸⁷ *Id.* at 309.

⁸⁸ *Id.* at 325.

⁸⁹ *Id.*

⁹⁰ *Hudgens v. N.L.R.B.*, 424 U.S. 507 (1976).

⁹¹ See *Flagg Bros., Inc. v. Brooks*, 436 U.S. 149 (1978); *Lloyd Corp. v. Tanner*, 407 U.S. 551 (1972); *Manhattan Cmty. Access Corp. v. Halleck*, 139 S. Ct. 1921 (2019); *Gallagher v. Neil Young Freedom Concert*, 49 F.3d 1442 (10th Cir. 1995).

the public⁹² banned the distribution of handbills made to protest the war in Vietnam.⁹³ The court distinguished *Logan Valley* by limiting its holding to acts “directly related in [their] purpose” to the function of the shopping center and where an injunction would allow no other opportunity for the picketers to express their view.⁹⁴ It similarly distinguished *Marsh* by noting that this shopping center did not exercise municipal power.⁹⁵ The Court explained that “property [does not] lose its private character merely because the public is generally invited to use it for designated purposes,” and held that the public use of this mall did not entitle its visitors to such First Amendment rights.⁹⁶

In *Flagg Bros Inc. v. Brooks*,⁹⁷ decided in 1978, the Court adopted Justice Black’s interpretation of *Marsh*’s limited reach, finding that private property is a public forum “when that property has taken on *all* the attributes of a town.”⁹⁸ The Eleventh Circuit declined to extend *Marsh*’s public function rationale to a concert at a state university in *Gallagher v. Neil Young Freedom Concert*.⁹⁹ In June 2019, the Supreme Court dealt the most recent blow to *Marsh*’s possible revival or expansion.¹⁰⁰ The Court held in *Manhattan Cmty. Access Corp. v. Halleck* that a private nonprofit corporation that operated a public access channel was not a state actor subject to First Amendment constraints.¹⁰¹ It explained that a private entity must exercise powers “traditionally exclusively reserved to the State” and stressed that “very few functions fall into that category.”¹⁰² Most importantly for social media companies, the Court stated that “merely hosting speech by others . . . does not alone transform private entities into state actors subject to First Amendment constraints.”¹⁰³

The evolution of public fora doctrine governing private owners is revealing. It is now unlikely that social media companies will be considered state actors subject to the speech constraints of the U.S. Constitution. Legislatures and the executive branch, however, are also

⁹² The center had private stores but open walkways which remained walkable after the shops were closed. In fact, the center encouraged the public to come and window shop. *Lloyd Corp.*, 407 U.S. 551.

⁹³ *Id.*

⁹⁴ *Id.* at 563.

⁹⁵ *Id.* at 569.

⁹⁶ *Id.* 569–70.

⁹⁷ 436 U.S. 149 (1978).

⁹⁸ *Id.* at 159 (emphasis in original).

⁹⁹ *Gallagher v. Neil Young Freedom Concert*, 49 F.3d 1442, 1457 (10th Cir. 1995).

¹⁰⁰ *Manhattan Cmty. Access Corp. v. Halleck*, 139 S. Ct. 1921 (2019).

¹⁰¹ *Id.*

¹⁰² *Id.* at 1928–29.

¹⁰³ *Id.* at 1930.

attempting to alter the free speech landscape on these platforms. As recent decisions show, they are having mixed success doing so.

B. Statutory Protection of Speech on Online Platforms

Congress enacted Title 47 Chapter 5 of the United States Code in order to regulate common carriers of telecommunication signals.¹⁰⁴ Tucked within this Chapter is 47 U.S.C. § 230, which protects platforms from liability for content created or posted by their users.¹⁰⁵ The relevant section of the code states that “[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”¹⁰⁶ “Interactive computer service” is defined broadly in subsection (f)(2) of the provision.¹⁰⁷ This definition includes social media companies.¹⁰⁸ The courts have also interpreted this section broadly,¹⁰⁹ applying the section’s protection to “claims for defamation, negligence, intentional infliction of emotional distress, privacy, terrorism support, and more.”¹¹⁰

Section 230 was enacted to enable innovation online and to protect free speech principles.¹¹¹ Since then, its popularity has declined.¹¹² Numerous legislators from both sides of the aisle now attack the law and offer differing views on how it should be altered or destroyed.¹¹³ Still, as the law currently stands, social media companies are shielded from liability based on user generated content and there is no consensus on whether or how to fix this section.¹¹⁴ In 2018, Congress amended the law to prevent it from providing immunity to platforms that promote or facilitate prostitution or

¹⁰⁴ 47 U.S.C. § 151 (1996).

¹⁰⁵ 47 U.S.C. § 230 (1996).

¹⁰⁶ 47 U.S.C. § 230(c)(1) (2018).

¹⁰⁷ These services are defined as follows: “‘interactive computer service’ means any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.” 47 U.S.C. § 230(f)(2).

¹⁰⁸ See, e.g., *Klayman v. Zuckerberg*, 753 F.3d 1354, 1357 (D.C. Cir. 2014).

¹⁰⁹ See Mark A. Lemley, *Rationalizing Internet Safe Harbors*, 6 J. TELECOMM. & HIGH TECH. L. 101 (2007).

¹¹⁰ Brown, *supra* note 62, at 42, (citing *Batzel v. Smith*, 333 F.3d 1018, 1020, 1026–27 (9th Cir. 2003)) (defamation); *Ben Ezra, Weinstein, & Co. v. Am. Online, Inc.*, 206 F.3d 980, 983–84 (10th Cir. 2000) (defamation & negligence claims); *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 330, 332 (4th Cir. 1997) (negligence claims); *Beyond Sys. v. Keynetics, Inc.*, 422 F. Supp. 2d 523, 525, 536 (D. Md. 2006) (claim under Maryland Commercial Electronic Mail Act).

