ANTIDEMOCRATIC EFFECTS OF THE INTERNET & SOCIAL MEDIA: A SURVEY

Taisa Goodnature
Professor Nathaniel A. Persily
LAW 805C: Policy Practicum
Spring Quarter 2016-2017
July 10, 2017

INTRODUCTION

Only five years ago, during the period known as the Arab Spring, social media was widely lauded as a democratizing tool.[1] During and after the 2016 U.S. presidential election, however, attention increasingly shifted to social media's power to undermine democracy.[2] This paper surveys popular and academic literature on, first, the challenges that social media and the Internet, more broadly, may present to democracy and, second, the unique vulnerabilities of several key social media platforms and their approaches to moderation, particularly in the face of increasing public pressure.

I. PROBLEMS

A. ECHO CHAMBERS

A prominent criticism of social media as a source of news is that users tend to selectively engage with news sources that reinforce their beliefs, while skipping over (either deliberately or inadvertently, as a result of social media platforms' algorithms) sources that challenge these beliefs.[3] In 2016, Facebook attracted criticism for the algorithm that it uses to select the stories

---

[1] *See, e.g.*, Heather Brown et al., *The Role of Social Media in the Arab Uprisings*, PEW RES. CTR. (Nov. 28, 2012), http://www.journalism.org/2012/11/28/role-social-media-arab-uprisings ("In covering what some deemed the Facebook or Twitter revolutions, the media focused heavily on young protesters mobilizing in the streets in political opposition, smartphones in hand."); Saleen Kassim, *Twitter Revolution: How the Arab Spring Was Helped by Social Media*, MIC (July 3, 2012), https://mic.com/articles/10642/twitter-revolution-how-the-arab-spring-was-helped-by-social-media#.nAakviCLn ("Through social networking sites, Arab Spring activists have not only gained the power to overthrow powerful dictatorship, but also helped Arab civilians become aware of the underground communities that exist and are made up of their brothers, and others willing to listen to their stories.").

[2] *See, e.g.*, Thomas B. Edsall, *Democracy, Disrupted*, N.Y. TIMES (Mar. 2, 2017), https://www.nytimes.com/2017/03/02/opinion/how-the-internet-threatens-democracy.html?_r=0 ("Even though in one sense President Trump's victory in 2016 fulfilled conventional expectations . . . it also revealed that the internet and its offspring have overridden the traditional American political system of alternating left-right advantage. They are contributing — perhaps irreversibly — to the decay of traditional moral and ethical constraints in American politics."); Vyacheslav W. Polonsky, *Is Social Media Destroying Democracy*, NEWSWEEK (Aug. 5, 2016), http://www.newsweek.com/social-media-destroying-democracy-487483 ("In political philosophy, the very idea of democracy is based on the principal of the general will, which was proposed by Jean-Jacques Rousseau in the 18th century . . . . The internet makes this an almost perpetual problem rather than an occasional obstacle. Only the most passionate, motivated and outspoken people are heard . . . .").

[3] *See, e.g.*, Amanda Hess, *How to Escape Your Political Bubble for a Clearer View*, N.Y. TIMES (Mar. 3, 2017), https://www.nytimes.com/2017/03/03/arts/the-battle-over-your-political-bubble.html?_r=0 ("The filter bubble describes the tendency of social networks like Facebook and Twitter to lock users into personalized feedback loops,

that users see in their News Feeds when they log in to the service.[4] Twitter launched an "algorithmic timeline" in February 2016.[5] And Google filters search results based on a user's location and previous searches and clicks.[6]

Some research has confirmed the echo chamber phenomenon. In 2016, a group of social scientists, including Cass Sunstein, uncovered evidence that "Facebook users are highly polarized," and "[t]heir polarization creates largely closed, mostly non-interacting communities centered on different narratives—i.e. echo chambers."[7] Within this structure, they found, "[t]he spreading of information tends to be confined to communities of like-minded people."[8]

When it comes to the causes of social network polarization, however, some scholars have found that the level of polarization in an individual's social media feed is driven more strongly by the news that user seeks out than by the social media platforms' algorithms. For example, a group of researchers analyzed 10.1 million U.S. Facebook users' engagement with socially shared news, to compare the effect of algorithmic filtering with users' choices about what

---

each with its own news sources, cultural touchstones and political inclinations."); *The Reason Your Feed Became an Echo Chamber—And What to Do About It*, NPR (July 24, 2016), http://www.npr.org/sections/alltechconsidered/2016/07/24/486941582/the-reason-your-feed-became-an-echo-chamber-and-what-to-do-about-it ("[A]lgorithms, like the kind used by Facebook, instead often steer us toward articles that reflect our own ideological preferences, and search results usually echo what we already know and like. As a result, we aren't exposed to other ideas and viewpoints, says Eli Pariser, CEO of Upworthy, a liberal news website. Pariser tells NPR's Elise Hu that as websites get to know our interests better, they also get better at serving up the content that reinforces those interests, while also filtering out those things we generally don't like.").

[4] *See, e.g.*, Farhad Manjoo, *Can Facebook Fix Its Own Worst Bug?*, N.Y. TIMES MAG. (Apr. 25, 2017), https://www.nytimes.com/2017/04/25/magazine/can-facebook-fix-its-own-worst-bug.html ("Trump had benefited from a media environment that is now shaped by Facebook—and, more to the point, shaped by a single Facebook feature, . . . [the] News Feed. . . . Every time you open Facebook, it hunts through the network, collecting every post from every connection . . . . Then it weighs the merits of each post before presenting you with a feed sorted in order of importance: a hyperpersonalized front page designed just for you.").

[5] Will Oremus, *Twitter's New Order*, SLATE (Mar. 5, 2017), http://www.slate.com/articles/technology/cover_story/2017/03/twitter_s_timeline_algorithm_and_its_effect_on_us_explained.html.

[6] Mostafa M. El-Bermawy, *Your Filter Bubble is Destroying Democracy*, WIRED (Nov. 18, 2016), https://www.wired.com/2016/11/filter-bubble-destroying-democracy.

[7] Walter Quattrociocchi et al., Echo Chambers on Facebook 14 (June 13, 2016) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795110.

[8] *Id.*

presented material they chose to engage.[9] Their study found that "individual choices . . . more than algorithms . . . limit exposure to attitude-challenging content in the context of Facebook."[10] Moreover, their work revealed, "[r]ather than people browsing only ideologically aligned news sources or opting out of hard news altogether . . . social media expose individuals to at least some ideologically cross-cutting viewpoints."[11]

A 2015 analysis by researchers at Facebook revealed that users are likely to click on stories that align with their existing beliefs, even when shown links from across the political spectrum.[12] The study found that if news were acquired from a random selection of Facebook users, approximately 45% of news seen by liberals and 40% of news seen by conservatives would be "cross-cutting" (i.e., conservative news for a liberal reader, and vice versa).[13] However, the effect of like-minded circles of Facebook friends reduced the proportion of cross-cutting stories displayed in a user's feed to 22% for liberals and 34% for conservatives.[14] According to the (admittedly self-funded) analysis, "Facebook's algorithm worsens the echo chamber, but not by much."[15] Finally, the proportion of cross-cutting stories on which users actually clicked was only 21% for liberals and 30% for conservatives.[16]

Some studies have found that some audiences are more vulnerable to social media's viewpoint-reinforcing than others. For example, according to a study published in the *Columbia Journalism Review*, news media polarization on social media outlets between April 1, 2015, and

---

[9] Eytan Bakshy et al., *Exposure to Ideologically Diverse News and Opinion on Facebook*, 348 SCIENCE 1130, 1131 (2015).
[10] *Id.* at 1131-32.
[11] *Id.*
[12] 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.
[13] *Id.*
[14] *Id.*
[15] *Id.*
[16] *Id.*

November 8, 2016, was asymmetric, with pro-Trump audiences more likely to devote the majority of their attention to polarized outlets.[17] Once again, the authors noted that their data suggest that social media platforms themselves are not the primary drivers of polarization; if they were, "we would expect to see symmetric patterns on the left and the right."[18] Their study analyzed 1.25 million news stories shared on Twitter and Facebook, to "map the ecosystem of campaign media."[19] The researchers found that "Breitbart became the center of a distinct right-wing media ecosystem, surrounded by Fox News, the Daily Caller, the Gateway Pundit, the Washington Examiner, Infowars, Conservative Treehouse, and Truthfeed."[20] Their analysis also revealed fewer center-right outlets than center-left outlets.[21] While these findings provide evidence that echo chambers exist, they tend to suggest that conservative media organizations have been more effective than their liberal counterparts in exploiting the opportunities that social media algorithms provide.[22]

Finally, in a paper that calls into question the conventional online echo chamber narrative, altogether, professors at Stanford University and Brown University found that between 1996 and 2012, "the growth in political polarization was most significant among older

---

[17] Yochai Benkler et al., *Study: Breitbart-Led Right-Wing Media Ecosystem Altered Broader Media Agenda*, COLUM. JOURNALISM REV. (Mar. 3, 2017), https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php.
[18] *Id.* "Moreover, the fact that these asymmetric patterns of attention were similar on both Twitter and Facebook suggests that human choices and political campaigning, not one company's algorithm, were responsible for the patterns we observe." *Id.*
[19] Ethan Zuckerman, *The Case for a Taxpayer-Supported Version of Facebook*, ATLANTIC (May 7, 2017), https://www.theatlantic.com/technology/archive/2017/05/the-case-for-a-taxpayer-supported-version-of-facebook/524037.
[20] Id.
[21] *Id.* ("Between the moderately conservative Wall Street Journal, which draws Clinton and Trump supporters in equal shares, and the starkly partisan sites that draw Trump supporters by ratios of 4:1 or more, there are only a handful of sites. . . . By contrast, starting at The Wall Street Journal and moving left, attention is spread more evenly across a range of sites whose audience reflects a gradually increasing proportion of Clinton followers as opposed to Trump followers.").
[22] Perhaps further contributing to the asymmetry of media bubbles, researchers who analyzed the ideological preferences of 3.8 million Twitter users found that "respect to both political and nonpolitical issues, liberals were more likely than conservatives to engage in cross-ideological dissemination." Pablo Barberá et al., *Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?*, 26 PSYCHOL. SCI. 1531, 1531 (2015).

Americans, who were least likely to use the internet."[23] A *New York Times* article covering the study emphasized the sharpness of the increase in political polarization of recent years.[24] The authors of the Brown/Stanford study emphasized that "any explanatory factor" of a shift of this magnitude "would have to make sense equally across demographics – something that the rise of social media failed to account for."[25]

B. FAKE NEWS

Propaganda long predates the Internet, of course. In the last war of the Roman Republic, "Octavian famously used a campaign of disinformation to aid his victory over Marc Anthony."[26] Fake news "took off" after the 1439 invention of the printing press, "at the same time that news began to circulate widely."[27] Sensationalized stories, including reports on "giant man-bats" that British astronomer John Herschel purportedly saw through a telescope directed at the moon, helped the *New York Sun* grow its readership from 8,000 to 19,000 in the mid-nineteenth century.[28] Yet as that century progressed, "impartiality and objectivity were increasingly venerated at the most prestigious newspapers."[29]

The development of the Internet only recently created a new medium for the age-old

---

[23] Jonah Engel Bromwich, *Social Media Is Not Contributing Significantly to Political Polarization, Paper Says*, N.Y. TIMES (Apr. 13, 2017), https://www.nytimes.com/2017/04/13/us/political-polarization-internet.html ("For instance, within the index of polarization trends they created, the authors found that among respondents age 75 and older, the increase in polarization between 1996 and 2012 was 0.38 points, compared to just 0.05 points for adults under the age of 40.").

[24] *Id.* (Last week, The Cook Political Report released the 20th edition of its partisan voter index . . . . The report found that there were only 72 districts in which politicians for both parties were competitive. The statistic represented 'a 20 percent decline from just four years ago, when there were 90 swing seats,' the report found.").

[25] *Id.*

[26] James Carson, *What Is Fake News? Its Origins and How It Grew in 2016*, TELEGRAPH (LONDON) (Mar. 16, 2017), http://www.telegraph.co.uk/technology/0/fake-news-origins-grew-2016. Carson traces the violent history associated with propaganda. *Id.*

[27] Jacob Soll, *The Long and Brutal History of Fake News*, POLITICO (Dec. 18, 2016), http://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535.

[28] Tom Standage, *The True History of Fake News*, ECONOMIST (June/July 2017), https://www.1843magazine.com/technology/rewind/the-true-history-of-fake-news.

[29] Id.

