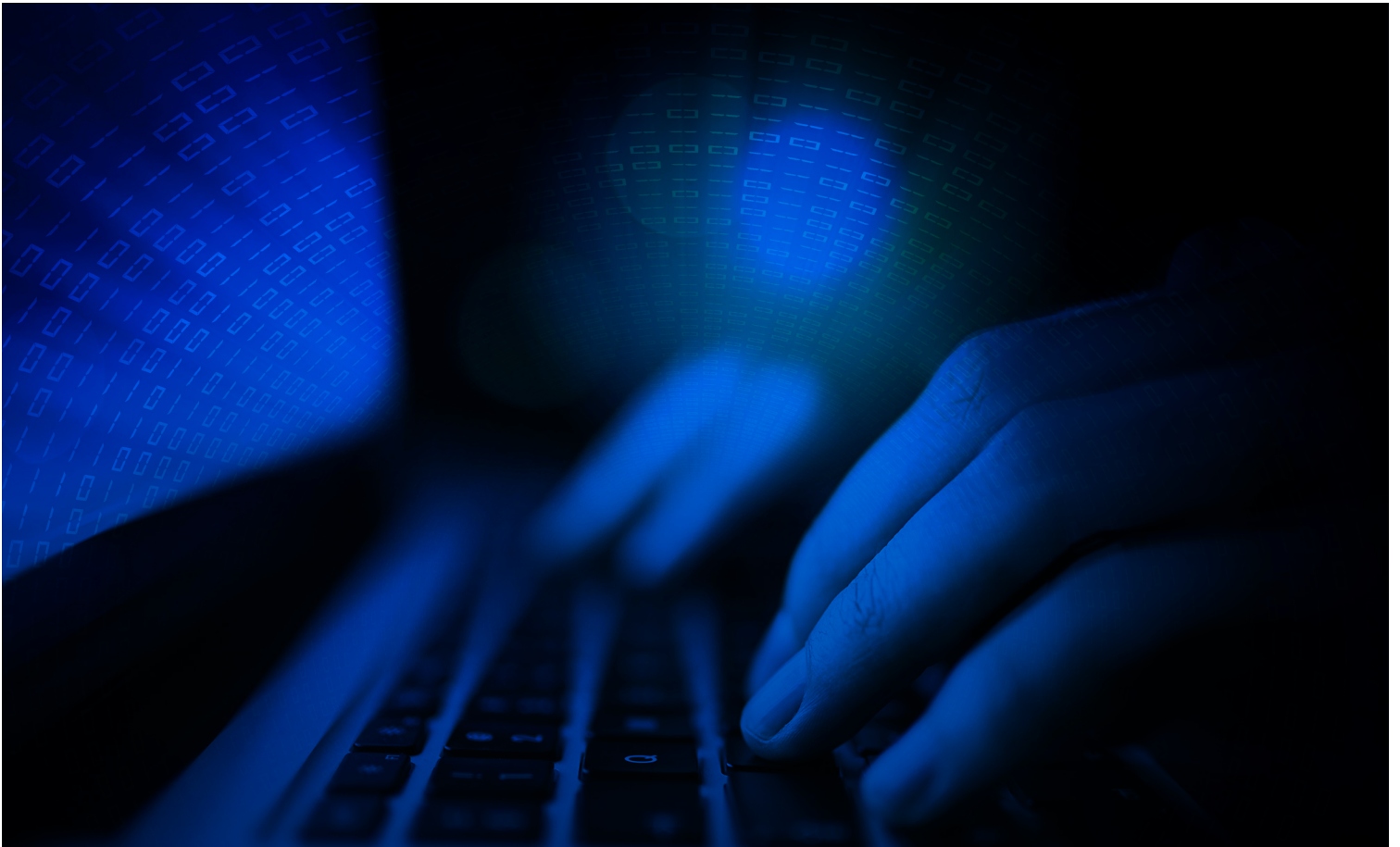


# Harmful Content: The Role of Internet Platform Companies in Fighting Terrorist Incitement and Politically Motivated Disinformation

