

Internet Immunity, Fake News, and Asymmetrical Action: A Survey of the World's Four Most Popular Internet Platforms and Disparate Regulatory Responses

Thomas Davidson
J.D. Candidate, 2019
Directed Research
Professor Nathaniel Persily
Fall 2017-18
Stanford Law School

Table of Contents

I. INTRODUCTION – FRAMING THE FAKE NEWS PHENOMENON 2

II. THE AMERICAN LEGAL FRAMEWORK FOR INTERMEDIARY LIABILITY AND GENESIS OF INTERNET PLATFORM SELF-REGULATION..... 3

A. THE 1996 COMMUNICATIONS DECENCY ACT (CDA) § 230 – THE CODIFICATION OF INTERNET PLATFORM IMMUNITY. 4

B. THE FIRST AMENDMENT AND FREE EXPRESSION ON THE INTERNET. 7

III. BALANCING THE BOTTOM LINE AND FREE SPEECH – THE SPECTRUM OF CONTENT MODERATION FROM REDDIT TO FACEBOOK 12

A. REDDIT – THE ORIGINS OF LAISSEZ-FAIRE CONTENT MODERATION. 12

B. WEAK GOVERNANCE AND A SUSCEPTIBLE PLATFORM – WHY FAKE NEWS ABOUNDS ON REDDIT. 15

C. THE PLATFORM PARADOX – THE LIMITATIONS OF REDDIT’S FREE SPEECH POSTURE..... 17

D. A LITTLE BIRDIE TOLD ME – TWITTER’S EARLY DAYS AND FREE SPEECH FUNDAMENTALISM. 21

E. RUSSIAN BOTS, MISINFORMATION, AND ACCOUNT ANONYMITY – TWITTER’S TRUMP-SIZE PROBLEM AND RUNTY RESPONSE. 24

F. GOOGLE – ORIGINS OF CENSORSHIP ON THE WORLD’S SEARCH GIANT. 27

G. AUTOCOMPLETE, ADS, & ALGORITHMS – HOW FAKE NEWS PERVADES GOOGLE & YOUTUBE..... 31

H. GOOGLE STRIKES BACK – THE SEARCH ENGINE’S WIDE-RANGING EFFORTS TO CURTAIL DECEPTIVE CONTENT AND ITS SHORTCOMINGS. 34

I. BE MY FRIEND? FACEBOOK’S RISE AND THE DAWN OF CENSORSHIP ON THE ‘SOCIAL NETWORK.’ 38

J. FAKE NEWS AND FILTER BUBBLES – WHY FRAUDULENT CONTENT CIRCULATES AND STICKS ON FACEBOOK. 41

K. FALLING SHORT – FACEBOOK’S FLAILING EFFORTS TO STOP FAKE NEWS. 45

IV. THE FRACTURED REGULATORY RESPONSE TO FAKE NEWS IN THE U.S. AND EUROPE..... 49

V. CONCLUSION – THE USER AS THE ARBITER OF TRUTH..... 55

I. Introduction – Framing the Fake News Phenomenon

Eleven minutes. No Twitter. No 140-character commentary. A contractor at Twitter deactivated President Trump’s Twitter account on November 2, 2017 for a matter of minutes and pandemonium broke loose on the Internet.¹ The event capped off a week of bad press about Twitter’s response to the Russian influence campaign during the 2016 U.S. election and raised questions of Twitter’s security protocol for V.I.P. accounts.² Headline’s read: “Twitter’s Panic After Trump’s Account Is Deleted Caps a Rough Week,”³ “Trump’s Twitter Takedown Reveals Another Tech Blind Spot,”⁴ and “Trump’s Temporary Disappearance from Twitter Was Very Scary. Here’s Why.”⁵

Most reports of the takedown focused on lax security protocol for employees and contractors to access high-profile accounts and Twitter’s lack of transparency. Yet almost no one questioned Twitter’s *authority* to delete or suspend Mr. Trump’s account. There are countless reasons why this would never happen, yet an errant contractor’s actions highlight a remarkable reality. Internet companies, like Twitter and Facebook, have incredible autonomy to determine what content is shared on their platforms and who can participate. With billions of people online, the decisions of these hyper-influential platforms reverberate throughout the world and civil society.

¹ Mike Isaac & Daisuke Wakabayashi, *Twitter’s Panic After Trump’s Account Is Deleted Caps a Rough Week*, N.Y. TIMES (Nov. 3, 2017), https://www.nytimes.com/2017/11/03/technology/trump-twitter-deleted.html?rref=collection%2Fsectioncollection%2Ftechnology&action=click&contentCollection=technology®ion=rank&module=package&version=highlights&contentPlacement=1&pgtype=sectionfront&_r=0 (last visited Nov. 4, 2017).

² Isaac and Wakabayashi write, “Mr. Trump’s deleted account showed how empty some of those promises [increased security protocols] seemed.” *Id.*

³ *Id.*

⁴ Issie Lapowsky, *Trump’s Twitter Takedown Reveals Another Tech Blind Spot*, WIRED (Nov. 3, 2017), <https://www.wired.com/story/trumps-twitter-takedown-reveals-another-tech-blind-spot/> (last visited Nov. 4, 2017).

⁵ Molly Roberts, *Trump’s Temporary Disappearance from Twitter Was Very Scary. Here’s Why*, WASH. POST (Nov. 3, 2017), https://www.washingtonpost.com/blogs/post-partisan/wp/2017/11/03/why-trumps-temporary-disappearance-from-twitter-was-so-scary/?utm_term=.14113d253def (last visited Nov. 4, 2017).

This paper examines Internet platform content moderation in the context of the fake news phenomenon. Fake news refers to intentionally deceptive content published for profit or political motive under the guise of legitimate, fact-based information.⁶ Four of the world’s most popular Internet platforms are surveyed – Reddit, Twitter, Google, and Facebook – after a review of America’s intermediary liability framework. Each platform’s institutional mechanisms to regulate content and response to the growth of fake news are examined in kind. Finally, the inadequacy and efficacy of each platform’s approach to deceptive content moderation are considered in light of new regulatory challenges Internet platforms face in the United States and abroad.

II. The American Legal Framework for Intermediary Liability and Genesis of Internet Platform Self-Regulation

The current state of Internet platform content moderation is best encapsulated by a recent exchange between Senator Amy Klobuchar (D-MN) and three legal representatives from Facebook, Twitter, and Google before the Senate Judiciary Subcommittee on October 31, 2017. Senator Klobuchar, referring to the platforms’ pledges to increase self-regulation efforts and transparency, questioned: “Just to clarify, while you are taking responsibility for a lot of what’s happened here and trying to make some changes, there wouldn’t be an outside enforcer of any of your policies, right? It would just be you. Is that true?” Her question was met with eight seconds of silence and nervous glances among the panel members. Senator Klobuchar prodded,

⁶ PEN America, *Faking News: Fraudulent News and the Fight for Truth*, PEN AMERICA 4 (Oct. 12, 2017), https://pen.org/wp-content/uploads/2017/10/PEN-America_Faking-News-Report_10.17.pdf. Fake news is also referred to as false content, fraudulent news, deceptive content, and disinformation throughout this paper.

“[c]an someone answer?” Finally, Sean Edgett, Acting General Counsel of Twitter, stated, “[t]hat’s correct.”⁷

Unburdened by regulation, both free expression and intentional deception have prospered on the Internet. This paper examines how non-state actors from Reddit to Facebook have developed content moderation policies in the absence of an Internet regulatory state. The resulting fake news free-for-all has left many regulators questioning: What needs to change?

a. The 1996 Communications Decency Act (CDA) § 230 – the codification of Internet platform immunity.

Today, it is widely understood that Internet service providers (ISPs) and Internet platforms, like Facebook and Twitter, are immune from liability as publishers or distributors of most content posted by third parties on their platforms. However, prior to the 1996 Communications Decency Act, ISPs (which include Internet platforms like online messaging boards) did not enjoy such broad protections. ISPs were treated by courts as publishers, distributors, or common carriers within a traditional First Amendment framework.⁸ Courts reached conflicting outcomes when left to their own devices. In *Cubby v. CompuServe*, 776 F. Supp. 135, 141 (S.D.N.Y. 1991), the district court declined to hold CompuServe, an ISP, liable for defamatory content hosted on its database. The court’s decision hinged upon the fact that CompuServe did not know and did not have reason to know of the defamatory content. In effect,

⁷ Kurt Wagner & Tony Romm, *Live Updates: Facebook, Google, and Twitter Testified Before Congress Today*, RECODE (Oct. 31, 2017), <https://www.recode.net/2017/10/31/16579072/live-update-facebook-google-twitter-testimony-russia-congress-hearing> (last visited Oct. 31, 2017) (referencing WASH. POST, *Facebook, Twitter, and Google Testify*, FACEBOOK, <https://www.facebook.com/washingtonpost/videos/10156918211492293/> (Oct. 31, 2017) (at time mark 1:32:37).

⁸ Ehrlich observes, “[p]rior to passage of the CDA, courts addressed this problem in the defamation context by placing the ISP in one of the traditional First Amendment categories of publisher, distributor, or common carrier.” Paul Ehrlich, *Communications Decency Act 230*, 17 BERKELEY TECH. L.J. 401, 403 (2002).

CompuServe was a distributor, not a publisher, of the defamatory content – a dispositive distinction that determined its standard of liability.⁹

The state court in *Stratton Oakmont v. Prodigy Services Co.*, 1995 WL 323710, *5 (N.Y. Sup. Ct. May 24, 1995) reached a different outcome than the *Cubby* court on the same principal question. The *Stratton Oakmont* court held Prodigy, an online message board company, liable as a publisher, meaning it was responsible for content posted by third parties.¹⁰ The company had a practice of filtering out offensive content and the court noted that Prodigy’s control over what was posted on the message board constituted a “principal-agent” relationship between Prodigy and its message board users.¹¹ The court explained that “[b]y actively utilizing technology and manpower to delete notes from its computer bulletin boards on the basis of offensiveness and ‘bad taste,’ for example, PRODIGY is clearly making decisions as to content, and such decisions constitute editorial control.”¹²

The *Cubby* and *Stratton Oakmont* decisions placed ISPs in a catch-twenty-two – (1) An ISP could filter content and become liable as a publisher or (2) An ISP could refuse to filter content and consequently inappropriate, defamatory content would remain unabated.¹³ In 1996, Congress stepped in and passed the Communications Decency Act (CDA). Section 230 of the CDA provides immunity to ISPs that voluntarily act in “good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy,

⁹ The court explained, “[a] computerized database is the functional equivalent of a more traditional news vendor, and the inconsistent application of a lower standard of liability to an electronic news distributor such as CompuServe than that which is applied to a public library, book store, or newsstand would impose an undue burden on the free flow of information. Given the relevant First Amendment considerations, the appropriate standard of liability to be applied to CompuServe is whether it knew or had reason to know of the allegedly defamatory Rumorville statements.” *CompuServe*, 776 F. Supp. at 140-41.

¹⁰ *Stratton Oakmont*, 1995 WL 323710 at *5.

¹¹ *Id.* at *6.

¹² *Id.* at *4 (citation omitted).

¹³ Klonick notes, “[r]ead together, the cases seemed to expose intermediaries to a wide and unpredictable range of tort liability if they exercised any editorial discretion over content posted on their site.” Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, HARV. L.R. 10 (2017).

excessively violent, harassing, or otherwise objectionable, *whether or not such material is constitutionally protected.*”¹⁴ Moreover, CDA § 230(c) states “[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.”¹⁵ CDA § 230 cleared the path for Internet platforms to operate largely immune from liability for the content posted on their platforms. However, it was not long before courts faced the question of what constituted a “publisher.”

The most significant precedent established after CDA § 230 was enacted came from the Fourth Circuit. In *Zeran v. America Online, Inc.*, 29 F.3d 327, 334 (4th Cir. 1997), the plaintiff sued American Online (AOL) as a distributor (not publisher) and claimed AOL should be liable for defamatory content posted on the site. The plaintiff contacted AOL to remove anonymous postings on the site which claimed the plaintiff sold t-shirts with offensive slogans about the 1995 Oklahoma City bombing.¹⁶ The posts included the plaintiff’s name and phone number, which led to death threats and violent calls from Oklahoma residents.¹⁷ The Fourth Circuit held AOL was not liable as a distributor because distributors fall within the scope of “publishers” protected by CDA § 230. The court cited Congressional intent for its rationale and noted:

The imposition of tort liability on service providers for the communications of others represented, for Congress, simply another form of intrusive government regulation of speech. Section 230 was enacted, in part, to maintain the robust nature of Internet communication and, accordingly, to keep government interference in the medium to a minimum.¹⁸

The court also noted Congress’s intent to establish incentives for ISPs to self-regulate and to afford ISPs an affirmative “good-Samaritan” defense.¹⁹ Further, the court denied the

¹⁴ 47 U.S.C. § 230(c) (1998) (emphasis added).

¹⁵ *Id.*

¹⁶ Ehrlich, *supra* note 8, at 406.

¹⁷ *Zeran*, 129 F.3d at 329.

¹⁸ *Id.* at 330.

¹⁹ *Id.* at 331 (“Congress enacted § 230 to remove the disincentives to self-regulation created by the *Stratton Oakmont* decision.”).

plaintiff’s alternative claim that AOL should be liable on a theory of notice (he notified AOL of the anonymous posts). The court emphasized the chilling effect notice liability would have on free speech stating, “[I]ike the strict liability imposed by the *Stratton Oakmont* court, liability upon notice reinforces service providers’ incentives to restrict speech and abstain from self-regulation.”²⁰ *Zeran* was adopted widely in the years following the decision, although its broad reading of CDA § 230 has receded modestly.²¹ Notwithstanding, the foundation of ISP immunity still applies so long as Internet platforms do not editorialize user content in excess and stay within the bounds of “good-Samaritan” content removal.²²

b. The First Amendment and free expression on the Internet.

Congress had dual intentions when it passed CDA § 230. First, as previously discussed, Congress provided ISPs good-Samaritan protection to incentivize ISPs to self-police offensive content. Second, Congress intended to protect free speech on the Internet in alignment with First Amendment values.²³ This binary set of priorities led to a familiar paradox: Prioritize the

²⁰ *Id.* at 332.

²¹ Kosseff writes that following the decision, “*Zeran* soon became the dominant interpretation of Section 230.... Courts across the nation quickly adopted *Zeran’s* broad reading of Section 230.” Jeff Kosseff, *The Gradual Erosion of The Law That Shaped The Internet: Section 230’s Evolution Over Two Decades*, 18 COLUM. SCI. & TECH. L.R. 1, 11 (2016). However, some limitations have been placed on the *Zeran* court’s broad reading of § 230 immunity. Notably, the Ninth Circuit’s decision in *Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC*, 521 F.3d 1157, 1172 (9th Cir. 2008) (en banc), held that Roommates.com was not immunized by CDA § 230 against claims of discrimination because “Roommate both elicits the allegedly illegal content and makes aggressive use of it in conducting its business.” The court noted Roommates.com was an interactive website that *required* user information to provide its service which therefore opened the platform to liability.

²² Matwyshyn observes that CDA § 230 provides Internet platforms “a modicum of legal protection for the content that exists on their platforms, as long as they don’t veer off too much into editorial functions.” Andrea Matwyshyn, *Fake News, Hate Speech, and Social Media Abuse: What’s the Solution?*, U. PENN., Knowledge@Wharton Podcast, (Nov. 21, 2016), <http://knowledge.wharton.upenn.edu/article/whats-the-solution-to-fake-news/> (last visited Oct. 31, 2017).

