Disinformation and the 2020 Election:

How the Social Media Industry Should Prepare

PAUL M. BARRETT

NYU | STERN
Center for Business and Human Rights

September 2019

Contents

Acknowledgments
This report benefited from the insights of many people, including Kevin Carroll of Wiggin and Dana, Joan Donovan of the Technology and Social Change Research Project at the Harvard Kennedy School, Yaël Eisenstat of the Center for Humane Technology, Lee Foster at FireEye, Dipayan Ghosh of the Platform Accountability Project at the Harvard Kennedy School, Brittan Heller of the Carr Center for Human Rights Policy at the Harvard Kennedy School, Justin Hendrix of the NYC Media Lab, Jeff Jarvis of the Craig Newmark Graduate School of Journalism at City University of New York, Darren Linvill and Patrick Warren of Clemson University, Filippo Menczer of Indiana University, Lisa-Maria Neudert of the Oxford Internet Institute, Nick Pickles of Twitter, Iain Robertson and Andy Carvin of the Atlantic Council's Digital Forensic Research Lab, Laura Rosenberger and Bret Schafer of the Alliance for Securing Democracy, Alexandria Walden and Clement Wolf at Google, and Candid Wüest of Symantec.

We are grateful for financial support from the Craig Newmark Philanthropies, the Open Society Foundations, Jonah Goodhart, and the John S. and James L. Knight Foundation.

Author
Paul M. Barrett is the Deputy Director of the New York University Stern Center for Business and Human Rights.

Executive Summary

This report assesses some of the forms and sources of disinformation likely to play a role during the presidential election campaign in 2020. While midterm Election Day in November 2018 did not feature much Russian interference, there is no guarantee that Russia, and possibly other U.S. antagonists, will refrain from digital meddling in the more consequential 2020 contest. What's more, in terms of sheer volume, domestically generated disinformation now exceeds malign content from foreign sources and will almost certainly be a factor in the next election.

> "In terms of sheer volume, domestically generated disinformation now exceeds malign content from foreign sources and will almost certainly be a factor in the next election."

(Executive Summary continued)

The report explores these risks and analyzes what the major social media companies—Facebook, Twitter, and YouTube (owned by Google)—have done to harden their defenses against disinformation. We also offer recommendations of additional steps the companies need to take to prepare for the almost-certain assault in 2020.

Terminology and Scope

By "disinformation," we mean a relatively broad category of false or misleading text or imagery which is intentionally or recklessly spread to deceive, propagandize, promote discord, or make money via "clickbait" schemes. For the sake of variety, we sometimes refer to "false content" or "false information."

We also refer to a narrower category of disinformation—"provably false content"—which we urge the social media companies to remove. To illustrate, consider these hypothetical examples: An article headlined, "The Sandy Hook Massacre Was Staged," can be proven definitively untrue and ought to be deleted for that reason. By contrast, a story headlined, "Journalists Really Are the Enemy of the People," may be unsubstantiated and misleading, but it isn't provably false.

This report doesn't cover cyberattacks — remote intrusions into computer networks, such as the Russian hack-and-release of Democratic email in 2016. Cyberattacks on campaigns and state voting systems are a distinct danger in 2020, but they are beyond the scope of this report. Also beyond the scope are the privacy and disclosure issues that led Facebook to agree in July 2019 to pay $5.1 billion in settlements to the federal government and to submit to new types of oversight.

Summary of our Recommendations to Social Media Companies

1. **Detect and remove deepfake videos.** Realistic but fraudulent videos have the potential to undermine political candidates and exacerbate voter cynicism.

2. **Remove provably false content in general.** The platforms already remove hate speech, voter suppression, and other categories of content; we recommend that they add one more.

3. **Hire a senior content overseer.** Each company needs an executive with clout to supervise the process of guarding against disinformation.

4. **Attend to Instagram's disinformation problem.** The photo-posting platform needs the concerted attention of its parent, Facebook.

5. **Limit the reach of WhatsApp.** Users should be restricted to forwarding content to one chat group at a time.

6. **Defend against for-profit disinformation.** The companies must prepare for false content generated by hired-gun firms.

7. **Back legislation regulating political ads and punishing voter suppression.** Narrowly tailored bills pending in Congress could help curb some forms of disinformation.

8. **Improve industry-wide collaboration on disinformation.** For example, when one platform takes down abusive accounts, others should do the same with affiliated accounts.

9. **Teach social media literacy in a more direct, sustained way.** Users have to take responsibility for recognizing false content, but they need more help to do it.

1. Introduction

> The Nancy Pelosi episode foreshadowed one type of disinformation that could disrupt the 2020 election: deliberately distorted video, amplified via social media.

In late May of 2019, two manipulated videos disparaging House Speaker Nancy Pelosi spread across social media. One, posted on a conservative Facebook page called Politics WatchDog, had been altered to slow Pelosi's speech during a public appearance, making it appear as if she drunkenly slurred her words. Rudolph Giuliani, one of President Donald Trump's personal lawyers, shared the video on Twitter, asking, "What is wrong with Nancy Pelosi?" (Giuliani later deleted the tweet.)

At roughly the same time, Trump himself tweeted a separate video of Pelosi—one edited by Fox Business Network. Fox interwove moments from a press conference to make it look like the Speaker continually garbled her words. "PELOSI STAMMERS THROUGH NEWS CONFERENCE," the president declared.[1]

The Pelosi episode foreshadowed one type of disinformation that is likely to disrupt the 2020 election: deliberately distorted video, amplified via social media. The most daunting threat comes from deepfakes, which are synthetic videos created with artificial intelligence (AI). Subjects of deepfakes appear to say or do things they never said or did. The crudely distorted Pelosi videos, made with long-available technology, are more appropriately described as "cheapfakes." They demonstrate that one doesn't need cutting-edge computer artistry to draw mass attention to a misleading portrayal of a prominent foe. The "drunken"-speech video alone received more than three million views in a matter of days. Deepfakes that are more difficult to refute could enjoy even wider circulation.

The Pelosi videos suggest another theme of this report: While Russian operatives and other foreign-based actors are all but certain to surface (or resurface) in 2020, a greater volume of disinformation probably will come from domestic U.S. sources. Some of these domestic sources will be obscure websites or social media accounts, such as Politics WatchDog; others will be better known. If he stays true to form, President Trump will use his Twitter account (62 million followers and counting) to issue statements at odds with the truth. As of early August, he had made more than 12,000 false or misleading claims during his presidency, many of them via Twitter, according to *The Washington Post's* Fact Checker.[2]

Deepfakes and domestic disinformation are two of our eight predictions for 2020. Here's the full list, in capsule form:

- **Deepfakes** threaten to cloud reality at a time when the existence of objective facts increasingly has been called into question.
- **Domestic disinformation** will come mostly from the political right, but the left will also indulge in social media truth-bending.
- **Unwitting Americans:** The Internet Research Agency or other Russian organizations may try to recruit actual Americans for real-life activities.
- **Instagram,** the Facebook-owned image and video service, may turn out to be more of a disinformation magnet than Facebook, Twitter, or YouTube.
- **WhatsApp,** Facebook's messaging service, could become a troubling source of false content in the U.S., as it has been during elections in Brazil and India.
- **Iran and China** are potential fonts of foreign interference, even as Russia continues to try to rile up the American electorate.
- **For-profit firms** based in the U.S. and abroad will offer their election-manipulation services.
- **Digital voter suppression,** a highly direct means of affecting election outcomes, will once again pose problems on social media.

It's impossible to say with certainty how these stratagems might interact or whether others will come into play as well. What seems safe to expect is that Russia's well-documented meddling in 2016—a multifaceted intrusion that included disguised social media personas, armies of automated bot accounts, and distribution of hacked Democratic emails—will not simply be replicated next year. As disinformation analyst Clint Watts has put it: "Last cycle's trickery won't work the same in 2020."[3]

Confusing Policies, Ad Hoc Enforcement

Beyond prognostication, this report examines how social media companies have reacted to politically oriented disinformation. In the case of the Pelosi "drunken"-speech video, there was a notable divergence among the platforms. YouTube promptly removed copies of the altered video. Facebook pointedly did not. Instead, Facebook labeled the video false and reduced its distribution to users. Twitter also declined to pull the video. It's difficult to generalize from this outcome except to say that the platforms' policies are confusing and often seem to be enforced in an ad hoc fashion.

We maintain as a general principle that when provably false content, like the Pelosi video, comes to the attention of social media platforms, they ought to take it down, rather than merely annotating or demoting it. The highest priority should be removing provably false content that affects politics or democratic institutions. Guarding against this abundance of disinformation is a tall order. Five hundred hours of video are uploaded to YouTube every minute.[4] Given this kind of volume and the velocity with which online content circulates, it would be unreasonable to assume that all disinformation can be removed. But that doesn't mean the companies should do nothing more than what they're already doing. They are best positioned to monitor their sites with algorithmic and human scrutiny, and we believe it's their responsibility to mitigate the damage that disinformation does to our public life.

To preserve a record of provably false content, the platforms each should maintain a searchable but cordoned-off archive of disinformation that they've removed. That way, scholars, journalists, and users generally would be able to view the false content for research or other purposes without its surfacing in ordinary search results, recommendations, or users' feeds. The archive would create a record of digital untruth without contributing to its amplification. And the archive's contents would provide the basis for appeals by users who believe their post or tweet was removed inappropriately.

Focus on the Companies

This is the fourth report by the NYU Stern Center for Business and Human Rights on online disinformation. Our most recent publication was "Tackling Domestic Disinformation: What the Social Media Companies Need to Do" (March 2019).[5] Before that, we published "Combating Russian Disinformation: The Case for Stepping Up the Fight Online" (July 2018).[6] We return to the topic because of the pernicious effects disinformation can have on elections. In many cases, it's designed to erode democratic values, heighten cynicism, and exacerbate political polarization.

We continue to focus on what the largest and most influential social media companies ought to do about disinformation. That's because the most obvious alternative—government regulation of online content—would raise immediate free speech concerns about official overreach and censorship. "Governments have only sledgehammers in their tool kits," Joan Donovan, director of the Technology and Social Change Research Project at the Harvard Kennedy School, said in an interview.

The First Amendment to the U.S. Constitution, however, doesn't restrict corporations, which means that social media companies may—and routinely do—moderate content on their platforms. In the absence of government regulation, it is incumbent on the companies to exercise more vigorous self-governance. That means taking a tougher line on disinformation. It is in the companies' enlightened self-interest to do so. Governments in Germany, France, Australia, and other countries lacking an equivalent to the First Amendment are enacting laws to police "hate speech" and "fake news." To reduce the probability of governmental content regulation in the U.S., the social media industry should show that it can close the governance gap when it comes to disinformation.