---



# Contents

Foreword ..... 1

Executive Summary ..... 3

Introduction: A More Proactive Approach  
to Governance of Digital Platforms..... 7

The Challenge: Harmful Content and Free Speech ..... 11

The Responsibilities of Digital Platforms:  
Algorithms, Advertisements, and Human Judgement..... 17

Conclusions and Recommendations ..... 25

Endnotes ..... 28

Acknowledgements ..... 29

NYU Stern Center for Business and Human Rights  
Leonard N. Stern School of Business  
44 West 4th Street, Suite 800  
New York, NY 10012  
+1 212-998-0722  
bhr@stern.nyu.edu  
bhr.stern.nyu.edu

© 2017 NYU Stern Center for Business and Human Rights  
All rights reserved. This work is licensed under the Creative  
Commons Attribution-NonCommercial 4.0 International  
License. To view a copy of the license, visit  
<http://creativecommons.org/licenses/by-nc/4.0/>.

# Foreword



Michael H. Posner

Jerome Kohlberg Professor of Ethics and Finance; Director, Center for Business and Human Rights, Stern School of Business, New York University



**Expecting to eradicate all harmful content is unrealistic.**

**But progress is possible.**



Founded in 2013, the NYU Stern Center for Business and Human Rights is the first human rights center established at a business school. The Center does research and advocacy aimed at promoting practical industry-wide solutions to human rights challenges. We describe our approach as pro-business, high standards. This report focuses on the human rights harms caused by online terrorist incitement and politically motivated false information. While there are many varieties of deleterious digital content, these two types are having a particularly detrimental effect on social and political discourse and are threatening human rights around the globe. We look closely at the obligations of major internet companies to address harmful content online.

Written and published by the Center, the report grows out of discussions among members of the World Economic Forum Global Future Council on Human Rights, which I co-chair. A number of Council members, especially Daniel Bross, Eileen Donahoe, and Andrew McLaughlin, provided invaluable insight and commentary. Given the timeliness of the topic, the Center proposed, and the World Economic Forum generously agreed, to publish the report now under the auspices of the Center. We will continue to work closely with the Forum to promote dialogue and collaboration on this important subject.

The Forum has pursued a broader study of how new technologies are “fundamentally changing the way we live, work, and relate to one another”.<sup>1</sup> In his 2016 book, *The Fourth Industrial Revolution*, World Economic Forum Founder and Executive Chairman Klaus Schwab calls for “ethical standards that should apply to emerging technologies”, which he rightly says are “urgently needed to establish common ethical guidelines and embed them in society and culture”.<sup>2</sup>

In the Center’s work on reducing the prevalence of harmful internet content, we look to the Universal Declaration of Human Rights and a dozen subsequent international treaties that provide legal standards for safeguarding free expression, media freedom, and the promotion of core political freedoms.

These international agreements are complemented by diplomatic actions, such as the 2012 United Nations Human Rights Council resolution on The Promotion, Protection, and Enjoyment of Human Rights on the Internet, which make clear that human rights must be protected online as they would be offline.<sup>3</sup>

Internet platform companies such as Facebook, Twitter, and Google have resisted national and international regulation imposing content restrictions on their products. They are justifiably concerned that many states would seek to suppress dissenting views, undermining free speech online. As an alternative to government regulation, however, the companies should assume a more active self-governance role. Corporate leaders need to take greater responsibility to vindicate such core societal interests as combating harmful online content and elevating journalistic reporting and civil discourse.

This report examines what companies can do to lessen the dangers, while still preserving free speech rights. We seek to offer worthwhile recommendations, not denigrate individual companies. We recognize the ever-changing and highly complex issues that confront internet platforms when navigating this difficult terrain. Expecting to eradicate all harmful content is unrealistic. But progress is possible. We hope this paper will advance internal corporate discussions, contribute to the larger public debate, and lead to constructive action.