²³ Klonick asserts, “[t]he [*Zeran*] court recognized two distinct Congressional purposes for granting immunity under § 230: (1) as a Good Samaritan provision written to overturn *Stratton Oakmont* and ‘to encourage intermediaries and users of such services to self-police the Internet for obscenity and other offensive material’ and (2) as a free speech protection for users meant to encourage the ‘unfettered and unregulated development of free speech on the Internet, and to promote the development of e-commerce.’” Klonick, *supra* note 13, at 13 (quoting *Blumenthal v. Drudge*, 992 F. Supp. 44, 52 (D.D.C. 1998)); *see also* *Batzel v. Smith*, 333 F.3d 1018, 1027 (9th Cir. 2003).

removal of harmful content or protect against erroneous removals?²⁴ A survey of four major Internet platforms below demonstrates the tradeoff between vigilant content moderation and near total deference to user expression. A brief review of the obligations imposed by the First Amendment on Internet platforms and the government lays the foundation for this later analysis.

Internet platforms are private corporations, not state actors. Ostensibly, this means that ISPs are not required to uphold First Amendment free speech protections. The First Amendment protects against *government* suppression or “coercive control” of speech, not the actions of private individuals or corporations.²⁵ In fact, CDA § 230 explicitly permits Internet platforms to censor content “whether or not such material is constitutionally protected.”²⁶ Congress understood that private actors could not be constitutionally compelled to permit all forms of speech on their online platforms, yet wanted to promote an Internet of free expression.²⁷ CDA § 230 sought to avoid “collateral censorship” whereby private online platforms, like Twitter, would censor user content as a proxy for government censorship.²⁸ As shown below, courts have given Internet platforms wide latitude in how to regulate speech on their platforms, largely motivated by principles of free expression. Ultimately, Congress and the courts handed the keys to ISPs who now control what information is shared, where it is shared, and who can share.

Seven out of ten American adults now use social media, leading some scholars to observe Facebook, Twitter, YouTube, and other online platforms are the “public square of the 21st

²⁴ See Klonick, *supra* note 13, at 14 (“These dual purposes can be seen as being at odds. If § 230 can be characterized as both (1) government created immunity to *encourage* platforms to remove certain kinds of content and (2) immunity to *avoid* the haphazard removal of certain content and the perils of collateral censorship to users, which interests do we want to prioritize?”).

²⁵ Lincoln Caplan, *Should Facebook and Twitter be Regulated Under the First Amendment?*, WIRED (Oct. 11, 2017), <https://www.wired.com/story/should-facebook-and-twitter-be-regulated-under-the-first-amendment/> (last visited Oct. 31, 2017).

²⁶ 47 U.S.C. § 230(c) (1998).

²⁷ See Ehrlich, *supra* note 8, at 406 (“While Congress tried to strike a balance between promoting decency and promoting free speech, recent court decisions interpreting Section 230 have emphatically tilted the scales in favor of free speech by removing distributor liability from ISPs entirely.”).

²⁸ Klonick, *supra* note 13, at 13-14.

century.”²⁹ The Supreme Court recently acknowledged this prevalence and influence in *Packingham v. North Carolina*, 137 S.Ct. 1730, 1732 (2017). In *Packingham*, the Court held that the right to access social media is protected by the First Amendment, even in the case of registered sex offenders. The Court notes, “[a] fundamental First Amendment principle is that all persons have access to places where they can speak and listen, and then, after reflection, speak and listen once more.”³⁰ Justice Kennedy, for the unanimous majority, expounds:

With one broad stroke, North Carolina bars access to what for many are the principal sources for knowing current events, checking ads for employment, speaking and listening in the *modern public square*, and otherwise exploring the vast realms of human thought and knowledge. Foreclosing access to social media altogether thus prevents users from engaging in the legitimate exercise of First Amendment rights.³¹

Klonick observes that the *Packingham* decision “might breathe new life into the application of state action doctrine to Internet platforms,” despite the longstanding status as non-state actors that Internet platforms have enjoyed.³²

Historically, when a company’s influence becomes so pervasive that the company itself takes on fundamental responsibilities of the state, the company may expose itself to constitutional liability. In *Marsh v. Alabama*, 326 U.S. 501, 509 (1946), the Court held that a company town, although privately owned and managed, was obligated to guarantee certain constitutional protections like those enshrined in the First and Fourteenth Amendments. The town provided many of the services of a municipality and the Court stated there is “nothing to distinguish them [the company town and shopping center] from any other town and shopping center except the fact that the title to the property belongs to a private corporation.”³³

²⁹ Caplan, *supra* note 25.

³⁰ *Packingham*, 137 S.Ct. at 1735.

³¹ *Id.* at 1732 (emphasis added).

³² Klonick, *supra* note 13, at 17.

³³ *Marsh*, 326 U.S. at 503.

While Internet platforms, like Google, may not provide state infrastructure at the scale of the company town in *Marsh*, the *Packingham* decision indicates that access to Internet platforms as a means to promote speech is a constitutional right – at least so far as state proscriptions against access are concerned. Scholars note, the distinction between state and non-state actors is blurred by the ubiquity of Internet platforms and the pervasive nature of online content sharing.³⁴ As noted, Facebook and Google can remove content (or abusive users) with immunity – effectively suppressing certain forms of speech.³⁵ Yet, the influence of Internet platforms in the context of free speech has led a growing chorus of academics and commentators to call for greater regulation of ISPs under the First Amendment.³⁶ Caplan writes that scholars calling for tighter ISP regulations “even suggest that it’s time to rewrite the Communications Decency Act—to impose responsibility on social media for the factual accuracy of content they host.”³⁷

Publishing false content raises further questions of liability and First Amendment guarantees. In *New York Times Co. v. Sullivan*, 376 U.S. 254, 279 (1964), the Court held that First Amendment protections extend to false statements of publishers made unknowingly. As a result, libel claims against publishers must be predicated on a publisher’s knowledge or reckless disregard for the truth in order to sustain judgment.³⁸ Justice Brennan for the unanimous Court wrote:

³⁴ Caplan writes, “[a] major premise of these recent proposals to reinterpret the First Amendment is what the Supreme Court said in *June*: Social media are today’s town halls and public parks, where ideas compete for influence on topics ‘as diverse as human thought.’” Caplan, *supra* note 25 (quoting *Packingham*, 137 S.Ct. at 1736).

³⁵ PEN America’s fake news report notes, “[t]he First Amendment does not—in all but the most limited of circumstances—preclude private parties from restricting speech. Thus, private companies—including social media companies, online service providers, and others—are legally permitted to act on their own initiative to ‘censor’ or otherwise restrict speech by those who use their services or platforms. For the Facebooks, Twitters, and Googles of the world, the organizations’ commercial relationships with its users, as well as its Codes of Conduct for how it will self-regulate as an organization, constitute the framework for its ability or responsibility to impede speech.” PEN America, *supra* note 6, at 25.

³⁶ Caplan, *supra* note 25.

³⁷ *Id.*

³⁸ *Sullivan*, 376 U.S. at 280.

A rule compelling the critic of official conduct to guarantee the truth of all his factual assertions—and to do so on pain of libel judgments virtually unlimited in amount—leads to a comparable ‘self-censorship.’ Allowance of the defense of truth, with the burden of proving it on the defendant, does not mean that only false speech will be deterred.³⁹

The *Sullivan* Court foresaw the danger of media self-censorship in the same manner in which Congress sought to avoid collateral censorship with CDA § 230. If any false content, whether or not published knowingly, exposed publishers to liability, the media may opt not to publish stories of public interest at all.

Applied in the age of social media, *Sullivan*’s outcomes are not so clear. Commentators note, “a lot of content produced by cranks, rogues, and hackers who don’t fit the model the Supreme Court had in mind in *Sullivan* finds an audience.”⁴⁰ This “breakdown” of traditional media has circumvented longstanding practices of fact-checking and journalistic discretion.⁴¹ PEN America, a free expression advocacy group, recently reported on the fake news phenomenon and observed the landscape for legal proscriptions against fake news. The report’s authors surmise that any state regulation to restrict fake news would likely face strict scrutiny by the courts.⁴² This means that to pass constitutional muster, such regulations would have to be narrowly tailored to a compelling government interest.⁴³ The report suggests that the prospect of constitutionally sound restrictions on fake news is dim. “In practice, the attempted government regulation of speech based on its content is almost never accepted as constitutional.”⁴⁴

This First Amendment framework helps explain how and why online platforms have evolved in their approach to content moderation. It is particularly enlightening in the context of

³⁹ *Id.* at 279.

⁴⁰ Caplan, *supra* note 25.

⁴¹ *Id.*

⁴² PEN America, *supra* note 6, at 25.

⁴³ See *Brown v. Ent. Merch. Ass’n*, 564 U.S. 786, 787 (2011). The Court struck down a California law which barred minors from purchasing violent video games noting, “[b]ecause the Act imposes a restriction on the content of protected speech, it is invalid unless California can demonstrate that it passes strict scrutiny, i.e., it is justified by a compelling government interest and is narrowly drawn to serve that interest.”

⁴⁴ PEN America, *supra* note 6, at 25.

greater state regulation internationally and competing jurisdictional norms that are discussed at the end of this paper. The mechanics of on-the-ground content moderation across four major platforms – Reddit, Twitter, Google, and Facebook – are analyzed before this comparative regulatory analysis.

III. Balancing the Bottom Line and Free Speech – The Spectrum of Content Moderation from Reddit to Facebook

The power and autonomy afforded to Internet platforms through CDA § 230 and the First Amendment have led to an array of content moderation approaches. This paper surveys four major Internet platforms on a spectrum – the least self-regulating platforms (Reddit) to the most heavy-handed censors (Facebook). Unsurprisingly, free rein to self-regulate within these domains has led to disparate outcomes and divergent means to address the proliferation of fake news. Each Internet platform is considered from a historical perspective foremost – how have issues of harassment, defamation, and general violations of terms of service been managed prior to the focus on fake news? Next, the susceptibility of each platform towards the proliferation of fake news is examined. Finally, each platform’s approach towards fake news content moderation is outlined, compared, and scrutinized.

a. Reddit – The origins of laissez-faire content moderation.

Facebook, Twitter, and Google receive the bulk of attention when it comes to self-policing content. This may be due, in part, to the significant disparity in revenue generated by Reddit (and its respective market capitalization) in comparison to its peers.⁴⁵ Yet regardless of

⁴⁵ In 2017, Reddit’s valuation was \$1.8 billion and the site reportedly generated \$8.3 million in ad revenue in 2015. Todd Spangler, *Reddit Has \$1.8 Billion Valuation After Chat-Room Site Banks \$200 Million in Funding*, VARIETY (Aug. 1, 2017), <http://variety.com/2017/digital/news/reddit-1-8-billion-valuation-funding-1202512082/> (last visited Oct. 31, 2017). This is quite modest in comparison to Twitter’s \$12 billion market cap, Facebook’s \$500 billion market cap, and Google’s \$700 billion market cap. Rani Molla, *Square is Worth Almost as Much as that Other*

revenue, since its inception in 2005 Reddit has become one of the most widely viewed sites on the Internet. As of June 2017, Reddit claimed more than 300 million active monthly users, living up to its self-proclaimed moniker as the “front page of the Internet.”⁴⁶ According to Alexa, a website traffic and marketing platform owned by Amazon.com, Reddit is the eighth most visited site in the world and fifth most visited site in the U.S.⁴⁷

Commensurate with its cash flow, Reddit has a starkly smaller staff than Twitter, Facebook, and Google. The Verge reported that in 2016, Reddit employed only 75 full-time staff, leaving much of the self-policing to the users themselves.⁴⁸ Perhaps more than any of its peers examined in this paper, Reddit is militantly committed to free and open expression, regardless of how offensive. This corporate ethos inherently conflicts with attempts to meaningfully moderate content on Reddit. Buni and Chemaly observe, “[t]he company continues to grapple with the paradox that to expand, Reddit (and other platforms) will likely have to regulate speech in ways that alienate a substantive percentage of their core customer base.”⁴⁹

In the fallout of “CelebGate,” where the private photos of more than 100 female celebrities were hacked and shared on the site in 2014, Reddit published some of its first

Company Jack Dorsey Runs, RECODE (Oct. 25, 2017), <https://www.recode.net/2017/10/25/16545776/square-twitter-earnings-worth-jack-dorsey-payments> (last visited Oct. 31, 2017); John Shinal, *Facebook Passes Amazon in Market Cap and Now Sits on the Brink of an Exclusive Club*, CNBC (July 28, 2017), <https://www.cnbc.com/2017/07/28/facebook-touches-500-billion-passes-amazon.html> (last visited Oct. 31, 2017); Ciara Linnane, *Google Parent Alphabet’s Market Cap Jumps Above \$700 Billion for First Time*, MARKETWATCH (Oct. 27, 2017), <https://www.marketwatch.com/story/google-parent-alphabets-market-cap-jumps-above-700-billion-2017-10-27> (last visited Oct. 31, 2017).

⁴⁶ Tom Huddleson Jr., *Reddit Will Sell Video Ads to Lure More Advertisers*, FORTUNE (June 23, 2017), <http://fortune.com/2017/06/23/reddit-to-add-video-ads-in-push-for-more-revenue/> (last visited Dec. 12, 2017).

⁴⁷ *The Top 500 Sites on the Web*, ALEXA, <https://www.alexa.com/topsites> (last visited Dec. 12, 2017); *Top Sites in the United States*, ALEXA, <https://www.alexa.com/topsites/countries/US> (last visited Dec. 12, 2017).

⁴⁸ Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, VERGE (Apr. 13, 2016), <https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech> (last visited Oct. 29, 2017).

⁴⁹ *Id.*

community standards on anti-harassment.⁵⁰ How did Reddit users (“Redditors”) react? A subsequent survey of more than 16,000 Redditors found that 35% of the complaints from “extremely dissatisfied users” were due to “heavy handed moderation and censorship.”⁵¹ Reddit’s foray into site-wide community standards was met with stiff backlash from users accustomed to total free expression.

Due to its small staff size, Reddit relies on an army of volunteer moderators to police content.⁵² These unpaid Redditors are tasked with moderating harassment on “subreddits” (forums dedicated to specific topics) and have many of the same powers as the paid contractors and employees at Reddit’s more profitable peer companies.⁵³ Moderators can remove content, mark content as unsuitable for work, ban users from commenting or viewing a subreddit, manage spam, and more.⁵⁴ This democratization and decentralization of moderation leads to a loose standards-based approach that is a far cry from the extensive rule-based culture at Facebook.⁵⁵

Adrian Chen of the *New York Times Magazine*, explains the problem posed by Reddit’s reliance on volunteer moderators who are inevitably armed with individual prerogatives and operate largely void of corporate governance. Chen writes, “[t]he idea of ‘moderating’ presupposes an outside vision of what is and isn’t acceptable in a conversation. But when moderators set their own rules, with no incentive to conform to anyone else’s standard, they can

⁵⁰ See *id.*; Dante D’Orazio, *Reddit Shuts Down Subreddit Tied to Nude Photos in ‘Celebgate’ Attack*, VERGE (Sept. 7, 2014), <https://www.theverge.com/2014/9/7/6117145/reddit-shuts-down-the-fapping-subreddit-tied-to-nude-photos> (last visited Nov. 2, 2017).

⁵¹ Buni & Chemaly, *supra* note 48.

⁵² Kim Renfro, *For Whom the Troll Trolls: A Day in the Life of a Reddit Moderator*, TECH INSIDER (Jan. 13, 2016), <http://www.businessinsider.com/what-is-a-reddit-moderator-2016-1/#crocker-is-a-graduate-research-assistant-for-the-emerging-media-studies-department-but-a-significant-chunk-of-her-time-is-spent-helping-moderate-the-science-subreddit-shes-been-a-moderator-for-over-a-year-and-a-half-now-1> (last visited Oct. 29, 2017).