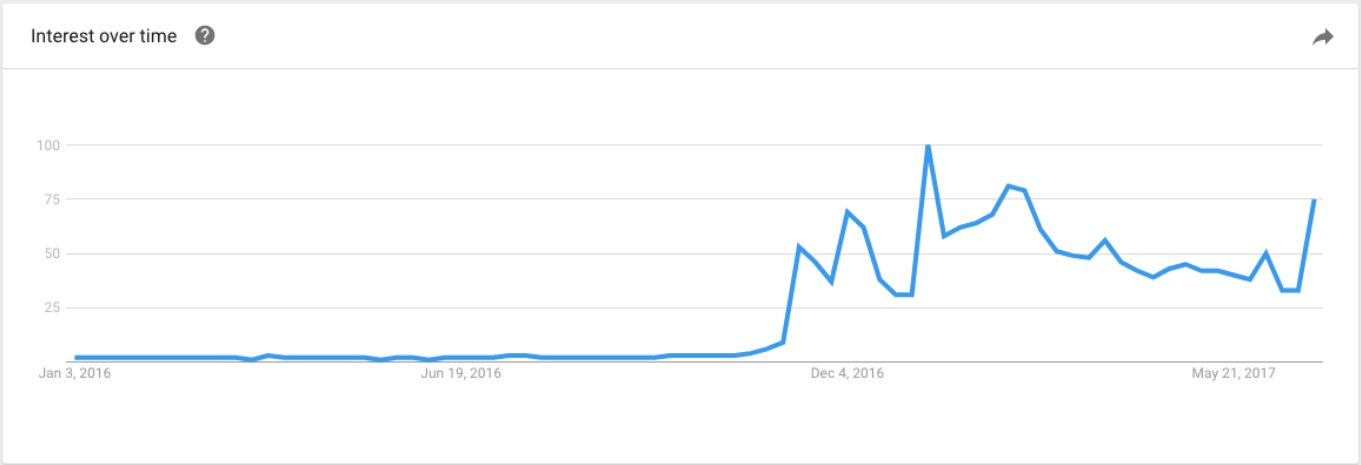
phenomena of misinformation and disinformation.[30] In 2014, the World Economic Forum named "[t]he rapid spread of misinformation online" one of its top ten trends of the year.[31] Likely contributing to the rise of fake news, overall confidence in news media has declined significantly in recent decades. In surveys by Pew Research Center, the number of respondents who agreed with the statement that "stories are often inaccurate" rose from 34% to 66% between 1985 and 2011.[32] Similarly, respondents who believed that "news organizations tend to favor one side" increased from 53% to 77% during the same time period.[33] In 2011, for the first time since the inception of the survey, as many Americans believed that news organizations hurt democracy as helped it.[34] Commentator Issie Lapowsky described the all-time low in public trust in the media as "the perfect petri dish in which a plague of misinformation could fester and bloom."[35] Of course, distrust of news media and the spread of false news are likely mutually reinforcing. According to University of Connecticut philosophy professor Michael Lynch, the proliferation of fake news makes it more difficult to convince people of objective facts.[36]

Fake news attracted even more attention during and after the 2016 U.S. presidential election. In 2016, "post-truth" was Oxford Dictionaries' word of the year.[37] A Google Trends

---

[30] *See, e.g.*, *id.* ("Thanks to internet distribution, fake news is again a profitable business.").
[31] *Top 10 Trends of 2014: 10. The Rapid Spread of Misinformation Online*, WORLD ECON. F., http://reports.weforum.org/outlook-14/top-ten-trends-category-page/10-the-rapid-spread-of-misinformation-online (last visited July 10, 2017) (citing a Twitter rumor during the 2011 U.K. riots that a children's hospital had been attacked by looters and a popular image of soldiers standing guard during a storm, which was shared in the context of Hurricane Sandy, in 2012, although it was taken during an entirely different storm).
[32] *Press Widely Criticized, But Trusted More than Other Information Sources*, PEW RES. CTR. (Sept. 22, 2011), http://www.people-press.org/2011/09/22/press-widely-criticized-but-trusted-more-than-other-institutions.
[33] Id.
[34] *Id.* (reporting both groups at 42% of those surveyed). "In the mid-1980s, about twice as many said that news organizations protect democracy rather than hurt democracy." *Id.*
[35] Issie Lapowsky, *2016: The Mainstream Media Melted Down as Fake News Festered*, WIRED (Dec. 26, 2016), https://www.wired.com/2016/12/2016-mainstream-media-melted-fake-news-festered.
[36] Sabrina Tavernise, *As Fake News Spreads Lies, More Readers Shrug at the Truth*, N.Y. TIMES (Dec. 6, 2016), https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html?_r=0 ("He described the thinking like this: 'There's no way for me to know what is objectively true, so we'll stick to our guns and our own evidence. We'll ignore the facts because nobody knows what's really true anyway.'").
[37] *Word of the Year 2016 Is...*, Oxford Dictionaries, https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016 (last visited July 10, 2017).

search of the term "fake news," with a window of January 1, 2016, to June 28, 2017, gives this trend line[38]:



Searches reached their highest point during the week of January 8-14, 2017, the week before President Trump's inauguration.[39]

In the wake of President Donald Trump's unexpected electoral victory over Secretary Hillary Clinton, many news stories hypothesized that online fake news may have contributed,[40] and scholars began examining that possibility more closely. In the *Columbia Journalism Review*,

[38] GOOGLE TRENDS, https://trends.google.com/trends/explore?date=2016-01-01%202017-06-28&q=%22fake%20news%22 (last visited July 10, 2017).

[39] Google Trends uses the following display scheme: "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. Likewise a score of 0 means the term was less than 1% as popular as the peak." *Id.* Searches for "fake news" never rose above 3 until October 2017; during the week of the election (November 6-12, 2017), they rose to 9, and they reached 100 the week of the inauguration. *Id.* The score has not dropped below 30 in the eight months, approximately, following the election.

[40] *See, e.g.*, Dewey, *supra* note **Error! Bookmark not defined.** (interviewing American fake news creator Paul H orner); Dana Milbank, *Trump's Fake-News Presidency*, WASH. POST (Nov. 18, 2016), https://www.washingtonpost.com/opinions/trumps-fake-news-presidency/2016/11/18/72cc7b14-ad96-11e6-977a-1030f822fc35_story.html?utm_term=.3c417f75db67 ("Not only is fake news getting more attention than actual news, but also the leading purveyor of fake news in the United States is now the president-elect."); Hannah Jane Parkinson, *Click and Elect: How Fake News Helped Donald Trump Win a Real Election*, GUARDIAN (Nov. 14, 2016), https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies ("The influence of verifiably false content on Facebook cannot be regarded as "small" when it garners millions of shares. And yes, it runs deep. The less truthful a piece is, the more it is shared.").

Claire Wardle identified six distinct types of "fake news" circulated during the 2016 U.S. presidential election cycle: (1) authentic material used in the wrong context; (2) imposter news sites designed to look like brands we already know; (3) fake news sites; (4) fake information; (5) manipulated content; and (6) parody content.[41]

Researchers Hunt Allcott and Matthew Gentzkow used web browsing data, a 1,200-person survey, and a database of 156 news stories related to the 2016 U.S. presidential election, to analyze consumption of fake news.[42] Their findings include:

- "Referrals from social media accounted for a small share of traffic on mainstream news sites, but a much larger share for fake news sites."[43]
- "[O]nly 14 percent of American adults viewed social media as their 'most important' source of election news."[44]
- "Education, age, and total media consumption are strongly associated with more accurate beliefs about whether headlines are true or false. Democrats and Republicans are both about 15 percent more likely to believe ideologically aligned headlines, and this ideologically aligned inference is substantially stronger for people with ideologically segregated social media networks."[45]
- "[F]ake news was both widely shared and heavily tilted in favor of Donald Trump."[46]

Allcott and Gentzkow also identified several reasons why the level of fake news saturation during the 2016 U.S. presidential election may not accurately reflect its persuasiveness in the

[41] Claire Wardle, *6 Types of Misinformation Circulated This Election Cycle*, COLUM. JOURNALISM REV. (Nov. 18, 2016), https://www.cjr.org/tow_center/6_types_election_fake_news.php.
[42] Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*. 31 J. ECON. PERSPS. 211, 212 (2017), https://web.stanford.edu/~gentzkow/research/fakenews.pdf.
[43] Id.
[44] Id.
[45] *Id.* at 213.
[46] *Id.* at 212.

minds of voters.[47]

According to a *BuzzFeed News* analysis, in the three months preceding the 2016 U.S. presidential election, the twenty top-performing fake news election stories generated more online engagement than the twenty top-performing election stories from major news outlets.[48] Until the three-month mark, content from major news outlets outperformed fake news, with the gap narrowing as the election approached.[49] Significantly, of the twenty top-performing fake news stories, seventeen were overtly pro-Trump or anti-Clinton.[50] The most prolific fake news publisher identified in the analysis, *Ending the Fed*, produced four of the top ten fake news stories examined, which together generated 2,953,000 Facebook engagements in the final three months of the election cycle.[51]

Researchers have examined possible causes for the anecdotal phenomenon that conservatives appear to be more vulnerable to fake news than liberals.[52] According to a study by anthropologist Daniel Fessler, conservatives are "hyper-attuned to hazards"; therefore, they are may be likely to err on the side of trusting reports of danger.[53] Conservative participants in the

---

[47] *Id.* at 232 ("We consider the number of stories voters read regardless of whether they believed them. We do not account for diminishing returns, which could reduce fake news' effect to the extent that a small number of voters see a large number of stories. Also, this rough calculation does not explicitly take into account the fact that a large share of pro-Trump fake news is seen by voters who are already predisposed to vote for Trump—the larger this selective exposure, the smaller the impact we would expect of fake news on vote shares.").

[48] 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=.vdmqwjb6Q#.fp5vVDKMe (showing that the stories generated 8,711,000 and 7,367,000 Facebook shares, reactions, and comments, respectively).

[49] *Id.*

[50] *Id.* ("Two of the biggest false hits were a story claiming Clinton sold weapons to ISIS and a hoax claiming the pope endorsed Trump, which the site removed after publication of this article. The only viral false stories during the final three months that were arguably against Trump's interests were a false quote from Mike Pence about Michelle Obama, a false report that Ireland was accepting American "refugees" fleeing Trump, and a hoax claiming RuPaul said he was groped by Trump.").

[51] *Id.* ("It was responsible for four of the top 10 false election stories identified in the analysis: Pope Francis endorsing Donald Trump, Hilary Clinton selling weapons to ISIS, Hillary Clinton being disqualified from holding federal office, and the FBI director receiving millions from the Clinton Foundation.").

[52] *See, e.g., supra* notes 46, 50 and accompanying text.

[53] Olga Khazan, *Why Fake News Targeted Trump Supporters*, ATLANTIC (Feb. 2, 2017), https://www.theatlantic.com/science/archive/2017/02/why-fake-news-targeted-trump-supporters/515433/.

study were more likely to believe false stories that reported risks, such as "terrorist attacks in the U.S. have increased since Sept 11, 2001," while liberal and conservative participants were equally likely to believe false stories that reported benefits, such as "[e]xercising on an empty stomach burns more calories."[54]

Some social scientists have found that conservatives are less critical of the information that they receive, overall; however, research on this point is inconclusive. Psychologist John Jost reviewed forty studies on the connection between political ideology and "need for cognition," or affinity for critical thinking.[55] Jost found that twenty-five of the studies revealed a "significant, negative" correlation between conservativism and need for cognition.[56] In contrast, Daniel Kahan, a professor of law and psychology, has not uncovered any correlation between biased information processing and political ideology.[57]

Closely related to fake news—and arising at the intersection of fake news and hate speech—is the purported research of fake think tanks, which often take small portions of peer-reviewed research out of context and cobble them together to create reports that reach a different conclusion than any of the source material.[58] Because think tanks are not registered, and because

---

[54] *Id.* Moreover, conservative participants were more likely to believe stories about dangers lurking in social issues, including abortion or same-sex marriage, rather than economic hazards. *Id.*

[55] Christopher Ingraham, *Why Conservatives Might Be More Likely to Fall for Fake News*, WASH. POST (Dec. 7, 2016), https://www.washingtonpost.com/news/wonk/wp/2016/12/07/why-conservatives-might-be-more-likely-to-fall-for-fake-news/?utm_term=.3cfeea610a2f.

[56] *Id.* In all but three of the others, there was a similar negative association but it wasn't statistically significant. *Id.*

[57] *Id.* ("Kahan used a measure of cognitive style called the 'cognitive reflection test.' It posed three quantitative questions that assess a person's ability to resist blurting out the first (wrong) answer that comes to mind. . . . Kahan's work administering these tests to liberals and conservatives didn't turn up any meaningful differences between the two.").

