Tackling Domestic Disinformation:

What the Social Media Companies Need to Do

PAUL M. BARRETT

Contents

Executive Summary

A growing amount of misleading and false content infests social media. A 2018 study by researchers at Oxford University found that 25 percent of Facebook and Twitter shares related to the midterm elections in the U.S. contained "junk news"— deliberately deceptive or incorrect information. A majority of this harmful content came not from Russia or other foreign state actors but from domestic U.S. sources.

> *The social media platforms are already making judgments about content when their algorithms rank and recommend posts, tweets, and videos.*

This report focuses on domestically generated disinformation in the U.S.: the nature and scope of the problem, what the social media platforms have done about it, and what more they need to do.

Domestic disinformation comes from disparate sources, including message boards, websites, and networks of accounts on Facebook, Twitter, and YouTube. This homegrown harmful content flows from both liberals and conservatives, but overall, it is predominantly a right-wing phenomenon. Social media researchers have said that highly partisan conservatives are more likely than highly partisan liberals to encounter and share disinformation.

Increasingly, social media platforms are removing disinformation from Russia and other foreign countries because of its fraudulent nature and potential to disrupt democratic institutions. In contrast, some commentators have argued that misleading content produced by U.S. citizens is difficult to distinguish from ordinary political communication protected by the First Amendment. According to this view, we shouldn't encourage the platforms to make judgments about what's true and untrue in politics.

But the platforms are already making similar judgments when their algorithms rank and recommend posts, tweets, and videos. They also remove certain categories of harmful content, such as harassment and hate speech. We urge them to add provably false information to the removal list, starting with content affecting politics or democratic institutions. The First Amendment, which precludes government censorship, doesn't constrict social media venues owned and operated by nongovernmental entities. The real question confronting the platforms is how to make their evaluations of factually questionable content more reasonably, consistently, and transparently.

Part One of this report provides an overview of the subject and our argument. We contend that the platforms ought to take a harder line on domestic disinformation, which pollutes the marketplace of ideas. Conspiracy theories, hate speech, and other untruths heighten popular cynicism and exacerbate political polarization. Given finite human attention, one scholar has noted, flooding social media with malign content actually suppresses the free exchange of ideas. Democracy suffers when policy decisions are based on fiction, rather than facts and rational argument.

Part Two describes various forms that domestic disinformation takes. We look at both right- and left-leaning websites connected to networks of Facebook and Twitter accounts. We also focus on the extraordinary case of President Donald Trump's use of Twitter, the right-wing affinity for YouTube, and conspiracy theories about philanthropist George Soros.

Part Three assesses steps the platforms have taken to address domestic disinformation. These include adjusting ranking algorithms to disfavor publishers generally; developing new artificial intelligence tools to identify potentially untrue material; hiring thousands of additional content monitors; introducing annotation features that give users more context when they encounter suspect content; battling malign bot networks; and, at times, removing disinformation and even banning its purveyors.

Part Four outlines our recommendations to the platforms and describes steps we think they need to take to intensify the fight against domestic disinformation.

A Word on Terminology

We refer to the domestically generated harmful content under scrutiny here as disinformation, by which we mean a relatively broad category of false or misleading "facts" that are intentionally or recklessly spread to deceive, radicalize, propagandize, promote discord, or make money via "clickbait" schemes. For the sake of variety, we also sometimes refer to false news, false information, and false content, intending these terms to reflect their ordinary meaning.

In our recommendations, we urge the social media companies to remove a narrower category of material—provably false content. Focusing on this more limited category will make the daunting task of identification and removal more feasible. Consider these hypothetical examples: A story consistent with the headline "The Holocaust Never Happened" is provably untrue and ought to be removed for that reason. By contrast, a story headlined "Democrats Secretly Favor Open Borders" may be unsubstantiated and misleading, but it isn't provably false.

Summary of Our Recommendations

1. **Remove false content, whether generated abroad or at home.**
   Content that's provably untrue should be removed from social media sites, not merely demoted or annotated.

2. **Clarify publicly the principles used for removal decisions.**
   The platforms need to explain the connection between facts, rational argument, and a healthy democracy.

3. **Hire a senior content overseer.**
   Each platform should bring in a seasoned executive who would have company-wide responsibility for combating false information.

4. **Establish more robust appeals processes.**
   The companies should provide a meaningful opportunity for appeal to a person or people not involved in the initial removal decision.

5. **Step up efforts to expunge bot networks.**
   The hunt for automated accounts that imitate human behavior online must be pursued with increased urgency.

6. **Retool algorithms to reduce the outrage factor.**
   Doing so would diminish the volume of falsehoods.

7. **Provide more data for academic research.**
   The platforms have an ethical and social responsibility to provide data they uniquely possess to facilitate studies of disinformation.

8. **Increase industry-wide cooperation.**
   No one company sees the problem in full, making it imperative for all of them to exchange data and analysis in an effort to address common challenges.

9. **Boost corporate support for digital media literacy**.
   Despite criticism of some literacy programs, teaching students and adults how to be more discriminating online should remain a priority.

10. **Sponsor more fact-checking and explore new approaches to news verification.**
    Fact-checkers don't provide a silver bullet, but their probing underscores the distinction between reality and unreality.

11. **Support narrow, targeted government regulation.**
    Sweeping content regulation would overreach; rules on political advertising and measuring the prevalence of disinformation would not.

1. Introduction

"
The amount of false material online is vast and growing. In the run-up to the 2018 midterms, fully 25% of Facebook and Twitter shares related to the elections contained 'junk news.'
"

On October 27, 2018, a Pittsburgh man named Robert Bowers posted his final tirade on Gab, a Twitter-like social media platform patronized by right-wing extremists. Boasting more than 700,000 users, Gab is an American-made venue catering to Americans who don't like liberals, Jews, or blacks. It served as an echo chamber for Bowers' view that immigrants pose a lethal threat, that Jews facilitate this threat, and that a caravan of Honduran migrants then moving toward the U.S. southern border constituted an "invasion" of the homeland.

Bowers fixated on an NGO called the Hebrew Immigrant Aid Society. "HIAS," he told his Gab audience, "likes to bring invaders in that kill our people. I can't sit by and watch my people get slaughtered. Screw your optics, I'm going in." Armed with a military-style rifle and three Glock handguns, he allegedly attacked the Tree of Life Congregation, a synagogue in suburban Pittsburgh, killing 11 people in the deadliest anti-Semitic assault in American history.[1]

The Tree of Life massacre took place in the immediate aftermath of another episode demonstrating the dangers of online radicalization. Cesar Sayoc Jr. had gone from posting on Facebook and Twitter about his meals and workout routines to declaring his indignation over immigration and Muslims. The south Florida resident posted stories that appeared originally on *Infowars, World Net Daily,* and other right-wing conspiracy sites that disseminate their paranoid ideas via social media. Scores of times, he retweeted a meme claiming that the February 2018 high school mass shooting in Parkland, Fla., had been staged by "crisis actors" and paid for by billionaire philanthropist George Soros. On October 26, Sayoc

was arrested and charged with mailing more than a dozen pipe bombs to Soros, other prominent Democrats, and CNN.[2]

Online fulmination grounded in phony information helped propel Bowers and Sayoc toward extreme action. This information didn't come from abroad; it wasn't part of a Russian campaign to sow discord. Instead, it was manufactured here at home in the United States—a small part of the domestic disinformation that spills across social media every day.

The amount of this false material is vast and growing, according to the Oxford Internet Institute, an arm of Oxford University. In a study released in November 2018, the institute found that in the 30-day run-up to U.S. midterm elections, fully 25% of Facebook and Twitter shares related to the midterms contained "junk news"—an increase of five percentage points from the 2016 U.S. presidential election season. By junk news, the Oxford team referred to deliberately "misleading, deceptive, or incorrect information purporting to be real news about politics, economics, or culture."[3]

> ❝ Domestic disinformation comes from disparate sources, including message boards, websites, and networks of Facebook pages and Twitter accounts, both human and automated. ❞

Addressing the source of junk news in the U.S. in the fall of 2018, Oxford researcher Lisa-Maria Neudert said: "It is domestic alternative-media outlets that are dominating the political debate on social media. What we are seeing is homegrown conspiracy theories and falsehoods."[4]

To be harmful, domestic content does not need to help stimulate crime or bloodshed. During a November 7, 2018, press conference, President Donald Trump sparred with CNN correspondent Jim Acosta over immigration. At one point, a White House intern tried to take the microphone from Acosta, who said, "Pardon me, ma'am," and kept it. That evening, Paul Joseph Watson, an editor of the *Infowars* conspiracy mill, tweeted an altered version of a video of Acosta's interaction with the intern. In the edited video, Acosta's arm movement was accelerated so he appeared to chop down forcefully on the White House aide's arm. Presidential Press Secretary Sarah Sanders subsequently tweeted a version of the video identical to Watson's, which presumably was meant to make Acosta seem more aggressive than he had been. (Watson said in a YouTube response that all he had done was "zoom in" and compress the video, which made it look "marginally different.") Sanders suspended Acosta's White House credentials, accusing him

of "putting his hands on a young woman." CNN went to court to get the suspension rescinded, but for several news cycles, the false accusation against Acosta reinforced the president's often-stated claim that the media—and especially CNN—are "the enemy of the people."[5]

The Acosta incident illustrates two salient themes about domestic disinformation: First, it often takes the form not of text articles, but memes—videos or still images, typically with punchy captions, designed to spread virally. Second, tweeting by President Trump, or in this instance, his spokeswoman, plays an extraordinary role in amplifying a wide array of misleading right-leaning content.

Turning to the Homefront

In July 2018, the NYU Stern Center for Business and Human Rights published "Combating Russian Disinformation: The Case for Stepping Up the Fight Online." The report examined the continuing threat of harmful content generated by proxies of Russian President Vladimir Putin and provided recommendations for how governments and the major social media platforms can do more to counter interference by the Kremlin.[6] Our latest report tackles another aspect of the harmful content problem—namely, falsehoods and distortions generated domestically. The two strains of untruth, one originating abroad, the other at home, bear some resemblance to one another but also have critical differences.

The Russian campaign—which has sought to inflame the electorate and undermine democracy—is, in a sense, easier to understand. One can represent it as a vector pointing directly from the St. Petersburg headquarters of the Internet Research Agency (IRA), a Kremlin-connected disinformation factory, toward Facebook, Twitter, YouTube, and Instagram. By now, the IRA may have morphed into other organizations using different names and different techniques. But Russian

disinformation continues to flow, and its purpose isn't difficult to comprehend: The Putin government seeks to destabilize democratic institutions, not just in the U.S., but in former Soviet republics like Ukraine and throughout Europe.[7] In contrast, domestic U.S. disinformation comes from disparate sources, including message boards, websites, and networks of Facebook pages and Twitter accounts, both human and automated. Domestic producers of false content don't have a unified aim comparable to the Russian mission.

Domestic disinformation comes from both the left and the right, but conservative Facebook and Twitter users are more likely than liberals to circulate false content. Using samples collected during a 90-day period in late 2017 and early 2018, the Oxford Internet Institute divided Twitter users into 10 groups, including Trump supporters, Democrats, and progressives. Trump supporters, the Oxford researchers found, "share[d] the widest range of known junk news sources and circulate[d] more junk news than all the other groups put together." Using slightly different groupings for their Facebook analysis, the Oxford team found that "extreme hard right pages—distinct from Republican pages—share the widest range of known junk news sources and circulate more junk news than all the other audiences put together."[8]

Researchers at the Berkman Klein Center for Internet & Society at Harvard University have come to similar conclusions about what they call "the central role of the radicalized right in creating the current crisis of disinformation and misinformation." In a 2018 book, the Berkman Klein team writes: "No fact emerges more clearly from our analysis of how four million political stories were linked, tweeted, and shared over a three-year period than that there is no symmetry in the architecture and dynamics of communications within the right-wing media ecosystem and outside of it."[9]

Alice Marwick, a social media scholar at the University of North Carolina, adds the important point that "it is not that Republicans are more credulous than Democrats" about disinformation. "It is that they are inhabiting an information system that is full of inaccurate information." Elaborating, Marwick writes: "Since the conservative media sphere is infested with disinformation, very partisan conservatives would then be more likely than very partisan liberals to share disinformation."[10]

Another difference between Russian and domestic U.S. false content is that once disguised Russian material is identified, the decision of what to do about it is not challenging: The social media platforms take it down. It seems unproblematic for these corporations to remove from their privately operated venues phony material generated by foreigners pretending to be American citizens and seeking to set American voters against one another. On at least three occasions in 2018, Facebook blocked hundreds of fake accounts originating in Russia and Iran. But the misleading output of U.S. citizens seems to present a thornier problem. The contentious domestic material "starts to look a lot like normal politics," says Alex Stamos, Facebook's chief security officer from 2015 through mid-2018. Now an adjunct professor at Stanford University, he adds: "I don't think we want to encourage the [platform] companies to make judgments about what's true and what's not in politics."[11]

As we'll see, however, the companies are already making judgments about falsehood, manipulativeness, and divisiveness. And they're making decisions that affect the viability of politically oriented websites and their affiliated social media accounts. So, the real question is how to make these evaluations more reasonably, consistently, and transparently.

We contend that the platforms ought to take a harder line on domestic disinformation, starting with false

content affecting politics and democratic institutions. This content doesn't fuel democracy; it contaminates democratic discourse. In extreme cases, such as those of Robert Bowers and Cesar Sayoc, it can create an environment that encourages violence. But even well short of those outlying situations, the combination of conspiracy theories, hate speech, and other un-truths heighten public cynicism and exacerbate the acute polarization that characterizes American politics today.

"Given our finite attention, flooding social media with junk is a way to *suppress* the free exchange of ideas," says Filippo Menczer, a professor of informatics and computer science at Indiana University who studies disinformation. "Democracy suffers as a consequence," he adds, "because critical policy decisions are based on fiction and emotion, rather than facts and rational arguments."[12]

What of the First Amendment? By its terms and according to judicial precedent, the First Amendment precludes government censorship—and wisely so. We wouldn't want the White House, Congress, or a regulator dictating what constitutes social media falsehood worthy of being demoted or deleted.

The First Amendment, however, doesn't inhibit organizations outside of government from making choices about what speech they sponsor. It doesn't prevent newspapers from selecting which articles to print or reject. And it doesn't inhibit the social media companies from choosing and ranking content.