⁵³ *Id.*

⁵⁴ *What Are Moderators Able to Do that Users Can’t?*, REDDIT, <https://www.reddit.com/wiki/moderation> (last visited Oct. 29, 2017).

⁵⁵ Buni and Chemaly explain, “[t]he sharp contrast between Facebook, with its robust and long-standing Safety Advisory Board, and Reddit, with its skeletal staff and dark pools of offensive content, offers a vivid illustration for how content moderation has evolved in isolated ways within individual corporate enclaves.” Buni & Chemaly *supra* note 48.

look a lot less like custodians and a lot more like petty tyrants.”⁵⁶ Chen goes on to note that Reddit’s extreme tolerance for hateful speech is largely a byproduct of the moderator class being “detached from its mediating role” and ultimately self-serving.⁵⁷ This laissez-faire approach not only contributes to hate speech and defamation, but also has made Reddit ground zero in fake news warfare shown below.

b. Weak governance and a susceptible platform – why fake news abounds on Reddit.

Reddit is a breeding ground for fake news for several reasons. First, as noted, Reddit’s volunteer content moderators act within a system largely void of governance and rules.⁵⁸ This gives moderators carte blanche to control what content is posted and who can participate. Second, Reddit’s system of up-votes and down-votes – a system designed to provide users the power to promote or demote discussion threads – can be easily gamed with a little money and willpower.

Journalist Jay McGregor explained how simple it is for “internet marketing agencies” to manipulate conversations on Reddit.⁵⁹ Agencies can be hired for very little money (as low as a couple hundred dollars) and deploy hundreds, or thousands, of bots or contractors to do someone else’s bidding. McGregor quotes one UK-based internet marketing agency, which boasted of its ability to covertly promote brands, as stating “[w]ell there’s different IP addresses, they have real emails behind them that aren’t anything to do with your company at all, different avatars, you know, if you can tell me roughly what they’re saying, we can rework it so it looks natural. So

⁵⁶ Adrian Chen, *When the Internet’s ‘Moderators’ are Anything But*, N.Y. TIMES MAG. (July 21, 2015), https://www.nytimes.com/2015/07/26/magazine/when-the-internets-moderators-are-anything-but.html?_r=0 (last visited Oct. 29, 2017).

⁵⁷ *Id.*

⁵⁸ Buni & Chemaly, *supra* note 48.

⁵⁹ Jay McGregor, *Reddit is Being Manipulated by Big Financial Services Companies*, FORBES (Feb. 20, 2017), <https://www.forbes.com/sites/jaymcgregor/2017/02/20/reddit-is-being-manipulated-by-big-financial-services-companies/2/#7e41dee924b5> (last visited Oct. 29, 2017).

we'll make an effort to make it look natural.”⁶⁰ Unapologetically, the UK marketing agency described the ease with which Reddit forums can be manipulated by companies or individuals masquerading as an army of average Reddit users.

To illustrate the point, McGregor was able to promote two completely fabricated stories to the second and first places, respectively, on two popular subreddit threads in 2016.⁶¹ The first story, titled “Brexit: Who needs the EU when we’ve got China?” authored by a fictitious Norwegian Professor from a made-up university, climbed to the second highest rank on the “r/unitedkingdom” subreddit (with 145,000 followers) at the peak of the pre-Brexit vote frenzy.⁶² The story was removed after several hours when enough Reddit users commented on the falsity of the story (there were many obvious clues), but only after it was viewed and possibly believed by countless others. The second story was posted to the “r/videos” subreddit and its 13 million followers. The story falsely claimed the second season of *Narcos* (a Netflix original TV show) was coming out in September 2016 and linked to an old trailer. Within two hours, the post reached the top spot on the forum and was not removed despite its lack of veracity. The cost for this double-barreled fake news ploy? \$200.⁶³

Actors with mal-intent (or financial prerogative) rely on sophisticated means to dupe the system. Have \$40? If so, you can buy 200 organic votes “manually logged by human labour somewhere inside Pakistan.”⁶⁴ Savvy hackers have even created computer programs to “roll[]

⁶⁰ Jay McGregor, *Reddit’s Inclusion in the Trump-Russia Investigation Makes Perfect Sense*, FORBES (Sept. 28, 2017), <https://www.forbes.com/sites/jaymcgregor/2017/09/28/reddits-inclusion-in-trump-russia-investigation-makes-perfect-sense/#24d47db41dc8> (last visited Oct. 29, 2017).

⁶¹ Jay McGregor, *Reddit for Sale: How We Made Viral Fake News for \$200*, FORBES (Dec. 14, 2016), <https://www.forbes.com/sites/jaymcgregor/2016/12/14/how-we-bought-reddit-for-200/2/#ac763979bdf0> (last visited Oct. 30, 2017).

⁶² *Id.*

⁶³ *Id.*

⁶⁴ *Id.*

through” thousands of IP addresses to give the appearance of authenticity when launching a “viral” campaign.⁶⁵

McGregor’s experiment exposed how simple it is for anyone with a brand to promote or an ideology to espouse, to influence conversations on one of the most popular message boards in the world. One online marketing agent reportedly told McGregor, “I’ve been doing viral marketing and reputation management since 2005. As I said above, I’ve worked over 100 of these types of campaigns.”⁶⁶ The agency representative went on to note that most of her clients were in the sports and entertainment industries, but the success of “shilling” agencies, like those unveiled by McGregor, arguably paved the way for more sinister forms of electorate manipulation via Reddit.⁶⁷ Fake up and down votes, fake comments, fake followers, and fake articles – all one needs is a little cash.

c. The platform paradox – The limitations of Reddit’s free speech posture.

Volunteer content moderators, barebone governance, and cheap marketing tricks to game the system are a toxic mix when it comes to combatting fake news. Facing this challenge, what resources do moderators have to tamp down false content? Very little. Moderators have the ability to review individual accounts, check account history, and view a user’s “karma” score – a system of reward points for posting content – yet those are essentially the only tools at their disposal.⁶⁸ This creates a massive problem of scale. Thousands of volunteer moderators

⁶⁵ *Id.*

⁶⁶ McGregor, *supra* note 59.

⁶⁷ *Id.*; see also McGregor, *supra* note 60 (“This investigation [McGregor’s Reddit exposé] all took places months after the 2016 election and, more importantly, Reddit doesn’t appear to have made any significant changes to stop this. It’s entirely feasible that anyone actor that wished to manipulate voters would include Reddit given its popularity and the ease with which it can be gamed.”).

⁶⁸ McGregor quotes an anonymous moderator of the politics subreddit: “As far as stopping everything, there’s nothing the moderators will ever be able to do. We can only see the user history. That’s going to have to come from the admin side of things. There’s just nothing we can do.” McGregor, *supra* note 59.

operating within subreddit silos, hundreds of millions of users, and incalculable numbers of discussion threads make the magnitude of potential manipulation hard to overstate.

In the lead-up to the 2016 U.S. election, Reddit is where several of the most infamous fake news stories spawned. Pizzagate?⁶⁹ Reddit. New Yorkers being bussed to Philadelphia to vote?⁷⁰ Reddit. Democrats intimidating voters at the polls?⁷¹ Reddit. And on and on. Unlike Facebook, Reddit's forums are open to everyone (even users without accounts) and conversations are well organized by topic, which makes targeting specific forums appealing for bad actors and foreign influencers.⁷²

In the aftermath of the 2016 U.S. election and Congressional inquiry into Russian meddling, Reddit has not faced the same scrutiny as its peers, like Facebook, Twitter, and Google, for selling advertisements to Russian trolls.⁷³ Instead, investigators have focused on the substance of the conversations promoted on the chat board. In September 2017, U.S. Senator Mark Warner (D-VA) of the Senate Intelligence Committee hinted that Congress might turn its focus towards the type of shilling McGregor exposed in his fake news experiment.⁷⁴ Breland for the *Hill* writes, "lawmakers and experts say they're more concerned with coordinated efforts by fake users and bots to sway public opinion organically."⁷⁵ Further illuminating the problem, Breland notes, "Russian actors using Reddit to spread certain types of information also wouldn't

⁶⁹ Pizzagate refers to the conspiracy theory that claimed Democratic Presidential candidate Hillary Clinton and her campaign manager John Podesta ran a child sex ring out of a pizzeria in affluent Northwest D.C. Lizzie Plaugic, *As it Spreads Online and Off, Pizzagate Gets Weirder and More Dangerous*, VERGE (Dec. 7, 2016), <https://www.theverge.com/2016/12/7/13870990/pizzagate-robotas-brooklyn-conspiracy-theory> (last visited Oct. 31, 2017); see also Ali Breland, *Warner Sees Reddit as Potential Target for Russian Influence*, HILL (Sept. 27, 2017), <http://thehill.com/policy/technology/352584-warner-sees-reddit-as-potential-target-for-russian-influence> (last visited Oct. 29, 2017).

⁷⁰ Breland, *supra* note 69.

⁷¹ *Id.*

⁷² Breland explains, "[u]nlike Facebook, Reddit's groups are open and accessible to anyone, even users who don't have accounts. And unlike Twitter, the content and conversations are neatly organized and can be arranged by popularity or time." *Id.*

⁷³ *Id.*

⁷⁴ *Id.*

⁷⁵ *Id.*

necessarily violate the company's rules, as long as real people are behind the accounts disseminating the information."⁷⁶

This scrutiny has forced Reddit to respond. Under Reddit's general features and policies, a 114-word article entitled "What about 'Fake News' on Reddit?" lays out the company's mantra. The article states:

Users on Reddit tend to call things like they see it. So if they see a link that is inaccurate, they'll call out fake news, they'll downvote it, and moderators have the option of blacklisting certain sources from being posted to their subreddits. Our community is very empowered to curate the content they want to see.⁷⁷

In essence, Reddit is leaving the fake news policing up to its users. The policy further notes, "[w]ith regard to promoted posts, Reddit has very clear policies on advertising that does not allow the promotion of fake news." As previously noted, direct ad sales on Reddit are not the foremost concern when it comes to the proliferation of fake news on the platform. In fact, Reddit's advertising-purchase process is arguably more adept at weeding out fake or misleading ads because ad buyers are required to work directly with a Reddit employee, unlike Facebook, Twitter, and Google.⁷⁸ Reddit's issue is more fundamental to the DNA of the message board – which users are real and who is artificially juicing the numbers?

Reddit's co-founder, Alexis Ohanian, places a great amount of trust and belief in a system of user self-regulation. In an interview with *The Australian*, Mr. Ohanian detailed the distinction between Reddit and other platforms' approaches. He states, "[t]he difference with Reddit is that we have a very active community and a team that are invested and engaged in the site. . . . [Reddit has] thwart[ed] a lot of the bad actors who were posting fake news, simply

⁷⁶ *Id.*

⁷⁷ *What about "Fake News" on Reddit?*, REDDIT, <https://www.reddithelp.com/en/categories/using-reddit/general-reddit-features-and-policies/what-about-fake-news-reddit> (last visited Oct. 31, 2017).

⁷⁸ Breland writes, "[u]nlike Facebook, Twitter and Google, though, Reddit's ad-buying process requires purchasers to interact with a person at the company, a step that could complicate attempts to buy election-influencing ads." Breland, *supra* note 69.

because they were notorious for trying to game our system.”⁷⁹ Mr. Ohanian emphasized his belief in the efficacy of Reddit’s approach claiming, “[m]ost of those fake news creators were trying to cheat Reddit but the nature of how Reddit works and our methods — like introducing a new blocking tool to minimise online harassment — have allowed us to thwart these evildoers and nab most of them. So they didn’t have much success.”⁸⁰ Mr. Ohanian may believe Reddit is largely immune from the woes of fake news that plague Facebook, Twitter, and Google, but the facts say otherwise.

Reddit’s culture of anonymity and reverence for unadulterated free speech has real consequences. In the wake of the 2013 Boston Marathon bombing, the name of an incorrect suspect surfaced on Reddit and was picked up by mainstream media.⁸¹ In December 2016, a rogue man with a rifle fired a shot into Comet Ping Pong, a popular pizzeria in Northwest Washington D.C., because of his belief the pizzeria was operating an underground child sex ring run by John Podesta, Hillary Clinton’s presidential campaign chair, and Hillary Clinton herself.⁸² The conspiracy started as a joke, but found an audience on Reddit and quickly jumped to other platforms.⁸³ To be sure, Reddit is not the only culprit in the fake news affair, yet institutionally it is one of the least-prepared platforms to genuinely confront the dilemma.

⁷⁹ Mitchell Bingemann, *Reddit: We can Fix Fake News, Unlike Google, Facebook*, AUSTRALIAN (Jan. 30, 2017), <http://www.theaustralian.com.au/business/media/reddit-we-can-fix-fake-news-unlike-google-facebook/news-story/e81b6f2b3194aaec2b564b859e9bbf6b> (last visited Oct. 31, 2017).

⁸⁰ *Id.*

⁸¹ Ben Strauss, *Misinformation Is the New Normal of Mass Shootings*, POLITICO MAG. (Oct. 2, 2017), <https://www.politico.com/magazine/story/2017/10/02/las-vegas-shooting-fake-news-guns-215670> (last visited Oct. 31, 2017).

⁸² Reddit moderators did attempt to contain Pizzagate propaganda on the site, but their efforts were unsuccessful. Plaugic writes, “the service closed the subreddit r/pizzagate for violating site rules. But Pizzagate theorists quickly found a new spot for their hive mind in the r/The_Donald and r/conspiracy subreddits.” Plaugic, *supra* note 69.

⁸³ Kaitlyn Tiffany, *If a Scientific Conspiracy Theory is Funny, That Doesn’t Mean It’s a Joke*, VERGE (Oct. 9, 2017), <https://www.theverge.com/2017/10/9/16424622/reddit-conspiracy-theories-memes-irony-flat-earth> (last visited Oct. 31, 2017).

d. A little birdie told me – Twitter’s early days and free speech fundamentalism.

In 2017, Twitter is synonymous with Trump. But before the 45th U.S. President took to Twitter, the company built a user base of hundreds of millions and established itself as the go-to source for fast breaking news and opinion.⁸⁴ Twitter was founded in 2006 as a messaging platform for users to Tweet personal status updates to their friends.⁸⁵ However, the power of instant communication with strangers quickly became apparent.⁸⁶ Today, Twitter has a reported 328 million active monthly users and analysts estimate more than 150 million people use the platform every day.⁸⁷ Twitter is the thirteenth most visited site in the world and eighth most popular in the United States.⁸⁸

Like Reddit, Twitter’s reverence for free speech runs deep. In 2013, Jeffrey Rosen of the *New Republic* wrote an article entitled “The Delete Squad” in which he profiled the origins of content moderation at Twitter, Facebook, and Google.⁸⁹ He noted, “[t]he company that has moved the furthest toward the American free-speech ideal is Twitter, which has explicitly concluded that it wants to be a platform for democracy rather than civility.” Rosen further observed, “[u]nlike Google and Facebook, it [Twitter] doesn’t ban hate speech at all; instead, it prohibits only ‘direct, specific threats of violence against others.’”⁹⁰ Notably, Twitter’s policies

⁸⁴ Kurt Wagner, *How Many People Use Twitter Every Day?*, RECODE (July 27, 2017), <https://www.recode.net/2017/7/27/16049334/twitter-daily-active-users-dau-growth-q2-earnings-2017> (last visited Nov. 4, 2017).

⁸⁵ Nicholas Carlson, *The Real History of Twitter*, BUS. INSIDER (Apr. 13, 2011), <http://www.businessinsider.com/how-twitter-was-founded-2011-4?op=1> (last visited Nov. 4, 2017).

⁸⁶ Carlson notes the information-sharing potential of Twitter in the platform’s infancy, writing “[i]n August [2006], a small earthquake shook San Francisco and word quickly spread through Twitter — an early ‘ah-ha!’ moment for users and company-watchers alike.” *Id.*

⁸⁷ Wagner, *supra* note 84.