[58] Emma Grey Ellis, *Fake Think Tanks Fuel Fake News—And the President's Tweets* (Jan. 24, 2017, 6:00 AM), https://www.wired.com/2017/01/fake-think-tanks-fuel-fake-news-presidents-tweets ("The Family Research Council does quite a bit of that [cherry-picking and taking material out of context], as does the homophobic American College of Pediatricians (which doesn't call itself a think tank). . . . Which is how FRC and ACPeds came to assert homosexuality is connected to pedophilia, even though none of their source material agreed.").

many of the illegitimate organizations use names that sound credible[59] and are registered as 501(c)(3) non-profits, it is sometimes difficult to distinguish them from their bona fide counterparts.[60] The most prominent fake think tanks include anti-immigration, anti-LGBT, and white supremacist groups, many of which appear on the Southern Poverty Law Center's list of hate groups.[61] Social media and Internet search algorithms amplify the reach of pseudoscientific think tanks' propaganda.[62]

Demonstrating the potential of fake news to incite violence, the promulgation of fake news led directly to a violent attack in at least one instance over the past year. In December 2016, a man fired a rifle inside a Washington, D.C., pizza restaurant; he told the police that he had come to the restaurant to investigate an online conspiracy theory that the restaurant was at the center of a child abuse ring.[63] The story spread on the online message board 4chan, after Wikileaks released hacked emails in which Podesta communicated with the owner of the restaurant about a fundraiser.[64] In July 2017, fliers appeared on doors in the neighborhood of the restaurant, insisting that the conspiracy theory is true and calling for its reinvestigation.[65] Almost 22 million users log on to 4chan each month; the anonymous message board has previously been

---

[59] "[T]he Employment Policies Institute [an anti-minimum wage increase public relations firm] practically steals its name from the Economic Policy Institute." *Id.*

[60] *Id.* ("'For the lay person who reads about these topics for 10 minutes a week, I don't think there is an easy way to see who's full of it,' says Alex Nowrasteh, an immigration policy analyst at the Cato Institute.").

[61] Id.

[62] *Id.* According to Heidi Beirich of the Southern Poverty Law Center, "Think-tanky white supremacist organizations have generated enough material that a search topic like 'black on white crime' is dominated by their propaganda. That's what happened to Dylann Roof, and how Trump ended up tweeting those false statistics." *Id.*

[63] Eric Lipton, *Man Motivated by "Pizzagate" Conspiracy Theory Arrested in Washington Gunfire*, N.Y. TIMES (Dec. 5, 2016), https://www.nytimes.com/2016/12/05/us/pizzagate-comet-ping-pong-edgar-maddison-welch.html. *BuzzFeed News* tracked the spread of the rumor, which began with an October 30, 2016, tweet from a white supremacist Twitter account posing as a Jewish lawyer in New York; the tweet claimed to that police were investigating "evidence that emails from Anthony Weiner's laptop contained evidence of Clinton involvement in an "international child enslavement ring." Craig Silverman, *How the Bizarre Conspiracy Theory Behind "Pizzagate" Was Spread*, BUZZFEED NEWS (Dec. 5, 2016), https://www.buzzfeed.com/craigsilverman/fever-swamp-election?utm_term=.it8jLpdNb#.dqe5Jlwj7.

[64] Id.

[65] *Fliers Backing "Pizzagate" Theory Appear in Neighborhood*, ABC (July 2, 2017), http://abcnews.go.com/US/wireStory/fliers-backing-pizzagate-theory-dc-neighborhood-48409014.

linked to other online rumors with physically dangerous consequences.[66]

Finally, researchers at the Oxford Internet Institute compared levels of "junk news" saturation on Twitter in recent elections across the globe.[67] They found:

- Among Michigan twitter users, between November 1 and November 11, 2016, "[f]irst, the proportion of professional to junk news is roughly one-to-one. Second, when the amount of junk news is added to the number of links to unverified WikiLeaks content, and Russian-origin news stories, it appears that fully 46.5 percent of all the content that is presented as news and information about politics and the election is of an untrustworthy provenance or falls under the definition of propaganda based on its use of language and emotional appeals."[68]
- Among German Twitter users, between February 11 and February 13, 2017, "[f]irst, the ratio of professional to junk news is roughly four to one. Second, when the amount of junk news is added to Russian-origin news stories, it appears that fully 12.8 percent of all the content that is clearly news and information about politics and the election is of an untrustworthy provenance."[69]

---

[66] *See* Caitlin Dewey, *Absolutely Everything You Need to Know to Understand 4chan, the Internet Bogeyman*, WASH. POST (Sept 25, 2014), https://www.washingtonpost.com/news/the-intersect/wp/2014/09/25/absolutely-everything-you-need-to-know-to-understand-4chan-the-internets-own-bogeyman/?utm_term=.1c84518fb786 (describing the cyberbullying of 11-year-old Jessi Slaughter, who later attempted suicide more than once; a meme that encouraged young Justin Bieber fans to cut themselves; and an Ebola "mascot" that became a hoax directed at West Africans).

[67] *See, e.g.*, Memorandum from Philip N. Howard, Oxford Univ., et al. (Mar. 26, 2017), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/03/What-Were-Michigan-Voters-Sharing-Over-Twitter-v2.pdf. They define "junk news" as:

various forms of propaganda and ideologically extreme, hyper-partisan, or conspiratorial political news and information. . . . This content is produced by organizations that do not employ professional journalists, and the content uses attention grabbing techniques, lots of pictures, moving images, excessive capitalization, ad hominem attacks, emotionally charged words and pictures, unsafe generalizations and other logical fallacies.

Id.

[68] Id.

[69] Memorandum from Lisa-Maria Neudert, Oxford Univ., et al. (Mar. 27, 2017), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/03/What-Were-German-Voters-Sharing-Over-Twitter-v6-1.pdf.

- Among French Twitter users, between April 27 and April 29, 2017, "[o]verall, junk news, Russian content, and religious content account for 10 percent of all links shared over Twitter (6 percent for junk news alone)."[70] Additionally, "French voters appear to be sharing political news and information of a lesser quality between the two rounds compared to before the first round of the election."[71]
- Among U.K. Twitter users, between May 1 and May 7, 2017, "the largest proportion of content being shared by Twitter users interested in UK politics comes from professional news organizations, which accounts for 43.3% of the total content shared. Junk news accounts for over a third of other political news and information and accounts for 10.2% of the total content shared."[72]

C. COMPUTATIONAL PROPAGANDA (BOTS)

"Computational propaganda" is defined as "the use of algorithms, automation, and human curation to purposefully distribute misleading information over social media networks," during elections, political crises, and national security incidents.[73] Bots are "integral to [its] spread,"[74] although they also perform countless other functions. Bots themselves are simply "application[s] that perform[] an automated task, such as setting an alarm, telling you the weather or searching online."[75] Overall, these automated scripts generate approximately sixty percent of total traffic on the Internet.[76]

---

[70] Memorandum from Clementine Desigaud, Oxford Univ., et al. (May 4, 2017), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/05/What-Are-French-Voters-Sharing-Over-Twitter-Between-the-Two-Rounds-v7.pdf.
[71] Id.
[72] Memorandum from John D. Gallacher, Oxford Univ., et al. (May 31, 2017), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Junk-News-and-Bots-during-the-2017-UK-General-Election.pdf.
[73] Id.
[74] *Id.* at 6.
[75] Sarah Mitroff, *What Is a Bot?: Here's Everything You Need to Know*, CNET (May 5, 2016), https://www.cnet.com/how-to/what-is-a-bot.
[76] Woolley & Howard, *supra* note 98.

Imperva Incapsula, a cloud-based web platform, analyzes global bot traffic annually. In 2015, it found that humans were, for the first time, responsible for the majority (51.5%) of online traffic, up from 38.5% in 2013.[77] In 2016, however, bot traffic once again accounted for the majority of web traffic overall.[78] Imperva further breaks down its analysis by "good bots," including feed fetchers,[79] search engine bots,[80] commercial crawlers,[81] and monitoring bots,[82] and "bad bots," including, most prominently, impersonator bots.[83] Imperva found that 94.2% of websites surveyed in 2016 experienced at least one bot attack during the course of the study and that, over the past five years, one-third of website visitors were attack bots.[84]

Douglas Guilbeault described how bots work in the *International Journal of Communication*:

> [B]ot designers equip bots with basic scripts for adapting to the user data that social media platforms make available. Bots can harness public data to target users who are likely to connect with bots, based on their connection forming habits . . . . Bots can also model user data and adapt to social norms, which enhances their persuasiveness . . . . Most daunting is the rise of botnets, which can use a central bot intelligence to coordinate hundreds of bots in data mining or denial of service attacks. Boshmaf, Muslukhov, Beznosov, and Ripeanu (2011, 2013) showed how botnets can penetrate Facebook with a success rate of more than 80%. Rodriguez-Gomez, Marcia-Fernandez, and Garcia-Teodoro (2013) further showed how botnets can evolve through a dynamic life cycle that

[77] Igal Zeifman, *2015 Bot Traffic Report: Humans Take Back the Web, Bad Bots Not Giving Any Ground*, IMPERVA INCAPSULA (Dec. 9, 2015), https://www.incapsula.com/blog/bot-traffic-report-2015.html ("The data presented herein is based on a sample of over 19 billion human and bot visits occurring over a 90-day period, from July 24, 2015 to October 21, 2015. It was collected from 35,000 Incapsula-protected websites having a minimum daily traffic count of at least 10 human visitors.").
[78] Igal Zeifman, *Bot Traffic Report 2016*, IMPERVA INCAPSULA (Jan. 24, 2017), https://www.incapsula.com/blog/bot-traffic-report-2016.html ("In 2015 we documented a downward shift in bot activity on our network, resulting in a drop below the 50 percent line for the first time in years. In 2016 we witnessed a correction of that trend, with bot traffic scaling back to 51.8 percent—only slightly higher than what it was in 2012.").
[79] *Id.* ("Bots that ferry website content to mobile and web applications, which they then display to users.").
[80] *Id.* ("Bots that collect information for search engine algorithms, which is then used to make ranking decisions.").
[81] *Id.* ("Spiders used for authorized data extractions, usually on behalf of digital marketing tools.").
[82] *Id.* ("Bots that monitor website availability and the proper functioning of various online features.").
[83] *Id.* ("[A]ttack bots masking themselves as legitimate visitors so as to circumvent security solutions.").
[84] Id.

challenges the ability to distinguish between bots and humans.[85]

Sociologists Alex Wilkie, Mike Michael, and Matthew Plummer-Fernandez identified three conceptual archetypes on which bots tend to draw and programmed their own experimental bots in these forms,[86] as the idiot,[87] the diplomat,[88] and the parasite.[89] In another tripartite typology, Samuel C. Woolley, of the Oxford Internet Institute, categorized political bots, in particular, as follower bots,[90] roadblock bots,[91] and propaganda bots.[92]

Woolley also examined use of political bots by states and political parties, between 2011 and 2014, across eighteen countries—some democratic and some autocratic.[93] His findings included:

- "Many cases of political bot use occur when governments target perceived cyber-security threats or political-cultural threats from other states. Several articles mention state-

[85] Douglas Guilbeault, *Growing Bot Security: An Ecological View of Bot Agency*, 10 INT'L J. COMMS. 5003, 5005 (2016).
[86] Alex Wilkie et al., Speculative Method and Twitter: Bots, Energy and Three Conceptual Characters, 63 SOC. REV. 79, 86, 88-97 (2015).
[87] *Id.* at 87 ("[T]he idiot is thus a figure who slows down thought by making no sense in its actions . . . . the idiot prompts us to consider the likelihood that "'there is something more important' . . . going on in the event which is yet to be understood. That is to say, we become open to the possibility of a dramatic redefinition of the parameters of the event.").
[88] *Id.* ("[T]he diplomat is a character that in the cosmopolitical event presents the 'voice of whose practice, whose mode of existence and whose identity are threatened by decision' . . . . In other words, diplomats serve to trouble the usual course of action (whether that be structured by necessity or 'progress', or by 'general interest', or by translation into monetary terms) by evoking how such activities might culminate in an 'act of war' . . . and thus a calamitous reordering of relations.").
[89] *Id.* at 88 ("[T]he parasite [is] an uninvited, and initially disruptive, guest at the dinner table, who exchanges stories for food. . . . In parallel to the unruly parasite, is the figure of Hermes . . . who operates to connect across disparate domains, such as science and myth. However, Hermes is untrustworthy, and the connections and communications he mediates might entail mistranslations and mischief—that is, inject disorder into apparent order.").
[90] CAN PUBLIC DIPLOMACY SURVIVE THE INTERNET?: BOTS, ECHO CHAMBERS, AND DISINFORMATION 16 (Shawn Powers & Markos Kounalakis, eds., May 2017), https://www.state.gov/documents/organization/271028.pdf ("[U]sed to boost political figures' follower numbers and passively like or re-tweet content.").
[91] *Id.* ("[U]sed to spam hashtags associated with activists or political opposition in order to shut down or interrupt dissent via non-traditional communication channels.").
[92] *Id.* ("[U]sed to mimic humans while sending out effusively positive information about an embattled government or politician or to propagate negative attacks against the opposition.").
[93] Samuel C. Woolley, *Automating Power: Social Bot Interference in Global Politics*, FIRST MONDAY (Apr. 2016), http://journals.uic.edu/ojs/index.php/fm/article/view/6161/5300.