2. What to Watch for in 2020

> If the U.S. military has the ability to take the Russian Internet Research Agency offline, at least for a day or two, does that mean the 2020 elections are secure? In a word, no.

In 2016, Russia's Internet Research Agency (IRA) engaged in what former Special Counsel Robert S. Mueller III called "multiple systemic efforts to interfere" with the U.S. presidential election.[7] The IRA has remained active since then, but when the November 2018 midterms rolled around, the amount of Russian meddling was negligible.[8] It's possible that the IRA decided to keep its powder dry for the 2020 presidential race. It's also possible that more aggressive removal of suspicious accounts by the social media companies hindered the Russians.

And there could have been another factor: U.S. Cyber Command, an arm of the Pentagon, reportedly used its offensive hacking capabilities to temporarily block the IRA from using the internet in early November 2018.[9]

So, does this mean the 2020 elections are secure? In a word, no. "The IRA is a small component of the overall Russian operation," which also includes Moscow's military intelligence service and possibly others, said Lee Foster, who leads the disinformation team at FireEye, a cybersecurity firm whose clients include some of the major social media companies. "All of these actors rework their approaches and tactics," he added.[10] As a result, said Kevin Carroll, a former senior counselor at the Department of Homeland Security, the relative calm during election season 2018 doesn't guarantee a repeat performance in 2020.[11]

President Trump refrains from acknowledging this risk, reportedly because he equates attention to Russian election meddling with skepticism about the legitimacy of his 2016 victory.[12] But the U.S. intelligence community isn't as reticent. In its 2019 Worldwide Threat Assessment, the Office of the Director of National Intelligence predicted that next year, Russia and other American adversaries "almost certainly will use online influence operations to try to weaken democratic institutions, undermine U.S. alliances and partnerships, and shape policy outcomes."[13]

Deepfakes

The Worldwide Threat Assessment specifically expressed concern about doctored videos. In 2020, it said, U.S. adversaries "probably will attempt to use deepfake or similar machine-learning technologies to create convincing—but false—image, audio, and video files to augment influence campaigns directed against the United States and our allies and partners."

The term deepfakes comes from a combination of "deep learning" and "fakes." It originated in 2017 with an anonymous Reddit user who called himself "deepfakes." This individual gained attention by using deep-learning algorithms to superimpose faces of celebrities onto the bodies of pornographic actors. Under outside pressure, Reddit banned the individual and porn deepfakes generally, but imitators proliferated elsewhere on the internet.[14]

> ‘Deepfakes can be made by anyone with a computer, internet access, and interest in influencing an election.’ — John Villasenor, Professor of Electrical Engineering, Public Policy, and Management at the University of California, Los Angeles

The movie industry has altered video footage for decades. But Hollywood’s tricks require skilled technicians, a lot of money, and ample time. Open-source AI democratized video fabrication. “Deepfakes can be made by anyone with a computer, internet access, and interest in influencing an election,” according to John Villasenor, a Professor of Electrical Engineering, Public Policy, and Management at the University of California, Los Angeles.[15]

A brief digression for some technical background: The broad term artificial intelligence (AI) refers to algorithms able to perform humanlike tasks, such as recognizing a human face. Deep learning, a subset of AI, refers to arrangements of algorithms that can learn and make intelligent decisions on their own. Deepfakes are falsified videos made by means of deep learning.

A deep-learning system can produce a persuasive counterfeit by studying photographs and videos of a target person, as well as video of an actor speaking and behaving the way the target will be depicted in the new, bogus video. The technology can, in essence, merge the target and the actor into one. Once a preliminary fake has been produced, a method known as GANs, or generative adversarial networks, makes it more believable. The GANs process seeks to detect flaws in the forgery, leading to improvements addressing the flaws. After multiple rounds of detection and improvement, the deepfake is completed.

On June 7, 2019, two British video artists posted a deepfake of Mark Zuckerberg, Facebook’s founder and CEO, on Instagram. The video had been concocted by an Israeli tech startup called Canny AI. The firm used as its target a video Zuckerberg released publicly in 2017. Canny AI said in an email exchange that it altered the facial movements of the original video to match those of an actor whose voice was substituted for Zuckerberg’s. The result is visually convincing, but the actor’s voice isn’t a close match for the tech executive’s. Moreover, the artists signaled their intention to satirize Zuckerberg by having him talk like a James Bond super-villain: “Imagine this for a second: one man with total control of billions of people’s stolen data, all their secrets, their lives, their futures.”[16]

It’s surprising that more serious deepfakes haven’t already surfaced in American politics. Experts testifying at a House Intelligence Committee hearing in June 2019 agreed that it’s just a matter of time. “Imagine that the night before the 2020 election, a deepfake showed a candidate in a tight race doing something shocking he never did,” suggested Danielle Keats Citron, a law professor at Boston University. “The damage would be irreparable. Elections cannot be undone.”[17] (An alleged deepfake of the president of Gabon, Ali Bongo, delivering a 2019 New Year’s address helped precipitate an unsuccessful coup in the African country.[18])

Citron and two co-authors, writing for the *Lawfare* website, noted that political defamation in the U.S. goes back to Alexander Hamilton’s rivalry with Thomas Jefferson. “What is different today,” they said, “is that the falsehoods involve visual and audio ‘evidence’ that our eyes and ears are deeply inclined to trust (not just written words that might more readily be dismissed).” And courtesy of social media, “the frauds can rapidly reach countless individuals.”[19]

Some scholars of disinformation downplay the potential significance of deepfakes, arguing that less-polished frauds—like the Pelosi videos or clumsily Photoshopped still images—could be enough to unsettle an election. “Deepfakes aren’t necessary. A lot of people don’t need ‘real’ evidence,” said Patrick Warren, an economist at

Clemson University who studies social media. In a politically polarized nation, he added, "it's about ideology and finding things that confirm what [groups of voters] already believe."[20]

Warren may have a point when it comes to some committed partisans. But his observation doesn't negate the more consequential dangers that could result from a series of incriminating and seemingly genuine videos introduced in the run-up to Election Day.

Unwitting Americans

Unlike cutting-edge, AI-driven deepfakes, some disinformation methods are about the old-fashioned goal of getting people onto the streets. "One thing to look for is more effort to recruit real Americans to unwittingly organize political rallies or create new groups online," Laura Rosenberger, director of the Alliance for Securing Democracy, said in an interview.

The March 2019 Mueller report fleshed out our understanding of how IRA operatives, posing as grassroots U.S. activists, mobilized Americans to participate in dozens of rallies in 2016 and thereafter. By expanding on this strategy in 2020, the Russians would accomplish one of their main goals—translating influence online into real-world discord.

In 2016, the technique entailed phony IRA social media personas or groups first announcing events, such as dueling anti-Muslim and pro-Muslim demonstrations that the Russians successfully organized in May 2016 in Houston. The IRA would send direct messages promoting events to their online American followers. From those who indicated interest in attending, the IRA would seek local people to coordinate events. The IRA operative would make an excuse for not attending but would seek out conventional media coverage. Attendance at events varied from a handful of participants to hundreds.[21]

An incident from the summer of 2018 illustrated how this stratagem can unfold. A counterfeit left-wing Facebook account called Resisters, which Facebook labeled as probably Russian, created an August 10 event to counter a planned white supremacist gathering in Washington, D.C. Resisters, the first Facebook page to promote the counter-protest, coordinated with five other apparently real pages dedicated to opposing the white supremacists, according to Facebook. "These legitimate pages unwittingly helped build interest in [the counter-demonstration] and posted information about transportation, materials, and locations so people could get to the protests," the company added. Before the demonstrations, however, Facebook determined that Resisters was a fraud and removed it. The company also notified 2,600 users who had indicated interest to Resisters about attending.[22]

In the end, only 40 white supremacists showed up for the rally near the White House, and they were greatly outnumbered by counter-protesters.[23] Even though it was unmasked, Resisters accomplished its presumed goal of amplifying division within American society.

Instagram

Instagram hasn't received as much attention in the disinformation context as Facebook, Twitter, and YouTube, but it played a much bigger role in Russia's 2016 election manipulation than most people realize. And it could become a crucial Russian instrument next year, according to a report commissioned by the Senate Intelligence Committee.[24]

Started in 2010, Instagram was acquired by Facebook 18 months later for $1 billion. Today, Instagram has 1 billion users (compared to 2.38 billion for Facebook, 2 billion for YouTube, and 330 million for Twitter). The photo- and video-posting platform is a common destination for younger users bored by Facebook.[25]

> **"That Instagram has outperformed Facebook as a Russian engagement machine may 'indicate its strength as a tool in image-centric memetic warfare,' according to a report commissioned by the Senate Intelligence Committee."**

Instagram's image-oriented service makes it an ideal venue for memes, which are photos combined with short, punchy text. Memes, in turn, are a popular vehicle for fake quotes and other disinformation.

In 2016, the IRA enjoyed more user engagement on Instagram than it did on any other social media platform. That was a finding of the Intelligence Committee report released in December 2018. "Instagram was a significant front in the IRA's influence operation, something Facebook executives appear to have avoided mentioning in Congressional testimony," the report said. "Our assessment is that Instagram is likely to be a key battleground on an ongoing basis." The report was written for the committee by experts at the Tow Center for Digital Journalism at Columbia University and two research firms, New Knowledge and Canfield Research.

Analyzing data from 2015 through 2018, the researchers found there were 187 million user engagements with IRA material on Instagram—more than twice as many as on Facebook (77 million) or Twitter (73 million). One meme posted on Instagram by the IRA's "Blacktivist" account showed a police officer half-clad in a Ku Klux Klan hood-and-sheet above the statement, "The KKK has infiltrated police departments for years."[26]

Many Americans say disinformation is a critical problem that needs to be fixed.

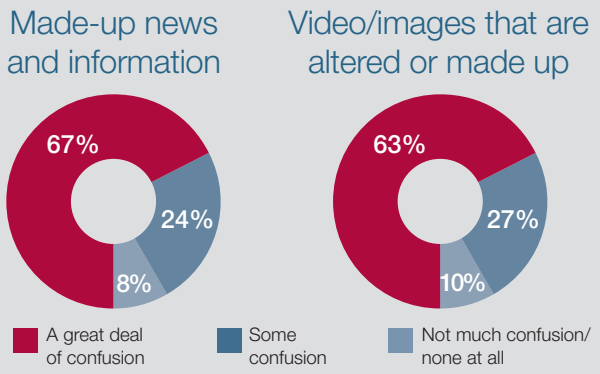
Percentage of Americans who say the issue in question is a very big problem in the country today

70%	Drug addiction
67%	Affordability of healthcare
52%	U.S. political system
51%	Gap between rich and poor
50%	**Made-up news/info**
49%	Violent crime
46%	Climate change
40%	Racism
38%	Illegal immigration
34%	Terrorism
26%	Sexism

Percentage who say made-up news and information have a very big impact on...