¹¹¹ Matt Laslo, *supra* note 19.

¹¹² *Id.*

¹¹³ *Id.*

¹¹⁴ *Id.*

sex trafficking.¹¹⁵ Even this amendment, however, has led to some backlash, and lawmakers and experts are urging restraint to avoid unintended consequences.¹¹⁶

C. Government Regulation of Online Speech

The judiciary has recently wrestled with both legislative and executive suppression of speech on social media. *Packingham v. North Carolina* is the most notable case in this context.¹¹⁷ Here, the Supreme Court examined a North Carolina statute that barred registered sex offenders from accessing commercial social networking sites.¹¹⁸ Justice Kennedy, writing for the majority, began the opinion by stressing the importance of social media as a new democratic public square.¹¹⁹ He explained that, because this was one of the first cases the Court had taken to address the First Amendment on this new public sphere, it “must exercise extreme caution before suggesting that the First Amendment provides scant protection for access to vast networks in that medium.”¹²⁰

The Court discussed the fundamental promise of the First Amendment—that all persons may access forums where they can speak and listen.¹²¹ The internet is such a forum and, therefore, the Court held that individuals are protected from unlawful government restrictions on access.¹²² Still, the internet and social media are tools that can be exploited for criminal purposes, and the Court noted the legitimate interest served by

¹¹⁵ See Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No. 115-164, 132 Stat. 1254, (codified as amended at 47 U.S.C. § 230(e)). This amendment created Subsection (e)(5) of the act which provides that “[n]othing in this section (other than subsection (c)(2)(A)) shall be construed to impair or limit—

(A) any claim in a civil action brought under section 1595 of Title 18, if the conduct underlying the claim constitutes a violation of section 1591 of that title;

(B) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 1591 of [T]itle 18, United States Code; or

(C) any charge in a criminal prosecution brought under State law if the conduct underlying the charge would constitute a violation of section 2421A of [T]itle 18, United States Code, and promotion or facilitation of prostitution is illegal in the jurisdiction where the defendant's promotion or facilitation of prostitution was targeted.”

¹¹⁶ Laslo, *supra* note 19. At the time of writing, a bill had been introduced in the House of Representatives to amend Section 230 to “provide that an owner or operator of a social media service that hinders the display of user-generated content shall be treated as a publisher or speaker of such content, and for other purposes.” H.R. 492, 116th Cong. (2019). The bill was sent to committee and its path forward is unclear.

¹¹⁷ *Packingham v. North Carolina*, 137 S.Ct. 1730 (2017).

¹¹⁸ *Id.* at 1731.

¹¹⁹ *Id.* at 1735.

¹²⁰ *Id.* at 1736.

¹²¹ *Id.* at 1735.

¹²² *Id.* at 1735–37.

the statute.¹²³ But even after the Court assumed the statute was content neutral, and subject to only intermediate scrutiny,¹²⁴ it found that the statute violated the First Amendment by banning substantially more speech than necessary.¹²⁵

Before the decision in *Packingham*, the Fourth Circuit reached a similar conclusion in *Liverman v. Petersburg*.¹²⁶ The court in *Liverman* was addressing a social networking policy promulgated by a police department which prohibited officers from posting discrediting or unfavorable speech about the department online.¹²⁷ Although the court weighed the policy against the governmental interest, paying attention to “the capacity of social media to amplify expressions of rancor and vitriol,” it ultimately struck down the policy as overbroad, noting its capture of speech that is undoubtedly in the public interest.¹²⁸ After these decisions, statutory and municipal restraints on speech over social media face a steep uphill battle in the courts.

Relatedly, although the Supreme Court has not yet addressed an analogous case, lower courts are split on whether governmental actors can ban users from their social media pages under the First Amendment.¹²⁹ In the most prominent case to date, *Knight First Amendment Institute at Columbia University v. Trump*, the Second Circuit Court of Appeals held that the President's blocking of users from his Twitter account was unconstitutional viewpoint discrimination under the First Amendment.¹³⁰ President Trump kept a public Twitter account, over which he often tweeted about official government business or otherwise interacted with the public.¹³¹ The court held that, because the President is a governmental actor acting in that capacity on a platform open to the public, he may not discriminate based on differing viewpoints.¹³² The court notably limited its holding to public officials' *public accounts* used for “*all manner of official*

¹²³ *Id.* at 1736 (explaining that “it is clear that a legislature ‘may pass valid laws to protect children’ and other victims of sexual assault ‘from abuse.’”).

¹²⁴ In order to survive intermediate scrutiny, a law must be “narrowly tailored to serve a significant governmental interest.” *McCullen v. Coakley*, 134 S.Ct. 2518, 2534 (2014).

¹²⁵ *Packingham*, 137 S.Ct. at 1737–38.

¹²⁶ *Liverman v. City of Petersburg*, 844 F.3d 400 (4th Cir. 2016).

¹²⁷ *Id.* at 404.

¹²⁸ *Id.* at 407–09.

¹²⁹ See, e.g., *Knight First Amendment Inst. at Columbia Univ. v. Trump*, 928 F.3d 226 (2d Cir. 2019) (holding President Trump's blocking of users from his social media account was unconstitutional viewpoint discrimination under First Amendment); but see *Morgan v. Bevin*, 298 F. Supp. 3d 1003 (E.D. Ky. 2018) (finding that a governor's

¹³⁰ *Knight First Amendment Inst.*, 928 F.3d 226.

¹³¹ *Id.* at 230.

¹³² *Id.* at 234–39.