Complicating this analysis, Facebook, Twitter, and YouTube—which is owned by Google—have historically insisted that they are mere platforms, not responsible for the content they display. In fact, they are somewhere in-between passive digital platforms and traditional publishers. They don't individually select each item they show users in the manner of *The New York Times.* But the algorithms they craft do sort billions of posts and tweets, inevitably making

choices about what content users see. And at times, applying a combination of software and human judgment, the companies already exclude content or ban users altogether. Our position is that they ought to take the next step and act more vigorously to diminish domestically generated false content.

Cast of Characters

A good deal of domestic disinformation bubbles up from ideological communication channels such as 4chan, a collection of message boards where users post anonymously. 4chan's Politically Incorrect board, notorious for its sexist and racist exchanges, often generates unsavory ideas that migrate elsewhere online. Individuals can also go to Gab, the right-wing Twitter equivalent, to express white nationalist ideas. The_Donald, a pro-Trump section of the Reddit website, serves as a particularly effective disseminator of inaccurate right-wing content. Disinformation from all of these sources frequently reaches wider audiences as it resurfaces on Facebook (2.3 billion monthly users), Instagram (800 million), Twitter (335 million), and YouTube (1.8 billion).

A similar pattern—material moving from the periphery to mainstream social media—holds for hyper-partisan right-wing websites like *Breitbart* and the even-more-extreme *Infowars*. Some of these outfits have employees, sell advertising, and hawk merchandise. *Breitbart* has enjoyed the financial backing of the billionaire hedge fund mogul Robert Mercer. Other hyper-partisan communities fester in private Facebook groups, where they can spew hateful attacks and conspiracy allegations with little, if any, outside scrutiny.

A closely related breed of conservative enterprises post sensationalistic "clickbait" headlines on Facebook and Twitter with the apparent goal of luring users to visit websites fueled by advertising. The clickbait headlines are often similar to those on more hardcore

> **We contend that the social media platforms ought to take a harder line on domestic disinformation. This content doesn't fuel democracy; it pollutes the marketplace of ideas.**

hyper-partisan sites. A site called *News Punch* exemplifies this type of operation, running articles such as "Trump: New Evidence Proves Bush & Clinton Orchestrated 9/11" and "Hillary Clinton Refuses to Deny Putin's Claim She Took $400 Million from Russia." Defending his approach, Sean Adl-Tabatabai, editor-in-chief of *News Punch,* says that "the use of clickbait headlines is a practice used by almost every major news outlet."[13]

Discerning the contours of the disinformation ecosystem isn't easy. There are hundreds of false news and conspiracy-oriented websites, most with links to Facebook pages and Twitter accounts. Many are run by one or a few individuals dabbling at the fringes of politics and/or trying to make a buck on advertising. There are also networks of "bots," or automated accounts programmed to post content and interact with each other as if they're human. Research published in November 2018 by Professor Menczer's team at Indiana University shows that botnets are effective at "amplifying low-credibility content" and are heavily used to promote domestic sources of disinformation.[14]

The emotionally negative tone of much disinformation makes it more likely to spread rapidly. Understanding why requires a brief digression on the social media business model. At its core, the model involves platforms selling their users' attention to advertisers. The platforms devise algorithms that determine what items go into a Facebook user's News Feed or what videos YouTube recommends a user watch next. Engineers design these algorithms to maximize engagement, which means keeping users on the site—liking, sharing, and commenting. High levels of engagement, or user attention, translate into more advertising revenue for the companies.

Studies show that social media users are drawn to material that elicits an emotional reaction. "One of the biggest issues social networks face," Facebook chief executive Mark Zuckerberg has written, "is that when left unchecked, people will engage disproportionately with more sensationalist and provocative content."[15] More specifically, users are prone to share material that makes them morally outraged. That was the finding in 2014 of researchers at Beihang University in Beijing who studied Weibo, a 500 million-user Chinese site similar to Twitter.[16]

A separate 2017 study of Twitter by researchers at New York University found that the presence of "moral-emotional" language in a politically oriented tweet makes it more likely to be retweeted among people with similar views. The NYU team concluded that the chances of a tweet being shared increased by 20 percent with each additional moral-emotional word associated, for example, with anger or love. Such words include "safe" and "faith," as well as "hate," "war," "greed," "evil," and "shame."[17] False and conspiratorial content, which tends to contain language designed to provoke anger and fear, seems custom-made to zoom around the Internet.

User age also helps determine the spread of disinformation. Older social media users, on average, share more false stories than their younger counterparts. Facebook users over 65 shared nearly seven times as many fake articles as those aged 18 to 29, according to a January 2019 study by researchers from Princeton University and NYU. One reason for the difference might be that senior citizens who didn't grow up immersed in digital culture are less skilled at discerning the credibility of online content.[18]

One might guess that sheer inaccuracy would slow the spread of disinformation. In fact, the opposite is true. In 2018, researchers at the Massachusetts Institute of Technology published the results of a study analyzing every English-language news story distributed on Twitter over 11 years that had been verified as either true or false—some 126,000 stories tweeted by 3 million users. The MIT group found that, on average, untrue news is 70 percent more likely to be retweeted than true news. One possible reason, the researchers suggested, is that Twitter users are drawn to novel information. And untrue content often seems newer than true material.[19]

As the 18th century satirist Jonathan Swift wrote, "Falsehood flies, and truth comes limping after it."[20]

2. A Landscape of Lies

Domestic disinformation first surged on social media in 2010, coinciding with the U.S. midterm elections that year. False articles alleged that then-President Barack Obama was a Muslim non-citizen with connections to Islamic terrorism.

Disinformation has a centuries-long history. Anti-Semitic blood libels circulated in 15th century Europe. Caustic *pasquinades* of 16th century Rome and *canards* of 17th century Paris used false content to score political and social points. Thomas Jefferson complained to a friend in 1807 that "the man who never looks into a newspaper is better informed than he who reads them." At the turn of the 20th century, yellow journalism filled with exaggeration and lies simultaneously produced profits and helped propel the U.S. into the Spanish-American War. As a government weapon, false information flourished on both sides of both World Wars and during the Cold War, when Russian *dezinformatsiya* and equivalent U.S. efforts produced spurious accounts damaging to the enemy.[21]

The Internet has provided a new and welcoming environment for fakery. In the late 1990s and early 2000s, online trolls deployed false news to upset targets of their ire. Political bloggers created websites to distribute news and opinion, some of it fact-based, some not. After Facebook started in 2004, YouTube in 2005, and Twitter in 2006, many website proprietors linked their sites to social media accounts, where they posted work of their own and that of their ideological allies. Professor Menczer of Indiana University recalls noticing a surge of domestic disinformation on social media in 2010, coinciding with the U.S. midterm elections that year. False articles alleged, among other things, that then-President Barack Obama was a Muslim non-citizen with connections to Islamic terrorism.[22]

False content, and especially right-leaning false content, has flourished for several reasons. One is that more traditional sources of information have atrophied. U.S. newspapers, having lost much of their advertising revenue to the Internet, shed more than half of their employees from 2001 through 2016.[23] Meanwhile, Republicans, to a striking degree, have lost faith in what remains of the mainstream media. Only 21 percent of Republicans polled by Gallup said they trust mass media outlets to report the news "fully, accurately, and fairly." In contrast, 76 percent of Democrats expressed trust in the mass media.[24]

As many long-established news sources have declined or disappeared, people have shifted to social media for information. Two-thirds of Americans get at least some of their news from social media, which offer a combination of mainstream journalism and less reliable sources. Even as they rely on Facebook and Twitter for news, more than half of these users expect the information to be inaccurate.[25] All of these developments have coincided with—and to some extent helped cause—intensified polarization in American public life.

The 2016 Election

Conspiracy theorist Alex Jones started *Infowars* in 1999 and over the years transformed it into a video-heavy website which relies on social media to distribute its dark messages. *Infowars* became perhaps the most notorious conspiracy shop online, contending, for example, that the September 11 attacks were an "inside job" and that the 2012 Sandy Hook elementary school massacre was staged in an attempt to promote gun control.[26] Jones developed a large and loyal audience. Between 2008 and 2018, his YouTube channel amassed 2.4 million subscribers and more than 1.6 billion views of the nearly 36,000 videos it featured. His various interlocking businesses related to *Infowars* reportedly brought in revenue of $20 million in 2014 from advertising and sales of health supplements, survivalist gear, and other merchandise.[27]

In December 2015, as candidate Donald Trump was jostling for position in a crowded GOP primary field, *Infowars* leapt into the internecine fray on Trump's behalf. The site's official Twitter account posted a preposterous story claiming that former Florida Governor Jeb Bush had "close Nazi ties."[28] Trump himself had implicitly introduced the calumny a few weeks earlier by retweeting a meme of Bush next to a swastika.[29]

Although better known and better compensated than most, Jones has had many right-wing rivals. Jim Hoft, a former corporate trainer, started *The Gateway Pundit* website in 2004, naming it after the Gateway Arch in his home city of St. Louis. Hoft became active on Facebook and Twitter, where he posted articles attacking Bill and Hillary Clinton and promoting wild conspiracy theories. "I am somewhat known in my business for my headlines," he has said.[30] By the presidential election season of 2016, *The Gateway Pundit* indeed had become a known brand on the political right, in part because it pushed stories suggesting that Hillary Clinton had health problems that precluded her from running the country. One unsubstantiated article alleging that Clinton suffered "seizures" was quickly picked up by Fox News host Sean Hannity, who invited some of his on-air guests to assess Clinton's supposed declining health—a narrative that bedeviled Clinton for much of the campaign.[31]

A study by Harvard's Berkman Klein Center for Internet & Society found that in 2016, *The Gateway Pundit* was among the most frequently shared media sources on Twitter and Facebook among Trump followers. Some of the others included *The Daily Caller* and *Washington Examiner*. "In this group, *The Gateway Pundit* is in a class of its own, known for publishing falsehoods and spreading hoaxes," the Harvard researchers wrote.[32] (Hoft responds that he consistently achieves his goal of being "more trustworthy than *The Washington Post*."[33])

The Harvard team analyzed millions of online news stories, together with Twitter and Facebook shares, broadcast television, and YouTube videos. They described the conservative online complex benefiting from Facebook and Twitter traffic as "a network of mutually reinforcing, hyper-partisan sites that revive what [historian] Richard Hofstadter called 'the paranoid style in American politics,' combining decontextualized truths, repeated falsehoods, and leaps of logic to create a fundamentally misleading view of the world."[34]

This network of falsehood appears at first glance to be sprawling, but actually, it's made up of relatively concentrated nodes. The Russian IRA provided one example of concentration: a discrete group of professional trolls pumping out fake news stories via thousands of social media accounts. Research funded by the Knight Foundation and published in October 2018 identified another aspect of concentration. It discovered that "just a few fake and conspiracy outlets dominated during the [2016] election—and nearly all of them continue to dominate today." Specifically,

the Knight Foundation study found that 65 percent of fake and conspiracy news links on Twitter traced back to just the 10 largest disinformation websites, including *Infowars*.[35]

While this report focuses primarily on domestic U.S. disinformation, it's worth noting that a certain degree of overlap exists between foreign information operations and American right-wing websites active on social media. "There isn't always a clear line between the two," says Claire Wardle the head of research at First Draft, a nonprofit dedicated to tackling falsehoods online. She notes, for example, that in 2016, writers for U.S. disinformation sites reportedly also contributed material aimed at American voters by a group of false content sites based in Macedonia.[36] Separately, Russian content has turned up in the false-information flow of at least several U.S. sites. Between 2014 and late 2017, *Infowars* republished more than 1,000 articles from RT, the Kremlin-controlled television and digital news organization considered by U.S. intelligence agencies to be a propaganda arm of the Putin government.[37]

Disinformation from the Left

While polarization exists on both sides of the political spectrum, it's not symmetrical. Liberal audiences pay some attention to extreme left-oriented websites, but they are more heavily influenced by traditional media outlets like the *New York Times, Washington Post,* and *Wall Street Journal,* which don't traffic in made-up stories. Conservative audiences, by contrast, pay more of their attention to extreme-right sources online and to Fox News, some of whose hosts echo disinformation and conspiracy theories.[38]

That said, there are a number of left-wing sites that also use Twitter and Facebook to project political falsehoods. One useful case study occurred in May 2017, when liberal Senator Ed Markey (D., Mass.) told CNN that a grand jury had been impaneled in New York to investigate the Trump campaign's alleged collusion with Russia. This was untrue. Markey, who apologized for his mistake, apparently picked up the false lead from anti-Trump social media, which had been bandying about the grand jury rumor for days. One possible source was a left-leaning dubious-content site called the *Palmer Report.*[39]

Proprietor Bill Palmer distributes his stories on Twitter (232,000 followers) and Facebook (110,000). He has written that he's "built a growing and loyal audience based on the timeliness and accuracy of our reporting." But many of *Palmer Report's* articles range from the unsubstantiated ("You're Darn Right Donald Trump Is a Russian Spy") to the sophomoric ("It's a Good Thing Donald Trump Is an Idiot"). In October 2017, the *Palmer Report* claimed that presidential son-in-law and senior adviser Jared Kushner "secretly" traveled to Saudi Arabia to avoid possible arrest amid the Trump-Russia investigation. The fact-checking organization *Snopes* branded the story "false."[40] A few days later, Palmer acknowledged in a follow-up that Kushner had returned home and was not arrested.

One the most intriguing—and ominous— illustrations of disinformation from the left came to light in December 2018, when *The New York Times* reported that Democratic operatives had used a Russian-like ploy during a special U.S. Senate election in Alabama a year earlier. Consultants working independently of the Democratic candidate, Doug Jones, created Facebook pages on which they posed as conservative Alabamians. They used one counterfeit page to promote a conservative write-in candidate to take votes away from the main Republican candidate, Roy Moore. The Democratic consultants also deployed thousands of Twitter accounts to make it seem as if Russian bots were supporting Moore. "We orchestrated an elaborate 'false flag' operation that planted the idea that the Moore campaign was amplified on social media by a Russian botnet," an internal report on the manipulation project said.