sanctioned Russian bot deployment. In these articles, Russian bots were allegedly used to promote regime ideals or combat anti-regime speech against targets abroad."[94]

- "Governments, politicians and contractors in their employ also use political bots to attack in-state targets on social media. . . . According to numerous sources, the Mexican government has used Twitter bot armies to stifle public dissent and effectively silence opposition through spam tactics."[95]
- "Political bots have been used during elections to send out pro-government or pro-candidate microblog messages."[96]
- "Political bots have also been used during elections to pad politicians' social media follower lists."[97]

In the United States, bots have been present in electoral politics since at least 2010.[98] Recently, according to researcher Emilio Ferrara, who developed a 95% accurate bot-detection tool, 400,000 bots generated at least 4 million election-related tweets between September 16 and October 21, 2016, amounting to at least 15% of all users discussing the U.S. presidential election.[99] Similarly, The Oxford Internet Institute analyzed the political content of Twitter bot

---

[94] Id.
[95] Id.
[96] Id.
[97] Id.
[98] Sam Woolley & Phil Howard, *Bots United to Automate the Presidential Election*, WIRED (May 15, 2016), https://www.wired.com/2016/05/twitterbots-2/ ("Researchers at Wellesley College found evidence that when Scott Brown successfully ran for senator in 2010, a conservative group used bots to attack his opponent, Martha Coakley. Gawker reported in 2011 that Newt Gingrich's campaign bought more than a million fake followers."); *see also* Woolley, *supra* note 93 ("Metaxas and Mustafaraj . . . effectively outline an exemplary, and U.S.-based, case of this sort of propaganda dissemination. Their research analyzes the role of social bots, or 'sock puppets', in spreading biased and flawed political information during the 2010 Brown and Coakley Massachusetts Senate race. They found that Astroturf political groups that supported Brown used Bots to carry out significant attacks on the Coakley campaign over social media.") (footnote omitted).
[99] Emilio Ferrara, *How Twitter Bots Affected the US Presidential Election*, CONVERSATION (Nov. 8, 2016), https://theconversation.com/how-twitter-bots-affected-the-us-presidential-campaign-68406?utm_content=buffer8bb03&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer.

activity during the third 2016 U.S. presidential debate.[100] It examined 10 million tweets collected between October 19 and October 22, 2016, screening for a set of pro-Clinton, pro-Trump, and neutral hashtags.[101] The researchers defined "highly automated accounts" as those posting at least fifty times per day.[102] Their findings included:

- "[T]he proportion of highly automated twitter activity increased slightly from debate to debate, rising 23 percent in the first to 27 in the final."[103]
- "Highly automated pro-Trump bots generated four tweets for every one that highly automated pro-Clinton accounts generated."[104]
- "Pro-Clinton highly automated accounts increased their activities from the first to final debate but still never reached the level of automation behind pro-Trump traffic."[105]

The Oxford Internet Institute also performed a similar analysis surrounding the 2017 U.K. general election.[106] Analyzing approximately 2,489,000 tweets collected between May 27 and June 2, 2017, with hashtags associated with political parties,

---

[100] Memorandum from Bence Kollanyi, Cornivus Univ., et al. (Oct. 27, 2016), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2016/10/Data-Memo-Third-Presidential-Debate.pdf.
[101] "Pro-Trump hashtags include[d] #50points, #AltRight, #AmericaFirst, #benghazi, #ClintonFoundation, #clintonscandals, #CrookedHillary, #DebateSideEffects, #deplorable, #DrainTheSwamp, #hillaryshealth, #ImWithYou, #LatinosForTrump, #LawAndOrder, #lockherup, #MAGA, #MakeAmericaGreatAgain, #MSM, #NeverHillary, #pepe, #ProjectVeritas, #realDonaldTrump, #RiggedSystem, #RNC, #tcot, #TeamTrump, #Trump, #TrumpDebateGuests, #TrumpPence16, #TrumpPence2016, #TrumpTrain, #TrumpWon, #Veritas, #VoterFraud, #VoteTrump, #WakeUpAmerica, #WikiLeaks. Pro-Clinton hashtags include[d] #Clinton, #ClintonKaine, #ClintonKaine16, #ClintonKaine2016, #CountryBeforeParty, #ctl, #dems, #DirtyDonald, #DNC, #Factcheck, #failedtaxaudit, #HillaryClinton, #HillarysArmy, #hillarywon, #ImWithHer, #lasttimetrumppaidtaxes, #LoveTrumpsHate, #NeverTrump, #OHHillYes, #p2, #p2b, #shareblue, #StrongerTogether, #TNTweeters, #TrumpedUpTrickleDown, #trumptape, #UniteBlue, #VoteDems, #WhyIWantHillary. Neutral hashtags include[d] #Debates2016, #Debates, #Debate, #Election2016, #POTUS." *Id.*
[102] Id.
[103] Id.
[104] Id.
[105] Id.
[106] Memorandum from Monica Kaminska, Oxford Univ., et al. (June 5, 2017), http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Social-Media-and-News-Sources-during-the-2017-UK-General-Election.pdf.

candidates, and the election, the Institute found:

- "By political party, it appears that the Conservative Party and the Labour Party have a higher number of highly automated accounts generating traffic about them when compared to the other three parties."[107]
- "[T]here was a large overlap between accounts that tweeted using Labour-related hashtags and accounts that tweeted using Conservative-related hashtags."[108]
- "These accounts, however, generated significantly more content when using Labour-related hashtags . . . ."[109]
- "On average, 16.5% of traffic about UK politics is generated by highly automated accounts that [the Institute was] able to track."[110]

As with fake news, which is often a central element of the computational propaganda that bots help spread,[111] social science suggests that some social media users may be more vulnerable to bots than others. Researchers with Florida Atlantic University and the Online Privacy Foundation found that Twitter users were more likely to follow or reply to a bot if they have higher "Klout scores"[112] and follow a larger number of Twitter users.[113]

D. HATE SPEECH

---

[107] Id.
[108] Id.
[109] Id.
[110] Id.
[111] *Id.* at 8 ("False news reports, widely distributed over social media platforms, can in many cases be considered to be a form of computational propaganda. Bots are often key tools in propelling this disinformation across sites like Twitter, Facebook, Reddit, and beyond.").
[112] Randall Wald et al., *Predicting Susceptibility to Social Bots on Twitter*, IIEE 1, 9 (2013), http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6642447 ("This is a metric calculated by the private company Klout.com, which collects information from a user's Facebook, Twitter, G+, LinkedIn, and other social networking profiles to determine their overall social influence.").
[113] *Id.* at 10 ("It makes sense that these two features are at the top, because they both reflect involvement with social networking in general . . . . Individuals who are highly engaged in social networking would seem to be more likely to interact with an unknown user (such as a social bot) even if this user might have imperfect grammar or word use.").

72% of American Internet users report that they have witnessed harassment online, while 47% report experiencing harassment personally.[114] Recently, news media reported a significant spike in hate speech online over the course of the 2016 presidential campaign.[115] The Anti-Defamation League released a report in October 2016, which documented "a significant uptick in anti-Semitic tweets in the second half (January-July 2016) of this study period."[116] More specifically, the ADL found, "At least 800 journalists received anti-Semitic tweets with an estimated reach of 45 million impressions. The top 10 most targeted journalists (all of whom are Jewish) received 83 percent of these anti-Semitic tweets."[117]

A team of social scientists at NYU's SMaPP (Social Media and Political Participation) Laboratory tested the hypothesis that Trump's candidacy in the 2016 U.S. presidential election led to an increase in online hate speech.[118] Using "over 125 million Tweets referencing Hillary Clinton, over 500 million tweets referencing Donald Trump, and over 300 million tweets collected from a random sample of 500,000 American Twitter users," the study conducted both a dictionary-based analysis supplemented with machine learning and "robustness tests using an alternative new semi-supervised non-dictionary based method."[119] SMaPP found the following:

> Contrary to the current prevailing narrative, in our core dictionary-based analysis we find no systematic evidence of a meaningful increase in hate speech on

---

[114] Erin Carson, *Google and Jigsaw Puzzle Out AI Fix for Toxic Comments*, CNET (Feb. 23, 2017), https://www.cnet.com/news/google-jigsaw-puzzle-perspective-toxic-comments-machine-learning-ai.
[115] *See, e.g.*, Jessica Guynn, *"Massive Rise" in Hate Speech on Twitter During Presidential Election*, USA TODAY (Oct. 23, 2016), https://www.usatoday.com/story/tech/news/2016/10/21/massive-rise-in-hate-speech-twitter-during-presidential-election-donald-trump/92486210 ("Leslie Miley, a former Twitter employee, says hate speech has always lurked on Twitter. But the alt-right, the community of activists that embrace white nationalism and supremacy, 'has its Twitter game on point right now,' he says.").
[116] *Anti-Semitic Targeting of Journalists During the 2016 Presidential Campaign*, ADL Rep. (Oct. 19, 2016), https://www.adl.org/sites/default/files/documents/assets/pdf/press-center/CR_4862_Journalism-Task-Force_v2.pdf.
[117] Id.
[118] *See* Alexandra Siegel et al., Trumping Hate on Twitter: Online Hate Speech and White Nationalist Rhetoric in the 2016 US Election Campaign and its Aftermath 6-7 (May 2017) (unpublished manuscript) (on file with author) ("This anecdotal evidence of increased hate crimes and hate speech suggest that the "Trump effect" has played a key role in legitimizing and mainstreaming extremist rhetoric. . . . [W]e test the extent to which hate speech on Twitter increased over the course of the campaign or following the election.").
[119] *Id.* at 2-3.

> Twitter over the course of Trump's campaign or following his election. Although our random sample of American Twitter users shows a one day spike on Election Day in the volume of hate speech and number of users tweeting it, this is largely driven by an uptick in misogynistic language, which is unsurprising given the sexism Hillary Clinton faced on the campaign trail as the nation's first female presidential candidate. Finally, while we do observe a positive significant increase in the popularity of white nationalist language over the course of the campaign across the Trump, Clinton, and random sample datasets, the absolute levels of this rhetoric remain quite low, representing a tiny fraction of our data, even on the most prolific days.[120]

Hate speech may cause "frequently targeted groups," including women and minorities, to opt out of participation in online discourse.[121] For example, Jaclyn Munson, a pro-choice activist, wrote a story in 2013 that documented her undercover experience at an anti-abortion crisis pregnancy center.[122] She received death threats in response, and, one year later, Munson abandoned her online writing career and deleted her Twitter account.[123] Similarly, Emily McCombs, executive editor of xoJane, a website geared toward young women that specializes in first-person narratives, reported that writers often receive their break on the website but quickly move on to a "safer" medium, "like print."[124] McCombs herself reported that "her next job won't be online."[125]

Social media platforms prohibit hate speech, as well as obscenity and other offensive content, through their community standards platforms.[126] Content moderation takes a toll on

---

[120] *Id.* at 4.
[121] Pierre M. Omidyar, Key Risks of Social Media for Democracy 8 (Apr. 10, 2017) (unpublished manuscript) (on file with author).
[122] Michelle Goldberg, *Feminist Writers Are So Beseiged by Online Abuse That Some Have Begun to Retire*, WASH. POST (Feb. 20, 2015), https://www.washingtonpost.com/opinions/online-feminists-increasingly-ask-are-the-psychic-costs-too-much-to-bear/2015/02/19/3dc4ca6c-b7dd-11e4-a200 c008a01a6692_story.html?tid=a_inl&utm_term=.85c3aedd967e.
[123] *Id.* ("'It was just becoming really emotionally overwhelming to be on the front lines all the time,' she says.").
[124] *Id.* ("'Part of that is definitely not being able to handle the harassment,' [McCombs said].").
[125] Id.
[126] YouTube prohibits "[h]ate speech" which the platform defines as speech that "refers to content that promotes violence or hatred against individuals or groups based on certain attributes, such as: race or ethnic origin, religion, disability, gender, age, veteran status, [and] sexual orientation/gender identity." *Policy Center*, YOUTUBE, https://support.google.com/youtube/answer/2801939?hl=en (last visited July 10, 2017). Facebook's Community Standards ban "hate speech, which includes content that directly attacks people based on their: [r]ace, [e]thnicity,

human moderators, however. Increasingly, these employees are based in the Philippines, where the former U.S. colony's "close cultural ties to the United States" help moderators identify content that Americans will find offensive, while market wages are "a fraction" of those in the U.S.[127] Most are required to sign non-disclosure agreements.[128] Burnout is common, especially among American moderators, many of whom enter the field as a career of "last resort."[129] In 2014, Hemanshu Nigam, MySpace's former chief security officer, estimated that "well over 100,000" content moderators are employed by social media sites, mobile apps, and cloud storage services.[130]