“

It's a new challenge for internet communities to have to deal with nation states attempting to subvert elections. But if that's what we must do, then we are committed to rising to the occasion.

— Mark Zuckerberg  
Chief Executive Officer  
Facebook

”

# Executive Summary

## Introduction: A More Proactive Approach to Governance of Digital Platforms

The internet does a lot of good for the estimated 3.5 billion people who use it today. But increasingly, harmful content contaminates the web, threatening democratic institutions and human rights around the globe.

In this white paper, the World Economic Forum Global Future Council on Human Rights focuses on two types of dangerous online content: terrorist incitement and politically motivated disinformation. Though emanating from different sources, both seek to distort the truth, discredit liberal institutions, and, in the words of the European Parliament, undermine “democratic values, human rights, and the rule of law”.

Internet companies have resisted government content regulations. They justifiably worry that many states would seek to suppress dissenting views, undermining free speech online. The danger in this regard is all too obvious. A number of governments have blocked Facebook, Twitter, and Google.

In the absence of government regulation, however, it is incumbent on the major platforms to assume a more active self-governance role. Corporate leaders should take responsibility to vindicate core societal interests, such as combating political disinformation and terrorist incitement, while elevating journalistic reporting and civil discourse.

## The Challenge: Harmful Content and Free Speech

ISIS and other terrorist groups have exploited social media in an unprecedented manner to recruit new members and encourage violence. The group’s digital propaganda has helped motivate an army of foreign fighters (once estimated to be as large as 40,000) to take up its cause in Syria and Iraq, and has inspired a range of devastating attacks by self-directed supporters overseas.

With regard to the 2016 presidential campaign, American intelligence agencies have concluded that Russian agents systematically spread false information online and in so doing, undermined “public faith in the democratic process”. “Trolls” posing as Americans used Twitter to push messages larded with made-up news and conspiracy theories. Facebook has revealed that fake accounts operated by Russians bought thousands of online advertisements on divisive social and political issues.

Confronted with the problem of harmful speech, some people invoke US Supreme Court Justice (1916 -1939) Louis Brandeis, who wrote in 1927 that “the remedy to be applied is more speech”. But the speed and scale of internet traffic has eroded the more-speech solution. Today, harmful online expression can spread so widely and quickly that rebuttal often becomes ineffectual.

But government regulation is often too blunt an instrument for dealing with terrorist incitement and political disinformation. Legislation may stifle the aspects of the internet that have made it so valuable. A law recently enacted in Germany—which heavily penalizes internet companies that fail to remove “hate speech”—may create an incentive for the platforms to err on the side of taking down massive amounts of content.



**Corporate leaders should take greater responsibility to vindicate such core societal interests as combating harmful online content and elevating journalistic reporting and civil discourse.**



## The Responsibilities of Digital Platforms: Algorithms, Advertisements, and Human Judgement

Internet platform companies assert that unlike traditional news outlets, they are neither editors nor publishers and have no practical ability to serve as arbiters of the truth. This long-standing position rests on an incorrect premise that either the platforms serve as fully responsible (and potentially liable) news editors or they make no judgements at all about pernicious content. We argue for a third way—a new paradigm for how internet platforms govern themselves.

The companies' own activities suggest that Facebook, Google, Twitter, and others can take effective action to counter the onslaught of false information and terrorist incitement, even if they

cannot be expected to develop impregnable defences. The best evidence comes from examples of what these companies are already doing:

- Google is experimenting with a new system that detects a user's interest in ISIS from search patterns. Once it has identified such an individual, the system targets them with videos that show terrorist brutality in an unflattering light.
- YouTube has toughened its stance toward videos that contain inflammatory religious or supremacist content but do not cross the line and violate company policies. Such material now comes with a warning and is not eligible for recommended status, endorsements, or user comments.
- Facebook has introduced a fact-checking function to its News Feed in some markets. Based on user reports and other signals, the company sends stories to third-party fact-checking

organizations. When fact checkers question a story, Facebook notifies users that it has been "disputed" and discourages sharing. Microsoft's Bing search engine has announced a similar function.