Americans' confidence in government	68%
Americans' confidence in each other	54%
Political leaders' ability to get work done	51%

Two-thirds of U.S. adults think made-up information and altered videos create a great deal of confusion about current issues and events.



Note: Those who didn't answer not shown. Source: Pew Research Center survey of 6,127 U.S. adults, conducted between February 19 and March 4, 2019

Percentage of Americans who have at least "some" trust in social media platform

42%	YouTube
29%	Facebook
23%	Instagram
22%	Twitter
16%	Snapchat

Percentage who believe the platform in question is at least "somewhat" responsible for spreading disinformation

64%	Facebook
55%	Twitter
48%	YouTube
46%	Instagram
39%	Snapchat

Source: Institute for Public Relations survey of 2,200 Americans between March 19 and March 24, 2019

That Instagram outperformed Facebook as an IRA engagement machine may "indicate its strength as a tool in image-centric memetic warfare," the committee report concluded. It's also possible, the researchers acknowledged, that "click farms" created some of the IRA's Instagram engagement. But that didn't erase the researchers' prediction that Instagram would play a major role in 2020.

Other observers have noted that, beyond Russian interference, domestically generated hoaxes and conspiracy theories increasingly are thriving on Instagram. One example is QAnon, a widely circulated right-wing conspiracy theory about a "deep state" plot to take down President Trump. In March 2019, *The Atlantic* predicted that Instagram "is likely where the next great battle against misinformation will be fought, and yet it has largely escaped scrutiny."[27] With the approach of the next election, scrutiny will be needed.

WhatsApp

Researchers at the Atlantic Council's Digital Forensic Research Lab suggested in an interview that another Facebook property, WhatsApp, deserves a place on our watch list. The case for worrying about WhatsApp rests on two observations: First, the messaging service is encrypted "end-to-end," meaning that even its corporate overseers can't see what users are up to. That's great for privacy but also a recipe for mischief or worse.

Second, WhatsApp has served as a powerful vehicle for disseminating false information during recent presidential elections in both Brazil and India. People in those countries used a WhatsApp forwarding feature to spread wildly misleading content—much of it generated by warring campaigns—to large groups of users. Content on WhatsApp may have more credibility in recipients' eyes because it often comes from senders with whom recipients are personally familiar. In Brazil, far-right presidential candidate Jair Bolsonaro triumphed in October 2018, "helped in part by a wave of toxic rumors and misinformation, much of it spread over WhatsApp," the *Financial Times* reported. Seven months later, Narendra Modi won reelection as India's prime minister in an election similarly marred by falsehoods and in which WhatsApp became "the central battleground."[28]

There are reasons to be skeptical, however, that WhatsApp will be used widely for stirring trouble in the U.S. Acquired by Facebook in 2014 for $22 billion, the messaging platform has gained tremendous popularity in many parts of the world. It has 1.5 billion users globally, 300 million of them in India and another 120 million in Brazil. But it hasn't caught on to a similar degree in the U.S., where it has fewer than 60 million adult users. Its smaller U.S. user base limits its potential to spread disinformation. In addition, Americans who use WhatsApp primarily do so for one-to-one or one-to-a-few communication, not one-to-many, as occurred during the Brazilian and Indian elections. With one-to-one messaging, disinformation (or any kind of content) is much less likely to go viral. Finally, WhatsApp has limited the number of groups to which a user can forward a piece of content at one time—a move designed to slow the distribution of untruths.

Still, the Digital Forensic Research Lab cautions that WhatsApp should not be ignored. "It's been a proven vector [for disinformation] in other countries," Iain Robertson, the lab's deputy managing editor, said. "There's room for that to expand here."[29]

Iran and China

Inevitably, other foreign rivals of the U.S. have imitated Russia's disinformation operation by launching their own. The most likely to surface during the 2020 election is Iran, especially if hostility between Tehran and Washington remains acute.

> **"'Iranians are now following the Kremlin's playbook,' Senator Mark Warner (D., Va.) said at the time of an August 2018 removal of 652 fake Facebook accounts, pages, and groups of Iranian and Russian origin."**

Iran has been one of the most consistent digital foes of the U.S. for the past decade, best known for cyberattacks on banks, hospitals, and government agencies. For the last several years, it has also been test-driving its U.S.-directed disinformation apparatus. In May 2019, FireEye, the cybersecurity firm, released a report identifying an Iran-based social-media network that used fake American personas espousing both conservative and liberal political views. Some of the phony accounts impersonated real Republican political candidates who ran for office in 2018. The Iranian network promoted anti-Saudi, anti-Israeli, and pro-Palestinian themes. It also expressed support for the 2015 nuclear deal, from which President Trump unilaterally withdrew. Some of the phony personas—using names such as "John Turner" and "Ed Sullivan"—had letters published on these themes in the *New York Daily News, Los Angeles Times,* and other mainstream media outlets.[30]

In response to FireEye's tip, Facebook removed nearly 100 accounts, pages,and groups from Facebook and Instagram.[31] Acting independently, Twitter announced that it had removed more than 2,800 "inauthentic accounts originating in Iran." [32] These actions followed two other Facebook takedowns of Iranian accounts in 2018. "Iranians are now following the

> One of the domestically generated conspiracy theories likely to gain traction in coming months is that the major social media companies are colluding with Democrats to defeat President Trump's bid for reelection.

Kremlin's playbook," Senator Mark Warner (D., Va.) said at the time of an August 2018 removal of 652 fake Facebook accounts, pages, and groups of Iranian and Russian origin.[33]

Iran and Russia aren't the only U.S. adversaries or strategic competitors that are candidates to deploy disinformation in 2020. Enmeshed in a bitter trade war with the U.S., China also could pose a digital threat.

Higher U.S. tariffs or other punitive measures aimed at the Chinese economy could provide sufficient motive for Beijing to unleash "an army of operators" to try to meddle online in 2020, according to Dipayan Ghosh, a former Facebook policy advisor who is co-director of the Platform Accountability Project at the Harvard Kennedy School.[34] Kevin Carroll, the former senior counselor at Homeland Security, agreed, saying, "If Sino-U.S. relations further deteriorate, the Chinese could impact U.S. social media simply through brute numbers, with or without the encouragement of the Chinese security services, as so many bilingual and fervently nationalistic Chinese are active online."[35]

One major difference between potential Russian and Chinese digital campaigns: The Russians have supported Trump, Ghosh noted, but "the Chinese would probably work to get him out of office because they don't like his policies or his rhetoric."

Domestic Disinformation

While foreign election interference has dominated discussion of contemporary disinformation, most purposely false content in the U.S. is generated by domestic sources. That's the consensus of social media executives, cybersecurity sleuths, and academic researchers.[36] Referring to the more plentiful home-grown variety, Justin Hendrix, executive director of NYC Media Lab, said in an interview, "I think we're going to see a ton of it" in 2020.

Matt Masterson, a senior cybersecurity adviser at the Department of Homeland Security, noted during a recent conference hosted by the Hewlett Foundation that for the federal government, domestically generated disinformation is "the hardest challenge that we have."[37] Foreign interference in a U.S. election is plainly illegal. The U.S. has indicted about two dozen operatives with the Russian IRA and military intelligence service, although it's highly unlikely Moscow will extradite any of the defendants. By contrast, misleading content created by U.S. citizens can be difficult to distinguish from ordinary political expression, which enjoys First Amendment protection from government regulation or prosecution.

Domestic disinformation emanates from both the political left and right, but as discussed in our March 2019 report, studies show that conservative social media users are more likely to circulate false content.[38] One of the domestically generated conspiracy theories likely to gain traction on the right in coming months is that the major social media companies are conspiring with Democrats to defeat President Trump's bid for reelection. *Breitbart News,* a leading far-right outlet (3.87 million followers on Facebook), has promoted the contention that "big tech companies are trying to make sure conservative voices are suspended ahead of the 2020 election cycle."[39] Over the past year, Facebook, Instagram, Twitter, and YouTube have banned a series of right-wing figures, including Alex Jones, proprietor of the *Infowars* conspiracy website. Explaining these ousters, the social media sites point to their policies against promoting hatred or violence and have denied any vendetta against conservatives.[40]

But President Trump doesn't buy the denial. In a series of tweets and interviews, he has lashed out at big tech. "I can tell you they discriminate against me," he told *CNBC*. "You know, people talk about collusion. The real collusion is between the Democrats and these companies."[41]

Introducing a right-wing "social media summit" at the White House in July 2019, Trump tweeted about "the tremendous dishonesty, bias, discrimination, and suppression practiced by certain companies." Then he threatened: "We will not let them get away with it much longer."[42] The president told attendees that he was directing his aides to explore "all regulatory and legislative solutions" for the alleged plot by Facebook, Google, and Twitter.[43]

The White House gathering seemed intended to rev up conservative disinformation purveyors who, even before Trump gave them his public stamp of approval, had begun to go after would-be Democratic presidential nominees. Some have raised spurious questions about whether Senator Kamala Harris of California is really an "American black."[44] Others have orchestrated false sexual assault allegations against South Bend, Ind., Mayor Pete Buttigieg.[45]

There has also been domestic disinformation from the left. A January 11, 2019, interview on the public radio station WNYC illuminated a liberal social media-distortion scheme from Alabama. "This was a false flag operation," Democratic activist Matt Osborne said. "The idea for doing this came from studying what the Russians had done."[46]

Osborne helped set up a fake "Dry Alabama" Facebook page and companion Twitter account which appeared to have been created by Baptist teetotalers supporting Republican Roy Moore in a special U.S. Senate race in December 2017. The goal was to associate Moore with a statewide alcohol ban and thereby disillusion moderate, business-oriented Republicans. The scheme was carried out on behalf of Democrat Doug Jones, who ultimately won the election but by all accounts didn't know about the digital skullduggery.[47] Democrats reportedly executed another false-flag maneuver employing more than a dozen misleading Facebook pages designed to appeal to conservatives nationally in the run-up to 2018 midterms.[48]

Contending that Republicans are using similar deceptive tactics, Osborne said Democrats have to fight fire with fire. In 2020, he predicted, "you're going to see a movement toward dark money spending on digital campaigns in the closing days of the race."[49] For example, he added in an email exchange: "Dissuasion can be as simple as showing, say, a Republican voter an image of a red wave with a triumphal statement that imbues them with a sense of inevitable victory: 'No need to bother voting. Trump has got it in the bag.'"[50]

For-Profit Firms

Another person who participated in the Alabama disinformation episode was Jonathon Morgan, CEO of New Knowledge, an Austin, Texas-based social media research firm. His role points to an additional danger in 2020: the involvement of private companies or their executives in crafting disinformation.