*purposes.*¹³³ Determining whether a public official's social media account is a public forum for the purposes of First Amendment protection is a fact-specific inquiry.¹³⁴ Other federal courts have come to similar conclusions.¹³⁵

Some courts, however, have found no cause of action when a government actor bans users from its social media page. One representative example, among many,¹³⁶ is *Morgan v. Bevin*.¹³⁷ Here, some citizens of Kentucky were blocked on their Governor's social media pages because of their viewpoints.¹³⁸ The court denied their motion for preliminary injunction against the Governor and held that, even though his accounts were used "to communicate his policies and visions, and to seek specific feedback," the accounts were private channels of communication that did not become public fora simply because a public official was using them.¹³⁹ Differing from *Knight*, where the court seemed concerned with the blocked users being able to *listen* to the President's speech, the court in *Bevin* seemed to be focused on the fact that the citizens did not have a right to a government audience.¹⁴⁰ In other words, it is important to frame the inquiry by considering which free speech right is being infringed—the right to be heard or the right to listen.¹⁴¹

Judges have been less hesitant to restrain First Amendment rights on social media as conditions of probation. Not all agree on the limits, however, and this issue has created a rift in the courts.¹⁴² Even in these

¹³³ *Id.* at 230 (emphasis added).

¹³⁴ *Id.* at 236.

¹³⁵ See *Leuthy v. LePage*, 2018 WL 4134628 (D. Me. 2018) (denying a governor's motion to dismiss an action against him, based on his excluding constituents from his Facebook page, and ruling that plaintiffs plausibly stated a claim for violation of First Amendment rights); *Dingwell v. Cossette*, 2018 WL 2926287 (D. Conn. 2018) (finding a viable claim against a police department for free speech violations after the department blocked a resident from its social media pages due to his criticism).

¹³⁶ See *Davison v. Plowman*, 247 F. Supp. 3d 767 (E.D. Va. 2017), *summarily aff'd*, 715 Fed. Appx. 298 (4th Cir. 2018) (holding that a resident was not entitled to First Amendment protection for posting critical comments on the social media page of the commonwealth's attorney); *Robinson v. Hunt County, Tx.*, 2018 WL 1083838 (N.D. Tex. 2018) (dismissing claim against county's sheriff's office that it violated the plaintiff's First Amendment rights by blocking him from the sheriff's office's Facebook page).

¹³⁷ *Morgan*, 298 F. Supp. 3d 1003.

¹³⁸ *Id.*

¹³⁹ *Id.* at 1012.

¹⁴⁰ *Id.* at 1011.

¹⁴¹ *Packingham v. North Carolina*, 137 S. Ct. 1730, 1735 (2017).

¹⁴² See *Manning v. Powers*, 281 F. Supp. 3d 953 (C.D. Cal. 2017) (granting a preliminary injunction against a parole condition that banned a parolee's access to social media as a violation of First Amendment rights); *Yunus v. Robinson*, 2018 WL 3455408 (S.D.N.Y. 2018) (magistrate judge recommending that the court grant an injunction against a condition of parole that banned social media

limited circumstances it is clear that the judiciary is suspicious of infringing on one's ability to speak online.

The state undoubtedly has a significant interest in preventing violence that is exacerbated or incited over social media.¹⁴³ But it has proven difficult to prevent these harms without trampling on First Amendment rights, even in the most limited circumstances.¹⁴⁴ Federal courts are struggling with the same issue, also without settling on a satisfactory answer.¹⁴⁵ Given the increasingly accepted view of social media as a new public square, entitled to robust First Amendment protections, governments may need to formulate alternative means to prevent the violence that these forums may incite.

III. INFORMATION FIDUCIARIES

Social media provides a forum for anyone to participate in the marketplace of ideas. Users' speech has proven difficult to regulate, presenting an arduous challenge: how can policymakers prevent unlawful speech and dangerous speakers without incidentally suppressing anything else? This conversation mainly focuses on the speech itself. However, this fails to consider the ways in which social media alters and incentivizes the content sent and received by its users. To avoid infringing on lawful speech, it may be more effective to reform the mechanisms used to disseminate communications that poison the public discourse in the first place.

A. Big Data Companies as Fiduciaries

The theory of social media companies as "information fiduciaries" was first put forth by Professor Jack Balkin.¹⁴⁶ A fiduciary duty is a special duty owed by one party to another based on a relationship of trust.¹⁴⁷ Many fiduciary duties have a long history in the common law, such as "trustee to beneficiary, agent to principal, lawyers to clients, doctors to patients, [and] personal representatives to the estates they represent."¹⁴⁸ Other, less

access as violative of the First Amendment); *but see* United States v. Carson, 924 F.3d 467 (8th Cir. 2019) (declining to extend *Packingham* to supervised release); United States v. Perrin, 926 F.3d 1044 (8th Cir. 2019) (holding that a special condition of supervised release prohibiting defendant from using computer or accessing online service did not violate First Amendment).

¹⁴³ *Packingham*, 137 S. Ct. at 1736

¹⁴⁴ *Id.*

¹⁴⁵ *See supra*, note 142.

¹⁴⁶ Balkin, *supra* note 22; Jack M. Balkin, *Information Fiduciaries in the Digital Age*, BALKINIZATION (Mar. 5, 2014, 4:50 PM), <http://balkin.blogspot.com/2014/03/information-fiduciaries-in-digital-age.html>

¹⁴⁷ DAN B. DOBBS ET AL., LAW OF TORTS § 267 (2d ed. 2020).