- In May 2017, Facebook Chief Executive Officer Mark Zuckerberg said his company would add 3,000 people to the 4,500 it already had screening for harmful videos and other posts. Four months later, in the wake of the Russian ad-buying revelations, Facebook said it would hire 1,000 more.

As these illustrations show, internet platforms are capable of taking varied steps to rid their websites of some, if not all, objectionable content. We believe they can do still more.

## Conclusions and Recommendations

<b>Enhance company governance</b>	Conduct cross-the-board internal assessments of vulnerabilities to terrorist content and political disinformation—and then act on the results.
<b>Refine the algorithms</b>	Identify new and more precise indicators of the credibility of content.
<b>Introduce more “friction”</b>	Adjust user interfaces to include warnings, notifications, and other forms of friction between suspicious content and individuals.
<b>Increase human oversight</b>	Devote a sufficient number of people to monitoring and evaluating content, while also giving users more tools to report harmful material.
<b>Reform advertising models</b>	Apply recommendations about governance, technology, and human oversight to advertising.
<b>Advance industry cooperation</b>	Share knowledge to maximize the benefits of combating problematic content.
<b>Identify government’s role</b>	Promote media literacy—a mission government can take on without threatening free speech.

“

Internet platforms  
are capable of taking  
varied steps to rid their  
websites of some,  
if not all, objectionable  
content. We believe they  
can do still more.

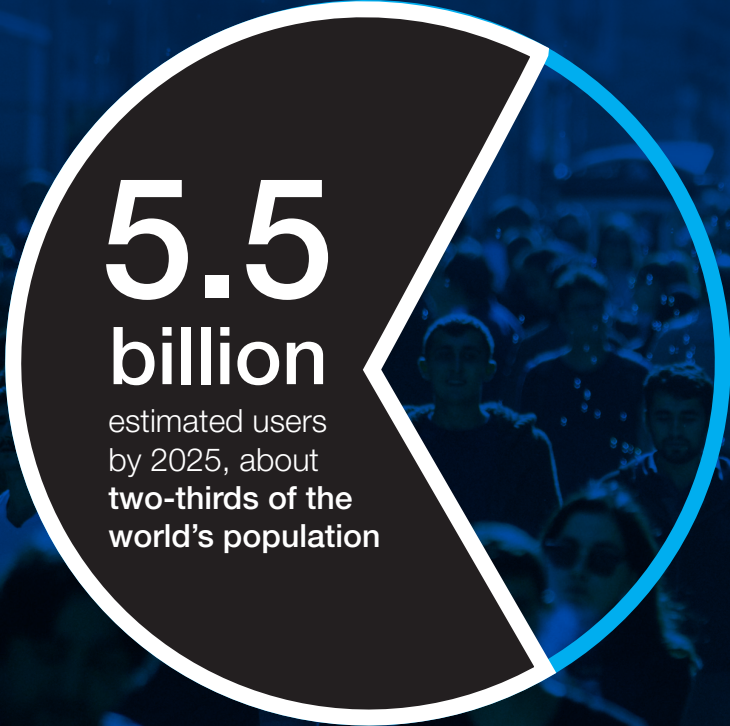
”

## Internet Users

---

**3.5 billion** consumers, business people, government officials, and others use the internet today.

---



**5.5  
billion**

estimated users  
by 2025, about  
**two-thirds** of the  
world's population



# Introduction: A More Proactive Approach to Governance of Digital Platforms

“  
Government intervention, as a general matter, is not the way to promote a free-flowing and beneficial internet.  
”