Additionally, social media platforms, which operate globally, must navigate the different standards of hate speech—from both a legal and a cultural perspective—in the many countries in which they operate. In the United States, the First Amendment protects speech that would violate the hate speech policies of any of the major platforms, discussed above at note 126.[131] On June 19, 2017, the Supreme Court of the United States held that social media implicates some free speech concerns: specifically, that a North Carolina statute that criminalized registered sex offenders' use of "commercial social networking Web site[s] . . . [that] permit[] minor children

---

[n]ational origin, [r]eligious affiliation, [s]exual orientation, [s]ex, gender, or gender identity, or [s]erious disabilities or diseases." *Community Standards*, FACEBOOK, https://www.facebook.com/communitystandards#hate-speech (last visited July 10, 2017). And Twitter's "Twitter Rules" prohibit "hateful conduct," defined as "promot[ing] violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease." *The Twitter Rules*, TWITTER, https://support.twitter.com/articles/18311 (last visited July 10, 2017). Twitter also does not permit "accounts whose primary purpose is inciting harm towards others on the basis of these categories." *Id.*

[127] Adrian Chen, *The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed*, WIRED (Oct. 23, 2014, 6:30 AM), https://www.wired.com/2014/10/content-moderation ("A brand-new American moderator for a large tech company in the US can make more in an hour than a veteran Filipino moderator makes in a day."

[128] Id.

[129] *Id.* ("'Everybody hits the wall, generally between three and five months,' says a former YouTube content moderator I'll call Rob.").

[130] *Id.* ("[T]hat is, about twice the total head count of Google and nearly 14 times that of Facebook.").

[131] U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech . . . ."); *see also, e.g.*, *Snyder v. Phelps*, 562 U.S. 443, 458 (2011) (holding that the First Amendment shielded Westboro Baptist Church from tort immunity for picketing a soldier's funeral with hateful signs).

to become members or to create or maintain personal Web pages" violated the First Amendment.[132] Yet many democracies, including Canada, Britain, France, Germany, the Netherlands, South Africa, Australia, and India, prohibit hate speech by law or by international treaty.[133] Nonetheless, Facebook, for example, aims to create global rules, although it sometimes makes content unavailable only in a country where it is illegal.[134] YouTube, in contrast, is more apt to prohibit content in some countries and not others, if reluctantly.[135]

Reflecting these cross-border legal differences, Europe has taken a more state-centric approach to reducing hate speech online than has the United States. In 2016, European officials demanded that tech companies remove at least 50% of hate speech upon notification.[136] In May 2017, the European Commission, the European Union's executive branch, found that Google and Facebook were in compliance, while Twitter was not.[137] On June 30, 2017, Germany passed the Network Enforcement Act, which authorizes fines of up to $57 million against social media companies that fail to remove hate speech, defamation, incitements to violence, and other "'obviously illegal' content" within twenty-four hours.[138] Internationally, "Germany has some of

---

[132] Packingham v. North Carolina, 582 U.S. ___ (2017).
[133] Adam Liptak, *Hate Speech or Free Speech?: What Much of the West Bans Is Protected in U.S.*, N.Y. TIMES (June 11, 2008), http://www.nytimes.com/2008/06/11/world/americas/11iht-hate.4.13645369.html.
[134] Julia Angwin & Hannes Grassegger, *Facebook's Secret Censorship Rules Protect White Men from Hate Speech But Not Black Children*, PROPUBLICA (June 28, 2017, 5:00 AM), https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms.
[135] *Id.* ("I don't love traveling this road of geo-blocking," [Google attorney Nicole] Wong said, but "it's ended up being a decision that allows companies like Google to operate in a lot of different places.").
[136] Mark Scott, *Twitter Fails E.U. Standard on Removing Hate Speech Online*, N.Y. TIMES (May 31, 2017), https://mobile.nytimes.com/2017/05/31/technology/twitter-facebook-google-europe-hate-speech.html.
[137] *Id.* ("Twitter removed hate speech from its network less than 40 percent of the time after such content had been flagged to the company. . . . [Nonetheless], it has improved significantly from a study published late last year, which found that it removed a mere 19 percent of hate speech when notified."). The findings reflected 2,500 instances of potential hate speech reported over seven weeks by nongovernmental organizations and twenty-four E.U. member states. *Id.* Facebook reported removing 66,000 posts that were reported as hate speech each week for the two months preceding June 27, 2017. Richard Allan, *Hard Questions: Hate Speech*, FACEBOOK NEWSROOM (June 27, 2017), https://newsroom.fb.com/news/2017/06/hard-questions-hate-speech.
[138] Amar Toor, *Germany passes controversial law to fine Facebook over hate speech*, VERGE (June 30, 2017), https://www.theverge.com/2017/6/30/15898386/germany-facebook-hate-speech-law-passed ("Germany has in recent years intensified efforts to crack down on hate speech, amid a rise in anti-migrant sentiment that has been fueled in part by the ongoing refugee crisis. Facebook, Twitter, and Google agreed to remove such content from

the world's toughest laws covering defamation, public incitement to commit crimes and threats of violence, with prison sentences for Holocaust denial or inciting hatred against minorities."[139]

Even in the absence of fines, failure to remove hate speech may be costly to social media platforms. In March 2017, over 200 brands, including the British government, pulled their advertisements from YouTube,[140] after discovering that they were running alongside "homophobic and hate-spewing videos" and "websites run by anti-semites, hate mongers, white supremacists, and pornographers."[141] Analysts estimated that the boycott could cost Google between $750 million and $1 billion in 2017.[142]

As an alternative to Europe's prohibitive approach to hate speech online, researcher Kevin Munger experimented with the use of Twitter bots to socially sanction racist harassment on the platform.[143] He found that sanctioning racist posts in the guise of an "in-group," i.e., white, account with a large number of followers was most effective.[144] Additionally, he found that the effects of social sanctioning persisted for the first month of the study, but not the second.[145]

Beyond social media, offensive and threatening content contributed to the decisions of a

---

their platforms within 24 hours, under a 2015 deal with the German government, but a 2017 report commissioned by the Justice Ministry found that the companies were still failing to meet their commitments.").

[139] *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.

[140] Sharyl Kaur, *Google faces major fallout as brands withdraw YouTube advertising*, TECH WIRE ASIA (Mar. 31, 2017), http://techwireasia.com/2017/03/watch-google-faces-major-fallout-brands-pull-youtube-advertising/#isfUh39QSqX8k94c.97.

[141] David Schrieberg, *U.S., U.K. Boycott of Google and YouTube Spreading Over Hate Content*, FORBES (Mar. 23, 2017), https://www.forbes.com/sites/davidschrieberg1/2017/03/23/u-s-u-k-boycott-of-google-and-youtube-by-major-advertisers-spreading-over-hate-content/#5f3a9a07eb93.

[142] Kaur, *supra* note 140. Pivotal Research Group downgraded the shares of YouTube owner Alphabet from "Buy" to "Hold" because "Google wasn't taking the problem seriously." *Id.*

[143] Kevin Munger, *Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment*, Pol. Behav. (Nov. 22, 2016), https://link.springer.com/content/pdf/10.1007%2Fs11109-016-9373-5.pdf.

[144] *Id.* ("When generating the bots, I chose handles that consisted of first and last names that were identifiably male and white or black. . . . The only treatment that significantly decreased the rate of racist language use was the In-group/High Follower treatment.").

[145] Id.

number of media outlets to disable user comments on their websites. In 2015, *Bloomberg*, *The Verge*, *The Daily Beast*, and *Motherboard* (*Vice*'s technology and science news site) all removed their sites' comments sections.[146] In April 2016, law blog *Above the Law* made the same decision, citing increasing "abuse and insult" in the section, including regular "racist, sexist and homophobic attacks.[147] The same month, the *Guardian* examined the comments directed toward its regular opinion writers and found that "[t]he 10 regular writers who got the most abuse were eight women (four white and four non-white) and two black men. Two of the women and one of the men were gay. And of the eight women in the 'top 10', one was Muslim and one Jewish."[148] All ten writers who received the fewest abusive comments were men.[149]

Under the Communications Decency Act, publishers cannot be liable in the United States for third-party comments on articles.[150] Yet moderating comments more carefully—or removing them altogether—might nonetheless be in the best interest of online publishers. An informal experiment by Adam Felder, of the *Atlantic*, found that "[r]espondents who saw comments evaluated the article as being of lower quality—an 8 percent difference."[151] Moreover, after the

---

[146] Klint Finley, *A Brief History of the End of the Comments*, WIRED (October 6, 2015), https://www.wired.com/2015/10/brief-history-of-the-demise-of-the-comments-timeline.
[147] David Lat, *Comments Are Making the Internet Worse. So We Got Rid of Them*, Wash. Post (Apr. 21, 2016), https://www.washingtonpost.com/posteverything/wp/2016/04/21/comments-are-making-the-internet-worse-so-i-got-rid-of-them.
[148] Becky Gardiner et al., *The Dark Side of Guardian Comments*, GUARDIAN (Apr. 12, 2016), https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments?CMP=share_btn_tw. The study analyzed 70 million comments posted between January 4, 1999, and March 2, 2016, and used comments that were blocked for violating the site's community standards as a proxy for "abuse and/or disruption." Mahana Mansfield, *How We Analysed 70m Comments on the* Guardian *Website*, GUARDIAN (April 12, 2016), https://www.theguardian.com/technology/2016/apr/12/how-we-analysed-70m-comments-guardian-website ("Although mistakes sometimes happen in decisions to block or not block, we felt the data set was large enough to give us confidence in the findings.").
[149] Id.
[150] 47 U.S.C. § 230(c)(1) ("No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."); *see also* Lat, *supra* note 147.
[151] Adam Felder, *How Comments Shape Perceptions of Sites' Quality—And Affect Traffic*, ATLANTIC (June 5, 2014), https://www.theatlantic.com/technology/archive/2014/06/internet-comments-and-perceptions-of-quality/371862 ("I ran a quick study using respondents from Amazon's crowdsourcing platform Mechanical Turk. I asked 100 Americans to read a snippet of a *National Journal* article from late April. Half of them saw the article alone. The other half saw the article along with a representative sample of actual comments.").