> " Democratic operatives orchestrated a 'false flag' operation in Alabama that used Twitter to plant the idea that Republican Senate candidate Roy Moore was supported on social media by Russian bots. "

In yet another stratagem, Democrats tried to alienate moderate Republicans by linking Moore to a fake campaign on Facebook and Twitter to impose a statewide ban on alcohol. It isn't clear whether these tactics affected the election, which Jones won narrowly.[41]

After the *Times* report, Facebook suspended five accounts involved in the Alabama episode, including that of the CEO of New Knowledge, a social media research firm. The executive, Jonathon Morgan, acknowledged the suspension and said he'd been running an experiment in Alabama on how online disinformation works, not trying to influence the outcome of the special election. The covert activity was funded by Democratic political donors, including Reid Hoffman, the co-founder of LinkedIn, who apologized and said he hadn't known about the underhanded tactics. But such tactics went beyond Alabama. Hoffman reportedly also provided financial backing to a separate organization called News for Democracy which helped create more than a dozen misleading Facebook pages designed to appeal to conservative voters nationally in the run-up to the 2018 midterms.[42]

Trump on Twitter

President Trump's active engagement on Twitter has helped shape the current online environment. His tweets directly reach his nearly 58 million followers (some of them, no doubt, bots) and often receive extensive mainstream media coverage. He has used Twitter to advance a range of conspiracy theories and fictional assertions that otherwise might not have made it to center stage.

In one representative case, the syndicated conservative radio host Mark Levin asserted in March 2017 that members of the Obama Administration had attempted to undermine Trump in a "silent coup." This concoction moved swiftly to *Breitbart* and then on to Fox News. On a Saturday morning, Trump responded: "Terrible! Just found out that Obama had my 'wires tapped' in Trump Tower just before the victory," adding, "This is McCarthyism!" Minutes later, he switched historical references, tweeting, "This is Nixon/Watergate." Trump's Fox News favorite, Sean Hannity, tweeted: "What did OBAMA know and when did he know it??"[43] Subsequently pressed to substantiate the wire-tapping accusation, Trump cited Fox News coverage. National security officials later testified before Congress that there was "no information" indicating Trump had been targeted.[44]

Since Robert Mueller's appointment as Special Counsel in May 2017, the topic that has most preoccupied President Trump on Twitter—and elicited from him numerous false statements—is the Russia investigation. In June 2018, the president asserted that Mueller's appointment "is totally UNCONSTITUTIONAL!" even though the Trump Justice Department selected Mueller and gave him his marching orders.

President Trump often takes peripheral stories circulating on the far right and injects them into the mainstream.

In August 2018, a conservative website called *PJ Media* said it had done a study showing that 96 percent of the results from a Google News search for "Trump News" returned web pages from "left-wing media." *PJ Media* acknowledged that its study was "not scientific," but the caveat didn't stop the President from repeating the 96 percent figure. "Google & others are suppressing voices of conservatives and hiding information," he tweeted. The fact-checking organization *Politifact* deemed the Trump tweet "false," in part because *PJ Media* categorized any media outlet not expressly conservative as being part of the "left," a category that included major wire services, broadcast networks, and newspapers.[45] Google search results actually depend on such factors as the freshness of material, what the user has searched for in the past, and what other sites link to a given search result.

President Trump's allegations of conspiracies hostile to his presidency elicit cheers at rallies and likes on Twitter. He has tweeted that the Federal Bureau of Investigation is part of "a criminal deep state" (May 2018) and that the mainstream press scrutinizing his administration deserves condemnation as the "Fake News Media, the true Enemy of the People" (October 2018). Repeated regularly, the president's assertions further corrode his supporters' faith in important national institutions.

Hate from the Alt-Right

The alt-right (alternative right) refers to a loose agglomeration of American white nationalists, anti-Semites, neo-Nazis, and other unsavory types. Although hard to pin down organizationally, the alt-right has produced a steady stream of false conspiracy theories, often intertwined with hate speech. For example, some within the alt-right promote the myth of "white genocide": a secret Jewish-led plot to eliminate white people in the U.S. as a racial group and replace them with non-whites.

Because of its dispersed nature, the alt-right depends heavily on websites and social media to get its message out. The Anti-Defamation League estimated that during the run up to the 2018 midterm elections, anti-Semitic "Twitter bombing" of Jews, and especially Jewish journalists, averaged 5 million tweets per day.[46]

To their credit, social media platforms have at times tried to marginalize or exclude alt-right figures. But the hate mongers tend to switch venues and resume spreading their noxious mixture of disinformation and ethnic animosity.

Alt-right exemplar Andrew Anglin has recounted in an online essay that he "got into Hitler" while participating on a 4chan message board that "was going full Nazi." He started his website, *Daily Stormer*, in 2013, naming it after Hitler's favorite newspaper, *Der Stürmer*. Anglin's site and associated social media accounts have provided inspiration to racists such as Dylann Roof, reportedly a reader and commenter. In June 2015, Roof massacred nine black worshipers in a Charleston, S.C., church.[47] *Daily Stormer* cheered on participants in the "Unite the Right" rally in Charlottesville, Va., in August 2017. Nominally protesting the planned removal of a Confederate war memorial, the right-wing demonstrators chanted, "Jews will not replace us" and waved swastika flags. A counter-protester was killed in the ensuing violence.

In the wake of Charlottesville, *Daily Stormer* and other alt-right outlets came under fire. Go Daddy and Google rescinded *Daily Stormer*'s web-hosting arrangements, forcing it into a state of limbo. Twitter, meanwhile, stiffened its hate speech rules and deleted the accounts of several neo-Nazi websites and organizations, including that of *Daily Stormer*. Facebook killed *Daily Stormer* links while also removing a series of organizations with names like *Right Wing Death Squad* and *White Nationalists United*.[48]

Conspiracies and Frauds
Screenshots of domestic disinformation

President Trump has used Twitter to spread false information.

On Facebook, Glenn Beck joined a chorus condemning philanthropist George Soros.

Conspiracy theorist Alex Jones has told his followers about phony "false flag" plots.
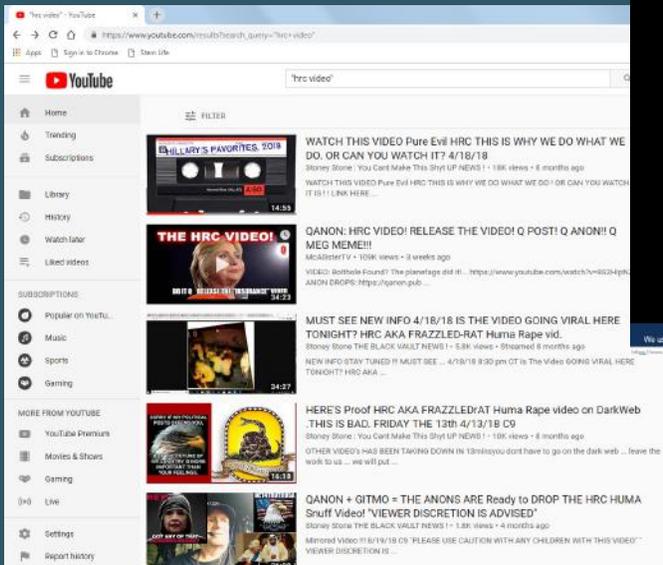
YouTube has offered users lurid made-up accusations against Hillary Clinton.

> **"Gab gives well-known alt-right figures a megaphone to influence people like Robert Bowers, who may be willing to act on extremist rhetoric."**

These actions did not disable the alt-right permanently. Many adherents merely shifted from Twitter to Gab. There, they received a warm greeting from Gab member Christopher Cantwell, who spent three months in jail after pepper spraying counter-protestors at the Unite the Right event. "For all of you who are new to Gab, don't worry about the racism," he said in his welcome post. "I know it can be a little weird at first, but pretty soon you're going to realize that racism is normal, and the only reason you haven't seen it before is because the Jews were censoring it."

Gab is where Anglin reintroduced *Daily Stormer*, now registered online with a hosting service in Hong Kong and sporting a ".name" domain. "Annnd we're back!" Anglin celebrated on Gab. He urged users to keep his site in mind "when you're ready to start shoving Jews onto a train."

A conservative programmer named Andrew Torba started Gab in 2016 as a response to what he saw as Silicon Valley political correctness. Gab drew alt-right luminaries such as Milo Yiannopoulos, who was kicked off Twitter for harassment, and Richard Spencer, a prominent white nationalist who coined the term "alt-right" and saw his Facebook pages shut down in 2018. Even before the

Pittsburgh massacre, Gab had its app rejected by both Google and Apple for failing to moderate hate speech.[49]

Gab, which describes itself as "The Home of Free Speech Online," was a natural social media gathering spot for Robert Bowers, the Pittsburgh synagogue shooter. He opened an account in January 2018. After the attack, Gab issued a statement saying that it "unequivocally disavows and condemns all acts of terrorism and violence." Torba said in interviews that there had been no basis to censor Bowers, and Gab had done nothing wrong. He told National Public Radio: "The answer to bad speech, or hate speech, however you want to define that, is more speech, and it always will be."[50]

Disclaimers notwithstanding, Gab gives well-known alt-right figures a megaphone to influence people like Bowers, who may be willing to act on extremist rhetoric. That was the conclusion of a joint study by the Network Contagion Research Institute, an inter-university academic collective, and the Southern Poverty Law Center, both of which track the spread of disinformation and hate speech online.[51] Gab has also won an international following, attracting right-wing users in Brazil and other countries.

Gab is not unique as an alt-right communications channel. Discord, a chat app catering to video game aficionados, has also served at times as a haven for the far right. Within hours of the Pittsburgh synagogue shooting, several Discord members were debating whether Bowers deserved criticism for endangering the neo-Nazi movement's long-term viability or praise for killing Jews.[52] In 2017, organizers of the Unite the Right rally used Discord to coordinate the march.[53] Alerted to such incidents, Discord has purged the accounts of some alt-right participants. But the problem persists. In January 2019, local police in upstate New York said that four young white men communicated on

Discord when they planned an attack on an African-American Muslim enclave near the Catskill Mountains. The defendants in the New York case allegedly stockpiled 23 guns and three homemade bombs before the authorities intervened.[54]

YouTube's Distinctive Role

When it comes to disinformation, Facebook and Twitter tend to receive most of the attention, obscuring the important role played by YouTube. The video network falls behind only Facebook as the social media site most popular for viewing news stories. Seventy-three percent of U.S. adults visit YouTube, with the percentage rising to 94% for 18-to-24-year-olds.[55] Some prominent figures in the alt-right seek to use YouTube to "attempt to reach young audiences by broadcasting far-right ideas in the form of news and entertainment," according to a 2018 study published by the Data & Society Research Institute.[56]

One example is Paul Joseph Watson, the *Infowars* editor who initially tweeted the altered video of CNN correspondent Jim Acosta which was apparently recycled by the Trump White House. Watson has made YouTube videos such as one called "Conservatism is the New Counter-Culture," in which he compares today's alt-right to punk rockers of the late 1970s. In May 2018, he tweeted a photo of himself holding a plaque YouTube sent him for surpassing 1 million subscribers. Watson added the caption, "YouTube secretly loves me."[57]

The affection, in fact, isn't clandestine. YouTube's algorithm, like Facebook's, seeks to maximize engagement. If a user shows interest in Watson's videos, the network's recommendation engine will serve up similar fare.[58] An investigation published by *The Wall Street Journal* in February 2018 found that "YouTube's recommendations often lead users to channels that feature conspiracy theories, partisan viewpoints, and misleading videos, even when those users haven't shown interest in such content."[59]

This phenomenon played out in January 2019, when stark disinformation dominated YouTube searches about the health of Supreme Court Justice Ruth Bader Ginsburg. At the time, Ginsburg was recovering from apparently successful cancer surgery. But *Washington Post* reporters found that a YouTube search for her initials, "RBG," directed users to false far-right conspiracy videos, some of which alleged that doctors were using mysterious illegal drugs to keep the 85-year-old jurist alive. Users who clicked on one of the conspiracy videos received recommendations to view other videos about a demonic "deep state" running the U.S. or a Jewish cabal controlling the world.[60]

Drawn by this kind of content, users of Gab and 4chan have demonstrated a distinct affinity for YouTube. They link to it more often than to any other single website—typically thousands of times a day. Beyond any sense of ideological kinship, YouTube is valuable for a technical reason: Gab and 4chan lack extensive video capacity of their own and essentially use YouTube as their backup video library.[61]

The YouTube collection of alt-right material is voluminous. Numerous YouTube videos echo a discredited 2016 conspiracy theory known as "Pizzagate," which posited a satanic child-sex ring involving Hillary Clinton and headquartered at a Washington, D.C., pizza restaurant. The original Pizzagate furor prompted a North Carolina man to travel to the capital in 2016 and fire rifle shots into the pizzeria in question. In early January 2019, some of the top results for a YouTube search for the seemingly harmless term "HRC video" were gruesome elaborations on the Clinton-child-sex-abuse delusion. Some of these videos allude to a murky "snuff film," code-named "Frazzledrip," in which Clinton and aide Huma Abedin are allegedly seen raping and mutilating a prepubescent girl. *Snopes* tracked down the supposed snuff film and branded it "demonstrably a hoax."[62]

On a related front, YouTube faces a major challenge in the form of "deep-fake" videos. Deepfake content uses cutting-edge artificial intelligence (AI) to combine altered imagery and audio for simulations intended to be undetectable by the human eye or ear. So, for example, political operatives using deepfake technology could create a video showing an opposing candidate giving a speech she never actually gave. The technology isn't perfected, but it's getting close. Just how close is illustrated by a *BuzzFeed News* video uploaded to YouTube in which former President Obama appears to deliver a talk on the dangers of deepfake. Obama's voice and facial expressions were provided by the actor-director Jordan Peele, who ventriloquized the ex-president referring profanely to President Trump. Deepfake has also been used to manufacture phony celebrity pornography. Unless it's countered by AI-driven detection tools, deepfake could power a whole new generation of disinformation.[63]

Conspiracy Theories about George Soros

Dark imaginings about George Soros litter the Internet. The Hungarian-born former hedge-fund manager and billionaire philanthropist, who is Jewish, exemplifies for certain people on the right the supposed secret clique of rich Jews who pull strings to influence world events. Invoking Soros has become a "dog whistle" to alert and mobilize anti-Semites.