Once the mainstream media brought the Alabama case to light, Morgan said he'd been acting in his own capacity as a researcher seeking to understand how digital disinformation works, not to affect the outcome of the Jones-Moore Senate race. Facebook nevertheless suspended Morgan's account and four others run by individuals involved in the episode. The activity in Alabama was funded by Democratic political donors, including Reid Hoffman, the co-founder of LinkedIn, who publicly apologized and said he hadn't known about the underhanded tactics.[51]

Contacted for comment, Morgan said that "it wasn't our intention to create a misleading page." His firm, New Knowledge, was mentioned earlier in this report because it helped write an assessment of Russian interference for the Senate Intelligence Committee.

In Russia, disinformation services are sold without excuse or apology. Jigsaw, a corporate think tank affiliated with Google, decided to test the Russian marketplace by retaining a firm called

> **"'This was a false flag operation. The idea for doing this came from studying what the Russians had done.' — Democratic activist Matt Osborne, describing a liberal social media-distortion plot from Alabama"**

SEO Tweet, which it discovered advertising online. The price for a modest two-week disinformation campaign aimed at a dummy website Jigsaw made especially for the purpose was a mere $250. The transaction, carried out in the spring of 2018, was an experiment designed to see how influence operations could be bought in Russia, Jigsaw later told *Wired*.[52] SEO Tweet didn't respond to a request for comment sent via Twitter.

Israel, home to a robust tech industry, could also be a source of for-profit disinformation. The satiric deepfake of Mark Zuckerberg was engineered by an Israeli start-up. Separately, a February 2019 article in *The New Yorker* described an Israeli intelligence firm called Psy-Group that reportedly specialized in "covertly spreading messages to influence what people believed and how they behaved." Among the firm's alleged activities were two digital campaigns in the U.S.: one that undermined American college students who advocated boycotting Israel and a second that tried to tilt a California hospital-board election. In 2016, Psy-Group unsuccessfully pitched its services to the Trump campaign. The company shut down in 2018, and people formerly associated with it didn't respond to requests for comment.[53]

Facebook announced in May 2019 that it was taking action against yet another Israeli outfit, a consulting and lobbying firm called Archimedes Group, whose employees allegedly ran political disinformation campaigns in Africa and around the globe. The operatives misrepresented themselves as locals, including local news organizations, and published allegedly leaked information about African politicians, Facebook said in connection with its ban of Archimedes and deletion of 265 accounts, pages, groups, and events. The deceptive activity allegedly focused on Angola, Niger, Nigeria, Senegal, Togo, and Tunisia and also extended to Latin America and Southeast Asia. Archimedes allegedly spent $812,000 on Facebook advertising paid for in U.S. dollars, Israeli shekels, and Brazilian reals. Facebook apparently didn't figure out whether Archimedes had an overarching agenda, and there is no evidence that the activity was linked to the Israeli government. "We don't want our services to be used to manipulate people," Facebook's head of cybersecurity, Nathaniel Gleicher, wrote in a company blog post.[54]

Efforts to reach Archimedes and individual executives with the firm were unsuccessful. Before its website disappeared in the wake of Facebook's action, the company reportedly described itself online as taking "every advantage available in order to change reality according to our client's wishes."[55]

Digital Voter Suppression

A final category of disinformation almost certain to surface in 2020 is voter suppression. There's nothing new about steering selected voters away from the polls, but the pervasiveness of social media makes this practice particularly pernicious when it goes online.

Recent examples of attempted voter suppression on social media are profuse. As part of their pro-Trump campaign in 2016, Russian IRA operatives tried to persuade African Americans that there was no point in voting because Democrats didn't care about them.[56] Facebook has said that in the weeks before the 2018 midterm election, it found and removed 45,000 pieces of voter suppression content.[57]

Researchers at the University of Wisconsin found several categories of social media suppression in the fall of 2018. Some posts on Twitter tried to deceive opponents of President Trump by providing incorrect information about the day to vote or announcing that "2018 is the first year citizens can vote by TEXT." Other tweets tried to intimidate liberal voters by telling National Rifle Association members and Republicans to bring their guns to the polls. Yet other social media posts attempted to deter African Americans in 2018 by arguing that voting is pointless or that there's no difference between the two major parties. Many of these messages used hashtags like #dontvote and #dontbeavoter.[58]

Some voter suppression efforts invoked more idiosyncratic arguments. Twitter has said it disabled thousands of anonymous voter-suppression bots in 2018, including one batch of more than 10,000 disguised to look like they were from Democrats. Some of the tweets tried to persuade male Democrats to skip the midterms, because otherwise men would drown out women's voices.[59]

Few behaviors strike as directly at the heart of democracy as confusing or bullying people who are entitled to vote. The social media companies will have to remain vigilant on this front.

"Few behaviors strike as directly at the heart of democracy as confusing or bullying people who are entitled to vote. The social media companies will have to remain vigilant on this front."

3. How Social Media Companies Have Responded to Disinformation

> ‘The companies are getting much better at detection and removal of fake accounts.’ — Dipayan Ghosh, co-director of the Harvard Kennedy School’s Platform Accountability Project

In a sense, the major social media companies have been preparing for the 2020 election since 2017, when they began to acknowledge the Russian disinformation problem. Since then, these companies have taken a wide range of general steps to clean up their sites and harden their defenses. They also have put in place a number of measures specifically aimed at protecting election integrity. Both categories—general and election-specific—should have a bearing on 2020.

To their credit, the companies are doing more—more communicating with each other, the government, and outside experts; more deleting of fraudulent accounts; more mobilizing of special teams focused on election irregularities. But there’s still much more to do.

General Changes

Removing Sham Accounts

“The companies are getting much better at detection and removal of fake accounts,” Dipayan Ghosh, co-director of Harvard’s Platform Accountability Project, said in an interview. Facebook uses improved AI to delete automated fake accounts by the billions—2.19 billion in just the first three months of 2019. Most of these accounts were blocked within minutes of their creation, preventing them from doing any harm, according to the company.[60]

Over the past two-and-a-half years, Facebook also has announced dozens of smaller, more-targeted takedowns comprising many thousands of accounts and pages which have demonstrated “coordinated inauthentic behavior.” [61] In Facebook’s opinion, the operators of these accounts and pages have worked together to deceive users about who runs them and what they’re doing. The company stresses that it punishes misleading behavior, not content. But the behavior in question often includes disseminating disinformation. For its part, Twitter has challenged and taken down millions of fake accounts for similar reasons. And YouTube has done so as well, if to a lesser degree.

Compared to the companies’ relative passivity in 2016, these actions demonstrate greater vigor. In addition to honing their AI-screening algorithms, the platforms have hired thousands of additional staff reviewers and outside contractors to hunt for accounts spewing problematic content. But the impressive numbers of fake or deceptive accounts eliminated also indicate the vast supply of such accounts and the near certainty that the companies aren’t catching them all—or perhaps even most of them. “Very many fake accounts are going undetected and can be used for manipulation,” according to Filippo Menczer, a professor of informatics and computer science at Indiana University who studies disinformation. These bot accounts, he added in an email exchange, “can be used to amplify the spread of misinformation, deepfakes, attacks, fear-mongering, voter suppression, or to fake support for candidates.”[62]

Demoting and Labeling False Content

Intertwined with the problem of fake accounts and inauthentic behavior is the widespread existence of false content that spreads rapidly on social media platforms. What should happen to disinformation once legitimate accounts and pages begin to share it? To deal with falsehoods that can distort the country's political discourse, the platforms have invested in more precise AI, expanded in-house reviewing, and added more third-party fact-checking capacity. Facebook now farms out potentially false stories and images to 53 fact-checking organizations fluent in 42 languages around the world. In particular, the social network has emphasized that it is "using both technology and people to fight the rise in photo and video-based misinformation"—a commitment that reflects growing concern about deepfakes.[63]

Facebook's Mark Zuckerberg recently said his company is considering banning and removing all deepfakes, or at least the ones it can identify.[64] But generally, the platforms don't remove content simply because it's been deemed to be false. Instead, they typically reduce the prominence of the untrue material, based on the notion that their executives and employees shouldn't be "arbiters of the truth."[65] Facebook, for example, "downranks" falsehoods in users' News Feeds by 80%, labels the content as untrue, and provides context on the topic. This is what Facebook did with the doctored video making Nancy Pelosi seem drunk.[66]

The companies don't apply these content-demotion procedures in a consistent way, however. Facebook has said that in certain contexts, it does remove false content. Examples include false information that creates a risk of physical violence or aims to keep voters from the polls or seeks to interfere with the U.S. census.[67] Twitter and YouTube have said that they, too, will eliminate voter-suppression content.[68]

In June 2019, YouTube announced that it would begin deleting some content based on its deceptiveness. In other words, in certain circumstances, it would act as an arbiter of the truth. Specifically, YouTube said it would take down videos "denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place."[69] Videos that perpetuate other falsehoods, such as the contention that the Apollo moon landings were staged or that a phony miracle cure will help patients with a serious illness, are no longer recommended to YouTube users, but they remain available on the platform.[70]

No obvious principle underpins these distinctions.

> **"Acknowledging that it needs to find additional ways to identify disinformation, Facebook has launched an initiative to crowdsource fact-checking."**

As mentioned earlier, the NYU Stern Center believes that when social media platforms discover provably false content, especially in the political realm, they should not merely reduce its visibility or have a recommendation algorithm cease to tee it up for users. The platform should remove the material altogether, keeping only a well-marked reference copy in an archive of deleted false content. The material in the archive should be publicly accessible to researchers and users generally but should not circulate via feeds or recommendations. Twitter has taken a step in this direction by releasing substantial datasets of tweets, which it has removed after linking the content to disinformation operations backed by Russia or other countries.[71]

Crowdsourcing Fact-Checks

Acknowledging that it needs to find additional ways to identify disinformation, Facebook has launched an initiative to crowdsource fact-checking. With 1 billion total posts a day, the volume of false content on the social network far exceeds what in-house reviewers and outside fact-checkers can handle. To deal with this challenge, Facebook says it is looking into how to harness "community-driven approaches to misinformation." By this, the company means "relying on groups of people who use Facebook to point to journalistic sources that can corroborate or contradict the claims made in potentially false content."[72]

CEO Mark Zuckerberg has spoken approvingly about this idea, but significant questions remain. One is whether Facebook users would handle the task responsibly. "You can't just have Joe Schmo, who thinks *The New York Times* is a liberal rag, [challenging the newspaper] just because Trump says it's the 'enemy of the people,'" said Brooke Binkowski, the former managing editor of *Snopes*, a professional fact-checking site that previously partnered with Facebook.[73] While such hesitations demand attention, it's still possible that Facebook could devise an algorithmic method to filter crowdsourced alerts that would diminish election disinformation—and do it in time for November 2020 voting.