National Journal removed its comments section, it experienced a notable increase in site traffic.[152] In February 2017, Google, alongside technology incubator Jigsaw, launched an AI tool, called Perspective, that uses machine learning to iteratively identify "toxic" comments."[153] In June 2017, the *New York Times* adopted the tool, announcing that it hoped to eventually allow comments on 80% of its online articles, up from 10% when it made the announcement.[154] Often, news sites that do close their comments sections direct conversation to social media, instead.[155]

E. Foreign Interference in Domestic Campaigns

Like echo chambers, fake news, bots, and hate speech, online interference in elections by foreign governments received increased attention in the context of the 2016 U.S. presidential election. In July 2016, WikiLeaks released approximately 20,000 hacked emails from the Democratic National Committee (DNC), revealing the committee's preference for Clinton over Senator Bernie Sanders, her rival in the Democratic primary election.[156] Later, in the weeks preceding the election, WikiLeaks released over 58,000 emails hacked from Clinton campaign chairman John Podesta's private email account.[157] WikiLeaks founder Julian Assange has denied that the emails came from the Russian government[158]; however, in January 2017, State Department spokesman John Kirby announced that the U.S. intelligence community is "100%

---

[152] *Id.* ("Pages views per visit increased by more than 10 percent. Page views per unique visitor increased 14 percent. Return visits climbed by more than 20 percent. Visits of only a single page decreased, while visits of two pages or more increased by almost 20 percent.").
[153] Carson, *supra* note 114.
[154] Benjamin Mullin, *The* New York Times *Is Teaming Up with Alphabet's Jigsaw to Expand Its Comments*, Poynter (June 13, 2017), http://www.poynter.org/2017/the-new-york-times-is-teaming-up-with-googles-jigsaw-to-expand-its-comments/463135.
[155] *See* Clothilde Goujard, *Why News Websites Are Closing Their Comments Sections*, Medium (Sept. 8, 2016), https://medium.com/global-editors-network/why-news-websites-are-closing-their-comments-sections-ea31139c469d.
[156] Tal Kopan, *WikiLeaks Releases More DNC Emails Near Eve of Election*, CNN (Nov. 6, 2016), https://www.washingtonpost.com/politics/clinton-blames-putins-personal-grudge-against-her-for-election-interference/2016/12/16/12f36250-c3be-11e6-8422-eac61c0ef74d_story.html?utm_term=.d09a5b25ef61.
[157] Id.
[158] Euan McKirdy, *WikiLeaks' Assange: Russia Didn't Give Us Emails*, CNN (Jan. 4, 2017), http://www.cnn.com/2017/01/04/politics/assange-wikileaks-hannity-intv/index.html.

certain in the role that Russia played" in hacking surrounding the election.[159] A secret Central Intelligence Agency (CIA) assessment concluded that "Russia intervened in the 2016 election to help Donald Trump win the presidency, rather than just to undermine confidence in the U.S. electoral system."[160]

According to a declassified version of the report, "We [the intelligence community] assess with high confidence that Russian President Vladimir Putin ordered an influence campaign in 2016 aimed at the US presidential election, the consistent goals of which were to undermine public faith in the US democratic process, denigrate Secretary Clinton, and harm her electability and potential presidency."[161] The report describes a "multifaceted" campaign of interference, combining cyber espionage against American political organizations, including the DNC; intrusion into state and local electoral boards, for the purpose of researching election processes, technology, and equipment; propaganda efforts, including stories by state-owned Russian media, such as RT and Sputnik, and amplification by professional trolls.[162] The report concluded that Russian efforts to interfere in the 2016 election represented "a significant escalation" compared to similar operations in the past.[163] It also determined that Putin "most

[159] Mike Krever, *US Administration "100% Certain" About Russian Hacking*, CNN (Jan. 4, 2017), http://www.cnn.com/2017/01/03/politics/russia-trump-hacking-john-kirby-amanpour/index.html ("'There's no question' about what Russia did to 'sow doubt and confusion, and getting involved through the cyber domain, into our electoral process,' he told Christiane Amanpour on Tuesday.").
[160] Adam Entous et al., *Secret CIA Assessment Says Russia Was Trying to Help Trump Win White House*, WASH. POST (Dec. 9, 2016), https://www.washingtonpost.com/world/national-security/obama-orders-review-of-russian-hacking-during-presidential-campaign/2016/12/09/31d6b300-be2a-11e6-94ac-3d324840106c_story.html?hpid=hp_hp-top-table-main_russiahack-745p%3Ahomepage%2Fstory&tid=a_inl&utm_term=.7161bfc09e8d; *see also U.S. Intelligence Report Identifies Russians Who Gave DNC Emails to Wikileaks*, TIME (Jan. 5, 2017), http://time.com/4625301/cia-russia-wikileaks-dnc-hacking ("In some cases, one official said, the material followed what was called 'a circuitous route' from the GRU, Russia's military intelligence agency, to Wikileaks in an apparent attempt to make the origins of the material harder to trace . . .").
[161] Natl. Intelligence Council, ASSESSING RUSSIAN ACTIVITIES AND INTENTIONS IN RECENT US ELECTIONS 1 (Jan. 6, 2017), https://www.dni.gov/files/documents/ICA_2017_01.pdf.
[162] *Id.* at 2-5.
[163] *Id.* at 5.

likely wanted to discredit Secretary Clinton" because he blamed her for inciting protests against his regime in 2011-2012.[164]

In June 2017, a report by the National Security Agency revealed that Russians attempted to gain access to voter-registration database information in advance of the 2016 election.[165] The report also concluded that hackers may have received access to vote-counting machinery, "particularly in states with electronic voting machines (which happen to include Pennsylvania and Wisconsin, two of the three states that decided the presidential election)."[166] To gain access, Russia used a "spear-fishing scheme"; hackers sent emails to state and local election offices prompting officials to provide their user credentials, and "at least one of the employee accounts was likely compromised."[167] The hackers then used information from the compromised account to launch another phishing attack against 122 local election officials.[168] According to intelligence officials, "Russia's attacks did not change any actual votes in the 2016 race"; nonetheless, voting security experts expressed concerns that if voting system vendors are hacked, malware could ultimately be spread to individual voting machines.[169] Moreover, even unsuccessful cyberattacks may undermine public confidence in the legitimacy of election results.[170]

The Obama administration learned of possible Russian cyberattacks in advance of the election, and on October 7, 2016, the White House formally blamed the Russian government for

---

[164] *Id.* at 1.
[165] Ed Kilgore, *Leaked NSA Report Suggests Russian Hacking Could Have Affected Election Day Itself*, N.Y. MAG. (June 5, 2017, 6:52 PM), http://nymag.com/daily/intelligencer/2017/06/leaked-nsa-report-suggests-russia-could-have-hacked-election.html.
[166] Id.
[167] Pam Fessler, *Report: Russia Launched Cyberattack on Voting Vendor Ahead of Election*, NPR (June 5, 2017, 9:00 PM ET), http://www.npr.org/2017/06/05/531649602/report-russia-launched-cyberattack-on-voting-vendor-ahead-of-election.
[168] Id.
[169] Id.
[170] *Id.* ("That could happen, for example, if voters showed up at the polls to find that their names were not listed or listed incorrectly.").

the DNC hacks.[171] The administration decided against more aggressive action, however, because it feared being perceived at trying to influence the outcome of the election in favor of Clinton.[172]

Trump himself has wavered in his assessment of the intelligence. Of December 2016, he had "consistently dismissed" it;[173] in January 2017, he seemed to acknowledge its veracity during a press conference,[174] but in a series of tweets in June 2017, he again dismissed the suggestion of interference as "a big Dem HOAX!"[175] In May 2017, former FBI director Robert Mueller was appointed as special counsel to investigate ties between the Trump campaign Russia.[176] In June 2017, former Federal Bureau of Investigation (FBI) director James Comey testified before Congress that Trump had asked him to "let[] . . . go" an investigation of former national security adviser Michael Flynn's ties to Russia.[177] Although Putin previously denied Russian involvement in the cyberattacks, on June 1, 2017, he acknowledged the possibility that "patriotically minded" Russian hackers may have been involved, while still denying that his government played any role.[178]

Since the 2016 U.S. presidential election, Russia has attacked France's computer networks during its May 2017 presidential election, attempting to undermine centrist candidate

---

[171] Emmarie Huetteman, *Obama White House Knew of Russian Election Hacking, but Delayed Telling*, N.Y. TIMES (June 21, 2017), https://www.nytimes.com/2017/06/21/us/politics/jeh-johnson-testimony-russian-election-hacking.html.

[172] *Id.* (""We were very concerned that we not be perceived as taking sides in the election, injecting ourselves into a very heated campaign or taking steps to delegitimize the election process and undermine the integrity of the election process," [former Secretary of Homeland Security Jeh Johnson] said.").

[173] Entous et al., *supra* note 160.

[174] David A. Graham, *Trump is a Russian-Interference Truther Once More*, ATLANTIC (June 22, 2017), https://www.theatlantic.com/politics/archive/2017/06/trumps-denial-of-russian-interference/531243.

[175] Id.

[176] Rebecca R. Ruiz & Mark Landler, *Robert Mueller, Former F.B.I. Director, Is Named Special Counsel for Russia Investigation*, N.Y. TIMES (May 17, 2017), https://www.nytimes.com/2017/05/17/us/politics/robert-mueller-special-counsel-russia-investigation.html.

[177] Stephen Collinson et al., *James Comey Testimony: Trump Asked Me to Let Flynn Investigation Go*, CNN (June 8, 2017), http://www.cnn.com/2017/06/07/politics/james-comey-testimony-released/index.html.

[178] Andrew Higgins, *Maybe Private Russian Hackers Meddled in Election, Putin Says*, N.Y. TIMES (June 1, 2017), https://www.nytimes.com/2017/06/01/world/europe/vladimir-putin-donald-trump-hacking.html?_r=0 ("'If [hackers] are patriotically minded, they start making their contributions — which are right, from their point of view — to the fight against those who say bad things about Russia,' Mr. Putin added, apparently referring to Hillary Clinton.").

(and now-President) Emmanuel Macron.[179] Admiral Mike Rogers, Director of the NSA, told the Senate Armed Services Committee that the U.S. became aware of the hack and alerted French officials.[180]

Of course, election hacking is not the exclusive province of Russia. In March 2016, *Bloomberg Businessweek* published a profile of Andrés Sepúlveda, a Colombian national who described his role in hacking presidential elections in Nicaragua, Panama, Honduras, El Salvador, Colombia, Mexico, Costa Rica, Guatemala, and Venezuela.[181] Sepúlveda is currently serving a ten-year prison sentence for charges relating to his interference in the 2014 Colombian presidential election.[182] For his first job, in 2005, Sepúlveda reports earing $15,000 to hack into the email database of Colombia's then-President Alvaro Uribe's opponent and "spamming the accounts with disinformation."[183] In 2012, he claims to have worked on behalf of Peña Nieto's successful presidential campaign in Mexico.[184] That operation reportedly involved tapping the phones and computers of opponents and their campaign staff, allowing Sepúlveda to view drafts of speeches and campaign schedules; the purchase of $50,000 software from Russia that allowed Sepúlveda to tap Apple, BlackBerry, and Android phones easily; 30,000 Twitter bots, which he used to instill fear that an opponent's election would depreciate the peso; and tens of thousands of 3 a.m., automated phone calls that purported to come from left-wing candidate in the swing

[179] *US Gave France "A Heads Up" Over Russian Hacking of Presidential Election*, TELEGRAPH (LONDON) (May 9, 2017), http://www.telegraph.co.uk/news/2017/05/09/us-gave-france-heads-russian-hacking-presidential-election ("The election commission said the leaked data apparently came from Macron's 'information systems and mail accounts from some of his campaign managers' - a data theft that mimicked Russian hacking of the Democratic National Committee in the 2016 U.S. presidential election.").
[180] Id.
[181] Jordan Robertson et al., *How to Hack an Election*, BLOOMBERG BUSINESSWEEK (March 31, 2016), https://www.bloomberg.com/features/2016-how-to-hack-an-election.
[182] *Id.* (describing charges that include "use of malicious software, conspiracy to commit crime, violation of personal data, and espionage").
[183] Id.
[184] Id.

state of Jalisco, annoying voters; and false Facebook profiles of gay men supporting a conservative Catholic candidate, offending his base.[185] Sepúlveda reportedly received a budget of $600,000 to fund his work in support of Peña Nieto.[186]

II. PLATFORMS

A. GOOGLE/YOUTUBE

In mid-2006, one year after YouTube's founding, its team of ten moderators, known as The SQUAD (the Safety, Quality, and User Advocacy Department) reviewed videos according to a single criterion: the moderators were instructed to ask themselves, "Can I share this video with my family."[187] In 2006, YouTube's internal counsel prepared a six-page booklet of guidelines for moderators.[188] In 2007, for the first time, YouTube promulgated "clearly articulated rules for users," prohibiting videos of pornography, criminal acts, gratuitous violence, threats, spam, and hate speech.[189] In 2009, during the Arab Spring, the SQUAD created an ad hoc newsworthiness exception, in response to the video of pro-government forces killing twenty-six-year-old Neda Agha-Soltan.[190]

Today, YouTube has over one billion users and reaches more eighteen- to forty-nine-year-olds than any American cable network.[191] Its content restrictions are more nuanced,[192] but the platform faces many of the same challenges as it did in its early years: as a platform that

---

[185] Id.
[186] Id.
[187] Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet: The Murky History of Moderation, and How It's Shaping the Future of Free Speech*, VERGE, https://www.theverge.com/2017/6/28/15886588/facebook-hate-speech-rules-broken-white-men-black-children (last visited July 10, 2017).
[188] Id.
[189] Id.
[190] *Id.* ("The maneuvers that allowed the content to stand took less than a day.").
[191] *YouTube for Press: Global Reach*, YOUTUBE, https://www.youtube.com/yt/about/press (last visited July 10, 2017).
[192] *Community Guidelines*, YOUTUBE, https://www.youtube.com/yt/policyandsafety/communityguidelines.html (last visited July 10, 2017).

exclusively hosts video content, YouTube is particularly vulnerable to terrorist videos, including videos of beheadings.[193] According to federal officials, many of the 47 Americans who were prosecuted in federal court between January and September 2015 for their support of ISIS were radicalized online, frequently after viewing recruitment videos.[194] As discussed above in Part I.D, YouTube's inability to prevent paid advertisements from running alongside terrorist propaganda videos cost the company an estimated sum in the hundreds of millions, if not the billions, and Europe threatens fines if YouTube falls out of compliance with its regulations.