¹⁴⁸ *Id.*

formal, confidential relationships may also give rise to fiduciary duties.¹⁴⁹ These relationships are based on undertakings of loyalty by one party which generate a beneficiary's confidence in them and may oblige the fiduciary to act affirmatively to protect its beneficiary.¹⁵⁰ A beneficiary has a cause of action against a fiduciary if: (1) a fiduciary relationship existed; (2) the fiduciary breached that duty; and (3) that breach was the proximate cause of the beneficiary's injury.¹⁵¹

These relationships, and their corresponding duties, can—and should—apply to social media companies' relationships with their end-users.¹⁵² Concepts of fiduciary duties have roots in early common law, but evolve over time to recognize new relationships.¹⁵³ Social trends may be augmenting the prevalence and importance of fiduciary relations in modern society.¹⁵⁴ As Professor Balkin argues, social media companies' relationship to their end-users constitutes such a legally-recognized relationship of trust for four main reasons: (1) Because big data companies have superior knowledge and expertise, and users do not, there is a vast difference in the ability of each party to monitor the other, forcing users to trust the companies to use their personal information in their best interest; (2) users depend on these companies, as they provide services that are now close to necessary in modern life; (3) these companies hold themselves out as experts, and in doing so, induce the exchange of users' personal information for their services; and (4) both parties know that the companies hold confidential and valuable information that may be used to an end-user's disadvantage.¹⁵⁵

Social media companies have superior knowledge and expertise, which, coupled with their extensive ability to monitor, forces users to put a lot of trust in them. A fiduciary relationship is formed when one party gains another's confidence and incidentally obtains a position of superiority and influence over that party.¹⁵⁶ Such relationships may arise

¹⁴⁹ *Id.*

¹⁵⁰ *Id.*

¹⁵¹ *See, e.g.,* French Broad Place, LLC v. Asheville Sav. Bank, S.S.B., 816 S.E.2d 886 (N.C. Ct. App. 2018). Herein lies one of the most important benefits of recognizing social media companies' fiduciary duties. As previously discussed, authorities worry about bestowing speech controlling power to either governments or the companies themselves. Fiduciary duties put the power in the hands of the users because enforcement takes the form of individual causes of action. Social media companies will be incentivized to avoid this liability just as they are incentivized to avoid other forms of potential legal liability.

¹⁵² *See* Balkin, *supra* note 22.

¹⁵³ *Id.*; *see* Tamar Frankel, *Fiduciary Law*, 71 CAL. L. REV. 795, 795 (1983).

¹⁵⁴ Frankel, *supra* note 153, at 979.

¹⁵⁵ Balkin, *supra* note 148, at 1222.

¹⁵⁶ *See* 19 ILL. PRAC., ESTATE PLAN. & ADMIN. § 200:9 (4th ed.); *Herbolsheimer v. Herbolsheimer*, 60 Ill.2d 574 (1975).

out of a diverse set of dealings, including “any type of legal, moral, social, domestic or personal situation.”¹⁵⁷ The average social media user does not know how the sites are using their information.¹⁵⁸ These companies can track each move that their users make, while the users only get a glimpse of the fraction of the interface tailored to them.¹⁵⁹ This power imbalance causes users to necessarily bestow the most fundamental element of a fiduciary relationship in these companies—trust. As social media becomes more ubiquitous, it also becomes more necessary to modern life. As early as 2007, 94% of first-year college students spent some time on social media each week.¹⁶⁰ Social media has also crept into the realm of employment. One recent study found that a comprehensive LinkedIn profile gave job seekers a 71% better chance at landing an interview.¹⁶¹ In some ways, users depend on social media companies to navigate the modern world.

Dependency in transactions and relations may give rise to a fiduciary relationship, even if the parties are otherwise on even terms.¹⁶² In this case, however, the consideration of dependency in the relationship is only multiplied by the previously discussed disparate power between the parties.

Similarly, when one party in a relationship holds itself out as an expert, it may create a fiduciary relationship. As society becomes more complicated, it may be necessary to have more expert-fiduciaries to help individuals navigate the world.¹⁶³ Lawyers are one of the oldest examples of expert-fiduciaries recognized by law.¹⁶⁴ The Restatement of the Law Governing Lawyers explains that their expectation of diligence,

¹⁵⁷ See 19 ILL. PRAC., ESTATE PLAN. & ADMIN. § 200:9 (4th ed.) (citing *Pepe v. Caputo*, 408 Ill. 321, 326 (1951)).

¹⁵⁸ Aaron Smith, *Many Facebook users don't understand how the site's news feed works*, Pew Research Center (Sept. 5, 2018), <https://www.pewresearch.org/fact-tank/2018/09/05/many-facebook-users-dont-understand-how-the-sites-news-feed-works/>

¹⁵⁹ See generally Rose-Stockwell, *supra* note 10.

¹⁶⁰ Paige Abe & Nickolas A. Jordan, *Integrating Social Media Into the Classroom Curriculum*, 18 ABOUT CAMPUS: ENRICHING THE STUDENT LEARNING EXPERIENCE 16 (Mar. 1, 2013), <https://journals.sagepub.com/doi/abs/10.1002/abc.21107?journalCode=acaa>

The economic phenomenon known as “network effects” explains why other sites cannot simply pop up and compete with current incumbent social media giants. A network effect occurs when the value of a good or service increases exponentially as more people join in its consumption. These are particularly strong in the communications industry. See Caroline Banton, *Network Effect*, INVESTOPEDIA (Oct. 15, 2019), www.investopedia.com/terms/n/network-effect.asp.

¹⁶¹ Peter Yang, *Resume Study: How LinkedIn Affects the Interview Chances of Job Applicants*, RESUMEGO (2019), <https://www.resumego.net/research/linkedin-interview-chances/>.

¹⁶² See *In re Daisy Sys. Corp.*, 97 F.3d 1171 (9th Cir. 1996) (finding that an investment banker retained by a corporation may owe a fiduciary duty even though both parties are sophisticated in business dealings because of the corporation’s dependence on the banker).

¹⁶³ See Edward D. Spurgeon & Mary Jane Ciccarello, *The Lawyer in Other Fiduciary Roles: Policy and Ethical Considerations*, 62 FORDHAM L. REV. 1357 (1994).