In mid-September 2018, pro-Trump participants in 4chan's Politically Incorrect message board began a campaign to discredit Christine Blasey Ford, a psychology professor at Palo Alto University who accused Supreme Court nominee Brett Kavanaugh of having sexually assaulted her when they were in high school. The anonymous 4chan users' goal was, as one put it, to "prove she is a liar." This effort generated

> **With 'deepfake' technology, disinformation artists could create a convincing video showing a politician giving a speech she never actually gave.**

a series of memes suggesting Ford was an anti-Trump activist linked in some way to Soros, who is a Democratic donor.[64]

One image that was shared widely on Facebook claimed in its caption to show Ford standing next to Soros. "The pieces of the puzzle are finally coming together," the caption added. But the woman in the photo with Soros was actually Lyudmyla Kozlovska, president of a Polish human rights organization.[65]

President Trump also played the Soros card during the Kavanaugh confirmation debate. He tweeted on October 5, 2018, that protestors opposing his Supreme Court nominee were "paid professionals only looking to make Senators look bad." The demonstrations, Trump added, were "paid for by Soros and others." This unfounded accusation ricocheted back to 4chan, where participants cheered. "Trump has officially named the Jew," one wrote.

Another made-up Soros conspiracy in 2018 concerned groups of Central Americans moving through Mexico toward south Texas. As one such caravan mobilized in the spring, right-wing websites and Facebook pages declared Soros was behind it. "SOROS FUNDING MIGRANT 'CARAVAN,'" long-time conservative host Glenn Beck shouted in all capital letters from his Facebook page.

The accusation continued to percolate and again came to a boil in the fall of 2018, when a new caravan of Hondurans headed toward the U.S. In October, Rep. Matt Gaetz, a far-right Republican from Florida, posted a video on Twitter of a man supposedly handing cash to migrants who Gaetz said were planning to "storm the U.S. border." "Soros?" the congressman suggested. It wasn't clear where the video had been shot or whether the people depicted were migrants. Still, the next day, President Trump tweeted the same video with a different caption: "Can you believe this, and what Democrats are allowing to be done to our Country?" After Soros denied funding the migrants, Gaetz tweeted, "Pardon me for not taking Mr. Soros' word about what Mr. Soros is doing."

A Facebook Misadventure

In November 2018, *The New York Times* published a wide-ranging exposé about Facebook management. Among the episodes the *Times* recounted was Facebook's having hired a Washington public relations firm that sought to discredit anti-Facebook activists by tying them to George Soros. This strategy suggested a certain parallelism between Facebook and the conspiracy theorists who claim—via Facebook—that they see Soros behind every controversy.[66]

Facebook argued that it merely intended to demonstrate that Soros, a past critic of the company, had supported what had been represented as spontaneous grassroots opposition. The company called "reprehensible and untrue" suggestions that there was something anti-Semitic about pointing journalists to the Soros connection.[67]

The story took another turn. Definers Public Affairs, the Washington PR firm Facebook retained, has what amounts to an affiliated disinformation shop. Specifically, a Definers co-founder, veteran Republican operative Joe Pounder, started and edits a website called *NTK Network,* which has an associated page on Facebook with more than 123,000 followers. *NTK* (Need to Know), which shares office space with Definers, presents itself as an ordinary conservative political website. It just happens to post friendly articles about Definers' clients—including, for a time, Facebook—and unfriendly articles about rivals of Definers' clients.[68] In a blog post in November 2018, *NTK* said, "We do not and did not work with Facebook. We share offices with a firm that does. Joe Pounder works with that firm, but Pounder has many separate projects."[69]

Definers thus goes beyond the traditional PR strategy of trying to persuade reporters to provide favorable coverage about clients. Courtesy of *NTK,* the firm generates pseudo-news stories that benefit Definers clients and often get picked up by other conservative outlets. The client-centric pieces are blended in with political articles, such as one from mid-December 2018 accusing former Federal Bureau of Investigation Director James Comey of leaking classified information.

Facebook cut ties with Definers after the *Times* article ran.

3. Company Responses

> According to a 2018 study by researchers at New York University and Stanford, Facebook users' engagement with 'fake news' websites declined after the 2016 election, while continuing to rise on Twitter.

It took the startling revelations of Russian interference in the 2016 presidential election to persuade the leadership of the most prominent social media platforms of the need for more vigorous oversight of their operations. Even then, the companies seemed to acknowledge their predicament only reluctantly and under pressure from Congress, private analysts, and journalists.[70]

Since 2017, Facebook CEO Mark Zuckerberg says he has focused more on "content governance and enforcement issues" than on any other topic.[71] Judging from public announcements, his company has made more changes than Twitter or YouTube, and the more vigorous approach may have made a difference. That was the conclusion of a study published in September 2018 by researchers from NYU and Stanford University.

The NYU-Stanford team assembled a list of 570 websites known for producing what the researchers called "fake news." Most of the sites in the study leaned right; some tilted left. Using a time frame of January 2015 to July 2018, the researchers measured the volume of Facebook and Twitter engagement (shares, likes, comments) with the dubious stories generated by the fake-news sites. The findings: Engagement with fake news rose steadily on both platforms through the end of 2016, just after the election. Then, engagement fell sharply on Facebook—by more than 50 percent—while continuing to rise on Twitter. The researchers observed no similar pattern for other news, business, or cultural websites, where Facebook and Twitter engagement was relatively

stable over time. "Some factor has slowed the relative diffusion of misinformation on Facebook," the researchers concluded. "The suite of policy and algorithmic changes made by Facebook following the election seems like a plausible candidate."[72]

One sobering qualification to the NYU-Stanford findings is that even after the marked drop-off in Facebook engagement following the 2016 election, Facebook interactions with fake news sites still average roughly 70 million per month—a testament, in part, to Facebook's sheer size. "The absolute level of interaction with misinformation remains high," the researchers observed, and "Facebook continues to play a particularly important role in its diffusion."[73]

Since Russian government interference became an issue, Facebook and the other major social media platforms have said they've hardened their disinformation defenses in a number of ways. They've refined their ranking and recommendation algorithms and improved artificial intelligence that identifies potentially harmful content. On the human side, they've hired thousands of additional content reviewers and contracted with platoons of outside fact-checkers.

> **Before the 2018 U.S. midterm elections, Twitter removed more than 10,000 bots used in a coordinated voter-suppression campaign aimed at male Democrats. The automated accounts pushed hashtags such as #LetWomenDecide.**

The upshot of this activity is the elimination of millions of fake accounts, the exclusion of some particularly prolific producers of falsehoods, and the reduction of how much phony material can circulate. But all of the platforms still have a long way to go in combating disinformation.

Arbiters of the Truth

"The single most important improvement in enforcing [Facebook] policies," according to Mark Zuckerberg, "is using artificial intelligence to proactively report potentially problematic content to our team of reviewers, and in some cases to take action on the content automatically as well."[74] The same statement about the significance of AI holds true for Twitter and YouTube. Given the daily flow of billions of posts, tweets, and video uploads, humans alone cannot police the platforms. AI offers the only realistic hope for cleaning up harmful content, including domestic disinformation, at scale.

AI refers to algorithms able to perform human-like tasks, such as understanding human language. Machine learning, a way of achieving AI, describes the "training" of algorithms, by feeding them huge amounts of data, so they can learn for themselves how to accomplish the task at hand.[75]

Facebook has used machine learning to improve its ability to identify potentially false stories, photographs, and videos. Without describing in detail how the technology works, the company has said that it compares characteristics of past false items to the material currently in question. An article or image singled out in this fashion typically goes to an in-house reviewer or a third-party fact-checker. Facebook has said that in 2018 it tripled to 30,000 the number of people it has working on "safety and security." Fifteen thousand of them are content reviewers, many of whom are outside contractors.

If the item under scrutiny is ultimately deemed untrue, however, it is not eliminated from Facebook. Instead, the company demotes it in News Feed, typically resulting in future views being reduced by more than 80 percent. Zuckerberg has said that Facebook has also started to demote "sensationalist and provocative content" that borders on violating the company's rules but doesn't quite cross the line.[76]

Explaining its thinking about what it calls "false news," Facebook says in its Community Standards: "We want to help people stay informed without stifling productive discourse. There is also a fine line between false news and satire or opinion. For these reasons, we don't remove false news from Facebook but instead, significantly reduce its distribution by showing it lower in News Feed."[77] Despite its caveats, Facebook acknowledges that, with the aid of third-party fact-checkers, it can identify at least some false content.

So does YouTube. In January 2019, the Google subsidiary announced that it would "begin reducing recommendations of borderline content and content that could misinform users in harmful ways—such as videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11." YouTube emphasized that "this will only affect recommendations of what videos to watch, not whether a video is available on YouTube." The 9/11 conspiracy theory videos aren't being removed; they're just going to be harder to find.[78]

Twitter does its own version of content demotion when it ranks tweets and search results. Tweets from accounts the company has identified with AI and/or human review as "bad-faith actors who intend to manipulate or divide the conversation" are ranked lower but aren't necessarily removed. Twitter maintains that it down-ranks content based strictly on user behavior, not on the substance of tweets. It uses a variety of indirect "signals" to determine whether users are acting in bad faith. These include whether users have a confirmed email address and have uploaded a profile

image; whom users follow and retweet; and who, in turn, follows, retweets, or blocks them.[79] In effect, Twitter's focus on behavioral signals allows it to remove some disinformation without focusing directly on the truth or untruth of the content in question.

Facebook, Twitter, and YouTube thus concede that they make certain kinds of problematic content less available to their users. Twitter focuses on behavioral indications of manipulation and divisiveness. Facebook and YouTube go further and look at whether content is substantively false. Having made these determinations, the companies marginalize the undesirable material. These practices undercut a mantra frequently recited by the social media industry—that the platforms aren't and shouldn't be "arbiters of the truth." As Sheryl Sandberg, Facebook's chief operating officer, has phrased it, "We definitely don't want to be the arbiter of the truth."[80] Twitter CEO Jack Dorsey has warned that it would be "dangerous" for the company's employees to act as "arbiters of the truth."[81]

But in varying ways, the companies do play an arbiter role when they relegate objectionable content to where users are less likely to see it. As we explain more fully in our Conclusions and Recommendations in Part Four, we believe that when the platforms encounter provably false content, they ought to go the full distance and remove it.

We recognize that these will not be easy decisions. But Facebook, Twitter, and YouTube already make a wide range of tough calls. Facebook presumably does careful research when designing the algorithms and human-review systems it uses to bury false news in the lower reaches of News Feed. YouTube, after being criticized for its dissemination of misleading and conspiratorial videos following mass shootings in 2017 and 2018, promised it would undertake the delicate task of adjusting its algorithm and human oversight to

promote more accurate content. It hasn't fully succeeded, as illustrated by the Justice Ruth Bader Ginsburg health-conspiracy videos mentioned earlier. But the company did commit to distributing more truthful content.[82]

Beyond the question of how to deal with domestic disinformation, all three major platforms have made the determination to exclude whole categories of other kinds of harmful content. These categories include child pornography, terrorist incitement, harassment, and hate speech. Facebook also removes

misinformation that leads to a risk of physical violence or to voter suppression. Alerted by AI tools, user complaints, or in-house moderation, the companies assess the substance of the detrimental content and, if it violates company policy, get rid of it. Such removal decisions surely require the sort of complicated and serious-minded appraisals we're advocating in connection with provable falsehood. In a sense, we're not urging anything brand new. We're calling for the platforms to classify provably false content as another category worth removing.

Neutrality Not Required

Section 230 of the Communications Decency Act of 1996 protects Internet platforms from most kinds of liability for what users post or tweet. A misunderstanding of the law has given rise to the notion that if Facebook, Twitter, and YouTube actively moderate what goes on their sites—as, in fact, they currently do, and as we believe they should do even more vigorously—they could lose their liability shield. It's worth clarifying why this view is wrong.

The Internet wouldn't have developed into its current robust form if online businesses such as the social media platforms had not been protected against lawsuits over what people say and do online. But a myth has grown up that to receive Section 230 protection, the platforms must operate as neutral public forums. Senator Ted Cruz (R., Texas) and others have entwined this argument with the claim that since the platforms allegedly censor conservatives, they ought to lose some or all of their Section 230 protection.

At multiple hearings in 2018, Republicans lectured executives from Facebook, Twitter, and YouTube about supposedly squelching conservative content—a charge the company representatives strenuously denied. Because the platforms aren't run as neutral public forums, the Republicans declared, they don't deserve Section 230 protection.

But this binary choice—neutrality or no liability shield—is a fallacy. Neither Section 230, nor any other statute, obliges the platforms to remain neutral. Indeed, a more reasonable interpretation of the provision is that it represents lawmakers' giving the tech companies discretion to moderate their platforms without fear of liability. Under relevant Supreme Court precedent, moreover, this moderation should be seen as a form of corporate expression protected by the First Amendment. In other words, social media companies may pick and choose the content they provide to users and enjoy the benefits of Section 230. It's not an either/or choice.

Short of deletion, the platforms have developed various features to annotate factually dubious items. These features ought to be retained to warn users about material that may not qualify for removal as an outright untruth but deserves to be regarded with skepticism. Facebook, for instance, has introduced "Related Articles," a feature that offers factually reliable context on dodgy stories. Related Articles replaced an earlier system of merely flagging stories that had been disputed by fact checkers. Facebook discovered that the disputed flags alone caused many users to click on items out of curiosity—the opposite of what was intended.[83]

YouTube has instituted a program similar to Related Articles. When a user searches for topics that the video site has identified as having "often been subject to misinformation, like the moon landing and the Oklahoma City bombing," YouTube now offers as a preface to its video results a link to information from reliable third parties, such as *Encyclopedia Britannica.* In the wake of breaking events, before trustworthy news sources have had time to upload video, YouTube is posting short bursts of text from such sources in hopes of preempting hastily cobbled-together video from unreliable outlets. Once vetted news sources do produce videos, YouTube says it is making that material easier to find on the site.[84]

The companies have all tried to reduce the financial incentives to spread disinformation. Down-ranking spurious content is one way to accomplish this. Another is Facebook's policy of blocking ads from pages that repeatedly share false news. The most effective method is removing accounts or pages that, in Facebook's words, engage in "coordinated inauthentic behavior," meaning that they try to "mislead others about who they are and what they are doing." This mostly refers to accounts that spread spam and clickbait, but it also includes purveyors of political misinformation (some of whom are in the clickbait racket, too). All told, Facebook says it removes millions of fake accounts every day.[85]

Battling Bots

Twitter appears to have focused considerable energy on detecting and eliminating automated bots, many of which tweet domestically generated false information. To accomplish a bot purge, Twitter had to take a hit to its monthly active user count, a key measure on Wall Street of a social media company's financial prospects. In the wake of the Russian-interference scandal, Twitter concluded it had no choice.[86] This imperative raised the question of whether moving aggressively against disinformation-spewing bots would impinge on free speech. The company tilted toward cleaning up the site. "Free expression doesn't really mean much if people don't feel safe," Del Harvey, Twitter's vice president for trust and safety, told *The Washington Post* in July 2018.[87]

Jack Dorsey, Twitter's CEO, explained his company's approach during congressional testimony in September 2018. Twitter, he said, uses AI to identify potential botnet activity, "such as exceptionally high-volume tweeting with the same hashtag or mentioning the same @ handle without a reply from the account being addressed." The company then requires confirmation that a human is controlling the account. Twitter has also stepped up its use of "challenges," using technology such as CAPTCHAs, which require users to prove they're human by identifying portions of an image or typing in letters or numbers. Sometimes, Twitter ferrets out bots by requesting password resets or, in the case of new accounts, demanding email or cell phone verification. As a result of these steps, Dorsey said, Twitter is challenging 8.5 million accounts per week and removing 214 percent more of them, year-over-year.[88]

In a colorful illustration of anti-bot enforcement, Twitter confirmed in November 2018, just before the midterm elections, that it deleted more than 10,000 automated accounts used in a coordinated voter-suppression campaign aimed at male Democrats.