Over the last 25 years, the internet has redefined the way we gain access to information, communicate, and do business. During the 1990s, internet traffic grew by 100 percent annually.<sup>4</sup> Today, some 3.5 billion users—including consumers, businesspeople, and government officials—benefit from the internet’s enormous capacity.<sup>5</sup> By 2025 that number will likely exceed 5.5 billion, about two-thirds of the people on the planet.<sup>6</sup>

The internet has accelerated economic development across the globe, in part by enhancing access to education and promoting cross-cultural engagement. It has become a vital source of news and helps generate discussions of political and social issues. Its accessibility and global scale make it a powerful engine to advance political movements. A new generation of activists has put governments and companies on notice that, courtesy of the internet, they are subject to ever-increasing public accountability.

A handful of technology companies dominate the digital realm:

- **Google** operates the world’s largest search engine with an average of 40,000 searches per second, or 3.5 billion per day.<sup>7</sup>
- **Google** also owns **YouTube**, the largest video-sharing website online, with over 4 billion videos viewed per day.<sup>8</sup>
- **Facebook** is the world’s largest social media platform, distributing content and communication generated by two billion monthly users worldwide.<sup>9</sup>

- **Twitter** offers 328 million monthly users—among them journalists, politicians, and other opinion leaders—the ability to communicate instantly with a global audience.<sup>10</sup>
- **LinkedIn**, a subsidiary of Microsoft, is the largest professional networking site, with over 500 million users in more than 200 countries and territories around the world.<sup>11</sup>

While they have public-spirited aims, most internet platforms derive a majority of their revenue from advertising. It is a lucrative business (see table, next page). Google, Facebook, and Microsoft are among the world’s most valuable publicly traded companies.<sup>12</sup> (Microsoft sells online ads via its search engine, Bing, but most of its revenue comes from software, hardware, and services.)

## Digital Advertising by the Numbers

Company	Prominent Products	Total 2016 Ad Revenue (in US\$ billions)	Ad Revenue as % of Company Revenue	% Share of Global Digital Ad Market
Google*	<ul style="list-style-type: none"> <li>• Search Engine</li> <li>• Advertising</li> <li>• YouTube (video sharing)</li> <li>• Google Cloud (cloud software)</li> <li>• Google Play (media subscription and rental service)</li> <li>• Chrome (web browser)</li> <li>• Hardware products (Google Pixel phone and Google Assistant)</li> </ul>	79.4	89%	44%
Facebook	<ul style="list-style-type: none"> <li>• Profile</li> <li>• News Feed</li> <li>• Advertising</li> <li>• Messenger (communication service)</li> <li>• Groups</li> <li>• Events</li> <li>• Instagram (social platform)</li> <li>• WhatsApp (communication service)</li> </ul>	26.9	89%	15%
Microsoft**	<ul style="list-style-type: none"> <li>• Windows (operating system)</li> <li>• Office Suite (word processing and other software)</li> <li>• Outlook (email)</li> <li>• Skype (video and audio service)</li> <li>• Cloud services</li> <li>• Internet Explorer (web browser)</li> <li>• Bing (search engine)</li> <li>• Bing Ads</li> <li>• Xbox (video games)</li> <li>• LinkedIn (professional networking service)</li> </ul>	6.1	7%	3%
Twitter	<ul style="list-style-type: none"> <li>• Timeline (tweet stream)</li> <li>• Advertising</li> <li>• Video</li> </ul>	2.2	89%	1%

Sources: Google, Facebook, Microsoft, Twitter, Statista

\*Google's corporate parent is a holding company called Alphabet.

\*\*Microsoft's 2016 results refer to the period from July 2015 to June 2016.



While they have public-spirited aims, internet platforms are largely advertising companies that sell space next to the content they or their users generate.