⁸⁸ *Twitter.com Traffic Statistics*, ALEXA (Dec. 11, 2017), <https://www.alexa.com/siteinfo/twitter.com> (last visited Dec. 12, 2017).

⁸⁹ Jeffrey Rosen, *The Delete Squad*, NEW REPUBLIC (Apr. 28, 2013), <https://newrepublic.com/article/113045/free-speech-internet-silicon-valley-making-rules> (last visited Nov. 4, 2017).

⁹⁰ *Id.*

have changed since 2013 and the company has taken a more assertive stance towards minimizing abusive content on the platform, including banning certain forms of targeted hate speech.⁹¹

However, the precept of free speech over censorship remains.

Historically, Twitter has relied on the power of the platform itself, absent company intervention, to drown out distasteful speech.⁹² Today, Twitter moderators work within the company’s “Trust & Safety Team” and an outside advisory council addresses thorny issues of censorship.⁹³ The company’s approach towards content moderation remains largely unpublicized and members of Twitter’s advisory board are forbidden from sharing information about internal policies.⁹⁴ Former employees and statements of company representatives have shed some light on Twitter’s free speech conundrum and highlight a patchwork of efforts to respond to global pressures and disaffected users.⁹⁵

Like Facebook and Google, Twitter has faced an increasing demand to block content based on local laws. To comply (and avoid heavy fines or government shut down), Twitter implements ‘geo-blocking’ technology that censors certain content in a specific region or country.⁹⁶ Twitter explains, “[w]ith hundreds of millions of Tweets posted every day around the world, our goal is to respect our users’ expression, while also taking into consideration

⁹¹ John Bowden, *Twitter Rolling Out New Rules on Hate Speech, Sexual Harassment*, HILL (Oct. 14, 2017), <http://thehill.com/policy/technology/355441-twitter-rolling-out-new-rules-on-hate-speech-sexual-harassment> (last visited Nov. 4, 2017).

⁹² Recalling an incident in 2012 in which Twitter users in France began Tweeting anti-Semitic hashtags, Rosen writes, “[w]ithin days, the bulk of the Tweets carrying the hash tag had turned from anti-Semitic to denunciations of anti-Semitism, confirming that the Twittersphere is perfectly capable of dealing with hate speech on its own, without heavy-handed intervention.” Rosen, *supra* note 89.

⁹³ Buni & Chemaly, *supra* note 48.

⁹⁴ Buni and Chemaly note that members of Twitter’s Trust & Safety Advisory Council operate under nondisclosure agreements and the council’s work “remain[s] behind the scenes.” *Id.*

⁹⁵ Buni and Chemaly interviewed Twitter’s former manager of the User Safety Policy Team and former engineer who provided insight into Twitter’s efforts which is highlighted throughout this paper. *Id.*

⁹⁶ *The Tweets Still Must Flow*, TWITTER (Jan. 26, 2012), https://blog.twitter.com/official/en_us/a/2012/tweets-still-must-flow.html (last visited Nov. 4, 2017).

applicable local laws.”⁹⁷ However, Twitter’s willingness to censor content locally has its limits. Biz Stone, one of Twitter’s co-founders, penned a widely-read blog post entitled “The Tweets Must Flow” in the wake of the Arab Spring and the Egyptian government’s shut down of the Internet in 2011.⁹⁸ In the post, Stone articulates Twitter’s ethos writing:

Our goal is to instantly connect people everywhere to what is most meaningful to them. For this to happen, freedom of expression is essential. Some Tweets may facilitate positive change in a repressed country, some make us laugh, some make us think, some downright anger a vast majority of users. We don’t always agree with the things people choose to tweet, but we keep the information flowing irrespective of any view we may have about the content.⁹⁹

Twitter implemented its country-based content blocking policies nearly a year after the Arab Spring, ceding some ground in its battle for undiluted free speech.¹⁰⁰ Since 2012, Twitter’s free speech stance has continuously adapted in the wake of criticism over abusive content and trolls. Twitter rolled out new policies in 2015 after a slew of celebrities quit Twitter and publically chastised the platform for how it handles abusive users.¹⁰¹ The new policies forbid users from making “threats of violence against others or *promot[ing]* violence against others” – an adaptation from the previous policy that only banned “direct, specific threats of violence against others.”¹⁰² Twitter vowed to kick repeat offenders off the platform and more

⁹⁷ Twitter’s policy further states “[m]any countries, including the United States, have laws that may apply to Tweets and/or Twitter account content. In our continuing effort to make our services available to users everywhere, if we receive a valid and properly scoped request from an authorized entity, it may be necessary to reactively withhold access to certain content in a particular country from time to time.” *Country Withheld Content*, TWITTER, <https://support.twitter.com/articles/20169222> (last visited Nov. 4, 2017).

⁹⁸ Courtney Boyd Myers, *Amid Oppression in Egypt, Twitter’s Founder Stands Up for Freedom of Expression*, NEXT WEB (Jan. 28, 2011), <https://thenextweb.com/twitter/2011/01/28/amid-oppression-in-egypt-twiters-founder-stands-up-for-freedom-of-expression/> (last visited Nov. 4, 2017).

⁹⁹ Biz Stone, *The Tweets Must Flow*, TWITTER BLOG (Jan. 28, 2011), https://blog.twitter.com/official/en_us/a/2011/the-tweets-must-flow.html (last visited Nov. 4, 2017).

¹⁰⁰ Stone, *supra* note 96.

¹⁰¹ David Auerbach, *If You Can’t Stop It, Hide It*, SLATE (Apr. 21, 2015), http://www.slate.com/articles/technology/bitwise/2015/04/twitter_s_new_abuse_policy_if_it_cant_stop_it_hide_it.html (last visited Nov. 5, 2017).

¹⁰² *Id.* (emphasis added).

aggressively pursue violators.¹⁰³ By some measure, Twitter has followed through on its promise and permanently banned notable figures like Milo Yiannopoulos, a caustic conservative commentator, and Roger Stone, an outspoken advisor of President Trump.¹⁰⁴ However, Twitter's disparate treatment of whom it kicks off the platform has led many to criticize the platform's efforts to minimize harassment.¹⁰⁵

e. Russian bots, misinformation, and account anonymity – Twitter's Trump-size problem and runty response.

Twitter's reactionary efforts to handle harassment and moderate content have been complicated by misinformation campaigns and efforts to hijack the platform for political gain. As of November 2017, Twitter estimates that 36,000 Russian-linked accounts generated 1.4 million automated Tweets about the 2016 U.S. election.¹⁰⁶ Three thousand of those accounts were associated with Russia's Internet Research Agency, a shadowy pro-Kremlin propaganda organization.¹⁰⁷ One commentator remarked on the magnitude of the problem stating "[b]ut few people believe this [the number of compromised accounts] is a definitive tally."¹⁰⁸

Automated bots on Twitter are not a new phenomenon. According to a study by the University of Southern California and Indiana University, between 9-15% of all Twitter accounts

¹⁰³ *Id.*

¹⁰⁴ Abby Ohlheiser, *Just How Offensive Did Milo Yiannopoulos Have to Be to Get Banned from Twitter?*, WASH. POST (July 21, 2016), https://www.washingtonpost.com/news/the-intersect/wp/2016/07/21/what-it-takes-to-get-banned-from-twitter/?utm_term=.e7eb5d9fa72b (last visited Nov. 5, 2017); Jacey Fortin, *Roger Stone Suspended From Twitter After Expletive-Laden Tweets*, N.Y. TIMES (Oct. 29, 2017), <https://www.nytimes.com/2017/10/29/us/politics/roger-stone-twitter-suspended.html> (last visited Nov. 5, 2017).

¹⁰⁵ Frum writes, "[w]hat Twitter is saying is that some and only some speech will be policed, by standards that can only be guessed at in advance." David Frum, *Twitter's Misbegotten Censorship*, ATLANTIC (Nov. 16, 2016), <https://www.theatlantic.com/politics/archive/2016/11/twitter-censorship-will-only-empower-the-alt-right/507929/> (last visited Nov. 5, 2017).

¹⁰⁶ Selina Wang, *Twitter Sidestepped Russian Account Warnings, Former Worker Says*, BLOOMBERG (Nov. 3, 2017), <https://www.bloomberg.com/news/articles/2017-11-03/former-twitter-employee-says-fake-russian-accounts-were-not-taken-seriously> (last visited Nov. 5, 2017).

¹⁰⁷ *Id.*

¹⁰⁸ *Id.*

are bots (not human users).¹⁰⁹ This amounts to an astounding 48 million accounts.¹¹⁰ Unlike Facebook, Twitter does not require real names for users and does not prohibit automated accounts.¹¹¹ Bots can be mobilized to manipulate trending news on Twitter and give hashtags and content the appearance of Internet virality.¹¹² Indiana University researcher Chengcheng Shao analyzed 14 million Twitter accounts and found “[a]ccounts that actively spread misinformation are significantly more likely to be bots.”¹¹³ Bots are deployed to target specific individual users who are in turn more likely to share false content organically.¹¹⁴ Automated accounts inflate the genuine number of followers and ‘Re-Tweets’ of real users as well – one study suggests that half of President Trump’s Twitter followers are not real people.¹¹⁵

The anonymous nature of Twitter, like Reddit, lends itself to deceit and user manipulation. But there is also the financial reality. Leslie Miley, Twitter’s former Engineering Manager of Product Safety and Security, explained Twitter’s internal battle between user growth and account authentication in a recent interview with Bloomberg News. Miley was tasked with preventing platform manipulation, but he asserts that his work was sidestepped in order to demonstrate user growth – a key factor in Twitter’s market valuation.¹¹⁶ In 2015, Miley discovered a slew of accounts generated by IP addresses in Russia and Ukraine that he suspected

¹⁰⁹ Onur Varol et al., *Online Human-Bot Interactions: Detection, Estimation, and Characterization*, 2017 INT’L CONF. WEB & SOC. MEDIA (Mar. 27, 2017), <https://arxiv.org/pdf/1703.03107.pdf> (last visited Nov. 5, 2017).

¹¹⁰ Michael Newberg, *As Many as 48 Million Twitter Accounts Aren’t People, Says Study*, CNBC (Mar. 10, 2017), <https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html> (last visited Nov. 5, 2017).

¹¹¹ Scott Shane, *The Fake Americans Russia Created to Influence the Election*, N.Y. TIMES (Sept. 7, 2017), <https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html> (last visited Nov. 5, 2017).

¹¹² *Id.*

¹¹³ *First Evidence That Social Bots Play a Major Role in Spreading Fake News*, M.I.T. TECH. R. (Aug. 7, 2017), <https://www.technologyreview.com/s/608561/first-evidence-that-social-bots-play-a-major-role-in-spreading-fake-news/> (last visited Nov. 5, 2017).

¹¹⁴ Shao’s study notes, “targeting these influential users is key. Humans can easily be fooled by automated accounts and can unwittingly seed the spread of fake news (some humans do this wittingly, of course).” *Id.*

¹¹⁵ Jacqueline Thomsen, *Clinton Questions Trump’s ‘Fake’ Twitter Followers*, HILL (May 31, 2017), <http://thehill.com/homenews/administration/335861-nearly-half-of-trumps-twitter-followers-are-fake-report> (last visited Nov. 5, 2017).

¹¹⁶ Wang, *supra* note 106.

were fake. He claims to have informed his superiors to no avail. Miley states “[w]hen I brought the information to my boss, the response was ‘stay in your lane. That’s not your role.’”¹¹⁷ Miley is not alone in his assessment. Clinton Watts, a former F.B.I. agent overseeing Russian online activity, asserted that while Facebook “has begun cutting out the tumors by deleting false accounts and fighting fake news, Twitter has done little and as a result, bots have only spread since the election.”¹¹⁸

Colin Crowell, Twitter’s Vice President of Public Policy, Government, and Philanthropy, responded to critics of Twitter’s approach to “misinformation” in a blog post in June 2017. Crowell advocated that bots can be a “positive and vital tool” for customer support and public safety and that Twitter “strictly prohibits” the use of bots to manipulate the platform.¹¹⁹ He further asserted, “Twitter’s open and real-time nature is a powerful antidote to the spreading of all types of false information. This is important because we cannot distinguish whether every single Tweet from every person is truthful or not. *We, as a company, should not be the arbiter of truth.*”¹²⁰ Crowell’s comments make clear that Twitter views itself as a conduit for sharing ideas – not a filter of what is appropriate and true.

Amid mounting pressure, Twitter has removed some Russia-linked accounts.¹²¹ Notwithstanding, legislators have decried Twitter’s response. Senator Mark Warner (D-VA) called Twitter’s response “frankly inadequate” and asserted that Twitter has displayed “an

¹¹⁷ *Id.*

¹¹⁸ Shane, *supra* note 111.

¹¹⁹ Colin Crowell, *Our Approach to Bots and Misinformation*, TWITTER BLOG (June 14, 2017), https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html (last visited Nov. 5, 2017).

¹²⁰ *Id.* (emphasis added).

¹²¹ Twitter suspended 201 Russia-linked accounts affiliated with 470 Facebook accounts allegedly used for Russian election propaganda. See Lauren Gambino, *Democrats Rebuke Twitter for ‘Frankly Inadequate’ Response to Russian Meddling*, GUARDIAN (Sept. 27, 2017), <https://www.theguardian.com/technology/2017/sep/28/twitter-congress-russian-election-interference> (last visited Nov. 5, 2017).

enormous lack of understanding.”¹²² Regulatory demand to step up Twitter’s response to fake news and foreign meddling is not likely to let up soon and will force Twitter to confront a crisis that cuts to the core of its identity. Google and Facebook have faced similar legislative rebuke, but as shown below, both platforms have taken more drastic and meaningful approaches to curb misinformation compared to their 140-character counterpart.

f. Google – Origins of censorship on the world’s search giant.

No Internet platform can claim the ubiquity of Google. Since it was founded in 1998, Google has grown far beyond a simple search engine.¹²³ In 2015, Google created a shell corporation, Alphabet, under which the search engine and many of its former subsidiaries reside.¹²⁴ Today, Google owns YouTube, the second most popular website in the world (after Google itself) with one and a half billion users, and has five other products with more than a billion users each.¹²⁵ Google Search processes trillions of queries every year and accounts for more than three-fourths of the market share among search engines.¹²⁶ Growth in search queries and revenue has increased scrutiny of how the company provides users information from both an

¹²² *Id.*

¹²³ *From the Garage to the Googleplex*, GOOGLE, <https://www.google.com/intl/en/about/our-story/> (last visited Nov. 8, 2017).

¹²⁴ Avery Hartmans, *The Company Formerly Known as Google Is Far Bigger Than Most People Realize*, BUS. INSIDER (Apr. 5, 2017), <http://www.businessinsider.com/alphabet-google-company-list-2017-4> (last visited Nov. 8, 2017).

¹²⁵ Lucas Matney, *YouTube Has 1.5 Billion Logged-In Monthly Users Watching a Ton of Mobile Video*, TECHCRUNCH (June 22, 2017), <https://techcrunch.com/2017/06/22/youtube-has-1-5-billion-logged-in-monthly-users-watching-a-ton-of-mobile-video/> (last visited Nov. 8, 2017); *youtube.com Traffic Statistics*, ALEXA (Nov. 7, 2017), <https://www.alexa.com/siteinfo/youtube.com> (last visited Nov. 8, 2017); *Alphabet 2016 Annual Report*, ALPHABET (Jan. 26, 2017), https://abc.xyz/investor/pdf/2016_google_annual_report.pdf (last visited Nov. 8, 2017).