The bots circulated a series of memes encouraging Democratic men to stay home from the polls so that women could have more say in politics. "Haven't White Men Done Enough Damage Already?" asked one headline over an image of men chanting and holding torches at the 2017 Unite the Right Rally in Charlottesville, Va. The removed accounts pushed hashtags such as #NoMenMidterms and #LetWomenDecide. Twitter acted after Democratic Party operatives alerted the company to the bot caper. Apparently, voters weren't fooled, as the stunt didn't have a discernible effect on the election. A perpetrator wasn't publicly identified.[89]

Algorithm Adjustments

The social media companies continually tinker with algorithms that determine where items end up in users' feeds based on the source or nature of the content. While this sort of ranking may not explicitly turn on the truthfulness of content, it can have the effect of limiting users' exposure to domestic disinformation.

One such change took place in January 2018, when Facebook prioritized posts from friends and family over material generated by publishers or brands. This downgrading affected legitimate news outlets, as well as more questionable ones. In a counterbalancing adjustment, Facebook said it would prioritize news from publications whose content was deemed "trustworthy, informative, and local."[90]

The net effect of these alterations came swiftly. By March 2018, "both liberal and conservative publishers of clickbait and highly polarizing content" experienced a significant drop in Facebook engagement, according to the website *The Outline,* which crunched data provided by BuzzSumo. *The Outline* made a further observation: that right-wing websites were hit harder in the weeks following Facebook's January 2018 changes. Jim Hoft's *The Gateway Pundit,* with a 55 percent drop in traffic, was among the most affected. By comparison, the liberal site *Shareblue* experienced a 27 percent falloff.[91]

Hoft decried the effect of the Facebook algorithm modifications, describing them as an attack on conservatives and free speech. In January 2017, 24 percent of his website traffic came from Facebook, he said in testimony before a House subcommittee. By June 2018, that figure had declined to 2 percent. "If Facebook were seeking to hold a book burning," he added, "they wouldn't have been half as successful."[92]

Facebook described its changes in more benign terms. It said the alterations were designed merely "to help bring people closer together by encouraging more meaningful connections" and the exchange of news from more reliable sources.[93]

Infowars: A Case Study

In the summer of 2018, the social media establishment, under mounting public pressure, went after a major distributor of domestic disinformation: Alex Jones and his *Infowars* organization. Invoking policies against hate speech and abusive behavior, Facebook, Twitter, and YouTube all essentially banned Jones. Facebook, for instance, deleted four of his pages, citing Jones' glorification of violence and use of dehumanizing language. Significantly, the companies didn't say they were punishing Jones for the hoaxes and lies central to his repertoire.

Whatever the justification, the social media removals appeared to have a swift impact. Jones was cut off from millions of followers and subscribers— an action he described as unjustified censorship of his ideological views. Within weeks, daily traffic to his core website declined by half, to about 715,000 visits and video views.[94]

The platforms acted knowing they would take flak from conservatives. As a candidate, Donald Trump appeared on Jones' show and told the host: "Your reputation is amazing. I will not let you down." And sure enough, within days of Jones' banishment from Facebook,

President Trump tweeted, "Social Media is totally discriminating against Republican/Conservative voices."

But Jones and *Infowars* never entirely disappeared from social media. It took only a few weeks after the highly publicized ban in August 2018 for much of the removed *Infowars* content to resurface on Facebook, according to Jonathan Albright, director of the Digital Forensics Initiative at the Tow Center for Digital Journalism at Columbia University. Slightly amended pages with the names *News Wars* and *Infowars* Stream "were being promoted by Facebook via its search and video recommendation algorithms for searches about conspiracies and politics," Albright pointed out. Reconfigured Jones pages popped up, for example, when Albright did a search for "Soros caravan." The two *Infowars* pages have only 51,000 followers, combined, but between August and November 2018, they reported almost 700,000 interactions—"not that far off from what the combined blue-check-marked Jones and *Infowars* pages were getting in the three months before they were removed," Albright said.[95]

Jones' influence persists on Facebook by means of yet another channel: *Infowars*-themed "groups," which Facebook didn't ban and where Jones fans gather to exchange news and conspiracy ideas. In December 2018, members of one such group, *Alex Jones – Infowars.com*, were posting anti-Muslim content and suggestions that a recent deadly terrorist attack in Strasbourg, France, by a suspected Islamist radical had been a government-sponsored "false flag" affair meant to distract attention from populist protests in Paris. In another group, *Infowars Media,* users could follow a link to an *Infowars.com* report on "sealed indictments" filed against President Trump "by the Deep State in an attempt to drag on the phony Russia collusion Witch Hunt." (There were no indictments, sealed or otherwise.)

> **Conservative protests of the banning of Alex Jones and Infowars obscured that the same social media sites that disciplined Jones had given him prominence in the first place.**

> **Facebook stood on solid constitutional ground. The First Amendment bans censorship by government, not by a corporation weeding out those who violate the rules of a privately organized forum.**

In September 2018, *The New York Times* reported on a "closed" *Infowars* group with 110,000 members. In closed groups, only members can see posts. There are also "secret" Facebook groups, which are entirely invisible to searches unless you're a member—an obvious invitation to mischief.[96] Looking beyond *Infowars*, Albright wrote in November 2018 that he'd "found disinformation and conspiracies being seeded across hundreds of different groups." Closed and secret Facebook groups, he added, "have become the preferred base for coordinated information influence activities."[97]

Purging Networks

Facebook followed the *Infowars* action by purging 810 domestic pages and accounts that amplified misleading political content. The targets of the October 2018 sweep included *Right Wing News,* a site linked to a network of Facebook pages and accounts that boasted more than 3.1 million followers. *Right Wing News* used its Facebook network to share stories widely and quickly—and to draw users back to its ad-supported website. But Facebook considered most of these pages and accounts to be

Potemkin operations—shams used "to generate fake likes and shares." This artificially created engagement exaggerated the popularity of the site's stories and inflated their ranking in News Feed, the company said. Facebook cited the same reasons for expelling networks linked to left-leaning sites such as *Reasonable People Unite* (2.3 million followers) and *Reverb Press* (816,000).[98]

Proprietors of the sites in question—left and right—protested that they'd followed the rules as they understood them and done nothing wrong. "Facebook never provided any proof whatsoever of their charges," said John Hawkins, founder of *Right Wing News.* Chris Metcalf, the head of *Reasonable People Unite,* accused Facebook of using "intentionally ambiguous rules and standards" to silence political speech.[99]

First Amendment defenders sounded alarms. "The shift toward domestic disinformation raises potential free speech issues when Facebook and Twitter find and curtail such accounts that originate in the United States," the literary group PEN America tweeted. Facebook conceded in a public statement that the stories and opinions shared by the purged pages and accounts "are often indistinguishable from legitimate political debate." But the company insisted that it had acted only because of the behavior of *Right Wing News, Reasonable People Unite,* and the others, not the content of their expression.

As we've noted, Facebook stood on solid constitutional ground. The First Amendment bans censorship by government, not by a corporation weeding out those who violate the rules of a privately organized forum. "This kind of moderation, which we are likely going to see a lot more of, is viewpoint agnostic," according to Renée DiResta, director of research at New Knowledge. "It's based on quantifiable evidence of manipulative activity."[100]

In other words, the removals muffled some alleged disinformation artists, not because they spread untruths, but because they gamed the system. The platforms left for another day the question of whether to ban false content for its own sake. We believe that day has come, as we discuss in our Conclusions and Recommendations in Part Four.

4. Conclusion and Recommendations

> "No combination of algorithm and human analysis will rid social media of all falsehoods. But the impossibility of perfection shouldn't be an impediment to improvement."

The major social media platforms need to do more to address domestic disinformation, which presents a growing threat to democratic discourse. While these companies have taken some steps to counteract false information and other harmful content, they continue to employ a piecemeal approach, rather than adopting a comprehensive strategy.

They cobble together a reliance on legal obligations—for example, prohibitions on child pornography—with enforcement of "community standards" and rejection of "inauthentic" posts, meaning those that fail to disclose their true source. These are all moves in the right direction, but they are not sufficient. Likewise, the companies are reducing the prominence of false content, instead of embracing a straightforward commitment to removing it.

Too often, the platforms cling to the outmoded notion that they are not "arbiters of the truth." Granted, these companies are not akin to editors of *The New York Times*, but neither are they mere caretakers of passive digital platforms. They fall somewhere in-between, and they need to acknowledge this hybrid role. The starting point for a new paradigm is a willingness to take down provably untrue content, especially in the political realm.

We aren't alone in adopting this view. Dipayan Ghosh, a former Facebook privacy and public policy advisor, says purposeful untruths ought to be removed from Facebook, Twitter, and YouTube. "If statements are meant intentionally to mislead, they should be taken down," says Ghosh, now the co-director of the

Platform Accountability Project at the Harvard Kennedy School.[101] Neither the First Amendment nor parallel international principles protect lies on social media. These are private companies, not governments, and they have ample latitude to remove disinformation without running afoul of free-speech standards.

When pressed to police their platforms more rigorously, the Internet companies have underscored, first, the enormous practical challenges they face in monitoring the huge volume of material posted every day on their sites. This difficulty remains real, even as artificial intelligence advances. No combination of algorithm and human analysis will rid social media of all falsehoods. But the impossibility of perfection shouldn't become an impediment to improvement. The leaders of Facebook, Twitter, and YouTube created these sprawling businesses, and now they need to accept the responsibilities that come with their influence and financial success.

A second potential obstacle is the concern that aggressive content moderation could endanger the protection from legal liability for the user-generated content on their sites, which the companies enjoy in the U.S. under the Communications Decency Act.

But as we point out in this report, these protections do not depend on absolute neutrality and would not be jeopardized if the companies commit to taking down false information.

A third potential hurdle is the reality that removing disinformation would disproportionately affect accounts attached to conservative sites, whose operators and followers would protest. "It's the type of action that would tick off big parts of [the platform companies'] constituency," Ghosh says.[102] But this is a potential hazard that the companies can address by developing objective and transparent policies and applying them fairly.

Some commentators warn that it would be dangerous to empower the platforms to remove content based on its lack of veracity. "I'm very afraid of what happens five or 10 years out, when the companies have machine-learning systems that understand human languages as well as humans," says Alex Stamos, another Facebook alumnus who until mid-2018 headed security for the company. "We could end up with machine-speed, real-time moderation of everything we say online. Do we want to put machines in charge that way?"[103]

We obviously don't want the machines running amok. Humans need to remain firmly in control of site moderation, closely overseeing the deployment and impact of AI in all its forms. The problem of disinformation cannot be solved by technology alone. The Internet platforms will need to rethink their business models, recognizing that significant additional personnel—on top of the thousands hired in 2018—will be needed to address current and future challenges posed by disinformation. And while this report focuses on content generated in the U.S., the companies also must beef up their content-reviewing teams in other countries—especially where false information online has been used to manipulate populations and spark mass violence.

The platforms would be wise to recall a warning from Senator Dianne Feinstein when she addressed senior lawyers from Facebook, Twitter, and Google during a hearing in November 2017. "You've created these platforms, and now they are being misused," the California Democrat said. "You have to be the ones to do something about it. Or we will."[104] Feinstein was specifically addressing Russian disinformation, but her point covers all forms of falsehood. If the platforms do not improve their self-governance, they risk government intervention that could overreach and raise essential free-speech concerns. With the exception of certain narrowly targeted regulation (see below), we favor the companies getting their own houses in order.

> "
> 'You've created these platforms, and now they're being abused,' Senator Dianne Feinstein has told the social media companies. 'You have to be the ones to do something about it. Or we will.'
> „

Recommendations to Social Media Companies

1 | ## Remove provably false content, whether generated abroad or at home.

Content that's provably untrue should be removed from social media sites, not merely demoted or annotated, as is the platforms' current practice. Provably false content is a narrower category of material than disinformation, as we've defined it. Focusing on provably false content will make the companies' task more feasible. We recommend that they start with falsehoods bearing on the political process and democratic institutions. The companies can use systems already in place that rely on a combination of AI and human review to make the often-difficult judgments they're currently making to exclude categories of expression such as hate speech and harassment. If Facebook can down-rank "false news" so that its visibility decreases by 80 percent, the company can take the next step and—with all due care—get rid of untruths altogether. Likewise, if YouTube can identify and annotate videos that promote notorious conspiracy theories, it can remove the videos altogether. At the same time, the platforms can retain the option of demoting other content that borders on violating their rules but doesn't quite cross the line.

As we've noted, none of this implicates the First Amendment or international protections of free speech, which forbid government censorship. As nongovernmental entities, the platforms have a right, and, in fact, a duty, to protect their users from rank falsehood, whether it's motivated by a desire to deceive voters or generate clickbait profits (or some combination of the two). Alex Jones' hoaxes and prevarication do not contribute anything valuable to the marketplace of ideas. The time has come for the platforms to block content from such sources, not only because it may constitute hate speech or harassment, but because it's manifestly false. Jones and his ilk would remain free to preach their paranoid gospel from their own websites and alt-right havens like Gab.