Election-Related Changes

War Rooms and Special Teams

Asked about election defenses in an April 2019 interview, Zuckerberg said: "I'm confident in where we are right now." Facebook, he added, has "probably some of the most-advanced systems of any company or government in the world for preventing the kind of tactics that Russia and now other countries, as well, have tried."[74]

He's correct, as we've seen, that the social media platforms have made progress, but there's a danger that his confidence will lead to complacency.

Facebook, Twitter, and YouTube all have established dedicated teams to prevent manipulation of elections. Twitter's "cross-functional analytical team" uses specially designed software to identify "anomalous and potentially malicious automated and human-coordinated activity"—for example, large numbers of bots tweeting simultaneously about the same topic.[75] Google's "threat analysis group" serves a similar function on behalf of the company's YouTube subsidiary.[76] Facebook has a 500-member team devoted to protecting election integrity.[77]

Facebook has taken to setting up "war rooms" in the weeks preceding elections, including those in India and the European Union in 2019. These arrangements temporarily bring together several dozen staff members from different departments of the company who focus intensively on evidence of suspicious behavior or content. The Facebook war room gets input from the election-integrity team, which, in turn, can draw on some 15,000 Facebook employees and outside contractors who generally review content.[78]

War rooms are susceptible to being lampooned as mere backdrops for photo ops. But the symbolism can be constructive if it signals, both internally and externally, that the company is serious about protecting elections.

Government Relationships

Since 2016, the three major social media companies also have established or improved election-related partnerships with government agencies. Facebook at times has played a coordinating role—for example, hosting an industry-wide meeting with federal officials at its Menlo Park, Calif., headquarters in May 2018. Zuckerberg has spoken publicly about tips Facebook receives from intelligence agencies, which have led to targeted take-downs of suspicious accounts.[79]

Twitter has described relationships with the FBI's Foreign Influence Task Force and the Department of Homeland Security's Election Security Task Force. On Election Day in 2018, Twitter said that it "virtually participated" in a Homeland Security operations center that brought together representatives of the FBI and other relevant agencies.[80] In addition, Twitter has set up a dedicated "portal" on its site for state election agencies to alert the company to emerging problems as quickly as possible.[81]

Political Advertisements

In 2016, Russian IRA operatives bought divisive online political advertisements, often paying in Russian rubles. Since then, the major platforms have increased the transparency of ads related to elections and national issues—chiefly by requiring that they be labeled with "paid for by" notices. Facebook, Twitter and Google/YouTube now also require that ad purchasers confirm their identities and locations within the U.S. All three have created searchable political-advertising databases, which allow anyone to look for patterns in this kind of spending.[82] Twitter has taken the extra step of banning all advertising by RT (formerly Russia Today) and Sputnik, Russian media outlets that have been described by U.S. intelligence authorities as propaganda arms of the Putin regime.[83]

Although salutary in theory, the advertising-transparency measures may not be as potent as the companies claim. First, while the Russians' use of advertising in 2016 was important to build U.S. audiences, the majority of disinformation was not spread by purchased ads. Instead, it was disseminated by unpaid-for "organic" activity—posts, shares, and the like. Second, U.S. news organizations have

"War rooms are susceptible to being lampooned as mere backdrops for photo ops. But the symbolism can be constructive if it signals, both internally and externally, that the company is serious about protecting elections."

scrutinized Facebook's "paid for by" notice requirement and found that it is easily manipulated. As a test, *Vice News* bought fake ads naming all 100 sitting U.S. senators as the buyers. The purchases sailed through the platform's approval process.[84]

A better model for ad transparency is provided in proposed federal legislation known as the Honest Ads Act, which would go further than the policies the companies have voluntarily adopted. First introduced in the fall of 2017, the bill would mandate disclosure for political ads online in a fashion comparable to already-existing requirements for traditional broadcast and print media. The Democrat-controlled House approved the bill in early 2019 as part of an omnibus political-reform package. In May 2019, a bipartisan group of senators, including Amy Klobuchar (D., Minn.), Mark Warner (D., Va.), and Lindsey Graham (R., S.C.), reintroduced the Honest Ads Act in the Senate. But it faces almost certain opposition from Republican Senate Majority Leader Mitch McConnell of Kentucky, who has said that he is not going to bring election-related legislation to the floor for a vote.[85]

Have These Steps Worked?

Enough time has passed since the 2016 election-disinformation debacle to allow scholars to begin to assess whether the steps the platforms have been taking are working. Several studies have compared the situation in the U.S. in 2018 to that in 2016, while other evaluations have considered the run-up to the May 2019 European Union Parliamentary election.

Studies of the United States

The Oxford Internet Institute, an arm of Oxford University, published research in November 2018 showing that disinformation had become more common since 2016. During a 30-day period before the U.S. midterms, 25% of Facebook and Twitter shares related to the election contained "junk news," the Oxford researchers found. That represented an increase of five percentage points from the 2016 election season. The Oxford team defined junk news as deliberately "misleading, deceptive, or incorrect information purporting to be real news about politics, economics, or culture." Most of the junk news they identified came from domestic U.S. sources, not Russia.[86]

But another study, published in February 2019, reached a more sanguine conclusion. Political scientists from four universities—Princeton, the University of Exeter, Washington University at St. Louis, and the University of Michigan—found that Facebook played a "dramatically reduced" role in leading users to "fake news" sites in 2018 than it had in 2016. More broadly, the researchers said, "the proportion of Americans consuming fake news and their total consumption of fake news websites declined by approximately 75% between the 2016 and 2018 campaigns."[87]

The four-universities study was roughly consistent with research by scholars from NYU and Stanford, who found that engagement with "fake news" on Facebook fell by more than 50% from the end of 2016 to mid-2018. Meanwhile, the researchers said, engagement on Twitter rose during the same period. While tentative about the reasons for this outcome, the NYU and Stanford researchers wrote: "The suite of policy and algorithmic changes made by Facebook following the [2016] election seems like a plausible candidate." Still, the scholars added, even after the reduction in Facebook fake news engagement, the absolute level of interaction remained high—roughly 70 million interactions per month.[88]

Studies showing a decrease in disinformation between 2016 and 2018 cannot, of course, predict with any certainty what might happen in 2020, when a much more momentous presidential contest is likely to attract intensified interference by both foreign and domestic U.S. actors.

Studies of the European Union

Just weeks after the E.U. elections in May 2019, the European Commission and the bloc's foreign policy and security arm published a preliminary review that found that Russian-linked organizations, among other actors, had sought to undermine the credibility of the election by means of disinformation on Facebook, Twitter, and YouTube. "The evidence collected revealed continued and sustained disinformation activity by Russian sources aiming to suppress turnout and influence voter preferences," the review said. While praising the platforms' greater vigilance about political ads, the Commission noted that more than 600 Facebook groups and pages operating across France, Germany, Italy, the United Kingdom, Poland, and Spain reportedly spread disinformation and hate speech to millions of users. "More needs to be done by the platforms to effectively tackle disinformation," the Commission said.[89]

In contrast to the European Commission's gloomy pronouncement, research on the E.U. elections by the Oxford Internet Institute produced mixed results. On Facebook, the Oxford researchers found, individual junk news stories can "hugely outperform even the best, most important professionally produced stories, drawing as much as four times the volume of shares, likes, and comments." But looking generally at sources shared on Twitter, the researchers described more heartening findings: "Less than 4% of sources circulating on Twitter during our data collection period were junk news or known Russian sources, with users sharing far more links to mainstream news outlets overall (34%)."[90] Although the performance of particular viral junk news stories remains a significant problem, the low overall prevalence of false content seems like a hopeful sign.

The Upshot

The upshot of this scholarship and governmental evaluation appears to be that some social media anti-disinformation measures may be working to some extent. Overall, though, the results argue for caution. And it bears repeating that disinformation actors and tactics will likely shift in 2020, meaning that all bets will be off.

Facebook, for one, has publicly celebrated some of the research cited here, which it claims shows that the company is "making strides against false news."[91] But as the literary and human rights organization PEN America has observed, at this point, "any self-congratulation is premature."[92]

4. Conclusions and Recommendations

Disinformation has many permutations. This report offers educated guesses about some of what may unfold in 2020, but disinformation artists likely have other tricks up their sleeves.

> "The platforms are playing better defense than they did in 2016. But that's an exceedingly low bar. Now is the time for these companies to step up their games in anticipation of 2020."

Candid Wüeest, a senior principal threat researcher at the cybersecurity firm Symantec, said via email that he and his colleagues are seeing ruses specifically designed to fool platform AI filters. These methods could be combined next year with approaches discussed in Part 2 of this report. One technique, according to Wüeest, is to have bot accounts amplify legitimate content that's divisive—a news report on the immigration controversy, say—which is more difficult to detect than bots trumpeting made-up material. Another twist: bots recycle real news articles that are months or years old, with the aim of reviving online disputes that may have settled down. Again, the idea is to inflame partisan users without activating platform filters.

Confronting these ever-changing strategies, the social media companies are playing better defense than they did in 2016. But that's an exceedingly low bar. Now is the time for them to step up their games in anticipation of 2020.

The recommendations that follow are intended as practical-minded encouragement to do the right thing for the sake of users and society at large. Protecting against disinformation is a way of protecting democracy, which depends for its survival on authentic communication of real facts.

The good news for the platforms is that fighting disinformation, even if there are short-term R&D and personnel costs, ultimately should help restore their brand reputations and slow demands for draconian government regulation. Over time, if the companies address their challenges more effectively, they will see a benefit to the bottom line, as follows: Advertisers pay for users' attention. Users will continue to offer their attention only if they trust a given platform. Building trust, therefore, should be a central part of any social media business plan. And countering disinformation offers an excellent way to do it.

Recommendations to the Social Media Companies

1 Detect and remove deepfakes.

Realistic but fraudulent videos have the potential to undermine political candidates and exacerbate voter cynicism. Former senior foreign policy officials have warned that deepfakes simulating military activity or acts of terrorism could even help spark a war.[93] Another danger is that the advent of convincing video (and audio) fabrication could so blur the line between sham and reality that dishonest public figures would successfully wave off genuine depictions as deepfakes—a phenomenon that some have called the "liar's dividend."[94]

Legislation has been introduced in Congress that would criminally punish those who make deepfakes without disclosing the modifications. A better approach, and one that avoids the danger of overreaching government censorship, would be for the social media platforms to improve their AI-screening technology, enhance human review, and remove deepfakes before they can do much damage. In one promising development, Google has been assisting outside research groups working on deepfake detectors.[95] More such efforts are necessary.