¹⁶⁴ *Id.*

competence, and loyalty, as well as the nature of their complex and technical work, which often takes place in the client's absence, contributes to the formation of this fiduciary relationship.¹⁶⁵ Social media companies' work—gathering and utilizing their users' data—requires expansive teams of highly-trained employees to craft sophisticated code and technology outside of the users' grasp.¹⁶⁶ They also use this expertise to bring users to their sites in the first place by curating pages and creating a desirable experience. The formation of this connection, based on one party's expertise, mimics the formation of many associations recognized by common law as fiduciary relationships.

Most importantly, social media companies hold immense amounts of confidential information about their users.¹⁶⁷ Facebook, for example, collects data based on a user's interactions with pages and posts on the site, which it then analyzes in order to curate the user's experience and target advertising to them based on, for example, their political views.¹⁶⁸ Doctors, like lawyers, gain fiduciary duties to their patients based on their access to confidential information.¹⁶⁹ In both cases, as is the case in many similar relationships, information entrusted to the fiduciary obtains special protection.¹⁷⁰ Social media companies gain not only personal information that is provided by the users, but also information about the users based on complex analysis of their behavior patterns.¹⁷¹ In fact, they have been shown to divulge this information for arguably sinister uses.¹⁷² These platforms' unique connection with end-users, and their possession of such private information, resembles other, currently legally-recognized relationships of confidentiality.

Social media companies and end-users do not have the same relationship as a doctor to a patient or an attorney to a client. Each fiduciary relationship has different characteristics and duties. They have in the aggregate, however, been expanding over time, exponentially so in the

¹⁶⁵ RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 16 (2000).

¹⁶⁶ Facebook's newsfeed algorithm, for instance, is "closely guarded [,] constantly shifting, . . . and stubbornly opaque." Will Oremus, *Who Controls Your Facebook Feed*, Slate (Jan. 2016), http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html

¹⁶⁷ *See id.* (noting that another important rationale for the fiduciary relationship between lawyer and client is that the lawyer deals with a client's confidential and vital information).

¹⁶⁸ Jeremy B. Merrill, *Liberal, Moderate or Conservative? See How Facebook Labels You*, N.Y. TIMES (Aug. 23, 2016), <https://www.nytimes.com/2016/08/24/us/politics/facebook-ads-politics.html>

¹⁶⁹ Frankel, *supra* note 155, at 796.

¹⁷⁰ *Hammonds v. Aetna Cas. & Sur. Co.*, 237 F. Supp. 96, 102 (N.D. Ohio 1965) (citing *Smith v. Driscoll*, 94 Wash. 441, 443 (1917)).

¹⁷¹ *See Merrill, supra* note 168.

¹⁷² *See Cadwalladr, et al., supra* note 55.

modern era.¹⁷³ Fiduciary duties vary in nature and scope, but are based on the original relationship of trust.¹⁷⁴ Broadly, social media companies' fiduciary duties, stemming from the trust relationship defined above, should include the duties of non-discrimination, non-disclosure, and non-manipulation.¹⁷⁵ While each of these duties deserves further treatment, the duty of non-manipulation is the most relevant to the danger of algorithmic coercion of users toward misinformation, inflammatory and conspiratorial content.

B. The Duty of Non-Manipulation

A duty of non-manipulation, though sometimes going by a different name, exists in many fiduciary relationships. It stems from the fact that a fiduciary must put its beneficiary's interest above its own and not abuse its unique position of power over the beneficiary to its advantage.¹⁷⁶ This concept underlies duties imposed on a broad range of legally recognized relationships of trust, which are breached when a fiduciary takes advantage of a beneficiary physically, emotionally, or financially.

Breach of a fiduciary duty occurs when a party in a position of trust physically or emotionally harms another party that is vulnerable to their abuse of that position.¹⁷⁷ In *Destefano v. Grabrian*, the Supreme Court of Colorado decided a case involving a priest that induced his parishioner, whom he was also serving as a marriage counselor, into a sexual relationship.¹⁷⁸ The court held that the plaintiff stated a viable claim because, if true, the priest used his position of trust to take advantage of his beneficiary's vulnerability which was known to him based on that special relationship.¹⁷⁹ The physician-patient relationship was similarly examined by the Supreme Court of Nevada in *Hoppes v. Hammargren*.¹⁸⁰ The court noted that physicians often hold a position of superior knowledge, skill, and information, putting them in a unique position to take advantage of a patient's vulnerabilities.¹⁸¹ The essence of these relationships, the court explained, was that there exists a condition of unequal power giving one

¹⁷³ Frankel, *supra* note 155, at 79.

¹⁷⁴ RESTATEMENT (THIRD) OF AGENCY § 8.01(c) (AM. LAW INST. 2006).

¹⁷⁵ Balkin, *supra* note 22, at 1233.

¹⁷⁶ DAN B. DOBBS, PAUL T. HAYDEN & ELLEN M. BUBLICK, *THE LAW OF TORTS* § 267 (2d ed. 2016); *Barbara A. v. John G.*, 193 Cal. Rptr. 422 (Cal. Ct. App. 1983).

¹⁷⁷ *See, e.g., Destefano v. Grabrian*, 763 P.2d 275 (Colo. 1988); *McDaniel v. Gile*, 281 Cal. Rptr. 242 (Cal. Ct. App. 1991); *Barbara A.*, 193 Cal. Rptr. 422; *Hoopes v. Hammargren*, 102 Nev. 425 (Nev. 1986); *Erickson v. Christenson*, 99 Or. App. 104 (Cal. Ct. App. 1989).

¹⁷⁸ 763 P.2d 275.

¹⁷⁹ *Id.* at 284.

¹⁸⁰ 102 Nev. 425.