¹²⁶ Danny Sullivan, *Google Now Handles at Least 2 Trillion Searches Per Year*, SEARCH ENGINE LAND (May 24, 2016), <https://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247> (last visited Nov. 8, 2017); Robert Allen, *Search Engine Statistics 2017*, SMART INSIGHTS (Apr. 13, 2017), <https://www.smartinsights.com/search-engine-marketing/search-engine-statistics/> (last visited Nov. 8, 2017).

antitrust and factual-accuracy perspective.¹²⁷ This paper focuses on the Google products most implicated by fake news – Google Search (including Google News) and YouTube.

Before Google became a \$700 billion company, it operated under the mantra “[d]on’t be evil.”¹²⁸ When the company restructured in 2015, it adopted a new slogan, “[d]o the right thing.”¹²⁹ This subtle shift indicates the responsibility the company carries as the gatekeeper to the world’s information. From the beginning, Google’s mission has been to “[o]rganize the world’s information and make it universally accessible and useful.”¹³⁰ Behind this organization is a purportedly “neutral algorithm” that effectively determines every fact, story, and image that surfaces with a user’s query.¹³¹ The vulnerabilities and shortfalls of this algorithm have significant implications on fake news, which are examined in this paper.

Democratizing access to the world’s information did not come without repercussions. Google’s eighteen years long information liberation has inevitably led to cultural conflicts and local law showdowns. Content moderation on Google was born from the understanding that Google would have to apply some measure of censorship, whether locally or globally, to lure advertising dollars and prevent regulatory suspension.¹³²

¹²⁷ Foo Yun Chee, *EU Fines Google Record \$2.7 Billion in First Antitrust Case*, REUTERS (June 27, 2017), <https://www.reuters.com/article/us-eu-google-antitrust/eu-fines-google-record-2-7-billion-in-first-antitrust-case-idUSKBN191108> (last visited Nov. 8, 2017); Janko Roettgers, *Google Has Punished 340 Fake News Sites Since November*, VARIETY (Jan. 25, 2017), <http://variety.com/2017/digital/news/google-adsense-fake-news-1201969512/> (last visited Nov. 8, 2017).

¹²⁸ David Mayer, *Why Google Was Smart to Drop Its Don’t Be Evil Motto*, FAST COMPANY (Feb. 9, 2016), <https://www.fastcompany.com/3056389/why-google-was-smart-to-drop-its-dont-be-evil-motto> (last visited Nov. 9, 2017).

¹²⁹ *Id.*

¹³⁰ *Our Company*, GOOGLE, <https://www.google.com/intl/en/about/our-company/> (last visited Nov. 9, 2017).

¹³¹ Jeffery Rosen, *Google’s Gatekeepers*, N.Y. TIMES MAG. (Nov. 28, 2008), http://www.nytimes.com/2008/11/30/magazine/30google-t.html?_r=1&partner=rss&emc=rss&pagewanted=all (last visited Nov. 9, 2017).

¹³² Speaking critically of Google’s approach, Professor Lawrence Lessing states, “[i]f your whole game is to increase market share . . . it’s hard to . . . gather data in ways that don’t raise privacy concerns or in ways that might help repressive governments to block controversial content.” *Id.*

Like Twitter, Google has developed a robust geo-blocking regime to filter content based on a user's IP address. In 2008, if a user logged into Google.fr (Google's search domain for France) and searched for Holocaust-denial websites, that user would come up empty-handed.¹³³ This is because Holocaust denial is illegal in France.¹³⁴ However, if that user logged onto Google's global domain, Google.com, and typed in that same query, they would have a smorgasbord of Holocaust-denial websites to peruse.¹³⁵ Since 2008, Google has modified its approach to some offensive searches. In late 2016, Google adjusted its algorithm to demote websites that promote Holocaust denial.¹³⁶ Today, if a user types "did the holo" into Google, the website's first autocomplete suggestion is not "did the Holocaust happen."¹³⁷ This change came after fierce criticism and illustrates Google's larger problem – how to demote obscene, offensive, and salacious content while still providing access to billions of irreverent websites.

Unlike Google, YouTube not only surfaces search results, but hosts the actual content (videos) users seek.¹³⁸ Since inception, YouTube has been a flashpoint for content moderation. The nature of the site, which allows anyone in the world to upload videos, has demanded tighter control and wider corporate discretion of allowable content compared to YouTube's bigger brother, Google.¹³⁹ Videos of Iranian revolutionaries, beheadings, pornography, and overt racism have led to a spectrum of responses by YouTube moderators.¹⁴⁰ Plainly illegal content,

¹³³ *Id.*

¹³⁴ *Id.*

¹³⁵ *Id.*

¹³⁶ Julian Chokkattu, *New Google Algorithm Lowers Search Rankings for Holocaust Denial Sites*, DIGITAL TRENDS (Dec. 25, 2016), <https://www.digitaltrends.com/web/google-search-holocaust/> (last visited Nov. 9, 2017).

¹³⁷ *Id.*

¹³⁸ Nicholas Carlson, *This is Where Google Keeps Your Email and YouTube Videos*, BUS. INSIDER (Jan. 27, 2013), <http://www.businessinsider.com/where-google-keeps-your-email-and-youtube-videos-2013-1?op=1/#-fact-theres-someone-on-site-24-hours-a-day-4> (last visited Nov. 9, 2017).

¹³⁹ Buni & Chemaly, *supra* note 48.

¹⁴⁰ *Id.*

like child pornography, is removed without debate.¹⁴¹ The question is where to draw the line with content that does not explicitly violate the law.

In its nascent days, YouTube provided a small team of moderators a six-page booklet which spelled out corporate policy on permissible content.¹⁴² Buni and Chemaly remark on YouTube’s early policies stating, “[n]o booklet could ever be complete, no policy definitive. This small team of improvisers had yet to grasp that they were helping to develop new global standards for free speech.”¹⁴³ YouTube has grown up in the past decade. The company does not publically disclose how many content moderators work at YouTube, but experts estimate that there are well north of 150,000 content moderators working across all Internet platforms worldwide.¹⁴⁴ It is safe to assume the second most popular website in the world employs a significant portion of this censorship army.

YouTube’s moderation capacity extends beyond paid moderators. In 2016, the video platform launched a “YouTube Heroes” program to provide volunteer YouTube users access to a dashboard where they could flag abusive or inappropriate content.¹⁴⁵ YouTube Heroes did not have the ability to actually remove content like Reddit’s volunteer moderators – that power remained in the hands of paid moderators.¹⁴⁶ However, YouTube’s move was an acknowledgement of the company’s need to address moderation at scale. After a shaky rollout, YouTube Heroes is now known as “YouTube Contributors” and there is a waiting list to join the

¹⁴¹ David Post, a scholar with the Center for Democracy and Technology, commented on CDA § 230 noting, “[t]hese 26 words [§ 230] put free speech decisions into private hands, effectively immunizing platforms from legal liability for all content that does not violate federal law, such as child pornography.” *Id.*

¹⁴² *Id.*

¹⁴³ *Id.*

¹⁴⁴ Sarah Roberts et. al, *Scrubbing the Net: The Content Moderators*, ALJAZEERA (May 28, 2017), <http://www.aljazeera.com/programmes/listeningpost/2017/05/scrubbing-net-content-moderators-170527124251892.html> (last visited Nov. 9, 2017).

¹⁴⁵ Jacob Kastrenakes, *YouTube’s New Plan to Deal with Awful Comments: Have Commenters Help Moderate*, VERGE (Sept. 21, 2016), <https://www.theverge.com/2016/9/21/13001520/youtube-heroes-comment-moderation-program-announced> (last visited Nov. 9, 2017).

¹⁴⁶ *Id.*

volunteer community.¹⁴⁷ YouTube Contributors can flag videos and comments that violate YouTube’s Community Guidelines and that content is escalated to YouTube’s paid moderators for final decision authority.¹⁴⁸

Google and YouTube’s content moderation policies are a reaction to the success of the platforms and demonstrate a tepid foray into online censorship. Shown below, these policies, which were often drafted in response to bad press or legislative prodding, were ill equipped to handle the onslaught of fake news born from the slanderous 2016 presidential campaign.

g. Autocomplete, ads, & algorithms – How fake news pervades Google & YouTube.

Results that appear at the top of a Google search are assumed to be more credible than results further down the page rank.¹⁴⁹ Google’s algorithm employs more than 200 factors to produce ranked results for billions of searches.¹⁵⁰ The search engine estimates that 0.25% of daily queries return offensive or misleading content which amounts to tens of thousands of inaccurate or repugnant results every day.¹⁵¹ That begs the question: How does fake news trend to the top of the results with all of Google’s algorithmic inputs?

One of Google Search’s most problematic features is autocomplete. When a user begins to type a query into the search bar, Google offers suggestions to complete the query. For example, type “Barack Obama is” into Google and the first three suggestions are: “Barack Obama is a sauce,” “Barack Obama is your new bicycle,” and “Barack Obama is Osama bin

¹⁴⁷ *Get Involved with YouTube Contributors*, YOUTUBE, <https://support.google.com/youtube/answer/7124236?hl=en> (last visited Nov. 9, 2017).

¹⁴⁸ *Policies and Safety*, YOUTUBE, <https://www.youtube.com/yt/about/policies/#community-guidelines> (last visited Nov. 9, 2017).

¹⁴⁹ PEN America, *supra* note 6, at 40 (“The very fact that they appear at the top of the results, and are what the algorithm has determined are the best answers to consumers’ queries, imbues them with credibility whether it is deserved or not.”).

¹⁵⁰ *Id.* at 41.

¹⁵¹ *Id.* at 41.

Laden.”¹⁵² PEN America notes that, while useful, autocomplete “can also point users in some troubling directions.”¹⁵³ Autocomplete suggestions are based on what other users search – a tempting way to manipulate search queries and the respective results for those with mal-intent.¹⁵⁴ In response, some organizations have tried to “flood” search results with credible, accurate information, however Google’s algorithm is the “ultimate determinant” of what autocomplete phrases are suggested and the corresponding results.¹⁵⁵

Another source of fake news on Google hits the company where it hurts – advertising.¹⁵⁶ AdWords is Google’s advertising platform through which third parties buy ad space for advertisements that are shown on Google search results and partner websites.¹⁵⁷ Google raked in more than \$24 billion in advertising revenue in the third quarter of 2017 alone.¹⁵⁸ Google’s cash haul makes any compromise of its advertising platform particularly damning. A *New York Times* investigation discovered fake publishers were buying ad space through Google’s AdWords system and posting misleading headlines including: “Melania Says Farewell to D.C. – She’s Leaving the Whitehouse” and “Why Melania Isn’t Staying At The White House. First Lady Tells All About Being In New York.”¹⁵⁹ The ads even included a “featured in” footer claiming the news was featured in *People*, *Time*, *Vanity Fair*, Yahoo, and others. The real kicker is the sites

¹⁵² Autocomplete suggestions based on Google search performed Dec. 12, 2017.

¹⁵³ PEN America, *supra* note 6, at 42.

¹⁵⁴ *Id.*

¹⁵⁵ The Yaqeen Institute for Islamic Research published reports online to provide accurate descriptions of Islam and its teachings. Despite the Institute’s efforts, the rank and manner in which these reports surface on Google is ultimately up to Google. *Id.*

¹⁵⁶ Daisuke Wakabayashi & Linda Qui, *Google Serves Fake News Ads in an Unlikely Place: Fact-Checking Sites*, N.Y. TIMES (Oct. 17, 2017), https://www.nytimes.com/2017/10/17/technology/google-fake-ads-fact-check.html?_r=0 (last visited Nov. 10, 2017).

¹⁵⁷ *The Difference Between AdWords and AdSense*, GOOGLE, <https://support.google.com/adsense/answer/76231?hl=en> (last visited Nov. 10, 2017).

¹⁵⁸ *Alphabet Announces Third Quarter 2017 Results*, ALPHABET (Oct. 26, 2017), https://abc.xyz/investor/pdf/2017Q3_alphabet_earnings_release.pdf (last visited Nov. 10, 2017).

¹⁵⁹ Wakabayashi & Qui, *supra* note 156.

these fake ads were placed on: PolitiFact and Snopes – two of the most prominent fact-checking websites on the Internet.¹⁶⁰

Google conducted an internal investigation after the 2016 election and found that \$53,000 worth of election-related ads were purchased from IP addresses in Russia and \$4,700 worth of ads were purchased from accounts directly associated with the Russian government.¹⁶¹ The ads covered the “political spectrum.”¹⁶² Some of the ads questioned whether Barack Obama should resign, while other ads promoted a documentary about Donald Trump’s cavalier efforts to build a golf course in Scotland.¹⁶³ The ads also ran on YouTube where Russia’s state-linked news agency, RT, has a strong presence.¹⁶⁴

Beyond autocomplete and deceptive ads on Google and YouTube, Google search results themselves have been criticized for promoting fake news. In the aftermath of the Las Vegas mass shooting on October 1, 2017, Google (and Facebook) displayed posts from “alt-right” message boards that inaccurately named the suspect and motive.¹⁶⁵ Similarly, in the wake of another mass shooting at a Texas church on November 5, 2017, Google prominently displayed false information about the gunman.¹⁶⁶ Google displayed popular Tweets in its search results carousel at the top of the page that claimed the gunman was a left-wing Muslim extremist.

¹⁶⁰ *Id.*

¹⁶¹ Daisuke Wakabayashi, *Google Finds Accounts Connected to Russia Bought Election Ads*, N.Y. TIMES (Oct. 9, 2017), <https://www.nytimes.com/2017/10/09/technology/google-russian-ads.html> (last visited Nov. 9, 2017).

¹⁶² *Id.*

¹⁶³ *Id.*

¹⁶⁴ Elizabeth Dwoskin, *Google Uncovers Russian-bought Ads on YouTube, Gmail and Other Platforms*, WASH. POST (Oct. 9, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/10/09/google-uncovers-russian-bought-ads-on-youtube-gmail-and-other-platforms/?utm_term=.c9dfb99241df (last visited Nov. 9, 2017).

¹⁶⁵ Kathleen Chaykowski, *Facebook And Google Still Have A 'Fake News' Problem, Las Vegas Shooting Reveals*, FORBES (Oct. 2, 2017), <https://www.forbes.com/sites/kathleenchaykowski/2017/10/02/facebook-and-google-still-have-a-fake-news-problem-las-vegas-shooting-reveals/#290f77347138> (last visited Nov. 10, 2017).

¹⁶⁶ *Google Promoted Texas Gunman Fake Tweets*, BBC (Nov. 8, 2017), <http://www.bbc.com/news/technology-41915065> (last visited Nov. 10, 2017).

Interestingly, these false reports began on Twitter, but quickly spread to Google – a sign of the interwoven nature of fake news throughout Internet platforms.¹⁶⁷

These accounts of fake news on Google and YouTube, deceptive ads, and vulnerable autocomplete suggestions are the tip of the iceberg. As noted, Google serves billions of search results every day and untold thousands across the globe unwittingly read false news reports and advertisements. Google has taken a series of measures to cramp this Goliath problem, an issue that threatens to undermine the credibility of the platform altogether. Google and YouTube’s response to this fake news cancer is outlined below.

h. Google strikes back – The search engine’s wide-ranging efforts to curtail deceptive content and its shortcomings.

In April 2017, Google’s Vice President of Engineering, Ben Gomes, announced a series of actions to diminish the frequency of fake news on Google Search.¹⁶⁸ First, Google updated its “Search Rater Quality Guidelines” to provide better examples of “low-quality” webpages for human reviewers to flag misleading and offensive content. Second, Google “adjusted [its] signals to help surface more authoritative pages and demote low-quality content.” These “signals” are the algorithmic inputs that determine which content surfaces after a user query and the content page rank. Third, Google added direct feedback features for users to report inappropriate autocomplete predictions to “speed up” content and suggested search removals. Finally, Google published more detailed explanations of its content policies on its help center to promote transparency.¹⁶⁹

¹⁶⁷ *Id.*

¹⁶⁸ Ben Gomes, *Our Latest Quality Improvements for Search*, GOOGLE (Apr. 25, 2017), <https://www.blog.google/products/search/our-latest-quality-improvements-search/> (last visited Nov. 12, 2017).