2 Remove provably false content in general.

Beyond deepfakes, the social media companies should take down provably false content in general. This idea isn't as drastic as it may sound. The platforms already identify and remove certain categories of content, ranging from hate speech to voter suppression. We're recommending that they add one more category: material that can be definitively shown to be untrue. Pinterest, a social network that focuses on recipes and home projects, has blazed a path in this regard. According to its community guidelines, Pinterest removes "harmful advice, content that targets individuals or protected groups, and content created as part of disinformation campaigns."[96]

We don't expect Facebook, Twitter, and YouTube ever to be swept clean of falsehoods. Instead, as algorithmic and human review flag potentially untrue items, the platforms would evaluate them and purge what's provably false. This may put only a modest additional dent in the volume of social media fakery, but it would be an important step in the right direction. As noted earlier, we urge the companies to prioritize false content related to democratic institutions, starting with elections. And we suggest that they retain clearly marked copies of removed material in a publicly accessible, searchable archive, where false content can be studied by scholars and others, but not shared or retweeted. Finally, as the platforms begin to delete provably false content, they will need to provide a straightforward appeals process allowing users to seek timely review and a remedy when mistakes are made.

3 Hire a content overseer.

Hand-in-hand with instituting a policy on provably false content goes the appointment at each company of a senior official to oversee the process of guarding against disinformation. As we argued in our last report on domestically generated disinformation, responsibility for content decisions now tends to be scattered among disparate teams within the social media companies. Centralization would streamline key processes and underscore their importance, internally and externally. This is more likely to happen if the new official reports directly to the COO or CEO.

Choosing the right people for these sensitive jobs will be critical. One option is hiring an experienced editor or executive drawn from the journalism business who has the right combination of savvy and commonsense about what's real and what's not. A stiff spine is another job requirement, as the content overseer's responsibility will include standing firm in the face of unfounded claims of bias by right-wing activists and President Trump. In light of the ongoing partisan onslaught, the companies should put people in these posts who are beyond (reasonable) reproach and who will make sure all users are treated even-handedly every time they log in.

4 Make changes at Instagram.

In March 2019, a headline on the tech news site *The Verge* warned: "Instagram is facing a reckoning over misinformation." The article's subhead added: "Anti-vaccine posts and misinformation are rotting the platform."[97] The message of the *Verge* piece combines ominously with the finding in the Senate Intelligence Committee report that Instagram generated more than twice as much engagement with Russian disinformation in 2016 as its parent, Facebook.

The problem isn't a lack of technology. It appears to be a lack of a clear strategy for addressing the serious problems inherent in Instagram's operating model. The platform currently is testing a system that uses image recognition and other tools to find potential misinformation, which is sent to Facebook fact-checkers. According to *Wired* magazine, material deemed false isn't recommended to new users, but Instagram doesn't remove or down-rank it in users' main feeds. The company chooses to focus "on making it harder for new users to be algorithmically exposed to misinformation, rather than stemming the reach of misinformation."[98] But the goals aren't mutually exclusive. Why not pursue both?

Instagram, to be sure, has made progress in certain areas. It has made it easier for users to identify suspicious accounts by disclosing such information as accounts' location and the ads they're running. In the past year, Instagram also has removed hundreds of accounts for displaying coordinated inauthentic behavior. But these steps haven't cured the platform's burgeoning reputation as a vehicle for false content. Instagram needs to move swiftly and aggressively to protect all of its users from disinformation.

5 Limit the reach of WhatsApp.

As an encrypted service meant for private communication, not public sharing, WhatsApp is very different from Instagram, Facebook, Twitter, and YouTube. That means different approaches are necessary to protect its users. WhatsApp has been moving in the right direction and now needs to go further.

Originally, the messaging service allowed users to forward content to up to 256 chat groups at a time. Each group could have up to 256 members. Simple multiplication reveals that one person could single-handedly deliver a message to 65,536 others. If political campaigns wanted a fake rumor to go viral, these numbers would look attractive. Seeking to diminish this potential for abuse, WhatsApp has gradually reduced the number of groups a user may forward to at one time. The current limit is five, which means that the maximum audience for one forwarded message has shrunk to 1,280. This reduction has lessened forwarding by 25% globally, according to Facebook.[99] To further reduce the potential for misuse, the company should restrict users to forwarding to one chat group, or 256 people, at a time. This would preserve WhatsApp's usefulness as a private messaging service while making it more difficult to exploit.

6 Defend against for-profit disinformation.

The social media companies must pay more attention to false content distributed by corporations, consultants, and public relations firms. Profit-driven disinformation isn't brand new. For years, some clickbait purveyors have used conspiracy theories and phony news to draw gullible social media users to their ad-laden websites. But for-profit methods are evolving and becoming more ambitious than mere clickbait schemes. Disinformation services are for sale on a large scale.

Facebook's shutdown of a network linked to the Israeli firm Archimedes Group revealed an operation similar to the Russian IRA but one apparently operated to make money, not accomplish an overarching ideological goal. The international reach of Archimedes—from Africa to Latin America to Southeast Asia—served notice that capitalist motives are turning the manufacture of false content into a global business.[100]

(continued on p. 20)

Recommendations to the Social Media Companies (continued)

7 Support legislation on political ads and voter suppression.

Political ads: The social media companies should throw their considerable Washington lobbying clout behind the Honest Ads Act. As noted earlier, this measure would extend political-ad disclosure standards to online ads that are comparable to those imposed on traditional print and broadcast media. Facebook and Twitter have endorsed the Honest Ads Act, but they do not seem to have made adoption of the legislation a top political priority. The House has approved the provision, and now the companies should do more to push for its approval in the Senate. To date, the Republican leadership in the Senate has prevented a vote on the Honest Ads Act, but active support by the major platforms could make a difference. Depending on the 2020 elections, the Republican roadblock could disappear. The sooner the companies put their shoulder behind mandatory online disclosure, the better.

Voter suppression: The three main platforms, to their credit, have strengthened policies in this area and vow to remove false content meant to confuse or intimidate voters on Election Day. But these voluntary policies don't go far enough to deter voter suppression, which typically has sought to keep Blacks and Latinos in particular from participating in our democracy. Legislation pending in Congress, known as the Deceptive Practices and Voter Intimidation Act, would clarify that all forms of intentional voter disinformation are prohibited and punishable by up to five years in prison. Given that much modern-day suppression takes place via social media, the platforms have a duty to support this measure, even as they continue to remove voter-suppression content themselves.

8 Improve industry-wide collaboration.

Cooperation among the major platforms has improved, but they still sometimes go their own way when addressing harmful content. In March 2019, researchers at the Alliance for Securing Democracy pointed out that after Facebook took down 2,600 pages, groups, and accounts engaged in coordinated information operations, "related accounts on Twitter, YouTube, and Instagram continued to operate, spreading falsehoods."[101] This lack of coordination represents a wasted opportunity to make greater progress against the forces of disinformation.

More broadly, the social media industry should foster coordination by forming a permanent inter-company task force devoted to fighting disinformation. The platforms have worked together in a variety of contexts. There's the Global Internet Forum to Counter Terrorism and the Global Network Initiative, which focuses on freedom of expression and privacy. The PhotoDNA Initiative addresses child pornography. And a database of "digital fingerprints" allows the platforms to take down violent extremist content more efficiently. These joint efforts have enjoyed varying degrees of success, but they share a spirit of cooperation that ought to infuse the push to limit disinformation.

9 Teach social media literacy in a more direct, sustained way.

We applaud the social media companies for the financial support and encouragement they have provided to a variety of digital-literacy campaigns. These programs—ranging from classroom simulations for middle and high school students to videos by popular online personalities—are designed to educate users about questionable content, so they're less likely to be fooled by it or share it. Building on these admirable efforts, the companies should do more on this front.

Specifically, the platforms ought to make digital literacy lessons a permanent and prominently available feature on each of their sites. Doubtless, the companies would prefer not to remind users every time they log in that disinformation casts a shadow over social media. But that's the reality. The more often users are reminded of this fact—and are taught how to distinguish real from fake—the less influence false content will wield. Concise, incisive instruction, possibly presented in FAQs format, should be just one click away for all users of all of the platforms, all of the time.

Endnotes

1 Sarah Mervosh, "Distorted Videos of Nancy Pelosi Spread on Facebook and Twitter, Helped by Trump," *The New York Times*, May 24, 2019 (https://www.nytimes.com/2019/05/24/us/politics/pelosi-doctored-video.html). Another widely circulated "cheapfake" video falsely made it look as if CNN correspondent Jim Acosta was physically aggressive with a female White House intern during a November 2018 presidential press conference. We discussed the Acosta episode in our last report: Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do," New York University Stern Center for Business and Human Rights, March 2019 (https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_domestic_disinformation_digital?e=31640827/68184927).

2 Glenn Kessler, Salvador Rizzo, and Meg Kelly, "President Trump Has Made 12,019 False or Misleading Claims Over 928 Days," *The Washington Post,* August 12, 2019 (https://www.washingtonpost.com/politics/2019/08/12/president-trump-has-made-false-or-misleading-claims-over-days/).

3 Clint Watts, "The Kremlin's Strategy for the 2020 U.S. Election: Secure the Base, Split the Opposition," *Daily Beast,* May 29, 2019 (https://www.thedailybeast.com/the-kremlins-strategy-for-the-2020-us-election-secure-the-base-split-the-opposition).

4 James Loke Hale, "More than 500 Hours of Content Are Now Being Uploaded to YouTube Every Minute," *Tubefilter,* May 7, 2019 (https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/).

5 Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do" (supra note 1).

6 Paul M. Barrett, Tara Wadhwa, and 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_domestic_disinformation_digital?e=31640827/68184927).

7 "Full Transcript: Robert Mueller's Statement on the Russia Investigation," *Politico,* May 29, 2019 (https://www.politico.com/story/2019/05/29/robert-mueller-statement-russia-investigation-text-transcript-1346453).

8 Dustin Volz, "No Significant Foreign Interference Seen on Midterm Vote," *The Wall Street Journal,* November 7, 2018 (https://www.wsj.com/articles/u-s-on-alert-for-election-interference-says-nothing-significant-spotted-1541521048).

9 Ellen Nakashima, "U.S. Cyber Command Operation Disrupted Internet Access of Russian Troll Farm on Day of 2018 Midterms," *The Washington Post,* February 27, 2019 (https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7ff322e9_story.html?utm_term=.85441f897537).

10 Interview with the author.

11 Interview with the author.

12 Eric Schmitt, David E. Sanger, Maggie Haberman, "In Push for 2020 Election Security, Top Official Was Warned: Don't Tell Trump," *The New York Times,* April 24, 2019 (https://www.nytimes.com/2019/04/24/us/politics/russia-2020-election-trump.html).