In the U.S., Twitter faces a particular challenge relating to inaccurate tweets by President Trump. Twitter reasonably asserts that it needs to provide special leeway to world leaders whose tweets are newsworthy.[105] The public should know what world leaders are thinking and saying. But to counterbalance assertions by the U.S. president or other world leaders that are at odds with the truth, Twitter needs to consider actively curating the many opposing comments these tweets provoke. This curation could lend prominence to fact-checking by professional organizations and to helpful correctives offered by ordinary Twitter users. The public then would have access to both newsworthy (dis)information and the context with which to interpret it.

2 | ## Clarify publicly the principles used for removal decisions.

The social media platforms also need to be more transparent and consistent in articulating the principles they are relying on to make decisions about problematic content. An episode from July 2018 underscores the current social media muddle over disinformation. In an interview with *Recode,* Mark Zuckerberg said that Holocaust deniers are factually wrong and offensive but shouldn't be removed from Facebook. "I don't believe that our platform should take that down because I think there are things that different people get wrong," Zuckerberg explained. Referring to Holocaust deniers, he added, "I don't think that they're intentionally getting it wrong."[106]

It's true, of course, that people get a lot of things wrong. We don't expect the platforms to patrol for every trivial mistake. And even some egregious falsehoods will always slip through; no enforcement system is perfect. We are concerned about what Facebook does with instances of falsehood that are flagged by users, AI, or the company's own reviewers—and that involve content that's wrong in a way that matters. Holocaust denialism provides a perfect illustration: It is a provable fallacy that reflects and fuels anti-Semitism, a lethal form of bigotry. Contrary to Zuckerberg's comments, the intent of the Holocaust denier is irrelevant. After careful analysis by a human reviewer, such objectively false content deserves not just demotion in News Feed, but removal from the site.

The principles the platforms need to explain include the connection between facts, rational argument, and a healthy democracy. Social media sites contaminated by disinformation erode core democratic institutions like free exercise of religion and the right to vote in fair elections. That's why the Russians mounted their digital disinformation campaign and why that continuing effort remains so dangerous. Domestically generated falsity can be just as damaging. Both forms of untruth deserve removal.

3 | ### Hire a senior content overseer.

Each platform should bring in a seasoned executive who would have company-wide responsibility for combating false information. The person holding this position should report to the chief executive officer or chief operating officer to ensure that this work receives the resources and internal support it needs to be effective. This individual would help define the principles outlined above and be responsible for applying them fairly and in a manner that strengthens democratic discourse.

Advocating that Facebook should make such a hire, Margaret Sullivan, *The Washington Post's* media columnist and the former public editor at *The New York Times*, has suggested that someone with long experience in serious journalism would fit the bill. We agree. Top editors make decisions every day about what's real and fake. The decisions aren't always perfect, but that's what follow-up articles and corrections are for. "It comes down to judgment—the kind that can't be done by complicated code or by relying on well-intentioned but vague 'community standards,'" Sullivan has written in the *Post*. Explicitly injecting that kind of judgment into the content-review process would improve decision-making and signal to the public a greater degree of earnest-ness on the part of the social media companies.[107]

4 | ### Establish more robust appeals processes.

A more vigorous disinformation-removal policy would necessitate a more thorough and transparent appeals mechanism. Erroneous takedowns are inevitable. Mark Zuckerberg has acknowledged that Facebook's review teams "make the wrong call in more than one out of every 10 cases."[108] There's no reason to think that Twitter or YouTube have a better average. While they seek to improve their initial error rates, the platforms must develop more effective appeals processes so that users can seek to have themselves and/or their content reinstated. First, the social media companies should provide notice to each user whose content is removed or account suspended, including the reason for the action. Then, the companies should provide a meaningful opportunity for appeal to a person or people not involved in the initial decision.[109] At present, Facebook says that its Community Operations team hears appeals of takedown decisions within 24 hours.[110] But the relationship between that team and the front-line reviewers isn't clear. More transparency is needed.

Facebook has said that it is setting up a new, independent body made up of non-employees—perhaps 40 in number—to hear appeals on especially "consequential" content decisions and render rulings that are transparent and binding.[111] This initiative holds promise. If such a review panel were given real authority, it could achieve a greater degree of fairness while simultaneously shedding light on some of the platform's inner workings. Twitter and YouTube should explore similar arrangements.

5 | ### Step up efforts to expunge bot networks.

The bot infestation remains acute. On Twitter, suspected bots account for an astounding 66 percent of tweeted links to news and current events sites, according to the Pew Research Center.[112] By imitating human behavior online, botnets can boost the spread of disinformation by orders of magnitude. Humans are extremely vulnerable to this manipulation, sharing considerable amounts of dubious content posted by bots.[113]

The platforms have made strides in bot detection, but on Twitter the problem may be getting worse. Bot producers are notoriously inventive when seeking to stay one step ahead of pursuers. The cat-and-mouse game will continue. For fear of tipping off their prey, the social media platforms are reluctant to explain publicly how they go about tracking bots. That's fair enough. But the hunt must be pursued with increased urgency.

6 | Retool algorithms to reduce the outrage factor.

As we've noted in this report and earlier work, the advertising-driven social media business model inadvertently provides a receptive environment for purveyors of disinformation. That's because it rewards content, including disinformation, that provokes negative emotional reactions. Platform algorithms seek to promote user engagement, which maximizes ad revenue. And since users are drawn to sensationalist and provocative content, that's what the algorithms favor.[114]

This state of affairs has caused a number of analysts to propose rethinking the economics of social media. One idea, put forward by digital financier Roger McNamee, is to switch away from an advertising model altogether to one based on user subscriptions, as is the practice in the cable television industry. McNamee, an early investor in Facebook, argues that the change would allow the company to stop relying on algorithms that boost engagement "by appealing to emotions such as fear and anger."[115]

We are not proposing that Facebook and other Internet platforms abandon their core advertising-based business models, which generate tens of billions of dollars in annual revenue. But these companies can and should retool their algorithms in a manner that ceases to reward emotionally inflammatory falsehoods.

7 | Provide more data for academic research.

Academic research groups, such as those at Indiana University, MIT, NYU, and Oxford, are eager to expand their studies of social media, as are civil society groups. But they often lack the raw data that only the platform companies can provide. In April 2018, Facebook announced plans to form a "commission" of academics who would develop research priorities on social media's effect on elections. Meanwhile, though, labs at various universities are ready to move forward with research projects right now. Some academics are developing machine-learning algorithms to detect bots more effectively. Because of the danger of false positives—algorithms make errors, too—academic researchers are champing at the bit to investigate anti-bot countermeasures that "take into account the complex interplay between the cognitive and technological factors that favor the spread of misinformation."[116]

Beyond bots, a group of academics from various fields and universities has sounded a call for broader disclosure of platform data about disinformation as a way of encouraging more scholarly investigation. "There is little research focused on fake news and no comprehensive data-collection system to provide a dynamic understanding of how pervasive systems of fake news provision are evolving," the academics said in an article published in *Science.* "Researchers need to conduct a rigorous, ongoing audit of how the major platforms filter information," they added. There are challenges to scientific collaboration, not least company fears of revealing trade secrets. "Yet," the academics said, "there is an ethical and social responsibility, transcending market forces, for the platforms to contribute what data they uniquely can to a science of fake news."[117]

8 | Increase industry-wide cooperation.

We have previously recommended enhanced industry-wide cooperation to combat Russian disinformation, and the same logic applies to the fight against domestic falsehood. Each social media company sees a different slice of the disinformation picture. No one company sees the problem in its entirety. It thus makes sense for them to exchange data and analysis in hopes of strengthening what ought to be a joint effort against a common set of foes.[118]

The companies have worked together through the Global Internet Forum to Counter Terrorism and the Global Network Initiative, which focuses on freedom of expression and privacy. They also collaborate on the PhotoDNA Initiative, which deals with child pornography, and on a database of "digital fingerprints," which allow them to take down violent extremist video more efficiently. The concerted energy animating these efforts should carry over to a new industry initiative devoted to countering disinformation, both foreign and domestic. One topic worthy of cooperative research is detection of deepfake video, a threat to users of all of the platforms.

9 Boost corporate support for digital media literacy.

Attempts to foster critical thinking about social media content—where it comes from, how it can mislead, whether to share it—lately have come in for criticism. The knocks on digital media programs in schools include that they're often superficial or skewed to the values of white progressives.[119] These critiques deserve serious attention, but they aren't fatal. While they don't offer a panacea, media literacy efforts should evolve and be strengthened. The flow of information online is too vast for social media platforms to catch every instance of disinformation (or hate speech or violent extremism). Users must bear responsibility for helping separate wheat from chaff. Media literacy training prepares them to do so.

The platforms have taken some steps in the right direction and need to do more. Facebook has made available a "digital literacy library," with ready-made lessons for teachers of students between the ages of 11 and 18.[120] Twitter has introduced its own educator's guide and has supported nonprofits that promote media literacy.[121] YouTube is participating in a Google-funded project called MediaWise, which combines research by the Poynter Institute and Stanford University with videos by YouTube personalities to teach teenagers how to be smarter consumers of online information.[122] It would make sense for the platforms also to underwrite rigorous academic research evaluating various literacy efforts with an eye toward identifying the most effective ones. National programs in Finland, Norway, and Sweden have received positive attention, and deserve close study.[123]

10 Sponsor more fact-checking and explore new approaches to news verification.

Facebook's collaboration with fact-checkers has not gone smoothly. The social network has partnerships with 35 outside fact-checking organizations in 24 countries. Some employees of these organizations have criticized the arrangement as inefficient and, given the scope of the disinformation problem, ineffective.[124] At times, fact checkers have come under attack for their alleged liberal bias. In January 2018, Google suspended a fact-checking experiment called "Reviewed Claims" when conservative websites alleged the search engine was singling them out for unfair scrutiny.[125] Generally, though, the partisan assault on fact-checking appears to be unwarranted—yet another symptom of our hyper-polarized politics.

Whatever improvements need to be made, fact-checking remains an important exercise. It will never keep up with all of the untruth sloshing around the Internet. But fact-checkers collectively do catch scores of online whoppers every day. By their very existence, they serve as a reminder that there is a difference between reality and unreality. In this sense, fact-checking underscores the importance of traditional shoe-leather journalism and the vital role that reporters and editors play in holding accountable those with political and corporate power. The *Washington Post's* Fact Checker has done this job well, as have the Annenberg Center's *FactCheck.org,* the Poynter Institute's *Politifact, Snopes,* and others. We urge Facebook to continue to expand its fact-checking partnerships; Twitter and YouTube should follow suit.

Scrutinizing individual articles, which is what most fact checkers do, isn't the only valid approach to promoting reality-based public life. NewsGuard Technologies, a small startup, offers ratings of entire online news sites. Its analysts compile findings across nine "journalistic integrity criteria" to produce the equivalent of a nutrition label from which users can determine whether it's safe to consume what a site publishes. NewsGuard also offers a corresponding reliability rating: green for read on, red for "proceed with caution." The for-profit company has signed up Microsoft as its first major client and hopes to license its analysis to the social media giants. In the meantime, it's offering the red/green signals and nutrition labels in a free browser extension.[126] Other entrepreneurs are developing new verification strategies, and the platforms should consider patronizing—or investing in—the ones that work best. Finally, the Knight Commission on Trust, Media and Democracy recently issued a report—"Crisis in Democracy: Renewing Trust in America"—that contains worthwhile proposals for restoring trust in democratic institutions, including social media.

11 | Support narrow, targeted government regulation.

Broad government regulation, as we've observed, risks official censorship. But that's not to say that lawmakers ought to do nothing. One idea that has floated around Capitol Hill since fall of 2017 is mandating the same degree of disclosure for online political advertising as currently exists for traditional broadcast media. Known as the Honest Ads Act and co-sponsored by Democratic Senators Mark Warner of Virginia and Amy Klobuchar of Minnesota, the bill would expand on existing transparency requirements the platforms say they voluntarily enforce. We favor a codified regulatory model that would deter disinformation artists, foreign and domestic, from using advertising as an instrument for distortion. Enforcement authority should be given to the Federal Trade Commission or the Federal Communications Commission, rather than the Federal Election Commission, to take advantage of the greater enforcement capacity of the FTC and FCC.

Another proposal for increased transparency comes from Facebook's Mark Zuckerberg, who has invited government to require that platforms "report the prevalence of harmful content on their services and then work to reduce that prevalence." Reporting these metrics would allow regulators and the public to assess which companies are improving and which are not. Zuckerberg has said that Facebook is already working with the French government on this idea and hopes to do the same with the European Commission.[127]

It's likely that Facebook will use a definition for prevalence of harmful content that yields a very small-sounding percentage, as the company has done in the past. In 2016, Facebook claimed that Russian disinformation amounted to only 0.004% of all content on the social network. But the company later acknowledged that at least 146 million Americans encountered fraudulent Russian content on Facebook and Instagram alone. The latter statistic speaks loudly. Nonetheless, a comparison of prevalence—defined in a meaningful manner—could still be helpful in determining which platforms are attacking disinformation with sufficient vigor.

Endnotes

1 Dara Lind, "The Conspiracy Theory that Led to the Pittsburgh Synagogue Shooting, Explained," *Vox*, October 29, 2018 (https://www.vox.com/2018/10/29/18037580/pittsburgh-shooter-anti-semitism-racist-jewish-caravan).

2 Casey Newton, "How Platforms Are Driving Users to Misinformation," *The Verge*, October 27, 2018 (https://www.theverge.com/2018/10/27/18029490/cesar-sayoc-mail-bombs-twitter-instagram-misinformation); Kevin Roose, "Cesar Sayoc's Path on Social Media: From Food Photos to Partisan Fury," *The New York Times,* October 27, 2018 (https://www.nytimes.com/2018/10/27/technology/cesar-sayoc-facebook-twitter.html).

3 Nahema Marchal, et al., "Polarization, Partisanship, and Junk News Consumption on Social Media During the 2018 U.S. Midterm Elections," Oxford Internet Institute, November 1, 2018 (https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/11/marchal_et_al.pdf).

4 "Junk News Dominating Coverage of U.S. Midterms on Social Media, New Research Finds," Oxford Internet Institute, November 1, 2018 (https://www.oii.ox.ac.uk/news/releases/junk-news-dominating-coverage-of-us-midterms-on-social-media-new-research-finds/).