¹⁶⁹ *Id.*

Google ran a series of tests under a program called “Project Owl” prior to instituting the aforementioned changes.¹⁷⁰ The project team conducted 150,000 search experiments, which led to 1,600 adjustments in the code and internal practices. As part of the project, Google disclosed that it employs more than 10,000 “raters” who do not have the authority to determine exact search rankings, but impact how the algorithm responds to future queries. The roles of these raters and secrecy of Google’s algorithm has led some observers to chide the platform for its “opacity.”¹⁷¹ David North, the Editorial Chairman of the World Socialist Web Site, credits Project Owl with decreasing traffic to his site by 70%. Mr. North claims the drop off is due to Google directing users to more mainstream news outlets and away from websites like his (which he claims does not publish patently false stories).¹⁷²

In addition to Google Search, Google News has undergone significant changes to combat fake news. In October 2016, Google rolled out “Fact Check” tags to enable publishers and fact-check organizations to identify and label fact-checked articles.¹⁷³ Google imposed criteria for fact check organizations, like Snopes.com, to receive the “Fact Check” demarcation on news articles. Verified fact-checked articles must include “straightforward claims of fact so that readers can easily understand what was researched in the fact check. The fact checkers’ sources and methods must also be disclosed.”¹⁷⁴ Critically, Google does not fact check articles itself – this work is reserved for non-partisan groups like Snopes.com and PolitiFact. In addition to the “Fact Check” tag, Google redesigned its Google News webpage to provide users related news

¹⁷⁰ Daisuke Wakabayashi, *As Google Fights Fake News, Voices on the Margins Raise Alarm*, N.Y. TIMES (Sept. 26, 2017), <https://www.nytimes.com/2017/09/26/technology/google-search-bias-claims.html> (last visited Nov. 12, 2017).

¹⁷¹ *Id.*

¹⁷² *Id.*

¹⁷³ Justin Kosslyn & Cong Yu, *Fact Check Now Available in Google Search and News Around the World*, GOOGLE (Apr. 7, 2017), <https://www.blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/> (last visited Nov. 12, 2017).

¹⁷⁴ PEN America, *supra* note 6, at 46.

sites and articles from a variety of publishers.¹⁷⁵ This change was made to arm users with more sources and context for news stories and to break the “filter bubble” that promotes information echo-chambers.¹⁷⁶

Google has also targeted fake advertisements – at least to a limited extent. The platform updated its policies prohibiting misleading ads in November 2016.¹⁷⁷ A spokeswoman for Google told the *New York Times*, “we will restrict ad serving on pages that misrepresent, misstate or conceal information about the publisher, the publisher’s content or the primary purpose of the web property.”¹⁷⁸ The search engine banned or suspended the AdSense accounts of hundreds of fake news websites in the months after the 2016 U.S. election.¹⁷⁹ However, the results of Google’s fake advertising curtailment efforts, as with its algorithmic adjustments, are not fully known.¹⁸⁰

Google and YouTube have looked beyond algorithms, ads, policies, and tags to fight deceptive content. YouTube launched an “Internet Citizens” initiative in the United Kingdom to teach teenagers how to spot fake news on the platform.¹⁸¹ Similarly, Google’s Canadian subsidiary donated \$500,000 to News Wise, a news literacy program aimed at elementary and

¹⁷⁵ *Id.*

¹⁷⁶ *Id.*

¹⁷⁷ Nick Wingfield et al., *Google and Facebook Take Aim at Fake News Sites*, N.Y. TIMES (Nov. 15, 2017), <https://www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html> (last visited Nov. 12, 2017).

¹⁷⁸ *Id.*

¹⁷⁹ Google permanently banned 200 sites from its advertising services and suspended an additional 140 accounts. See Roettgers, *supra* note 127.

¹⁸⁰ PEN America writes that “Google should consider how it can offer more information to researchers and the public that enhances understanding of how its online searches are being conducted and how to process the information they produce.” PEN America, *supra* note 6, at 48.

¹⁸¹ Jon Fingas, *YouTube Will Fight Fake News by Offering Workshops to Teen*, ENGADGET (Apr. 23, 2017), <https://www.engadget.com/2017/04/23/youtube-fake-news-teen-workshops/> (last visited Nov. 12, 2017).

high school students.¹⁸² The workshops for both programs are designed to encourage teenagers to escape “social bubbles” and take action against ‘bogus’ content.¹⁸³

Observers have acknowledged Google’s willingness to make changes, but they assert that the platform needs to take a more proactive approach.¹⁸⁴ It is a delicate balancing act. Commentators have cautioned Google, along with other platforms, to focus attention strictly on the “purveyors of demonstrably false information being presented as news with the intention of deceiving the public.”¹⁸⁵ Casting too wide of a net may hamper legitimate public discourse. On the other hand, advocates say that Google should expand its fact checking efforts to a larger range of news results and partner organizations.¹⁸⁶ Despite Google’s efforts, the vast majority of articles on Google News and Search today are not fact-checked.¹⁸⁷ It is also worth noting that Google is part of a much larger online advertising ecosystem. Cutting off fake news publishers from one website or one search engine has a limited effect if those publishers can jump to other platforms.¹⁸⁸ That said, Google is the largest financial player in the system. Google’s position makes the platform the strongest catalyst for change and explains regulators’ penchant for tighter control and increased transparency.

¹⁸² PEN America, *supra* note 6, at 44.

¹⁸³ Fingas, *supra* note 181.

¹⁸⁴ PEN America, *supra* note 6, at 47.

¹⁸⁵ *Id.*

¹⁸⁶ PEN America writes, “Google should explore more specific partnerships with dedicated fact-checking organizations to expand upon these efforts, including financial support to allow them to do so.” *Id.* at 48.

¹⁸⁷ *Id.*

¹⁸⁸ *Id.*

i. Be my friend? Facebook’s rise and the dawn of censorship on the ‘Social Network.’

Fake news and Facebook are two peas in a pod. This sentiment, now widely-held, has not always prevailed.¹⁸⁹ A brief review of the company’s meteoric rise explains how Facebook became synonymous with fake news. The social networking site, born out of a Harvard dorm room in 2004 as the brainchild of founder Mark Zuckerberg, now dominates all other social media platforms.¹⁹⁰ Facebook has more than two billion active monthly users, six times that of Twitter, and two-thirds of its users (1.3 billion) log in daily.¹⁹¹ Beyond capturing more than one quarter of the world’s population, Facebook hauls in advertising cash hand over fist. Facebook reported \$10.1 billion in advertising revenue in the third quarter of 2017, a staggering 49% year-over-year increase.¹⁹² Facebook’s ascension to Internet domination has been fueled by CEO Mark Zuckerberg’s philosophy of “move fast and break things,” an ideology that explains the fake news predicament the social network faces today.¹⁹³

Like YouTube and Twitter, Facebook’s incursion into content control was haphazard and ill-defined at first. Dave Willner joined Facebook’s content team in 2008 at a time when the company’s content removal policies consisted of a one-page primer.¹⁹⁴ Willner recalls the list of

¹⁸⁹ PEN America observes, “[e]xperts believe that Facebook, more than any other social media platform, has facilitated the spread of fraudulent news because of its vast number of users and the many mechanisms that it offers for sharing information quickly.” *Id.* at 29-30.

¹⁹⁰ Matt Weinberger, *33 Photos of Facebook’s Rise from a Harvard Dorm Room to World Domination*, BUS. INSIDER (Sept. 7, 2017), <http://www.businessinsider.com/facebook-history-photos-2016-9/#whatever-happens-next-the-company-remains-driven-by-zuckerbergs-mission-to-connect-everybody-in-the-world-31> (last visited Nov. 14, 2017).

¹⁹¹ Josh Constine, *Facebook Now Has Two Billion Monthly Users ... And Responsibility*, TECHCRUNCH (July 27, 2017), <https://techcrunch.com/2017/06/27/facebook-2-billion-users/> (last visited Nov. 14, 2017).

¹⁹² *Facebook Reports Third Quarter 2017 Results*, FACEBOOK INVESTOR RELATIONS (Nov. 1, 2017), <https://investor.fb.com/investor-news/press-release-details/2017/Facebook-Reports-third-Quarter-2017-Results/default.aspx> (last visited Nov. 14, 2017).

¹⁹³ Weinberger, *supra* note 190.

¹⁹⁴ Julia Angwin & Hannes Grassegger, *Facebook’s Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children*, PROPUBLICA (June 28, 2017), <https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms> (last visited Nov. 14, 2017).

banned content included nudity and images of Hitler. For all other content, Facebook advised moderators: “Take down anything else that makes you feel uncomfortable.”¹⁹⁵ At first, Willner took it upon himself to draft a new set of rules to define content that violated Facebook’s terms of use. His rules were modeled after university anti-harassment policies, but it was soon apparent that these vague policies could not be implemented consistently across the globe.¹⁹⁶

In 2009, Facebook hired an outside firm to draft a 15,000-word “all encompassing” rulebook for live content moderators.¹⁹⁷ The rules, which were subsequently leaked, drew distinctions between “ordinary people” and “public figures” and advised moderators to “review only the reported content, not unreported adjacent content.”¹⁹⁸ Most importantly, the content policy rules and subsequent revisions led to Facebook’s decision to “ban attacks on groups, but not on institutions.”¹⁹⁹ Thus, an attack on the Pope does not violate Facebook’s policies (as long as it is not a direct physical threat), while an attack on Catholics as a whole does.²⁰⁰

Since 2009, Facebook has had to address everything from videos of suicide and murder to images of sexual assault.²⁰¹ To support Facebook’s endeavor to cleanse the platform of objectionable content, Facebook mobilized 10,000 employees to monitor the social network’s safety and security generally – a number set to double by the end of 2018.²⁰² Content moderators work around the clock and remove fake and abusive accounts. Facebook’s treatment of some pages and profiles has led to state-initiated countrywide shutdowns, including in Pakistan and

¹⁹⁵ *Id.*

¹⁹⁶ Rosen, *supra* note 89.

¹⁹⁷ Buni & Chemaly, *supra* note 48.

¹⁹⁸ *Id.*

¹⁹⁹ Rosen, *supra* note 89.

²⁰⁰ *Id.*

²⁰¹ Olivia Solon, *Facebook Is Hiring Moderators. But Is the Job too Gruesome to Handle?*, GUARDIAN (May 4, 2017), <https://www.theguardian.com/technology/2017/may/04/facebook-content-moderators-ptsd-psychological-dangers> (last visited Nov. 14, 2017).

²⁰² Anita Balakrishnan, *Facebook Pledges to Double its 10,000-Person Safety and Security Staff by End of 2018*, CNBC (Nov. 1, 2017), <https://www.cnbc.com/2017/10/31/facebook-senate-testimony-doubling-security-group-to-20000-in-2018.html> (last visited Nov. 14, 2017).

Bangladesh in 2010 after the site refused to remove pages for an event called “Everybody Draw Muhammad Day.”²⁰³

Like Twitter and Google, Facebook blocks content in specific countries while allowing the content to remain on its global platform. Facebook explains, “[g]overnments sometimes ask companies like Facebook to restrict access to content that they believe violates local law. If, after careful legal review, we find that the content is illegal under local law, then we make it unavailable only in the relevant country or territory.”²⁰⁴ In 2016, Facebook geo-blocked content (removed in a specific country) in 16,500 instances.²⁰⁵ France, India, Brazil, and Germany had the most requests for content removal respectively.²⁰⁶ Even with Facebook’s robust country-specific content removal regime, Facebook users (and the platform) do not operate with immunity worldwide. In Thailand, at least 90 people were arrested after the country’s 2014 military coup for violating the country’s lèse-majesté laws (defaming the royal family). Most of these arrests were spurred posts on social media, including Facebook and Twitter.²⁰⁷

Despite the impressive size of Facebook’s operations and now expansive internal content policies, a chorus of regulators and users alike proselytize that Facebook’s efforts are inadequate. Facebook has invested significant human and capital resources to develop and enforce self-created content policies, however there is more to be done. Facebook’s deficiencies are especially apparent in the context of viral fake news surveyed below.

²⁰³ Miguel Helft, *Facebook Wrestles with Free Speech and Civility*, N.Y. TIMES (Dec. 12, 2010), <http://www.nytimes.com/2010/12/13/technology/13facebook.html> (last visited Nov. 14, 2017).

²⁰⁴ *Government Requests Report, FAQs*, FACEBOOK, <https://govtrequests.facebook.com/faq/> (last visited Nov. 14, 2017).

²⁰⁵ *Government Requests Report*, FACEBOOK, <https://govtrequests.facebook.com/> (last visited Nov. 14, 2017).

²⁰⁶ *Id.*

²⁰⁷ Christina Larson, *Facebook Can’t Cope with the World It’s Created*, FOREIGN POL’Y (Nov. 7, 2017), <http://foreignpolicy.com/2017/11/07/facebook-cant-cope-with-the-world-its-created/> (last visited Nov. 14, 2017).

j. Fake news and filter bubbles – why fraudulent content circulates and sticks on Facebook.

In the days after the 2016 U.S. election, critics, who claimed Facebook’s platform helped Donald Trump win the election by inadequately addressing fraudulent news, castigated the social media site.²⁰⁸ CEO Mark Zuckerberg responded, “[p]ersonally I think the idea that fake news on Facebook, which is a very small amount of the content, influenced the election in any way — I think is a pretty crazy idea.”²⁰⁹ His comments were not well received by many and he has since retracted the statement.²¹⁰ Yet, setting aside the question of electoral influence momentarily, the abundance of fake news on Facebook is well documented.

A joint New York University and Stanford University study examined three months of debunked news stories from fact checking websites prior to the November 2016 U.S. election. The study identified 156 fake news stories that were shared 38 million times on Facebook – more than three-fourths of which were pro-Trump or anti-Hillary Clinton.²¹¹ Another media study found that the top twenty fraudulent news stories, from August 2016 until the election, were shared 8.7 million times on Facebook, while the top twenty mainstream news articles were

²⁰⁸ PEN America, *supra* note 6, at 30.

²⁰⁹ Casey Newton, *Zuckerberg: The Idea that Fake News on Facebook Influenced the Election is ‘Crazy,’* VERGE (Nov. 10, 2016), <https://www.theverge.com/2016/11/10/13594558/mark-zuckerberg-election-fake-news-trump> (last visited Nov. 14, 2017)

²¹⁰ Condliffe writes, “*New York* magazine claims that Trump won because of fake news on Facebook. Recode says that ‘if you’re blaming Facebook for the presidential election results, you’re an idiot. . . . It is undeniably the case that such stories will have shaped some opinions.’” Jamie Condliffe, *Regardless of its Influence on the Election, Facebook Needs to Change*, MIT TECH. R. (Nov. 14, 2016), <https://www.technologyreview.com/s/602851/regardless-of-its-influence-on-the-election-facebook-needs-to-change/> (last visited Nov. 14, 2017); *see also* Kurt Wagner, *Mark Zuckerberg Admits He Should Have Taken Facebook Fake News and the Election More Seriously: ‘Calling that Crazy Was Dismissive and I Regret It,’* RECODE (Sept. 27, 2017), <https://www.recode.net/2017/9/27/16376502/mark-zuckerberg-facebook-donald-trump-fake-news> (last visited Nov. 14, 2017).