¹⁸¹ *Id.* at 431–32.

party unique influence over the other.¹⁸² Taking advantage of this unequal power, as in this case by inducing a sexual relationship, constitutes an abuse of power and violates the duties that this special relationship imposes.¹⁸³

There are numerous examples of this kind of abuse of power throughout the common law. Courts have also found similar duties to not take advantage of unequal power in attorney-client,¹⁸⁴ pastor-congregation,¹⁸⁵ and psychotherapist-patient¹⁸⁶ relationships. All of these cases involve a fiduciary using a confidential relationship, and the information obtained from it, to further its own interests at the expense of the beneficiary's well-being.

Breach of the duty of non-manipulation or non-coercion also occurs when a fiduciary takes advantage of a beneficiary financially.¹⁸⁷ For instance, courts are sufficiently concerned about an attorney's unequal position of power over a client that many find a presumption of undue influence when an attorney transacts with their client.¹⁸⁸ The Illinois Supreme Court addressed this issue in *Klaskin v. Klepak*.¹⁸⁹ The court explained that there was a strong presumption of undue influence anytime an attorney transacts with a client and is benefited thereby.¹⁹⁰ The attorney in this case was left a condominium unit as part of one of his former client's estates.¹⁹¹ Although the client in this case was a savvy entrepreneur and "mentally alert at the time he executed the trust," the attorney did not provide evidence: "(1) that he or she made a full and frank disclosure of all relevant information; (2) that adequate consideration was given; and (3) that the client had independent advice before completing the transaction" sufficient to overcome the significant undue influence presumption.¹⁹² Due to their position of trust and unequal possession of knowledge and

¹⁸² *Id.*

¹⁸³ *Id.*

¹⁸⁴ *See, e.g.,* Barbara A. v. John G., 193 Cal. Rptr. 422, 432 (Cal. Ct. App. 1983).

¹⁸⁵ *See, e.g.,* Erickson v. Christenson, 99 Or. App. 104 (Or. Ct. App. 1989) (plaintiff stated a claim when pastor "mentally manipulated" her to have sexual relations with him using his position of trust).

¹⁸⁶ *See, e.g.,* Malone v. Sewell, 168 S.W.3d 243 (Tex. App. 2005) (assuming that therapist "hugging and caressing [patient], giving [her] a book that advocated sexual relationships between therapists and patients, and asking [her] on a date" violated the therapist's fiduciary duty).

¹⁸⁷ *See, e.g.,* Klaskin v. Klepak, 126 Ill. 2d 376 (Ill. 1989); Fair v. Bakhtiari, 125 Cal. Rptr. 3d 765 (Cal Ct. App. 2011).

¹⁸⁸ *Klaskin*, 126 Ill. 2d at 379–80, 386.

¹⁸⁹ 126 Ill. 2d 376.

¹⁹⁰ *Id.* at 386–87.

¹⁹¹ *Id.* at 379.

¹⁹² *Id.* at 387, 390.

information, attorney-fiduciaries¹⁹³ are closely scrutinized when dealing with their clients.¹⁹⁴

The case law is clear: parties that are in a position of trust—holding superior knowledge and confidential information—may not abuse the power entrusted in them. Although courts discuss undue influence, non-coercion, and the like, this duty can reasonably be called a duty of non-manipulation. Fiduciary duties alter and develop overtime in response to the changing social landscape.¹⁹⁵ Based on the common law developments discussed in this section, a general duty of non-manipulation for fiduciaries possessing superior knowledge and confidential information may be formulated as follows: *A fiduciary may not, by exercising its influence over a beneficiary, use that beneficiary’s confidential information, obtained as a result of the relationship, for its own gain and to the beneficiary’s detriment.*

C. Application to Social Media Companies

Social media companies are in a relationship with their end-users that entails the duty of non-manipulation. These companies possess vast collections of data and know an extensive array of confidential information about their users.¹⁹⁶ The users, on the other hand, know little to nothing about how these companies collect, store, and use their data, let alone how their algorithms control content with which their users engage.¹⁹⁷ As a result of this gulf between the parties’ knowledge, skill, and information, the end-users are “uniquely vulnerable” in this relationship.¹⁹⁸ When companies use this disparity in power to their advantage, and to the detriment of the end-user, they are violating a duty created by their position of trust.¹⁹⁹

¹⁹³ Courts have also recognized similar fiduciary duties for accountants based on confidential information. *See Miller v. Harris*, 985 N.E.2d 671 (Ill. App. Ct. 2013) (recognizing cause of action in Illinois against accountant for breach of fiduciary duty due to relationship including confidential information).

¹⁹⁴ *Klaskin*, 126 Ill. 2d at 386, 390.

¹⁹⁵ *See Frankel*, *supra* note 155, at 796.

¹⁹⁶ *See, e.g., Copeland*, *supra* note 41 (noting Google’s collection of massive amounts of health data); *THE ECONOMIST*, *supra* note 38 (describing data as surpassing oil as the world’s most valuable resource).

¹⁹⁷ *See, e.g., Josh Constine, How Facebook News Feed Works*, TECH CRUNCH (Sept. 6, 2016, 3:07 PM), <https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/> (explaining roughly how Facebook’s newsfeed algorithm operates).

¹⁹⁸ The reason court’s presume undue influence in some fiduciary relationships is because of a disparity in information and knowledge that puts the beneficiary in a vulnerable position. *See, e.g., Klaskin v. Klepak*, 126 Ill. 2d 376 (Ill. 1989); *Fair v. Bakhtiari*, 125 Cal. Rptr. 3d 765 (Cal. Ct. App. 2011).

¹⁹⁹ *Destefano v. Grabrian*, 763 P.2d 275, 284 (Colo. 1988).