5 Drew Harwell, "White House Shares Doctored Video to Support Punishment of Journalist Acosta," *The Washington Post*, November 8, 2018 (https://www.washingtonpost.com/technology/2018/11/08/white-house-shares-doctored-video-support-punishment-journalist-jim-acosta/?utm_term=.09f707a06a89). For Paul Joseph Watson's denial, see his YouTube video, "The Jim Acosta Controversy," November 8, 2018 (https://www.youtube.com/watch?v=zo7ORobbXPw).

6 Paul M. Barrett, Tara Wadhwa, Dorothée Baumann-Pauly, "Combating Russian Disinformation: The Case for Stepping Up the Fight Online," New York University Stern Center for Business and Human Rights, July 2018 (https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_stern_cbhr_combating_russian_di?e=31640827/63115656).

7 Id.

8 Vidya Narayanan et al., "Polarization, Partisanship, and Junk News Consumption Over Social Media in the U.S.," Oxford Internet Institute, February 6, 2018 (https://pdfs.semanticscholar.org/f41d/cb3a73094a5028e9ab841a1708c9451551b5.pdf).

9 Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford University Press, 2018 (https://global.oup.com/academic/product/network-propaganda-9780190923631?cc=us&lang=en&).

10 Alice E. Marwick, "Why Do People Share Fake News? A Sociotechnical Model of Media Effects," *Georgetown Law Technology Review*, July 2018 (https://georgetownlawtechreview.org/why-do-people-share-fake-news-a-sociotechnical-model-of-media-effects/GLTR-07-2018/).

11 Interview with author.

12 Interview with author.

13 Interview with author.

14 Cheng Cheng Shao et al., "The Spread of Low-Credibility Content by Social Bots," *Nature Communications,* November 2018 (https://www.nature.com/articles/s41467-018-06930-7).

15 Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement," Facebook, November 15, 2018 (https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/).

16 Rui Fan et al., "Anger is More Influential than Joy: Sentiment Correlation in Weibo," *PLoS One*, October 2014 (https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0110184&type=printable).

17 William J. Brady et al., "Emotion Shapes the Diffusion of Moralized Content in Social Networks," *Proceedings of the National Academy of Sciences of the United States of America,*" July 11, 2017 (https://www.pnas.org/content/pnas/114/28/7313.full.pdf).

18 Andrew Guess, Jonathan Nagler, and Joshua Tucker, "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook," *Social Advances,* June 9, 2019 (http://advances.sciencemag.org/content/5/1/eaau4586#F1).

19 Soroush Vosoughi, Deb Roy, Sinan Aral, "The Spread of True and False News Online," *Science,* March 9, 2018 (http://science.sciencemag.org/content/359/6380/1146).

20 Jonathan Swift, "Political Lying," *The Examiner*, 1710, quoted by Bartleby.com (https://www.bartleby.com/209/633.html).

21 Michael Schudson and Barbie Zelizer, "Fake News in Context," Understanding and Addressing the Disinformation Ecosystem (conference), December 2017 (https://firstdraftnews.org/wp-content/uploads/2018/03/The-Disinformation-Ecosystem-20180207-v2.pdf).

22 Interview with author.

23 "Newspaper Publishers Lose Over Half Their Employment from January 2001 to September 2016," Bureau of Labor Statistics, April 3, 2017 (https://www.bls.gov/opub/ted/2017/newspaper-publishers-lose-over-half-their-employment-from-january-2001-to-september-2016.htm).

24 Jeffrey M. Jones, "U.S. Media Trust Continues to Recover from 2016 Low," Gallup, October 12, 2018 (https://news.gallup.com/poll/243665/media-trust-continues-recover-2016-low.aspx).

25 Katerina Eva Matsa and Elisa Shearer, "News Use Across Social Media Platforms 2018," Pew Research Center, September 10, 2018 (http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/).

26 Elizabeth Williamson, "Truth in a Post-Truth Era: Sandy Hook Families Sue Alex Jones, Conspiracy Theorist, *The New York Times,* May 23, 2018 (https://www.nytimes.com/2018/05/23/us/politics/alex-jones-trump-sandy-hook.html).

27 Elizabeth Williamson and Emily Steel, "Conspiracy Theories Made Alex Jones Very Rich. They May Bring Him Down," *The New York Times*, September 7, 2018 (https://www.nytimes.com/2018/09/07/us/politics/alex-jones-business-infowars-conspiracy.html).

28   Jason Murdock, "Infowars Old Twitter Posts Probably Broke Policy—Then They All Disappeared," *Newsweek*, August 10, 2018 (https://www.newsweek.com/infowars-old-twitter-posts-probably-broke-policy-then-they-mysteriously-1067414).

29   Cassandra Vinograd, "Donald Trump Tweets a Picture of Jeb Bush Next to a Swastika," *NBC News*, November 4, 2015 (https://www.nbcnews.com/politics/2016-election/donald-trump-tweets-picture-jeb-bush-next-swastika-n456976).

30   James Hoft, "The State of Intellectual Freedom in America," Testimony Before the House Subcommittee on the Constitution and Civil Justice, September 27, 2018 (https://judiciary.house.gov/wp-content/uploads/2018/06/Supplemental-Testimony-Jim-Hoft-09.27.2018.pdf).

31   David Gilmour, "What Is the *Gateway Pundit*?" *Daily Dot*, March 26, 2018 (https://www.dailydot.com/layer8/gateway-pundit/).

32   Rob Faris et al., "Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election," Berkman Klein Center for Internet & Society at Harvard University, August 16, 2017 (https://cyber.harvard.edu/publications/2017/08/mediacloud).

33   Interview with author.

34   Yochai Benkler et al., "Breitbart-Led Right-Wing Media Ecosystem Altered Broader Media Agenda," *Columbia Journalism Review*, March 3, 2017 (https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php).

35   Matthew Hindman and Vlad Barash, "Disinformation, Fake News, and Influence Campaigns on Twitter," Knight Foundation, October 2017 (https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/238/original/KF-DisinformationReport-final2.pdf).

36   Interview with author.

37   Jane Lytvynenko, "Infowars Has Republished More Than 1,000 Articles from RT Without Permission," *BuzzFeed News*, November 8, 2017 (https://www.buzzfeednews.com/article/janelytvynenko/infowars-is-running-rt-content). Infowars did not respond to several requests for comment.

38   Benkler, Faris, and Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (supra note 9).

39   McKay Coppins, "How the Left Lost Its Mind," *The Atlantic*, July 2, 2017 (https://www.theatlantic.com/politics/archive/2017/07/liberal-fever-swamps/530736/); Chris Cassidy, "Ed Markey Issues Mea Culpa for Grand Jury Claim," *Boston Herald*, May 11, 2017 (https://www.bostonherald.com/2017/05/11/ed-markey-issues-mea-culpa-for-grand-jury-claim/).

40   Bethania Palma, "Did Jared Kushner Go to Saudi Arabia Because It Doesn't Have an Extradition Treaty With the US?" *Snopes,* October 31, 2017 (https://www.snopes.com/fact-check/jared-kushner-go-saudi-arabia-doesnt-extradition-treaty-us/).

41   Scott Shane and Alan Blinder, "Secret Experiment in Alabama Senate Race Imitated Russian Tactics," *The New York Times*, December 19, 2018 (https://www.nytimes.com/2018/12/19/us/alabama-senate-roy-jones-russia.html); Scott Shane and Alan Blinder, "Posing as Prohibitionists, 2nd Effort Used Online Fakery in Alabama Race," *The New York Times*, January 7, 2019 (https://www.nytimes.com/2019/01/07/us/politics/alabama-senate-facebook-roy-moore.html).

42   Tony Romm and Craig Timberg, "Facebook Suspends Five Accounts, Including that of a Social Media Researcher, for Misleading Tactics in Alabama Election," *The Washington Post*, December 22, 2018 (https://www.washingtonpost.com/technology/2018/12/22/facebook-suspends-five-accounts-including-social-media-researcher-misleading-tactics-alabama-election/?noredirect=on&utm_term=.be9a72b4e7bf); Tony Romm, Elizabeth Dwoskin, and Craig Timberg, "Facebook Is Investigating the Political Pages and Ads of Another Group Backed by Reid Hoffman," *The Washington Post*, January 7, 2019 (https://www.washingtonpost.com/technology/2019/01/07/facebook-is-investigating-political-pages-ads-another-group-backed-by-reid-hoffman/?utm_term=.9c7e25085689).

43   Brian Stelter, "Birth of a Conspiracy Theory: How Trump's Wiretap Claim Got Started," *CNN,* March 6, 2017 (https://money.cnn.com/2017/03/06/media/mark-levin-joel-pollak-breitbart-trump-obama/index.html); David Smith, "Obama Spokesman Dismisses Trump's Wiretap Outburst as 'Simply False,'" *The Guardian*, March 5, 2017 (https://www.theguardian.com/us-news/2017/mar/04/donald-trump-wiretap-barack-obama-coup).

44   Derek Hawkins, "Andrew Napolitano Reportedly Pulled from Fox News Over Debunked Wiretapping Claims," *The Washington Post*, March 21, 2017 (https://www.washingtonpost.com/news/morning-mix/wp/2017/03/21/andrew-napolitano-reportedly-pulled-from-fox-news-over-debunked-wiretapping-claims/?utm_term=.1655506c3501).

45   "All False Statements Involving Donald Trump," *Politifact*, undated, (https://www.politifact.com/personalities/donald-trump/statements/byruling/false/).

46   Jonathan A. Greenblatt, "When Hate Goes Mainstream," *The New York Times*, October 28, 2018 (https://www.nytimes.com/2018/10/28/opinion/synagogue-shooting-pittsburgh-anti-defamation-league.html).

47   Andrew Anglin, "Andrew Anglin Exposed," *Daily Stormer*, March 14, 2015 (https://dailystormer.name/andrew-anglin-exposed/); "About Andrew Anglin," Southern Poverty Law Center, undated (https://www.splcenter.org/fighting-hate/extremist-files/individual/andrew-anglin).

48   Keith Collins, "A Running List of Websites and Apps that Have Banned, Blocked, Deleted, and Otherwise Dropped White Supremacists," *Quartz*, August 16, 2017 (https://qz.com/1055141/what-websites-and-apps-have-banned-neo-nazis-and-white-supremacists/).

49   Kevin Roose, "On Gab, an Extremist-Friendly Site, Pittsburgh Shooting Suspect Aired His Hatred in Full," *The New York Times*, October 28, 2018 (https://www.nytimes.com/2018/10/28/us/gab-robert-bowers-pittsburgh-synagogue-shootings.html); Craig Timberg et al., "From Silicon Valley Elite to Social Media Hate: The Radicalization that Led to Gab," *The Washington Post*, October 31, 2018 (https://www.washingtonpost.com/technology/2018/10/31/silicon-valley-elite-social-media-hate-radicalization-that-led-gab/?utm_term=.3474c1b5b922).

50   Jasmine Garsd, "After Synagogue Attack, Web-Hosting Sites Suspend Gab," *National Public Radio,* October 29, 2018 (https://www.npr.org/2018/10/29/661676103/after-synagogue-attack-web-hosting-sites-suspend-gab).

51   Alex Amend and the Network Contagion Research Institute, "On Gab, Domestic Terrorist Robert Bowers Engaged With Several Influential Alt-Right Figures," *Southern Poverty Law Center Hatewatch,* November 1, 2018 (https://www.splcenter.org/hatewatch/2018/11/01/gab-domestic-terrorist-robert-bowers-engaged-several-influential-alt-right-figures); see also Savvas Zannettou et al., "What Is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber," *arXiv: 1802.05287,* February 2018 (https://arxiv.org/abs/1802.05287).

Endnotes (continued)

52 Kevin Roose, "On Gab, an Extremist-Friendly Site, Pittsburgh Shooting Suspect Aired His Hatred in Full" (supra note 49).

53 April Glaser, "White Supremacists Still Have a Safe Space Online," *Slate*, October 9, 2018 (https://slate.com/technology/2018/10/discord-safe-space-white-supremacists.html).

54 Meaghan M. McDermott, "Greece Man Accused of Muslim Bombing Plot Posted Alt-Right Conspiracies on Twitter," *Rochester Democrat and Chronicle*, January 27, 2019 (https://www.democratandchronicle.com/story/news/2019/01/27/islamberg-ny-attack-greece-twitter-youtube-alt-right-conspiracy-andrew-crysel-vincent-vetromile/2679674002/).

55 Aaron Smith and Monica Anderson, "Social Media Use in 2018," Pew Research Center, March 1, 2018 (http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/).

56 Rebecca Lewis, "Alternative Influence: Broadcasting the Reactionary Right on YouTube," Data & Society Research Institute, September 2018 (https://datasociety.net/output/alternative-influence/).

57 Id.

58 Paul Lewis, "'Fiction is Outperforming Reality': How YouTube's Algorithm Distorts Truth," *The Guardian*, February 2, 2018 (https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth).

59 Jack Nicas, "How YouTube Drives People to the Internet's Darkest Corners," *The Wall Street Journal*, February 7, 2018 (https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478).

60 Tony Romm and Drew Harwell, "Searching for News on RBG? YouTube Offered Conspiracy Theories about the Supreme Court Justice Instead," *The Washington Post*, January 11, 2019 (https://www.washingtonpost.com/technology/2019/01/11/searching-news-rbg-youtube-offered-conspiracy-theories-about-supreme-court-justice-instead/?utm_term=.c1b2930c95cb).

61 Craig Timberg et al., "Two Years After #Pizzagate Showed the Dangers of Hateful Conspiracies, They're Still Rampant on YouTube," *The Washington Post,* December 10, 2018 (https://www.washingtonpost.com/business/technology/hateful-conspiracies-thrive-on-youtube-despite-pledge-to-clean-up-problematic-videos/2018/12/10/625730a8-f3f8-11e8-9240-e8028a62c722_story.html?utm_term=.d358815fb5a7).

62 Ibid; David Emery, "Is Hillary Clinton 'Snuff Film' Circulating on the Dark Web?" *Snopes*, April 16, 2018 (https://www.snopes.com/fact-check/hillary-clinton-snuff-film/).

63 Craig Silverman, "How to Spot a Deepfake Like the Barack Obama-Jordan Peele Video," *BuzzFeed News*, April 17, 2018 (https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed).

64 Dan Evon, "Is This a Photograph of Christine Blasey Ford Holding a 'Not My President' Sign?" *Snopes*, September 19, 2018 (https://www.snopes.com/fact-check/christine-blasey-ford-not-my-president/).