²¹¹ Hunt Allcott and Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. ECON. PERSP. 211, 212 (2017), <https://www.aeaweb.org/articles?id=10.1257/jep.31.2.211> (last visited Nov 14, 2017).

shared 7.4 million times in the same time period.²¹² Put simply, fake news outperformed real news in terms of user engagement on Facebook in the months prior to the 2016 election.

The genesis of fraudulent news on Facebook, like Google and YouTube, is partially driven by deceptive advertising. In October 2017, Facebook disclosed 3,000 election-related ads purchased by Russian operatives prior to the 2016 election.²¹³ It is widely known that many of Russia's Facebook (and Google) ads were designed to support then-candidate Trump or undermine Hillary Clinton, but the ad campaigns were more sophisticated than playing two-sided politics. Russian ads sought to "exploit divisive social issues" and target precise demographic groups with the help of Facebook's advertising technology. The Russian ad interference campaign directly reached 11.4 million Facebook users through 80,000 posts. Real Facebook users then shared these posts and pages, which ultimately disseminated the false content to 126 million people.²¹⁴ The influence of Russia's election-meddling ads on actual voter decision-making is difficult to measure, but the penetration of the misinformation campaign is undeniable.²¹⁵

Facebook's News Feed, like Google Search's autocomplete feature, is driven in part by third-party user interest and engagement in a topic (or post). The News Feed algorithm considers how many "friends" have engaged in a post, the post's author, recency of the content, and type of

²¹² Craig Silverman, *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook*, BUZZFEED NEWS (Nov. 16, 2016), https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.wyQRmvZZR#.av6Pa3YYP (last visited Nov. 14, 2017).

²¹³ Leslie Shapiro, *Anatomy of a Russian Facebook Ad*, WASH. POST (Nov. 1, 2017), https://www.washingtonpost.com/graphics/2017/business/russian-ads-facebook-anatomy/?utm_term=.417ac32a0092 (last visited Nov. 14, 2017).

²¹⁴ *Id.*

²¹⁵ Allcott & Gentzkow conclude, "the new evidence we present clarifies the level of overall exposure to fake news, and it can give some sense of how persuasive fake news would need to have been to have been pivotal. We estimate that the average US adult read and remembered on the order of one or perhaps several fake news articles during the election period, with higher exposure to pro-Trump articles than pro-Clinton articles. How much this affected the election results depends on the effectiveness of fake news exposure in changing the way people vote." Allcott & Gentzkow, *supra* note 211, at 232.

post (status update, video, webpage link, etc.).²¹⁶ Some technologists note that this algorithm is anything but neutral. Zeynep Tufekci, a professor with the School of Information and Library Science at the University of North Carolina, writes that Facebook’s “goal is to maximize the amount of engagement you have with the site and keep the site ad-friendly.”²¹⁷ Tufekci continues, “[t]he newsfeed algorithm also values comments and sharing. All this suits content designed to generate either a sense of oversize delight or righteous outrage and go viral, hoaxes and conspiracies as well as baby pictures, happy announcements (that can be liked) and important news and discussions.”²¹⁸

Every Facebook user has a unique experience on the platform. This experience is based on the user’s previous engagement on posts, page likes, and clicks. Facebook’s News Feed algorithms tailor content to show posts and stories that a user is prone to like.²¹⁹ Commentators label this ‘unique experience’ as a “filter bubble” or echo chamber in which users only see the posts and opinions they want to see – the agreeable content.²²⁰ This poses a thorny challenge for a platform that wants to satiate users’ whims and desires to feed its advertising appetite, while serving as a legitimate source of news. Kartik Hosanagar, a professor at The University of Pennsylvania’s Wharton School, studied how Facebook users have been segregated into like-minded cohorts. Hosanagar writes, “[a]ll the empirical research to date suggests that the reason is not the use of personalization algorithms per se. Algorithms can easily expose us to diverse perspectives. It is the data being fed to the personalization algorithms and the actions we take as

²¹⁶ Josh Constine, *How Facebook Newsfeed Works*, TECHCRUNCH (Sept. 6, 2016), <https://techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed/> (last visited Nov. 14, 2017).

²¹⁷ Zeynep Tufekci, *The Real Bias Built In at Facebook*, N.Y. TIMES (May 19, 2016), <https://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html> (last visited Nov. 14, 2017).

²¹⁸ *Id.*

²¹⁹ PEN America, *supra* note 6, at 33.

²²⁰ *Id.*

end users.”²²¹ Hosanagar suggests that even if algorithms were truly impartial, the outputs may be anything but neutral if end users feed biased data into the system (e.g. engage in polarizing ideological posts).

In combination, Facebook’s News Feed algorithm and financial prerogative provide advertisers a tailored toolset to reach a desired audience. This advertising approach is known as “microtargeting.” PEN America defines microtargeting as “the ability to use large sets of data, including internet users’ personal information, to serve them with ads that are specifically tailored to their personal experiences and characteristics.”²²² The success of Barack Obama’s 2012 campaign, as well as Donald Trump’s 2016 campaign, are in part attributed to effective microtargeting efforts.²²³ Like shilling agencies on Reddit, a little money for a microtargeting campaign on Facebook can go a long way. Facebook reports that Russian operatives spent \$100,000 on election-related advertisements for the 2016 election.²²⁴ Through microtargeting, a relatively small sum of money can be used to test hundreds of Facebook posts and monitor user engagement, ultimately placing the most salient content in front of millions.²²⁵

Facebook’s fake news challenge is unique in that the company has unrivaled demographic and user data that it effectively provides to advertising agencies in exchange for cash. In addition, the nature of user profiles on Facebook lends itself towards information silos, perhaps more so than other platforms that do not so visibly promote the number of “likes” a post receives. These characteristics aside, Facebook faces the same fundamental problems as its

²²¹ Kartik Hosanagar, *Blame the Echo Chamber on Facebook. But Blame Yourself, Too*, WIRED (Nov. 25, 2016), <https://www.wired.com/2016/11/facebook-echo-chamber/> (last visited Nov. 14, 2017).

²²² PEN America, *supra* note 6, at 38.

²²³ *Id.*

²²⁴ Scott Shane, *Purged Facebook Page Tied to the Kremlin Spread Anti-Immigrant Bile*, N.Y. TIMES (Sept. 12, 2017), https://www.nytimes.com/2017/09/12/us/politics/russia-facebook-election.html?hp&action=click&pgtype=Homepage&clickSource=story-heading&module=second-column-region®ion=top-news&WT.nav=top-news&_r=1 (last visited Nov. 14, 2017).

²²⁵ *Id.*

aforementioned peers (Reddit, Twitter, and Google) when it comes to fake news. Like Google, the 2016 election forced Facebook off of the sidelines and into the political theater, leading to a variety of self-imposed measures to abrogate the scourge of fraudulent content on the social network. Facebook's new *modus operandi* is outlined below.

k. Falling short – Facebook's flailing efforts to stop fake news.

Facebook's Community Standards – the policies it imposes on users – only cryptically mention fraudulent content. Under a section entitled “Keeping your account and personal information secure” Facebook has a short “Fraud and spam” subsection. Within these guidelines, Facebook asserts, “[u]sing misleading or inaccurate information to artificially collect likes, followers, or shares is not allowed.”²²⁶ This statement says nothing about how Facebook enforces its policy against misleading content and does not call out fake news by name. Facebook has, however, taken several proactive approaches to minimize false content.

Facebook's Vice President of News Feed, Adam Mosseri, published a blog post on April 6, 2017 entitled “Working to Stop Misinformation and False News.” The post, which has been continually updated, outlines three priorities to stem the tide of fake news: (1) “**disrupting economic incentives** because most false news is financially motivated,” (2) “**building new products** to curb the spread of false news” and (3) “**helping people make more informed decisions** when they encounter false news.”²²⁷

On the first point, economic incentives, Facebook has partnered with third-party fact checking organizations like Snopes.com, Factcheck.org, ABC News, and Politifact, to check the

²²⁶ *Community Standards*, FACEBOOK, <https://www.facebook.com/communitystandards/> (last visited Nov. 16, 2017).

²²⁷ Adam Mosseri, *Working to Stop Misinformation and False News*, FACEBOOK (Apr. 6, 2017), <https://newsroom.fb.com/news/2017/04/working-to-stop-misinformation-and-false-news/> (last visited Nov. 16, 2017) (emphasis in original).

veracity of content that is reported by users.²²⁸ Facebook users can flag content by selecting “I don’t think this should be on Facebook” and then selecting “It’s a fake news story” from the drop-down menu of options.²²⁹ It is not clear at what point a post is escalated for content verification. However, if the content is reviewed by one of the four fact-checking organizations and deemed to be untrue, Facebook adds a flag to the post or page which reads, “[d]isputed by 3rd [p]arties. . . . Before you share this story you might want to know that independent fact-checkers disputed its accuracy.”²³⁰ Facebook then prohibits the promotion of disputed articles through its advertising platform.²³¹ If an advertiser repeatedly shares stories that are verified to be false, Facebook claims, “repeat offenders . . . will no longer be allowed to advertise on Facebook.”²³²

As mentioned, Facebook as also made technological improvements to its products to curb false content. Improvements include better user reporting tools and third-party fact-checkers (outlined above), in addition to adjusting the News Feed ranking algorithm. Mosseri notes that “if reading an article makes people significantly less likely to share it, that may be a sign that a story has misled people in some way.”²³³ Facebook’s algorithm may use this as a signal and demote stories that are not shared by users who actually read the post.²³⁴ Facebook’s algorithm also monitors users that post a lot – 50 or more times per day – and found there is a correlation

²²⁸ Casey Newton, *Facebook Partners with Fact-Checking Organizations to Begin Flagging Fake News*, VERGE (Dec. 15, 2016), <https://www.theverge.com/2016/12/15/13960062/facebook-fact-check-partnerships-fake-news> (last visited Nov. 16, 2017); *see also* Mosseri, *supra* note 227.

²²⁹ Newton, *supra* note 228.

²³⁰ *Id.*

²³¹ *Id.*

²³² Mosseri, *supra* note 227.

²³³ *Id.*

²³⁴ Mosseri writes, “[w]e’re continuing to test this signal and others in News Feed ranking in order to reduce the prevalence of false news content.” *Id.*

between super-posters and the dissemination of false content.²³⁵ Facebook’s efforts are targeted only at individuals, not pages, which can also post deceptive content with great frequency.²³⁶ These algorithmic changes are not exhaustive, but illustrative of Facebook’s endeavor to utilize technology, like machine learning, to nip fake news in the bud.²³⁷

On the final point, Facebook has taken similar measures to Google and YouTube to help users make informed decisions and partnered with civil society organizations to promote news literacy. The Facebook Journalism Project brings together experts and academics to release public service announcements in collaboration with The News Literacy Project (NLP).²³⁸ NLP develops classroom and after-school programming to teach students, “how to discern verified information from spin, opinion and propaganda.”²³⁹ Facebook also joined and contributes to the News Integrity Initiative, a consortium of technology and industry leaders that funds research and workshops.²⁴⁰

Despite Facebook’s three-pronged approach, many observers and journalists claim the platform’s efforts fall far short of adequately addressing the pandemic of false content on the network.²⁴¹ Foremost, some of Facebook’s third-party fact-checking organizations claim that Facebook has refused to provide data on how many stories, which are verified as untrue, are actually tagged as such on Facebook. User experience on the platform suggests that it is “rare” to see false content flags and journalists have raised questions about the independence of the

²³⁵ Kurt Wagner, *Facebook Found a New Way to Identify Spam and False News Articles In Your News Feed*, RECODE (June 30, 2017), <https://www.recode.net/2017/6/30/15896544/facebook-fake-news-feed-algorithm-update-spam> (last visited Nov. 16, 2017).

²³⁶ *Id.*

²³⁷ Mosseri, *supra* note 227.

²³⁸ *Id.*; see also *Our Mission*, NEWS LITERACY PROJECT, <http://www.thenewsliteracyproject.org/about/our-mission> (last visited Nov. 16, 2017).

²³⁹ *Our Mission*, *supra* note 238.

²⁴⁰ Mosseri, *supra* note 227.

²⁴¹ Sam Levin, ‘Way too Little, Way too Late’: Facebook’s Factcheckers Say Effort Is Failing, GUARDIAN (Nov. 13, 2017), <https://www.theguardian.com/technology/2017/nov/13/way-too-little-way-too-late-facebooks-fact-checkers-say-effort-is-failing> (last visited Nov. 16, 2017).

fact-checking organizations, noting that some of the partnerships are paid.²⁴² One anonymous journalist who fact-checks for Facebook remarked, “I don’t feel like it’s working at all. The fake information is still going viral and spreading rapidly.”²⁴³

Fact-checker independence and data transparency are important, but Facebook’s fake news flagging efforts are meaningless if they do not change user behavior and understanding of what is true and false. A Yale University study surveyed 7,500 Facebook users and found that Facebook’s ‘content disputed’ flag has virtually no benefit.²⁴⁴ The study found that only 3.7% of users changed their perception of whether content was inaccurate based on article tags.²⁴⁵ More disconcertingly, some users, notably President Trump supporters under 26 years old, were *more prone* to believe a false headline if it was marked as disputed.²⁴⁶ Gordon Pennycook and David G. Rand, the study’s authors, conclude, “[t]hese results suggest that the currently deployed approaches are not nearly enough to effectively undermine belief in fake news, and new (empirically supported) strategies are needed.”²⁴⁷

To address the continued criticism, Facebook added new features, including “Trust Indicators.”²⁴⁸ Users can select the little “i” icon next to an article and Facebook will show more information about the news organization that published the article. The idea spawned from the Trust Project, an international consortium of news companies, and it is in its nascent stages (only nine media partners currently use the icon). Moreover, commentators note, “[i]t’s doubtful

²⁴² *Id.*

²⁴³ *Id.*

²⁴⁴ Gordon Pennycook & David G. Rand, *Assessing the Effect of ‘Disputed’ Warnings and Source Salience on Perceptions of Fake News Accuracy*, YALE U. (Sept. 16, 2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035384 (last visited Nov. 16, 2017); *see also* Jason Schwartz, *Tagging Fake News on Facebook Doesn’t Work, Study Says*, POLITICO (Sept. 11, 2017), <https://www.politico.com/story/2017/09/11/facebook-fake-news-fact-checks-242567> (last visited Nov. 16, 2017).

²⁴⁵ Pennycook & Rand, *supra* note 244, at 1.

²⁴⁶ *Id.*

²⁴⁷ *Id.*

²⁴⁸ Kerry Flynn, *Facebook’s Latest Step to Fight Fake News Is Another Disappointment*, MASHABLE (Nov. 16, 2017), <http://mashable.com/2017/11/16/facebook-trust-indicators-fake-news-problem/#qEi0m57g0Zqo> (last visited Nov. 16, 2017).

anyone will actually bother clicking that little icon.”²⁴⁹ Again, the efficacy of this product change can only be determined in time and through empirical analysis.

Facebook was admittedly slow to acknowledge its responsibility and account for its culpability in disseminating false content.²⁵⁰ It has made strides towards providing users more information about publishers, indicating the veracity of specific articles, and removing abusive advertisers, however these moves will not uproot fake news altogether. Not even close. If Facebook is serious about fake news, it will need to provide researchers and media partners more data to assess the merits of new approaches. Regardless of Facebook’s intentions, regulators in the U.S. and abroad may force Facebook and its cohorts to take more responsibility for the blight of fake news on the Internet. The mounting regulatory pressure is outlined in the final section of this paper below.