A counter-argument to this theory is that, if social media companies are viewed as some kind of fiduciary, then many large companies²⁰⁰ may begin to look like fiduciaries as well. But social media companies' relationships with their users are entirely different than those of most modern corporations. Of course, other companies may hold their consumers' contact and credit card information, know their general preferences, or perhaps even know their social security numbers. Social media companies, on the other hand, know their users on an entirely different level. They collect a much broader array of personal and behavioral data to create a holistic view of their users' identities.

Facebook²⁰¹ now gets over 90% of its advertising revenue from its mobile applications.²⁰² This is at a time when the average person spends about four hours per day on their phone, about half of that on social media.²⁰³ Even this does not account for the additional time that people spend on social media on computers or other devices. Facebook not only has all of its users' contact and personal information shared on or held by the site, but continually monitors users' every click and view to build individual psychological profiles.²⁰⁴ These profiles can be used to "nudge" users toward content that is promoted by Facebook's algorithms (or by the party that it sold this data to).²⁰⁵ Facebook's product is its users, whose data it can mine to provide eyes for highly targeted advertising or for the highest bidding third-party data purchaser.²⁰⁶ But it is also the new public square, where people go to chat with local elected officials, communicate their ideas to their peers and neighbors, and gather their news. In fact, a recent Pew Research study revealed that two-thirds of Americans get at

²⁰⁰ Some companies, like *23andMe*, which collects and analyzes genetic data, may also require some similar duties. Holding individuals' genetic information requires the utmost sensitivity. Many similar emerging companies in the world of big data deserve closer scrutiny through further research. Those companies are outside of the scope of this paper.

²⁰¹ Here, Facebook is used as an example, but Twitter, Instagram, or YouTube could have easily replaced it.

²⁰² Emil Protalinski, *Over 90% of Facebook's Advertising Revenue Now Comes from Mobile*, VENTUREBEAT (Apr. 25, 2018), <https://venturebeat.com/2018/04/25/over-90-of-facebooks-advertising-revenue-now-comes-from-mobile/>.

²⁰³ Melanie Curtin, *Are You on Your Phone Too Much? The Average Person Spends This Many Hours on it Every Day*, INC. (Oct. 30, 2018), <https://www.inc.com/melanie-curtin/are-you-on-your-phone-too-much-average-person-spends-this-many-hours-on-it-every-day.html>

²⁰⁴ Coppins, *supra* note 6.

²⁰⁵ *Id.*

²⁰⁶ See Matthew Rosenberg, Nicholas Confessore & Carole Cadwalladr, *How Trump Consultants Exploited the Facebook Data of Millions*, N.Y. TIMES (Mar. 17, 2018), <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>.

least some of their news on social media.²⁰⁷

Social media is becoming an institution in American life, taking over part of the role of the press, the telecommunication industry, the town hall, and more. In doing so, it has captured vast amounts of data from its users which it uses to create almost all of its revenue.²⁰⁸ It therefore has a very different relationship with its end-users than traditional corporations that simply sell consumers their products or services.

Algorithms employed by these companies that have the purpose of increasing user engagement with no regard for propagating fear, misinformation, and inflammatory content are an example of abuse based on a position of trust. Indeed, even if it is not their purpose, these algorithms tend to spread lies faster than truth, and discourage reasoned, civil conversation while encouraging echo-chambers and polarization.²⁰⁹ The purpose of these algorithms is to keep users engaged for longer periods of time, thus increasing profit for the companies.²¹⁰ The unintended consequence is that user behavior can be affected, often in negative ways, by the prioritized false and extreme content.²¹¹ Social media companies, because of their position of trust, are benefiting at the expense of their beneficiaries' well-being.

Social media companies' relationship with their end-users then requires them to correct or alter their algorithms. To the extent possible, they must aim to prevent these harms, even if such action would reduce the benefit they receive from the relationship. Just as lawyers must go to great lengths to prove to courts that they did not exercise undue influence over their clients based on a fiduciary relationship,²¹² social media companies should be required to make their algorithms more transparent. Like psychotherapists and clergy, who cannot manipulate their clients' position

²⁰⁷ Elisa Shearer & Jeffrey Gottfried, *News Use Across Social Media Platforms 2017*, PEW RESEARCH Ctr. (Sept. 7, 2017), <https://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/>.

²⁰⁸ Matthew Johnston, *How Facebook Makes Money*, INVESTOPEDIA (June 25, 2019), <https://www.investopedia.com/ask/answers/120114/how-does-facebook-fb-make-money.asp>

²⁰⁹ See Marcia Stepanek, *The Algorithms of Fear*, STANFORD SOC. INNOVATION REV. (June 14, 2016), https://ssir.org/articles/entry/the_algorithms_of_fear; Soroush Vosoughi, et al., *The Spread of True and False News Online*, 359 SCIENCE 1146 (Mar. 9, 2018), available at <https://science.sciencemag.org/content/359/6380/1146>

²¹⁰ Rose-Stockwell, *supra* note 10.

²¹¹ See Fisher & Taub, *supra* note 4 (explaining the radicalizing effect of YouTube video algorithms on Brazil, in part responsible for death threats against public health officials).

²¹² Courts may even require affirmative disclosures from lawyers and alternate professional opinions for clients before the parties engage in transactions that benefit the attorney. *Klaskin v. Klepak*, 126 Ill. 2d 376, 387 (1989).

of vulnerability for their personal gain,²¹³ social media companies should not be able to use their ever-increasing knowledge of users' psychological profiles to keep them engaged for longer periods, especially by promoting inflammatory and false content.²¹⁴

Algorithms now prioritize engagement, but recognizing a duty of non-manipulation would require social media companies to be more conscious of the effects of the algorithms on their users' behavior. They would have to be more transparent and better control for the algorithmic manipulation that can occur when engagement is prioritized over credibility. They would have to correct for polarizing and inciting content being given the most value. Of course, this is not an antidote to what is in part a product of human nature, but this duty would require algorithms to tame humanity's worst impulses, or at least to not exploit them.