Likely driven, at least in part, by this profit motive,[195] YouTube has created new initiatives for content moderation. In June 2017, the platform announced a four-step plan to fight online terror.[196] First, the company will invest more in developing artificial intelligence (AI) software, "to identify and remove extremist content."[197] This step may prove particularly important, given the rate at which new content is uploaded to the site.[198] Even before the announcement, YouTube used AI software to identify videos that were previously posted and removed.[199] Second, YouTube will expand its Trusted Flagger program, adding fifty expert non-

---

[193] For example, a single pro-ISIS media group, Fursan al-Rafa, or The Upload Knights, posted 515 links across Google's services (YouTube, Google Drive, and Google Photos) in just ten days. Rita Katz, *How Terrorists Slip Beheading Videos Past YouTube's Censors*, Motherboard (May 26, 2017), https://motherboard.vice.com/en_us/article/xyepmw/how-terrorists-slip-beheading-videos-past-youtubes-censors.
[194] Jaweed Kaleem, *YouTube's Battle Against ISIS*, Huffington Post (Sept. 9, 2015), http://www.huffingtonpost.com/entry/youtube-battle-against-isis_us_55d61416e4b0ab468da037a6.
[195] *See* Arjun Kharpal, *Google Outlines 4 Steps To Tackle Terrorist-Related Content on YouTube*, CNBC (June 19, 2017), http://www.cnbc.com/2017/06/19/google-youtube-tackles-terrorist-videos.html ("It comes after internet firms have received criticism from politicians about not dealing with extremist content quick enough.").
[196] Travis M. Andrews, *YouTube Announces Plan "To Fight Online Terror," Including Making Incendiary Videos Difficult to Find*, WASH. POST (June 19, 2017), https://www.washingtonpost.com/news/morning-mix/wp/2017/06/19/youtubes-announces-plan-to-fight-online-terror-including-making-incendiary-videos-difficult-to-find/?utm_term=.be575e44a53b.
[197] Kharpal, *supra* note 195.
[198] *See* Andrews, *supra* note 196 ("[A]s of 2012, one hour of content is uploaded to the platform each second . . . [which] makes a century of video every 10 days.").
[199] Sam Schechner, *Facebook Boosts AI to Block Terrorist Propaganda*, WALL STREET J. (June 15, 2017, 3:57 PM ET), https://www.wsj.com/articles/facebook-boosts-a-i-to-block-terrorist-propaganda-1497546000 ("[M]ore than half of the content removed for terrorism in the last six months was removed at least in part using such technology.").

governmental organizations.[200] Third, edge cases—videos that have inflammatory or supremacist content but fall short of violating YouTube's content restrictions—"will appear behind a warning, will not be monetized, recommended, or even eligible for users to make comments on."[201] Finally, YouTube will partner with Jigsaw to use ad targeting to redirect users who search for ISIS videos to anti-terrorist videos, with the objective of "chang[ing] their mind[s] about joining extremist organizations."[202]

YouTube's parent company, Google, faces its own set of challenges. Just as European governments have more aggressively policed fake news and hate speech on social media, western democracies outside the United States have taken a more aggressive approach, overall, to regulating misleading or embarrassing Google search results. In 2011, an Italian court affirmed a decision ordering Google "to filter out libellous [sic] search suggestions."[203] And in 2012, an Australian court held Google liable for producing a photo of the plaintiff as a Google image search result in a context that implied that he was connected with organized crime.[204]

On the solution side, Google News Lab partnered with First Draft News in February 2017 to create CrossCheck, a "collaborative journalism verification project."[205] In advance of the 2017 French presidential election, CrossCheck solicited seventeen French newsrooms to help fact-

---

[200] Kharpal, *supra* note 195 ("Google said Trusted Flagger reports are accurate over 90 percent of the time.").
[201] *Id.*
[202] *Id.* ("Google said that in previous trials of this system, potential recruits have clicked through on the ads at an 'unusually high rate' and watched over half a million minutes of video content that 'debunks terrorist recruiting messages.'").
[203] David Meyer, *Google loses autocomplete defamation case in Italy*, ZDNET (Apr. 5, 2011), http://www.zdnet.com/article/google-loses-autocomplete-defamation-case-in-italy. The anonymous plaintiff brought suit because his name was accompanied with autocomplete suggestions meaning "con man" and "fraud" in Italian. *Id.* However, the same court ruled in favor of Google in a case presenting similar facts in 2013. Marco Bellezza & Federica De Santis, *Google Not Liable for Autocomplete and Related Search Results, Italian Court Rules*, CGCS MEDIA WIRE (Apr. 22, 2013), http://www.global.asc.upenn.edu/google-not-liable-for-autocomplete-and-related-search-results-italian-court-rules.
[204] *Trkulja v Google Inc. LLC [No. 5]* (2012) VSC 533 (Austl.). The plaintiff was awarded $200,000 in damages. *Id.*
[205] *French Newsrooms Unite to Fight Election Misinformation with the Launch of CrossCheck*, FIRST DRAFT NEWS (Feb. 6, 2017), https://firstdraftnews.com/crosscheck-launches.

check.[206]

B. FACEBOOK

In May 2016, Facebook attracted criticism for allegedly manipulating its News Feed algorithm to suppress conservative news[207]—although this rumor was not substantiated.[208] Later that year, the platform came under scrutiny for its role in Trump's electoral victory, particularly through the proliferation of fake news on the platform.[209] The controversy brought a fundamental question to the fore: what were Facebook's roles and corresponding responsibilities? Those of a social media platform, that merely facilitates its users' sharing of personal content, or those of a news organization, with a duty to present balanced information?[210]

Certainly, Facebook plays a significant role in providing news to its users. According to a

---

[206] *Id.*

[207] *See* Michael Nunez, *Former Facebook Workers: We Routinely Suppressed Conservative News*, GIZMODO (May 9, 2016), http://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006 ("'It was absolutely bias. . . . It just depends on who the curator is and what time of day it is,"'said the former curator. 'Every once in awhile [sic] a Red State or conservative news source would have a story. But we would have to go and find the same story from a more neutral outlet that wasn't as biased.'"); *see also* Philip Bump, *Did Facebook bury conservative news? Ex-staffers say yes.*, WASH. POST (May 9, 2016), https://www.washingtonpost.com/news/the-fix/wp/2016/05/09/former-facebook-staff-say-conservative-news-was-buried-raising-questions-about-its-political-influence/?utm_term=.e3996ffb12be (citing the *Gizmodo* blog post).

[208] Snopes flagged the claim (that "Facebook routinely suppresses conservative news in favor of liberal content") as "unverified." *The Algorithm Is Gonna Get You*, SNOPES (May 13, 2016), http://www.snopes.com/is-facebook-censoring-conservative-news.

[209] *See, e.g.*, Issie Lapowsky, *Here's How Facebook* Actually *Won Trump the Presidency*, WIRED (Nov. 15, 2016), https://www.wired.com/2016/11/facebook-won-trump-election-not-just-fake-news (quoting Trump's digital director, Brad Parscale, saying, "Facebook and Twitter were the reason we won this thing. . . . Twitter for Mr. Trump. And Facebook for fundraising."); Parmy Olson, *How Facebook Helped Donald Trump Become President*, FORBES (Nov. 9, 2016), https://www.forbes.com/sites/parmyolson/2016/11/09/how-facebook-helped-donald-trump-become-president/#6deaf6cc59c5 ("[Facebook] has helped divide family, friends and acquaintances into increasingly-concrete silos of opinion, stoked to irrational levels of fear and anger with fake news and conspiracy theories from sites like Briebart. 'A cloud of nonsense,' as Pres. Obama himself put it. The reason is simple: Facebook's hyper-personalized News Feed."); Rich McCormick, *Donald Trump says Facebook and Twitter 'helped him win,'* VERGE (Nov. 13, 2016), https://www.theverge.com/2016/11/13/13619148/trump-facebook-twitter-helped-win ("Mark Zuckerberg has spent the days since Donald Trump was voted in as 45th president of the United States downplaying Facebook's role in the election, but that position may be harder to maintain now, as Trump himself has identified Facebook as a key element in helping him secure victory.").

[210] *See Facebook's Fake News Problem: What's Its Responsibility?*, CHI. TRIBUNE (Nov. 15, 2016), http://www.chicagotribune.com/bluesky/technology/ct-facebook-fake-news-20161115-story.html ("Facebook is under fire for failing to rein in fake and biased news stories that some believe may have swayed the presidential election. Its predicament stems from this basic conundrum: It exercises great control over the news its users see, but it declines to assume the editorial responsibility that traditional publishers do.").

2016 survey analysis by Pew Research Center, approximately two-thirds of American adults use Facebook, and roughly two-thirds of Facebook users receive news through the site; thus, around 44% of American adults use Facebook as a news source.[211] Across twenty-six countries, the percentage of surveyed people who reported using Facebook for news was the same: 44%.[212]

In the immediate aftermath of the election, Facebook was reluctant to accept any responsibility for influencing its outcome—even indirectly. On November 12, 2016, Mark Zuckerberg, Facebook's founder, posted from his Facebook account, "Of all the content on Facebook, more than 99% of what people see is authentic. . . . The hoaxes that do exist are not limited to one partisan view, or even to politics. Overall, this makes it extremely unlikely hoaxes changed the outcome of this election in one direction or the other."[213] Moreover, in a subsequent post, on November 19, 2017, Zuckerberg emphasized the difficulty of accurately identifying fake news on the platform.[214] Earlier that month, however, four undergraduate students developed a Chrome browser extension to tag news stories on Facebook as "verified" or "not verified," using artificial intelligence.[215] They developed the extension, called "FiB: Stop living a lie," during a thirty-six-hour hackathon hosted by Princeton University.[216] The open-source extension categorizes all posts; for links, it "take[s] into account the website's reputation, also quer[ies] it against malware and phishing websites database and also take[s] the content, search[es] it on

---

[211] Jeffrey Gottfried & Elisa Shearer, *News Use Across Social Media Platforms 2016*, PEW RES. CTR. (May 26, 2016), http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016. The percentage of Facebook users who receive news through the site was up from only 47% in 2013. *Id.*
[212] Omidyar, *supra* note 121, at 3.
[213] Mark Zuckerberg, FACEBOOK (Nov. 12, 2016), https://www.facebook.com/zuck/posts/10103253901916271.
[214] Mark Zuckerberg, FACEBOOK (Nov. 19, 2016), https://www.facebook.com/zuck/posts/10103269806149061 ("The problems here are complex, both *technically* and philosophically." (emphasis added)).
[215] Julie Bort, *It Took Only 36 Hours for These Students to Solve Facebook's Fake-News Problem*, BUS. INSIDER (Nov. 14, 2016), http://www.businessinsider.com/students-solve-facebooks-fake-news-problem-in-36-hours-2016-11.
[216] Id.