13 Daniel R. Coats, "Worldwide Threat Assessment of the U.S. Intelligence Community," Office of the Director of National Intelligence, January 29, 2019 (https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf).

14 Tom Van de Weghe, "Six Lessons from My Deepfakes Research at Stanford," *Medium,* May 29, 2019 (https://medium.com/jsk-class-of-2019/six-lessons-from-my-deepfake-research-at-stanford-1666594a8e50) and Leo Kelion, "Reddit Bans Deepfake Porn Videos," *BBC,* February 7, 2018 (https://www.bbc.com/news/technology-42984127).

15 John Villasenor, "Deepfakes, Social Media, and the 2020 Election," *TechTank,* Brookings Institution, June 3, 2019 (https://www.brookings.edu/blog/techtank/2019/06/03/deepfakes-social-media-and-the-2020-election/).

16 Samantha Putterman, "Zuckerberg Video about 'Billions of People's Stolen Data' Is a Deepfake," *Politifact,* June 12, 2019 (https://www.politifact.com/facebook-fact-checks/statements/2019/jun/12/instagram-posts/zuckerberg-video-about-billions-peoples-stolen-dat/).

17 Danielle Keats Citron, "Prepared Written Testimony and Statements for the Record," Hearing on the National Security Challenges of Artificial Intelligence, Manipulated Media, and "Deep Fakes," U.S. House Permanent Select Committee on Intelligence, June 13, 2019 (https://intelligence.house.gov/uploadedfiles/citron_testimony_for_house_committee_on_deep_fakes.pdf).

18 Ali Breland, "The Bizarre and Terrifying Case of the 'Deepfake' Video that Helped Bring an African Nation to the Brink," *Mother Jones,* March 15, 2019 (https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/).

19 Robert Chesney, Danielle Keats Citron, Quinta Jurecic, "About That Pelosi Video: What to Do About 'Cheapfakes' in 2020," *Lawfare,* May 29, 2019 (https://www.lawfareblog.com/about-pelosi-video-what-do-about-cheapfakes-2020).

20 Interview with the author. It's worth noting that the mainstream media contribute to the dissemination of distorted videos.

21 Robert S. Mueller III, "Report on the Investigation into Russian Interference in the 2016 Presidential Election," Department of Justice, March 2019 (https://assets.documentcloud.org/documents/5955118/The-Mueller-Report.pdf).

22 "Removing Bad Actors on Facebook," Facebook, July 31, 2018 (https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/).

23 Joe Heim et al., "White-Supremacist Rally Near White House Dwarfed by Thousands of Anti-Hate Protesters," *The Washington Post,* August 12, 2018 (https://www.washingtonpost.com/local/washington-readies-for-todays-planned-white-supremacist-rally-near-white-house/2018/08/12/551720c4-9c28-11e8-8d5e-c6c594024954_story.html?utm_term=.569627792071).

24 Renee DiResta et al., "The Tactics and Tropes of the Internet Research Agency," New Knowledge for the U.S. Senate Select Committee on Intelligence, December 17, 2018 (https://www.intelligence.senate.gov/press/new-reports-shed-light-internet-research-agency%E2%80%99s-social-media-tactics).

25 Stan Schroeder, "U.S. Users Are Leaving Facebook, New Study Claims," *Mashable,* March 7, 2019 (https://mashable.com/article/facebook-losing-users-us/).

Endnotes (continued)

26 Alex Pasternack, "It's Not Over: Russia's Divisive Internet Memes Are Still Racking Up Likes," *Fast Company,* December 19, 2018 (https://www.fastcompany.com/90283167/russia-instagram-war-facebook-memes).

27 Taylor Lorenz, "Instagram is the Internet's New Home for Hate," *The Atlantic,* March 21, 2019 (https://www.theatlantic.com/technology/archive/2019/03/instagram-is-the-internets-new-home-for-hate/585382/).

28 Madhumita Murgia, Stephanie Findlay, and Andres Schipani, "India: The WhatsApp Election," *Financial Times,* May 5, 2019 (https://www.ft.com/content/9fe88fba-6c0d-11e9-a9a5-351eeaef6d84); Vindu Goel and Sheera Frenkel, "In India Election, False Posts and Hate Speech Flummox Facebook," *The New York Times,* April 1, 2019 (https://www.nytimes.com/2019/04/01/technology/india-elections-facebook.html), and Philippa Williams and Lipika Kamra, "Technology Could Torpedo India's First WhatsApp Election," *Quartz,* March 4, 2019 (https://qz.com/india/1563318/indias-2019-election-is-threatened-by-fake-news-on-whatsapp/).

29 Interview with the author.

30 Alice Revelli and Lee Foster, "Network of Social Media Accounts Impersonates U.S. Political Candidates, Leverages U.S. and Israeli Media in Support of Iranian Interests," FireEye, May 28, 2019 (https://www.fireeye.com/blog/threat-research/2019/05/social-media-network-impersonates-us-political-candidates-supports-iranian-interests.html).

31 Nathaniel Gleicher, "Removing More Coordinated Inauthentic Behavior From Iran," Facebook, May 28, 2019 (https://newsroom.fb.com/news/2019/05/removing-more-cib-from-iran/).

32 Yoel Roth, Twitter, May 28, 2019 (https://twitter.com/yoyoel/status/1133448387657736192). All told, Twitter reportedly removed more than 7,000 phony accounts tied to Iran during the first seven months of 2019. See Craig Timberg and Tony Romm, "It's Not Just the Russians Anymore as Iranian and Others Turn Up Disinformation Efforts Ahead of 2020 Vote," *The Washington Post,* July 25, 2019 (https://www.washingtonpost.com/technology/2019/07/25/its-not-just-russians-anymore-iranians-others-turn-up-disinformation-efforts-ahead-vote/?utm_term=.4408c57407ff).

33 Sheera Frenkel and Nicholas Fandos, "Facebook Identifies New Influence Operations Spanning Globe," *The New York Times,* August 21, 2018 (https://www.nytimes.com/2018/08/21/technology/facebook-political-influence-midterms.html).

34 Interview with the author.

35 Interview with the author.

36 See, e.g., comments of Lisa-Maria Neudert: "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/?).

37 Cat Zakrzewski, "The Technology 202: Disinformation Spread by Americans Is 'the Hardest Challenge that We Have,' DHS Official Says," *The Washington Post,* April 12, 2019 (https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2019/04/12/the-technology-202-disinformation-spread-by-americans-is-the-hardest-challenge-that-we-have-dhs-official-says/5caf9cf91ad2e567949ec16c/?utm_term=.ecec91773e5f).

38 See, e.g., Vidya Narayanan et al., "Polarization, Partisanship, and Junk News Consumption Over Social Media in the U.S.," Oxford Internet Institute, February 6, 2018 (https://comprop.oii.ox.ac.uk/research/polarization-partisanship-and-junk-news/) and Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and the Radicalization of American Politics,* Oxford University Press, 2018 (https://global.oup.com/academic/product/network-propaganda-9780190923631?cc=us&lang=en&). See also, Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do" (supra note 1). As discussed in "Tackling Domestic Disinformation," a good deal of false content bubbles up from secondary social media sites, such as Reddit and Gab, to Twitter, YouTube, and Facebook.

39 "Candace Owens: Big Tech Trying to Ban Conservatives Ahead of 2020 Election," *Breitbart,* June 12, 2019 (https://www.breitbart.com/politics/2019/06/12/watch-candace-owens-big-tech-trying-to-ban-conservatives-ahead-of-2020-election/).

40 See, e.g., Kurt Wagner, "Facebook Bans Alex Jones, Milo Yiannopoulos, Other Far-Right Figures," *Bloomberg,* May 2, 2019 (https://www.bloomberg.com/news/articles/2019-05-02/facebook-bans-alex-jones-yiannopoulos-other-far-right-figures).

41 Lauren Feiner, "Trump Says Facebook, Amazon, and Google Were Colluding With Democrats Against Him," *CNBC,* June 10, 2019 (https://www.cnbc.com/2019/06/10/trump-sounds-off-on-facebook-huawei-and-5g.html).

42 Donald J. Trump, Twitter, July 11, 2019 (https://twitter.com/realDonaldTrump/status/1149279675660869632).

43 Ben Brody, "Trump Says He'll Summon Companies After Social Media 'Summit,'" *Bloomberg,* July 11, 2019 (https://www.bloomberg.com/news/articles/2019-07-11/trump-s-social-media-summit-brings-fringe-voices-to-white-house).

44 Caroline Kelly, "'It Won't Work': Kamala Harris' Campaign Slams Online Attacks on Her," *CNN,* June 30, 2019 (https://www.cnn.com/2019/06/29/politics/kamala-harris-responds-donald-trump-jr-race-african-american-black/index.html).

45 Adam K. Raymond, "Right-Wing Troll Jacob Wohl Allegedly Behind Fake Buttigieg Sexual-Assault Claim," *New York,* April 30, 2019 (http://nymag.com/intelligencer/2019/04/right-wing-troll-set-up-fake-buttigieg-sexual-assault-claim.html).

46 "A Progressive Activist Defends His Deceptive Tactics," *WNYC,* January 11, 2019 (https://www.wnycstudios.org/story/progressive-activist-defends-his-deceptive-tactics).

47 Scott Shane and Alan Blinder, "Democrats Faked Online Push to Outlaw Alcohol 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).

48 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=.e1f7350447ac).

49 "A Progressive Activist Defends His Deceptive Tactics" (supra note 46).

50 Interview with the author.

51 Tony Romm and Craig Timberg, "Facebook Suspends Five Accounts, Including That of a Social Media Researcher, for Misleading Tactics in Alabama Elections," *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/?utm_term=.6fd96ea070dd) and Scott Shane and Alan Blinder, "Democrats Faked Online Push to Outlaw Alcohol in Alabama Race" (supra note 47).

52 Andy Greenberg, "Alphabet-Owned Jigsaw Bought a Russian Troll Campaign as an Experiment," *Wired,* June 12, 2019 (https://www.wired.com/story/jigsaw-russia-disinformation-social-media-stalin-alphabet/).

53 Adam Entous and Ronan Farrow, "Private Mossad for Hire," *The New Yorker,* February 11, 2019 (https://www.newyorker.com/magazine/2019/02/18/private-mossad-for-hire) and Mark Mazzetti, et al., "Rick Gates Sought Online Manipulation Plans from Israeli Intelligence Firm for Trump Campaign," *The New York Times,* October 8, 2018 (https://www.nytimes.com/2018/10/08/us/politics/rick-gates-psy-group-trump.html).

54 Nathaniel Gleicher, "Removing Coordinated Inauthentic Behavior from Israel," Facebook, May 16, 2019 (https://newsroom.fb.com/news/2019/05/removing-coordinated-inauthentic-behavior-from-israel/).