Facebook (and other social media companies) have the ability to change their algorithms.²¹⁵ As far back as 2012, Facebook was able to demote content made by two independent companies that was annoying its users.²¹⁶ Their visibility fell greatly, and shortly after, both companies shut down.²¹⁷ In 2013, two viral news companies were flooding Facebook users' timelines.²¹⁸ After Facebook tweaked its newsfeed algorithm, traffic to these sites fell dramatically.²¹⁹ When Facebook has been faced with agents spreading content that hurts its bottom-line, it has successfully thwarted their efforts. Facebook as a fiduciary would have to control its algorithms not only to maximize profit, but to protect users from dangerous behavior modification.

It could accomplish this in a few ways. One example, put forward by Alexis Madrigal of *The Atlantic*, is to make algorithms "antiviral."²²⁰ Social media companies could slow the spread of content calculated to go viral. Such content is currently highly valued by the algorithms, as it is of the kind that keeps users engaged and sharing.²²¹ Antivirality would devalue posts that may spread uncontrollably in the same way that Facebook demoted annoying independent companies' apps and the

²¹³ See, e.g., *Destefano v. Grabrian*, 763 P.2d 275 (Colo. 1988); *McDaniel v. Gile*, 281 Cal. Rptr. 242 (Cal. Ct. App. 1991); *Barbara A. v. John G.*, 193 Cal. Rptr. 422 (Cal. Ct. App. 1983); *Hoopas v. Hammargren*, 102 Nev. 425 (Nev. 1986); *Erickson v. Christenson*, 99 Or. App. 104 (Or. Ct. App. 1989).

²¹⁴ See, e.g., *Rose-Stockwell*, *supra* note 10; see, *Matthew Rosenberg et al.*, *supra* note 206.

²¹⁵ Erin Griffith, *Facebook Can Absolutely Control Its Algorithm*, WIRE (Sept. 26, 2017), <https://www.wired.com/story/facebook-can-absolutely-control-its-algorithm/>.

²¹⁶ The companies were SocialCam and Viddy. *Id.*

²¹⁷ *Id.*

²¹⁸ *Id.*

²¹⁹ Facebook also fairly successfully prevents posts with nudity and complies with alcohol advertisement regulation. *Id.*

²²⁰ Thompson, *supra* note 7.

²²¹ *Id.*

visibility of viral news sites.²²² Another example, currently offered by a bill in the House of Representatives, is to remove Section 230 immunization from companies whose algorithms present content in users' feeds based on any ordering other than chronological.²²³ A final example²²⁴, presented by internet entrepreneur Ryan Holmes, is to value contrary views highly for individuals,²²⁵ or to devalue patently fake or salacious material.²²⁶

These companies can—and do—alter their algorithms to control the dissemination of certain posts. There are many options that would prevent the spread of extremist and sensationalist content, or combat it with opposing viewpoints. If these platforms were required to value algorithms based on users' best interests, rather than profit alone, many different methods could be tried and evaluated. Each of the previously discussed dangers is better served, albeit imperfectly, by enforcing fiduciary duties on large data-companies to alter algorithms in favor of different, non-manipulative values, rather than limiting their users' free speech.²²⁷

IV. CONCLUSION

Certainly, Voltaire was not considering social media when he observed that “those who can make you believe absurdities can make you commit atrocities.”²²⁸ But these words seem to take on new, perhaps more urgent meaning in the modern world. Social media is becoming an unavoidable necessity of life. The algorithms utilized by these large companies currently value user engagement above all else, incentivizing dangerous, false, and inflammatory content. This exposure harms users at an individual level and can even damage long-cherished institutions at a larger scale.

Despite these harms, courts are hesitant to suppress speech on these platforms for good reasons. Social media has a documented history of

²²² *Id.*

²²³ H.R. 492, 116th Cong. (2019).

²²⁴ This list of examples is not exhaustive of the different options available.

²²⁵ This practice would force users to confront their own biases much more often.

²²⁶ Ryan Holmes, *supra* note 70.

²²⁷ The Supreme Court has previously entertained (though it did not ultimately adopt) a theory of newspapers as fiduciaries to the public due to their unique democratic responsibility. *See* *Miami Herald Pub. Co. v. Tornillo*, 418 U.S. 241, 251 (1974); *see also* *Red Lion Broad. Co., Inc. v. F.C.C.*, 395 U.S. 367, 389 (1969). As this paper lays out at great length, social media companies have a different relationship with democracy and individual users that much more closely represents a traditional fiduciary relationship. The fiduciary rationale that seemed to the *Tornillo* court worthy of further discussion is more present and important in social media's relationship to the public and its end-users. 418 U.S. at 251.

²²⁸ VOLTAIRE, QUESTIONS ON MIRACLES TO M. CLAPAREDE, PROFESSOR OF THEOLOGY IN GENEVA, BY A PROPONENT: OR EXTRACT FROM VARIOUS LETTERS FROM M. DE VOLTAIRE (1764).

democratizing information, fostering political participation, and connecting people and groups otherwise foreign to each other. It also appears that these companies are unlikely to want, or be required, to police false and inciting content on their sites. Imposing fiduciary duties may be more effective than regulating online speech. Social media companies operate from a position of trust and superior information to control their vulnerable users' engagement. As such, their relationships with their users resemble many legally regulated associations of trust.

Social media companies are undoubtedly different in many ways from paradigmatic fiduciaries. Fiduciary duties, however, evolve over time in common law to meet the changing social landscape. A derivative duty of non-manipulation would require social media companies to modify the mechanisms that push users to accept and act on dangerous ideas in the first place. Such a reform is not a cure-all for what ails society and social media, but it provides an alternative to choosing between suppressing the free flow of ideas and inflaming our most abhorrent tendencies.