65 Saranac Hale Spencer, "Viral Photo Doesn't Show Soros with Ford," *FactCheck.org*, September 28, 2018 (https://www.factcheck.org/2018/09/viral-photo-doesnt-show-soros-with-ford/).

66 Sheera Frenkel et al., "Delay, Deny, and Deflect: How Facebook's Leaders Fought Through Crisis," *The New York Times*, November 14, 2018, (https://www.nytimes.com/2018/11/14/technology/facebook-data-russia-election-racism.html?module=inline).

67 "*New York Times Update*," Facebook, November 15, 2018 (https://newsroom.fb.com/news/2018/11/new-york-times-update/).

68 Jack Nicas and Matthew Rosenberg, "A Look Inside the Tactics of Definers, Facebook's Attack Dog," *The New York Times*, November 15, 2018 (https://www.nytimes.com/2018/11/15/technology/facebook-definers-opposition-research.html https://www.nytimes.com/2018/11/15/technology/facebook-definers-opposition-research.html).

69 Joe Pounder, Fran Brennan, and Jeff Bechdel, "About *NTK Network*," *NTK Network*, undated (https://ntknetwork.com/about-ntk-network/).

70 Philip Howard et al., "The IRA, Social Media, and Political Polarization in the United States, 2012-2018," Oxford Internet Institute, December 2018 (https://comprop.oii.ox.ac.uk/research/ira-political-polarization/); Renee DiResta, et al., "The Tactics and Tropes of the Internet Research Agency," New Knowledge, December 2018 (https://disinformationreport.blob.core.windows.net/disinformation-report/NewKnowledge-Disinformation-Report-Whitepaper.pdf).

71 Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement" (supra note 15).

72 Hunt Allcott, Matthew Gentzkow, and Chuan Yu, "Trends in the Diffusion of Misinformation on Social Media," web.Stanford.edu, September 2018 (https://web.stanford.edu/~gentzkow/research/fake-news-trends.pdf).

73 Ibid.

74 Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement" (supra note 15).

75 Calum McClelland, "The Difference Between Artificial Intelligence, Machine Learning, and Deep Learning," *Medium*, December 4, 2017 (https://medium.com/iotforall/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning-3aa67bff5991).

76 Tessa Lyons, "Hard Questions: How Is Facebook's Fact-Checking Program Working?" Facebook, June 14, 2018 (https://newsroom.fb.com/news/2018/06/hard-questions-fact-checking/); Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement" (supra note 15).

77 "Community Standards," Facebook, undated (https://www.facebook.com/communitystandards/introduction).

78 "Continuing Our Work to Improve Recommendations on YouTube," YouTube, January 25, 2019 (https://youtube.googleblog.com/).

79 Vijaya Gadde and Kayvon Beykpour, "Setting the Record Straight on Shadow Banning," Twitter, July 26, 2018 (https://blog.twitter.com/official/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning.html); Del Harvey and David Gasca, "Serving Healthy Conversation," Twitter, May 15, 2018 (https://blog.twitter.com/official/en_us/topics/product/2018/Serving_Healthy_Conversation.html).

80 Arjun Kharpal, "Facebook Doesn't Want to Be the 'Arbiter of the Truth,' Top Exec Sheryl Sandberg Says, Amid Fake News Criticism," *CNBC*, April 24, 2017 (https://www.cnbc.com/2017/04/24/facebook-fake-news-sheryl-sanberg.html).

81 Brian Hiatt, "Twitter CEO Jack Dorsey: The *Rolling Stone* Interview," *Rolling Stone*, January 23, 2019 (https://www.rollingstone.com/culture/culture-features/twitter-ceo-jack-dorsey-rolling-stone-interview-782298/).

82 Tony Romm and Drew Harwell, "Searching for News on RBG? YouTube Offered Conspiracy Theories about the Supreme Court Justice Instead" (supra note 60).

83 Tessa Lyons, "Replacing Disputed Flags with Related Articles," Facebook, December 20, 2017 (https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/).

84 Neal Mohan and Robert Kyncl, "Building a Better News Experience on YouTube, Together," YouTube, July 9, 2018 (https://youtube.googleblog.com/2018/07/building-better-news-experience-on.html); Alex Hern, "YouTube to Crack Down on Fake News, Backing 'Authoritative' Sources," *The Guardian*, July 9, 2018 (https://www.theguardian.com/technology/2018/jul/09/youtube-fake-news-changes).

85 Nathaniel Gleicher and Oscar Rodriguez, "Removing Additional Inauthentic Activity from Facebook," Facebook, October 11, 2018 (https://newsroom.fb.com/news/2018/10/removing-inauthentic-activity/).

86 Simone Stolzoff, "The Problem with Social Media Has Never Been About Bots. It's Always Been About Business Models," *Quartz*, November 16, 2018 (https://qz.com/1449402/how-to-solve-social-medias-bot-problem/).

87 Craig Timberg and Elizabeth Dwoskiin, "Twitter Is Sweeping Out Fake Accounts Like Never Before, Putting User Growth at Risk," *The Washington Post*, July 6, 2018 (https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk/?utm_term=.7c28ea3af9a9).

88 Jack Dorsey, Testimony Before the U.S. House Committee on Energy and Commerce, September 5, 2018 (https://docs.house.gov/meetings/IF/IF00/20180905/108642/HHRG-115-IF00-Wstate-DorseyJ-20180905.pdf).

89 Christopher Bing, "Exclusive: Twitter Deletes Over 10,000 Accounts that Sought to Discourage U.S. Voting," *Reuters*, November 2, 2018 (https://www.reuters.com/article/us-usa-election-twitter-exclusive/exclusive-twitter-deletes-over-10000-accounts-that-sought-to-discourage-u-s-voting-idUSKCN1N72FA).

90 Adam Mosseri, "Bringing People Closer Together," Facebook, January 11, 2018 (https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/); Adam Mosseri, "Helping Ensure News on Facebook Is From Trusted Sources," Facebook, January 19, 2018 (https://newsroom.fb.com/news/2018/01/trusted-sources/).

91 Paris Martineau, "Conservative Publishers Hit Hardest by Facebook News Feed Change," *The Outline*, March 5, 2018 (https://theoutline.com/post/3599/conservative-publishers-hit-hardest-by-facebook-news-feed-change?zd=1&zi=i6mwmzky).

92 James Hoft, "The State of Intellectual Freedom in America" (supra note 30).

93 Adam Mosseri, "Helping Ensure News on Facebook is From Trusted Sources" (supra note 90).

94 Kevin Roose, "Alex Jones Said Bans Would Strengthen Him. He Was Wrong," *The New York Times*, September 4, 2018, (https://www.nytimes.com/2018/09/04/technology/alex-jones-infowars-bans-traffic.html).

95 Jonathan Albright, "Facebook's Failure to Enforce Its Own Rules," *Medium*, November 6, 2018 (https://medium.com/s/the-micro-propaganda-machine/the-2018-facebook-midterms-part-iii-granular-enforcement-10f8f2d97501).

96 Kevin Roose, "Facebook's Private Groups Offer Refuge to Fringe Figures," *The New York Times*, September 3, 2018 (https://www.nytimes.com/2018/09/03/technology/facebook-private-groups-alex-jones.html).

97 Jonathan Albright, "The Shadow Organizing of Facebook Groups," *Medium*, November 4, 2018 (https://medium.com/s/the-micro-propaganda-machine/the-2018-facebook-midterms-part-ii-shadow-organization-c97de1c54c65).

98 Elizabeth Dwoskin and Tony Romm, "Facebook Purged Over 800 U.S. Accounts and Pages for Pushing Political Spam," *The Washington Post*, October 11, 2018 (https://www.washingtonpost.com/technology/2018/10/11/facebook-purged-over-accounts-pages-pushing-political-messages-profit/?utm_term=.d42a09222d60); Sheera Frenkel, "Facebook Tackles Rising Threat: Americans Aping Russian Schemes to Deceive," *The New York Times*, October 11, 2019 (https://www.nytimes.com/2018/10/11/technology/fake-news-online-disinformation.html); Nathaniel Gleicher and Oscar Rodriguez, "Removing Additional Inauthentic Activity from Facebook" (supra note 85).

99 Interviews with author.

100 Renee DiResta, "Free Speech in the Age of Algorithmic Megaphones," *Wired*, October 12, 2018 (https://www.wired.com/story/facebook-domestic-disinformation-algorithmic-megaphones/).

101 Interview with author.

102 Interview with author.

103 Interview with author.

104 Craig Timberg, Hamza Shaban, and Elizabeth Dwoskin, "Fiery Exchanges on Capitol Hill as Lawmakers Scold Facebook, Google, and Twitter," *The Washington Post*, November 1, 2017 (https://www.washingtonpost.com/news/the-switch/wp/2017/11/01/fiery-exchanges-on-capitol-hill-as-lawmakers-scold-facebook-google-and-twitter/?utm_term=.ce958b61d616).

105 "World Leaders on Twitter," Twitter, January 5, 2018 (https://blog.twitter.com/official/en_us/topics/company/2018/world-leaders-and-twitter.html).

106 Kara Swisher, "Zuckerberg: The *Recode* Interview," *Recode*, July 18, 2018 (https://www.recode.net/2018/7/18/17575156/mark-zuckerberg-interview-facebook-recode-kara-swisher).

107 Margaret Sullivan, "Call It a 'Crazy Idea,' Facebook, But You Need an Executive Editor," *The Washington Post*, November 20, 2016 (https://www.washingtonpost.com/lifestyle/style/call-it-what-you-want-facebook-but-you-need-an-executive-editor/2016/11/20/67aa5320-aaa6-11e6-a31b-4b6397e625d0_story.html?utm_term=.74b54d5a2acb).

108 Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement" (supra note 15).

109 "The Santa Clara Principles," Content Moderation at Scale Conference, May 7, 2018 (https://newamericadotorg.s3.amazonaws.com/documents/Santa_Clara_Principles.pdf).

Endnotes (continued)

110 Monika Bickert, "Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process," Facebook, April 24, 2018 (https://newsroom.fb.com/news/2018/04/comprehensive-community-standards/).

111 Nick Clegg, "Charting a Course for an Oversight Board for Content Decisions," Facebook, January 28, 2019 (https://newsroom.fb.com/news/2019/01/oversight-board/).

112 Stefan Wojcik, "5 Things to Know About Bots on Twitter," Pew Research Center, April 9, 2018 (http://www.pewresearch.org/fact-tank/2018/04/09/5-things-to-know-about-bots-on-twitter/).

113 David M.J. Lazer, et al., "The Science of Fake News," *Science*, March 3, 2018, (http://science.sciencemag.org/content/359/6380/1094); Cheng Cheng Shao et al., "The Spread of Low-Credibility Content by Social Bots" (supra note 14).

114 Paul M. Barrett, Tara Wadhwa, Dorothée Baumann-Pauly, "Combating Russian Disinformation: The Case for Stepping Up the Fight Online" (supra note 6).

115 Roger McNamee, "How to Fix Facebook: Make Users Pay for It," *The Washington Post*, February 21, 2018 (https://www.washingtonpost.com/opinions/how-to-fix-facebook-make-users-pay-for-it/2018/02/20/a22d04d6-165f-11e8-b681-2d4d462a1921_story.html?utm_term=.30442925f95b).

116 Cheng Cheng Shao et al., "The Spread of Low-Credibility Content by Social Bots" (supra note 14).

117 David M.J. Lazer, et al., "The Science of Fake News" (supra note 113).

118 Paul M. Barrett, Tara Wadhwa, Dorothée Baumann-Pauly, "Combating Russian Disinformation: The Case for Stepping Up the Fight Online" (supra note 6); Clint Watts, "Extremist Content and Russian Disinformation Online: Working with Tech to Find Solutions," Testimony Before the U.S. Senate Judiciary Committee, October 31, 2017 (https://www.fpri.org/article/2017/10/extremist-content-russian-disinformation-online-working-tech-find-solutions/).

119 Danah Boyd, "You Think You Want Media Literacy…Do You?" Data & Society Research Institute, March 9, 2018 (https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2).

120 Antigone Davis and Karuna Nain, "A New Resource for Educators: Digital Literacy Library," Facebook, August 2, 2018 (https://newsroom.fb.com/news/2018/08/digitalliteracylibrary/).

121 Lucia Gamboa and Patricia Cartes, "Twitter's Contribution to Media Literacy Week," Twitter, November 10, 2017 (https://blog.twitter.com/official/en_us/topics/events/2017/medialiteracyweek2017.html).

122 "Poynter Receives $3 Million from Google to Lead Program Teaching Teens to Tell Fact from Fiction Online," Poynter Institute, March 20, 2018 (https://www.poynter.org/news-release/2018/poynter-receives-3-million-from-google-to-lead-program-teaching-teens-to-tell-fact-from-fiction-online/).

123 "Three Nordic Countries to Increase MIL Among All Citizens," Nordicom, Goteborgs Universitet, September 27, 2018 (http://www.nordicom.gu.se/en/latest/news/three-nordic-countries-increase-mil-among-all-citizens).

124 Sam Levin, "'They Don't Care': Facebook Fact-Checking in Disarray as Journalists Push to Cut Ties," *The Guardian*, December 13, 2018 (https://www.theguardian.com/technology/2018/dec/13/they-dont-care-facebook-fact-checking-in-disarray-as-journalists-push-to-cut-ties).

125 Daniel Funke, "Google Suspends Fact-Checking Feature Over Quality Concerns," Poynter Institute, January 19, 2018 (https://www.poynter.org/fact-checking/2018/google-suspends-fact-checking-feature-over-quality-concerns/).

126 "NewsGuard Now Available on Microsoft Edge Mobile Apps for iOS and Android," NewsGuard Technologies, January 16, 2019 (https://www.newsguardtech.com/press/newsguard-now-available-on-microsoft-edge-mobile-apps-for-ios-and-android/).

127 Mark Zuckerberg, "A Blueprint for Content Governance and Enforcement" (supra note 15).

128 Michael Cappetta, Ben Collins, and Jo Ling Kent, "Facebook Hired Firm with 'In-House Fake News Shop' to Combat PR Crisis," *NBC News*, November 15, 2018 (https://www.nbcnews.com/tech/tech-news/facebook-hired-firm-house-fake-news-shop-combat-pr-crisis-n936591).