Google/Bing, [and] retrieve[s] searches with high confidence."[217]

In December 2016, Facebook announced a new program designed to combat the spread of fake news on the platform.[218] The system allows users to flag possible fake news stories, which are then sent to fact-checking organizations—including ABC News, AP, FactCheck.org, Politifact, and Snopes—for verification.[219] Stories that fail the fact check receive a flag, with a link to a page that explains to users why the source is disputed.[220] Users who attempt to share flagged stories receive a second warning.[221] Finally, Facebook's News Feed algorithm may deprioritize flagged stories.[222] Deliberately, Facebook stopped short of blocking the debunked stories outright, "to avoid . . . censorship."[223]

In some instances, however, Facebook's fact-checking feature has had the opposite effect of that which the platform intended. When Facebook flagged a Newport Buzz article that falsely reported that "hundreds of thousands of Irish people were brought to the US as slaves," conservative groups publicized the post deliberately, leading site traffic to "skyrocket[]."[224] According to Jestin Coler, a publisher of fake news, "A far-right individual who sees it's been disputed by Snopes, that adds fuel to the fire and entrenches them more in their belief."[225] Indeed, social science supports the notion that "subsequent rebuttals may actually work to

---

[217] Id.
[218] Amber Jamieson & Olivia Solon, *Facebook To Begin Flagging Fake News in Response to Mounting Criticism*, GUARDIAN (Dec. 15, 2016), https://www.theguardian.com/technology/2016/dec/15/facebook-flag-fake-news-fact-check.
[219] *Id.* The fact checking organizations are not compensated for this service. *Id.*
[220] Id.
[221] Id.
[222] Id.
[223] David Pogue, *What Facebook Is Doing to Combat Fake News*, SCI. AM. (Feb. 1, 2017), https://www.scientificamerican.com/article/pogue-what-facebook-is-doing-to-combat-fake-news/.
[224] Sam Levin, *Facebook Promised to Tackle Fake News. But the Evidence Shows It's Not Working*, GUARDIAN (May 16, 2017), https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working.
[225] Id.

reinforce the original misinformation, rather than to dissipate it."[226] Moreover, stories that have been debunked by fact-checking groups still appear on Facebook without the disputed tag, and when Facebook does apply the tag, it often does so after the story has already "gone viral."[227]

In the context of hate speech, ProPublica recently revealed aspects of Facebook's censorship rules, culled from internal documents, that may undermine the platform's effectiveness in curbing hate speech.[228] While protected categories, including "race, sex, gender identity, religious affiliation, national origin, ethnicity, sexual orientation and serious disability/disease," are protected, subgroups of these categories are not.[229] A training document plainly shows that this translates into protection of white men but not of female drivers or black children.[230]

As of June 27, 2017, Facebook employed 4,500 moderators; it plans to add 3,000 within the next year.[231] But technology writer Adi Robertson wrote in response to the announcement, "that's still minuscule for a platform so big."[232] As of March 31, 2017, Facebook reported 1.94

---

[226] CAN PUBLIC DIPLOMACY SURVIVE THE INTERNET?: BOTS, ECHO CHAMBERS, AND DISINFORMATION 8 (Shawn Powers & Markos Kounalakis, eds., May 2017), https://www.state.gov/documents/organization/271028.pdf (citing Christopher Paul & Miriam Matthews, *The Russian 'Firehood of Falsehood' Propaganda Model*, RAND CORP. PERSPS. (2016), http://www.rand.org/pubs/perspectives/PE198.html). Paul and Matthews also found that "people tend to believe something when it is repeated" and that "propagandists gain the advantage when they get to make the first impression." *Id. See also* Michela Del Vacario et al, *The spreading of misinformation online*, 113 PROC. OF THE NAT'L ACAD. OF SCI. OF THE U.S. 554, 558 (2015), http://www.pnas.org/content/113/3/554.full.pdf ("Many mechanisms cause false information to gain acceptance, which in turn generate false beliefs that, once adopted by an individual, are highly resistant to correction . . . .").

[227] *Id.* For example, a story from BeforeItsNews.com claiming that Obama practiced Islam in the White House and includes purportedly leaked photographs, which Snopes debunked, remains on Facebook without the disputed tag. *Id.*

[228] Angwin & Grassegger, *supra* note 134

[229] Id.

[230] *Id.* Similarly, poet and activist Didi Delgado's account was deactivated for seven days after she posted, "All white people are racist," while U.S. Representative Clay Higgins of Louisiana was permitted to post a call to violence against "'radicalized' Muslims": "Kill them all. For the sake of all that is good and righteous. Kill them all." *Id.* Under Facebook policy, white people are a protected group, while radicalized Muslims are a subset of a group and, therefore, are not protected.

[231] Id.

[232] Adi Robertson, *Facebook Explains Why It's Bad at Catching Hate Speech*, VERGE (June 27, 2017), https://www.theverge.com/2017/6/27/15879232/facebook-hate-speech-moderation-hard-questions.

billion monthly active users.[233]

Facebook has recently expanded its use of AI technology for reviewing content.[234] Facebook cites advances in the technology as a factor in its increased willingness to turn to automation, although AI remains limited in its capacity.[235] Recently, for example, Facebook's AI technology developed the ability to distinguish between content that warrants a definitive response and content that requires escalation to a human reviewer.[236] Nonetheless, Richard Allan, Facebook's Vice President of Public Policy for Europe, the Middle East, and Africa wrote, "we're a long way from being able to rely on machine learning and AI to handle the complexity involved in assessing hate speech."[237]

C. TWITTER

Compared to Facebook, Twitter is particularly susceptible to bots, because, unlike Facebook, it does not require users to provide their real names.[238] A 2017 study by researchers at Indiana University and the University of Southern California estimate that between 9% and 15% of active, English-speaking Twitter accounts are bots.[239] This percentage translates to as many as 50 million bot accounts.[240] The number of Twitter bots has grown "exponential[ly]" since

---

[233] *Stats*, FACEBOOK NEWSROOM, https://newsroom.fb.com/company-info (last visited July 10, 2017).
[234] Schechner, *supra* note 199 ("The firm's moves reflect a growing willingness to trust machines to help enen in part with thorny tasks like distinguishing inappropriate content from satire or news coverage—something firms resisted after a spate of attacks just two years ago as a potential threat to free speech.").
[235] *Id.* ("While an Isis-propaganda photo posted without a caption may be an easy removal for an algorithm, the same image with a caption might for instance require human review, said Monika Bickert, Facebook's head of global policy management.").
[236] Id.
[237] Allan, *supra* note 137
[238] Id.
[239] Onur Varol et al., *Online Human-Bot Interactions: Detection, Estimation, and Characterization*, 280 AAAI CONF. ON WEB & SOC. MEDIA 288 (May 3, 2017), https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817. "Bot behaviors" include features linked to "friends, tweet content and sentiment, network patterns, and activity time series" and "reciprocity of friendship ties," given that "[h]umans tend to interact with more human-like accounts than bot-like ones, on average. *Id.* at 280-81.
[240] Omidyar, *supra* note 121, at 6.

2008.[241] However, observers have perceived the platform as unwilling to address its bot problem; some hypothesize that Twitter lacks an incentive to do so, because actively shutting down nonhuman accounts would dramatically reduce the size of its user base.[242]

Although Facebook receives greater attention in the realm of fake news, *New York Times* columnist Farhad Manjoo wrote that although Twitter "is far smaller than Facebook,"[243] it plays a disproportionately significant role in the spread of fake news online.[244] According to researcher Alice Marwick, "When journalists see a story getting big on Twitter, they consider it a kind of responsibility to cover it, even if the story may be an alternate frame or a conspiracy theory . . . That's because if they don't, they may get accused of bias."[245]

Like Facebook, Twitter has so far been reluctant to delve deeply into fact-checking.[246] Nonetheless, in June 2017, the *Washington Post* reported that Twitter was "exploring" a feature that would allow users to flag tweets containing links to fake news.[247] The announcement,

---

[241] Bence Kollanyi, *Where Do Bots Come From?: An Analysis of Bot Codes Shared on GitHub*, 10 INT'L J. COMM. 4932, 4937 (2016) ("In GitHub's first two years, 2008 and 2009, almost 100 different bot codes were published. The number has been growing ever since, and by 2013 it reached 1,000. In recent years, the number of bots has quadrupled, and at the time of this study (April 2016) there are more than 4,000 repositories.").
[242] *See, e.g.*, Zach Whittaker, *Twitter Has a Spam Bot Problem—And It's Getting Worse*, ZDNet (Apr. 23, 2017), http://www.zdnet.com/article/twitter-spam-bot-problem-on-the-rise ("Call us cynical, but it's not unreasonable to assume Twitter -- of which its entire worth and value is based on its reported number of users -- wouldn't want to tackle the spam bot problem, for fear that it dramatically cuts a large swathe of its perceived user base.").
[243] As of June 30, 2016, Twitter reported approximately 328 million monthly active users. *Twitter Usage: Company Facts*, Twitter, https://about.twitter.com/company (last visited July 10, 2017).
[244] Farhad Manjoo, *How Twitter Is Being Gamed to Feed Misinformation*, N.Y. TIMES (May 31, 2017), https://mobile.nytimes.com/2017/05/31/technology/how-twitter-is-being-gamed-to-feed-misinformation.html?_r=0&referer=https://t.co/jc2AuhKN6a ("[I]n many of the biggest misinformation campaigns of the past year, Twitter played a key role.").
[245] Id.
[246] *Compare* Colin Crowell, *Our Approach to Bots & Misinformation*, TWITTER BLOG (June 14, 2017), https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html ("We, as a company, should not be the arbiter of truth."), *with* Zuckerberg, *supra* note 213 ("I believe we must be extremely cautious about becoming arbiters of truth ourselves.").
[247] Elizabeth Dwoskin, *Twitter Is Looking for Ways to Let Users Flag Fake News, Offensive Content*, WASH. POST (June 29, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/06/29/twitter-is-looking-for-ways-to-let-users-flag-fake-news/?utm_term=.4da5d5dd6682.

however, sparked an intense backlash among users.[248] Commentators emphasized that Facebook's fake news flagging feature has been less than completely successful.[249] The day after the *Washington Post* article was published, Twitter's spokesperson told *Business Insider* that the company "has 'no current plans to launch anything along the lines described.'"

As discussed above in Part I.D, Twitter has been less successful than its peer companies in removing hate speech from the platform—at least from the perspective of European regulators. Historically, the platform has stood firmly for free speech.[250] Critics have noted, however, that "not policing for abuse has a chilling effect on speech."[251] Like YouTube, content posted on Twitter has also played a role in terrorist recruitment. ISIS and its supporters post approximately 90,000 tweets per day,[252] and the group uses hashtags to focus on "group messaging and branding concepts."[253]

Rather than blocking more content outright, the company has tended toward a compromise approach, recently announcing new tools to "mute" notifications from new users,

---

[248] Maya Kosoff, *Twitter Users Officially Hate the Idea of a "Fake-News Button*, VANITY FAIR HIVE (June 30, 2017), http://www.vanityfair.com/news/2017/06/twitter-users-officially-hate-the-idea-of-a-fake-news-button ("Critics, many of them journalists, immediately panned the idea of a similar "fake-news button" on Twitter.").
[249] Jay McGregor, *Twitter Will Face Big Hurdles in Effort to Fight Fake News*, FORBES (June 30, 2017), https://www.forbes.com/sites/jaymcgregor/2017/06/30/twitter-will-face-big-hurdles-in-effort-to-fight-fake-news/#4f38ef502d2e ("[T]he [*Washington Post*] report also points out that Twitter has some reasonable concerns about how the system could be 'gamed' by users. This is something I touched on earlier this year when Facebook hurriedly launched its own anti fake news initiative.").
[250] Josh Halliday, *Twitter's Tony Wang: "We Are the Free Speech Wing of the Free Speech Party,"* GUARDIAN (Mar. 22, 2012), https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech ("The general manager of Twitter in the UK has said that the social network sees itself as 'the free speech wing of the free speech party.'").
[251] Kate Klonick, *You'll Never Guess This One Crazy Thing Governs Online Speech*, SLATE (Aug. 24, 2016), http://www.slate.com/articles/technology/future_tense/2016/08/free_speech_is_the_wrong_way_to_think_about_twitter_and_facebook.html (citing Victoria Turk, *The Chilling Effect of Misogynistic Trolls*, MOTHERBOARD (Aug. 22, 2014), https://motherboard.vice.com/en_us/article/ezvbpn/the-chilling-effect-of-misogynistic-trolls ("The 'ignore it' strategy essentially tells women to just shut up—which is exactly the underlying attitude of their attackers, and the cause of the problem in the first place.")).
[252] Kaleem, *supra* note 194.
[253] *How Terrorists Are Using Social Media*, TELEGRAPH (LONDON) (Nov. 4, 2014), http://www.telegraph.co.uk/news/worldnews/islamic-state/11207681/How-terrorists-are-using-social-media.html.

non-followers, and those whom the primary user does not follow.[254] Earlier in 2016, the platform unveiled similar tools for muting users without a profile picture and those who have not verified their email address or phone number.[255]

From a methodological perspective, Twitter, like YouTube and Facebook, relies in part on AI technology to flag content, particularly in the context of terrorist propaganda.[256] Between July and December 2016, the platform reported that "internal tools flagged 74% of the 376,890 accounts it removed."[257]

CONCLUSION

Echo chambers, fake news, computational propaganda, hate speech, and foreign interference in elections demonstrate clearly that the Internet and social media are far from purely democracy-supporting forces in society. It remains an open question how democracies can best address these challenges while maintaining their commitment to free speech and expression—allowing the Internet to function as a robust marketplace of ideas.[258]

---

[254] Josh Costine, *Twitter Lets You Avoid Trolls by Muting New Users and Strangers*, TECHCRUNCH (July 10, 2017), https://techcrunch.com/2017/07/10/twitter-mute/?ncid=mobilenavtrend.
[255] *Id.*
[256] Schechner, *supra* note 199.
[257] Id.
[258] *See Abrams v. United States*, 250 U.S. 616, 630 (Holmes, J., dissenting) (1919) ("[T]he best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out. That, at any rate, is the theory of our Constitution.").