IV. The Fractured Regulatory Response to Fake News in the U.S. and Europe

The popularity and power, amassed by the four Internet platforms surveyed above, have led to greater regulatory scrutiny and proposals to sanction Internet companies for fake news, hate speech, and market dominance. In America, regulations of platforms like Facebook and Google have a high hurdle to overcome due to precedent, political paralysis, and constitutional protections. Nevertheless, U.S. lawmakers have vocalized their displeasure of perceived inadequacies to address the spread of disinformation online and introduced modest legislative proposals to increase transparency. In Europe, a different regulatory regime has developed to corral Internet platforms – state mechanisms to punish technology companies with hefty fines for the failure to remove disinformation. The mixed American and European approaches are reviewed below.

²⁴⁹ *Id.*

²⁵⁰ Wagner, *supra* note 210.

In the U.S., CDA § 230 and the First Amendment have historically provided immense protection to Internet platforms against liability for hosted content.²⁵¹ However, there are signs that this decades-old immunity may not last in its current form forever. In October 2017, U.S. Senators Amy Klobuchar (D-MN) and Mark Warner (D-VA) introduced a bill, the Honest Ads Act, aimed at increasing advertising transparency on Internet platforms with more than 50 million monthly visitors and at least \$500 in yearly political ads.²⁵² The bill is in its nascent stages, yet according to Senator Klobuchar, the chief objective is to let Americans “know who’s paying for the online [political] ads.” Senator John McCain (R-AZ) has joined as a co-sponsor, lending bi-partisan credibility to the legislation.²⁵³ Political ad transparency online is a far cry from liability for false content, however *any* congressional action to mandate Internet platform transparency would be a departure from the extreme latitude the platforms currently enjoy.

The overwhelming market share of companies like Google and Facebook has also led some lobbyists to advocate for greater regulation of the platforms under American anti-trust laws. Sally Hubbard, a senior editor of tech antitrust enforcement at the Capitol Forum (a nonpartisan legislative analysis company), asserts that the prevalence of fake news on Facebook and Google can be attributed in part to their monopoly power.²⁵⁴ Hubbard explains, “because Google and Facebook lack competition, two dominant algorithms control the flow of information. So purveyors of fake news only have to exploit the weaknesses of one algorithm to

²⁵¹ See [Section II. The American Legal Framework for Intermediary Liability and Genesis of Internet Platform Self-Regulation](#); see also PEN America, *supra* note 6, at 25; Klonick, *supra* note 13, at 13-14.

²⁵² Colin Lecher, *Senators Announce New Bill that Would Regulate Online Political Ads*, VERGE (Oct. 19, 2017), <https://www.theverge.com/2017/10/19/16502946/facebook-twitter-russia-honest-ads-act> (last visited Nov. 21, 2017).

²⁵³ *Id.*

²⁵⁴ Sean Illing, *Why ‘Fake News’ Is An Antitrust Problem*, VERGE (Sept. 23, 2017), <https://www.vox.com/technology/2017/9/22/16330008/facebook-google-amazon-monopoly-antitrust-regulation> (last visited Nov. 21, 2017).

potentially deceive hundreds of millions of people.” Hubbard views regulation of major Internet platforms as “inevitable unless we get some more robust competition.”²⁵⁵

Hubbard’s fake news anti-trust angle is novel. Yet it assumes, without empirical support, that a proliferation of Internet platforms could weed out fake news more effectively than a few major players. Hubbard posits, “[i]magine there were five Facebooks and five Googles, all with different algorithms that competed against each other to be the best. A purveyor of fake news perhaps couldn’t have as much of an impact if it had to figure out how to game many more algorithms.”²⁵⁶ As appealing as it may sound, Hubbard’s imaginary fractured market overlooks a key issue: Splintering social media and search engines could arguably lead to more echo chambers and greater insulation from a free market of ideas. One could select the search engine or social network that best suits one’s political leanings, further incentivizing partisan, deceptive information. Hubbard herself admits there is little political appetite in the U.S. to breakup online information monopolies, but suggests public sentiment is shifting.²⁵⁷

Congress and regulatory agencies have avoided talk of regulating content itself, presumably understanding that regulating free speech is a third rail of American politics.²⁵⁸ Some commentators also suggest that “Big Tech’s” influence has gone unbridled for so long that any legislation to substantially regulate Internet platforms is a nonstarter. They assert weak regulatory agencies that have not kept pace with the growth of the Internet, massive lobbying

²⁵⁵ *Id.*

²⁵⁶ *Id.*

²⁵⁷ Hubbard suggests, “I also just think there’s a lack of political will. It’s really only been very recently that people even think the tech platforms are a problem.” *Id.*

²⁵⁸ Free speech advocates assert, “[r]ecognizing fraudulent news as a threat to free expression should not be employed as a justification for broad new government or corporate restrictions on speech, measures whose effects would be far more harmful to free speech.” PEN America, *supra* note 6, at 76.

expenditures, and a dysfunctional Congress, have created a trifecta of inaction.²⁵⁹ Still, observers acknowledge “[t]here are growing efforts in Washington to do *something* about big tech.”²⁶⁰ The question is what.

On the other side of the Atlantic, the regulatory state has not been so sluggish to respond. In June 2017, German parliament passed a law which will require large social media sites to pay up to €50 million in fines for hate speech, criminal material, and fake news not removed within 24 hours or one week of notification.²⁶¹ The law was passed in the run-up to Germany’s September 2017 election due to concerns about the proliferation of fake news and hate speech online, particularly about migrants. The legislation also requires social media companies to publish semi-annual reports about the number of complaints received under the law and the resolution status of those complaints.²⁶² Like the European Court of Justice’s “Right to be Forgotten” ruling, the arbiter of what content should be removed under Germany’s new law is presumably the social media companies themselves – an inevitable complication of state content regulation.²⁶³

Germany is not alone. After the 2016 U.S. election, other European countries worried about the influence foreign governments might exert on democratic processes through disinformation campaigns. In February 2017, a legislative measure was introduced in the Italian

²⁵⁹ See Heather Timmons, *Washington Failed to Regulate Big Tech – And Now It’s About to Discover that It Can’t*, QUARTZ (Oct. 3, 2017), <https://qz.com/1089907/why-washington-dc-is-incapable-of-regulating-the-worlds-tech-giants/> (last visited Nov. 21, 2017).

²⁶⁰ Farhad Manjoo, *Can Washington Stop Big Tech Companies? Don’t Bet on It*, N.Y. TIMES (Oct. 25, 2017), https://www.nytimes.com/2017/10/25/technology/regulating-tech-companies.html?_r=0 (last visited Nov. 21, 2017) (emphasis in original).

²⁶¹ The time allotted for content removal depends on the type of content – “obviously illegal hate speech” must be removed within 24 hours, while other offensive content must be blocked within seven days. *Germany Approves Plans to Fine Social Media Firms Up to €50m*, GUARDIAN (June 30, 2017), <https://www.theguardian.com/media/2017/jun/30/germany-approves-plans-to-fine-social-media-firms-up-to-50m> (last visited Nov. 21, 2017).

²⁶² *Id.*

²⁶³ See Julia Powers & Enrique Chaparro, *How Google Determined Our Right to Be Forgotten*, GUARDIAN (Feb. 18, 2015), <https://www.theguardian.com/technology/2015/feb/18/the-right-be-forgotten-google-search> (last visited Nov. 21, 2017) (“Google has acted as judge, jury and executioner in the wake of Europe’s right to be forgotten ruling.”).

Senate to criminalize and fine the authors and distributors of fake news. Under the proposal, “ISPs must regularly monitor content, paying particular attention to any content that generates a substantial degree of interest among users, in order to assess the reliability and truthfulness of this content.”²⁶⁴ If an ISP fails to promptly remove content, it may be fined up to €5,000.²⁶⁵ The Italian government is also instituting mandatory instruction on identifying fake news at more than 8,000 high schools across the country.²⁶⁶ The move comes amid strong distrust of the country’s political class in the lead-up to nationwide elections in 2018.²⁶⁷

The French President, Emmanuel Macron, also made his intentions well known prior to his election in May 2017. In a television interview before the French vote, Macron said, “I want to stop fake news. . . . It pollutes, it degrades the political debate.”²⁶⁸ Macron further advocated that politicians must “regulate the Internet because today certain players are activists and have a very important role in the campaign.”²⁶⁹ Macron went on to win two-thirds of the vote.²⁷⁰ As of November 2017, Macron has yet to make good on his word.²⁷¹ Notwithstanding, Macron is less than six months into his five-year term, which leaves the possibility of regulatory action open for the foreseeable future.

In the Czech Republic, a government task force “charged with scouring the Internet and social media, fact-checking, then flagging false reports to the public” was created prior to

²⁶⁴ Ernesto Apa & Marco Bassini, *Legislative Proposal on Fake News*, PORTOLANO CAVALLO STUDIO LEGALE & BOCCONI U. (Feb 7, 2017), <http://merlin.obs.coe.int/iris/2017/5/article27.en.html> (last visited Nov. 21, 2017).

²⁶⁵ *Id.*

²⁶⁶ Jason Horowitz, *In Italian Schools, Reading, Writing, and Recognizing Fake News*, N.Y. TIMES (Oct. 18, 2017), <https://www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html> (last visited Nov. 21, 2017).

²⁶⁷ *Id.*

²⁶⁸ John Follain & Helene Fouquet, *France’s Macron Says He’ll Stop Fake News Hurting His Campaign*, BLOOMBERG (May 4, 2017), <https://www.bloomberg.com/news/articles/2017-05-04/france-s-macron-says-he-ll-stop-fake-news-hurting-his-campaign>, (last visited Nov. 21, 2017).

²⁶⁹ *Id.*

²⁷⁰ Angelique Chrisafis, *Emmanuel Macron Vows Unity After Winning French Presidential Election*, GUARDIAN (May 8, 2017), <https://www.theguardian.com/world/2017/may/07/emmanuel-macron-wins-french-presidency-marine-le-pen> (last visited Nov. 22, 2017).

²⁷¹ A Google News search conducted on November 21, 2017 provided no indication that President Macron or the French parliament has introduced measures to sanction Internet providers or curb fake news in another manner.

national elections in October 2017.²⁷² The agency did not remove fake news from the Internet and rarely posted on its official Twitter account created to debunk fake news.²⁷³ The agency's role was not to sanction Internet platforms, but to directly combat false news reports.²⁷⁴ Despite the fifteen-person agency's efforts, it is estimated that 25% of Czechs read false reports on the Internet before the October 2017 elections.²⁷⁵ Finland also tasked more than 100 government officials with studying and understanding the spread of disinformation.²⁷⁶ Like the Czech Republic, the primary aim is not to regulate content per se, but to combat inaccurate information with alternative narratives and strengthen institutional trust.²⁷⁷

The European approach has been fragmented, yet notably more pointed than the American approach. European distinctions between hate speech and “offensive content” (which may include disinformation) and a general lack of research on the impact of fake news in Europe, have slowed a collective response.²⁷⁸ However, the absence of coordinated European fake news regulations is not for lack of political zeal to go after Internet platforms. The European Union (EU) has shown an appetite for American technology companies' profits with

²⁷² Anthony Faiola, *As Cold War Turns to Information War, A New Fake News Police Combats Disinformation*, WASH. POST (Jan. 22, 2017), https://www.washingtonpost.com/world/europe/as-cold-war-turns-to-information-war-a-new-fake-news-police/2017/01/18/9bf49ff6-d80e-11e6-a0e6-d502d6751bc8_story.html?utm_term=.8fe3b24992c3 (last visited Nov. 21, 2017).

²⁷³ Rick Noack, *Czech Elections Show How Difficult It Is to Fix the Fake News Problem*, WASH. POST (Oct. 20, 2017), https://www.washingtonpost.com/news/worldviews/wp/2017/10/20/czech-elections-show-how-difficult-it-is-to-fix-the-fake-news-problem/?utm_term=.8dc0b87b5820 (last visited Nov. 21, 2017).

²⁷⁴ Benedikt Vangeli, the commander of the task force stated, “[w]e just tweet them to the public as false reports. . . . That’s how we fight back. We don’t take them down. We don’t censor.” Faiola, *supra* note 272.

²⁷⁵ Noack, *supra* note 273.

²⁷⁶ Reid Standish, *Why Is Finland Able to Fend Off Putin’s Information War?*, FOREIGN POL’Y (Mar. 1, 2017), <http://foreignpolicy.com/2017/03/01/why-is-finland-able-to-fend-off-putins-information-war/> (last visited Nov. 21, 2017).

²⁷⁷ Jed Willard, a consultant to the Finnish government and Director of the Franklin Delano Roosevelt Center for Global Engagement at Harvard, advocates, “[t]he best way to respond is less by correcting the information, and more about having your own positive narrative and sticking to it.” *Id.*

²⁷⁸ Anya Schiffrin, *How Europe Fights Fake News*, COLUM. JOURNALISM R. (Oct. 26, 2017), <https://www.cjr.org/watchdog/europe-fights-fake-news-facebook-twitter-google.php> (last visited Nov. 21, 2017).

significant antitrust and tax evasion fines in recent years.²⁷⁹ For the time being, it is more likely that individual European nations will follow in Germany's footsteps and require Internet platforms to remove objectionable content and hate speech in a specified time window.²⁸⁰

There are notable distinctions between the American and European regulatory states' approaches. American reverence for free speech, relatively weak regulatory bodies, and notorious legislative inaction, have prevented significant policies to combat disinformation from being offered. European regulators have not faced the same political challenges, but remain fractured in how to curb fake news. This analysis may suggest that Internet platforms can breathe relatively easy for the time being, however the lack of congruous regulation is not a result of political or social acceptance of the status quo. As shown, individual European nations have taken it upon themselves to fill the continent-wide regulatory void and will continue to do so. Additionally, politicians on both sides of the Atlantic have shown greater willingness to call out the failure of Internet platforms to shake the fake news influenza.

V. Conclusion – The User as the Arbiter of Truth

This survey of four of the world's most influential Internet platforms shows that online content moderation is still largely in the hands of the private companies themselves. The scourge of fake news has yet to fundamentally disrupt the immunity Internet platforms have enjoyed in the U.S. for decades, despite widespread distrust of social media.²⁸¹ There are promising signs of change from Google and Facebook, particularly the platforms' emphasis on news literacy

²⁷⁹ See Nick Statt, *Google Appeals Record €2.4 Billion Antitrust Fine Over Manipulated Search Results*, VERGE (Sept. 11, 2017), <https://www.theverge.com/2017/9/11/16291482/google-alphabet-eu-fine-antitrust-appeal> (last visited Nov. 21, 2017); see also Jennifer Rankin, *EU to Find Ways to Make Google, Facebook and Amazon Pay More Tax*, GUARDIAN (Sept. 21, 2017), <https://www.theguardian.com/business/2017/sep/21/tech-firms-tax-eu-turnover-google-amazon-apple> (last visited Nov. 21, 2017).

²⁸⁰ Schiffrin, *supra* note 278.

²⁸¹ Jeff Dunn, *Most Americans Do Not Trust News They Read on Social Media, Survey Says*, BUS. INSIDER (Jan. 25, 2017), <http://www.businessinsider.com/facebook-fake-news-reader-trust-survey-chart-2017-1> (last visited Nov. 22, 2017).

education. Yet at its core, the fake news fight is about who gets to call the shots. Who is the arbiter of truth and how should this power be wielded? Facebook's 10,000 content moderators were not able to prevent the massive disinformation campaign during the 2016 U.S. election, so what would suggest that 20,000 content moderators will make the difference?

All four companies, even Reddit to a limited extent, have formalized standards and community guidelines to address everything from hate speech to fake news. Each company has drawn a line in the sand and each platform has shown their penchant for profits is the ultimate determinant of where that line is drawn. European regulators have recognized this proclivity for profit and are slowly circling the wagons to hold Internet platforms accountable for the lies and bile spread online. Yet short of outright state censorship, fake news will prevail for the foreseeable future. The fractured nature of the World Wide Web and the patchwork of regulations make any change slow going. For now, it is up to the user to decipher fact from fiction.