55 Isabel Debre and Raphael Satter, "Facebook Busts Israel-based Campaign to Disrupt Elections," *Associated Press,* May 16, 2019 (https://www.apnews.com/7d334cb8793f49889be1bbf89f47ae5c) and Simona Wineglass, "Who Is Behind Israel's Archimedes Group, Banned by Facebook for Election Fakery?" *The Times of Israel,* May 19, 2019 (https://www.timesofisrael.com/who-is-behind-israels-archimedes-group-banned-by-facebook-for-election-fakery/).

56 Robert S. Mueller III, "Report on the Investigation into Russian Interference in the 2016 Presidential Election" (supra note 21).

57 "Facebook Press Call," Facebook, April 26, 2019 (https://fbnewsroomus.files.wordpress.com/2019/04/transcript-9549757_final.pdf).

58 Young Mie Kim, "Voter Suppression has Gone Digital," New York University Brennan Center for Justice, November 20, 2018 (https://www.brennancenter.org/blog/voter-suppression-has-gone-digital).

59 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-seeking-to-discourage-voting-idUSKCN1N72FA?feedType=RSS&feedName=technologyNews&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+reuters%2FtechnologyNews+%28Reuters+Technology+News%29).

60 Guy Rosen, "An Update on How We Are Doing at Enforcing Our Community Standards," Facebook, May 23, 2019 (https://newsroom.fb.com/news/2019/05/enforcing-our-community-standards-3/).

61 Nathaniel Gleicher, "Coordinated Inauthentic Behavior Explained," Facebook, December 6, 2018 (https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/).

62 Interview with the author.

63 Henry Silverman, "The Next Phase in Fighting Misinformation," Facebook, April 10, 2019 (https://newsroom.fb.com/news/2019/04/tackling-more-false-news-more-quickly/).

64 "Zuckerberg Says It's 'Really Important' for Facebook to Form Deepfake Video Policy," *CBS News,* June 27, 2019 (https://www.cbsnews.com/news/facebook-zuckerberg-says-its-really-important-to-form-policy-on-deepfake-videos/).

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

66 "Cooper Grills Facebook VP for Keeping Pelosi Video Up," *CNN,* May 25, 2019 (https://www.cnn.com/videos/tech/2019/05/25/facebook-monika-bickert-pelosi-video-cooper-intv-sot-ac360-vpx.cnn).

67 Jessica Leinwand, "Expanding Our Policies on Voter Suppression," Facebook, October 15, 2018 (https://newsroom.fb.com/news/2018/10/voter-suppression-policies/) and Sheryl Sandberg, "A Second Update on Our Civil Rights Audit," Facebook, June 30, 2019 (https://newsroom.fb.com/news/2019/06/second-update-civil-rights-audit/).

68 Kevin Kane, "Securing U.S. Election Infrastructure and Protecting Political Discourse," U.S. House Committee on Oversight and Reform, May 22, 2019 (https://docs.house.gov/meetings/GO/GO06/20190522/109538/HHRG-116-GO06-Wstate-KaneK-20190522.pdf) and "Policies and Safety," YouTube, undated (https://support.google.com/youtube/answer/2801973?hl=en&ref_topic=9282365).

69 "Our Ongoing Work to Tackle Hate," YouTube, June 5, 2019 (https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html).

70 "Continuing Our Work to Improve Recommendations on YouTube," YouTube, January 25, 2019 (https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html).

71 Yoel Roth, "Information Operations on Twitter: Principles, Process, and Disclosure," Twitter, June 13, 2019 (https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html).

72 Henry Silverman, "The Next Phase in Fighting Misinformation" (supra note 63).

73 Sam Levin and Julia Carrie Wong, "'He's Learned Nothing': Zuckerberg Floats Crowdsourcing Facebook Fact-Checks," *The Guardian,* February 20, 2019 (https://www.theguardian.com/technology/2019/feb/20/facebook-fact-checking-crowdsourced-mark-zuckerberg). Snopes said it ended its partnership with Facebook because the arrangement was too time-consuming and inefficient. See, e.g., Daniel Funke, "Snopes Pulls Out of its Fact-Checking Partnership with Facebook," *Poynter,* February 1, 2019 (https://www.poynter.org/fact-checking/2019/snopes-pulls-out-of-its-fact-checking-partnership-with-facebook/).

74 "Interview with Facebook CEO Mark Zuckerberg," *ABC News,* April 4, 2019 (https://abcnews.go.com/Business/interview-facebook-ceo-mark-zuckerberg-transcript/story?id=62152829).

75 Kevin Kane, "Securing U.S. Election Infrastructure and Protecting Political Discourse" (supra note 68).

Endnotes (continued)

76 Richard Salgado, "Securing U.S. Election Infrastructure and Protecting Political Discourse," U.S. House Committee on Oversight and Reform, May 22, 2019 (https://docs.house.gov/meetings/GO/GO06/20190522/109538/HHRG-116-GO06-Wstate-SalgadoR-20190522.pdf).

77 Katie Harbath and Samidh Chakrabarti, "Expanding Our Efforts to Protect Elections in 2019," Facebook, January 28, 2019 (https://newsroom.fb.com/news/2019/01/elections-2019/).

78 Samidh Chakrabarti, "Fighting Election Interference in Real Time," Facebook, October 18, 2018 (https://newsroom.fb.com/news/2018/10/war-room/).

79 Sheera Frenkel and Matthew Rosenberg, "Top Tech Companies Met with Intelligence Officials to Discuss Midterms," *The New York Times,* June 25, 2018 (https://www.nytimes.com/2018/06/25/technology/tech-meeting-midterm-elections.html) and "Interview with Facebook CEO Mark Zuckerberg" (supra note 74).

80 Kevin Kane, "Securing U.S. Election Infrastructure and Protecting Political Discourse" (supra note 68).

81 Del Harvey and Yoel Roth, "An Update on Our Elections Integrity Work," Twitter, October 1, 2018 (https://blog.twitter.com/en_us/topics/company/2018/an-update-on-our-elections-integrity-work.html).

82 See, e.g., Del Harvey and Bruce Falck, "Announcing New U.S. Issue Ads Policy," Twitter, August 30, 2018 (https://blog.twitter.com/en_us/topics/company/2018/Announcing-new-US-issue-ads-policy.html). Researchers have reported technical problems with Facebook's advertising database: See, Matthew Rosenberg, "Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised," *The New York Times,* July 25, 2019 (https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html).

83 "Announcement: RT and Sputnik Advertising," Twitter, October 26, 2017 (https://blog.twitter.com/en_us/topics/company/2017/Announcement-RT-and-Sputnik-Advertising.html).

84 William Turton, "We Posed as 100 Senators to Run Ads on Facebook. Facebook Approved Them All," *Vice News,* October 30, 2018 (https://news.vice.com/en_us/article/xw9n3q/we-posed-as-100-senators-to-run-ads-on-facebook-facebook-approved-all-of-them).

85 Jordain Carney, "Democrats Detail New Strategy to Pressure McConnell on Election Security Bills," *The Hill,* June 18, 2019 (https://thehill.com/blogs/floor-action/senate/449174-democrats-detail-new-strategy-to-pressure-mcconnell-on-election) and Maggie Miller, "Bipartisan Group of Senators Seeks to Increase Transparency of Online Political Ads," *The Hill*, May 8, 2019 (https://thehill.com/homenews/senate/442598-bipartisan-group-of-senators-seek-to-increase-transparency-of-online).

86 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 (http://blogs.oii.ox.ac.uk/comprop/wp-content/uploads/sites/93/2018/11/marchal_et_al.pdf).

87 Andrew Guess, et al., "Fake News, Facebook Ads, and Misperceptions," University of Michigan, February 2019 (http://www-personal.umich.edu/~bnyhan/fake-news-2018.pdf).

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

89 "Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions," European Commission, June 14, 2019 (https://eeas.europa.eu/sites/eeas/files/joint_report_on_disinformation.pdf).

90 Nahema Marchal, et al., "Junk News During the EU Parliamentary Elections: Lessons from a Seven-Language Study of Twitter and Facebook," Oxford Internet Institute, May 21, 2019 (https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/05/EU-Data-Memo.pdf).

91 "New Research Shows Facebook Making Strides Against False News," Facebook, February 7, 2019 (https://newsroom.fb.com/news/2018/10/inside-feed-michigan-lemonde/).

92 "Truth on the Ballot: Fraudulent News, the Midterm Elections, and Prospects for 2020," PEN America, March 13, 2019 (https://pen.org/wp-content/uploads/2019/03/Truth-on-the-Ballot-report.pdf).

93 Daniel Benjamin and Steven Simon, "How Fake News Could Lead to a Real War," *Politico Magazine,* July 5, 2019 (https://www.politico.com/magazine/story/2019/07/05/fake-news-real-war-227272).

94 Robert Chesney and Danielle Keats Citron, "Deepfakes: A Looming Challenge for Privacy, Democracy, and National Security," *California Law Review,* forthcoming, 2019 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954).

95 Daisy Stanton, "Advancing Research on Fake Audio Detection," Google, January 31, 2019 (https://www.blog.google/outreach-initiatives/google-news-initiative/advancing-research-fake-audio-detection/).

96 "Community Guidelines," Pinterest, undated (https://policy.pinterest.com/en/community-guidelines).

97 Casey Newton, "Instagram Is Facing a Reckoning Over Misinformation," *The Verge,* March 22, 2019 (https://www.theverge.com/interface/2019/3/22/18276479/instagram-reckoning-anti-vaxx-misinformation-conspiracy).

98 Paris Martineau, "Instagram Can Find Misleading Posts—but Won't Take Them Down," *Wired,* May 8, 2019 (https://www.wired.com/story/instagram-find-misleading-posts-wont-take-down/).

99 "More Changes to Forwarding," WhatsApp, January 21, 2019 (https://blog.whatsapp.com/10000647/More-changes-to-forwarding).

100 Luiza Bandeira, et al., "Inauthentic Israeli Facebook Assets Target the World," Digital Forensic Research Lab, May 17, 2019 (https://medium.com/dfrlab/inauthentic-israeli-facebook-assets-target-the-world-281ad7254264).

101 Jessica Brandt and Bradley Hanlon, "Online Information Operations Cross Platforms. Tech Companies' Responses Should, Too," *Lawfare,* April 26, 2019 (https://www.lawfareblog.com/online-information-operations-cross-platforms-tech-companies-responses-should-too).

NYU Stern Center for Business and Human Rights
Leonard N. Stern School of Business
44 West 4th Street, Suite 800
New York, NY 10012
+1 212-998-0261
bhr@stern.nyu.edu
bhr.stern.nyu.edu

Center for Business
and Human Rights