The internet will continue to grow, expanding to previously under-represented regions and playing an increasingly important role in the global economy. For many purposes, it has become the public square of the 21<sup>st</sup> century, a place where people go to listen and be heard. But it is a space that can be devoted to good or ill, to edifying expression or insidious untruth. We believe that only an open internet, one that promotes free speech and transcends national boundaries, will fulfill the technology's promise.

Government intervention, as a general matter, is not the way to promote a free-flowing and beneficial internet. Official regulation poses an immediate threat to the basic human right to free speech. A number of countries around the world constrain or censor online activity. Beyond censorship, some authoritarian regimes have deployed surveillance technologies that effectively convert the internet from an instrument of empowerment and expression into a mechanism of state control.

Given the peril of heavy-handed government interference, we advocate that major internet platforms ought to exercise a kind of governance authority of their own.

By virtue of their centrality to the flows of information and communication worldwide, these companies are in a unique position—and bear a unique responsibility—to advance democratic principles and the cause of human rights. In so doing, they would live up to their ambitions to be positive forces in the world while continuing to run successful businesses.

The two categories of troubling internet content we have chosen to frame this discussion are especially salient today. ISIS and other terrorist groups have exploited social media in an unprecedented manner to recruit new members and incite violence. ISIS' digital propaganda, for instance, has helped motivate an army of foreign fighters (once estimated to be as large as 40,000<sup>13</sup>) to take up the group's cause in Syria and Iraq, and has inspired a range of devastating attacks by self-directed supporters overseas.<sup>14</sup>

Digital disinformation activities, meanwhile, have become the new weapon in international conflict, with state-sponsored false news spread through the online platforms undermining political processes across borders. In a recent resolution, the European Parliament described Russian "information operations" and calls to violence by the Islamic State of Iraq and Syria (ISIS) as similar dangers. Both seek to distort the truth, discredit liberal institutions, and undermine "democratic values, human rights, and the rule of law", the Parliament said.<sup>15</sup>

During the past year, major internet platforms have taken a number of constructive steps to address the spread of harmful content, many of which we discuss in this white paper. But these actions often focus on maintaining the companies' protection from legal liability for the content they host. The platforms assert that unlike traditional news outlets, they are neither editors nor publishers and have no practical ability to serve as "arbiters of the truth". This long-standing position rests on the incorrect premise of a binary choice: Either the platforms

serve as fully responsible (and potentially liable) news editors, or they make no judgements at all about the placement of pernicious content. Recognizing the differences among these digital platforms and the services they provide, as well as the potential costs and risks associated with taking a more proactive approach, we nonetheless argue for a third way—a new paradigm for how internet platforms govern themselves.<sup>16</sup>



**Recognizing the differences among these digital platforms and the services they provide, as well as the potential costs and risks associated with taking a more proactive approach, we nonetheless argue for a third way—a new paradigm for how internet platforms govern themselves.**



Pro-ISIS Social Media Message Estimate

---

**200,000** messages per day,  
including those generated by members of the  
group, supporters, and computer programmes

---



# The Challenge: Harmful Content and Free Speech

A wide range of potentially harmful content pollutes the internet. We focus on what can be done about terrorist incitement and politically motivated false information because of the threat that these two types of content pose to democracies—and the complexity of controlling this kind of information.

Some comparisons illuminate the challenge. At one pole, consider material depicting or advocating the sexual abuse of children. Such content is universally regarded as harmful and, in most places, illegal. Internet platforms deploy combinations of human and automated capabilities to detect and remove it from their networks. The companies share information with one another to prevent the spread of this content and cooperate with law enforcement agencies tracking child pornography.

At the other pole are controversial political perspectives. While such content may cause certain users offense or distress, most of it clearly falls under the umbrella of free speech protected by international human rights law. The legal systems of healthy democracies also recognize the value of political opinion. Accordingly, the major internet platforms have taken a permissive, even protective approach to this material, creating an invaluable space for political discourse to flourish and unpopular opinions to be expressed.

In contrast to these relatively easy cases, no consensus exists for dealing with terrorist content or politically motivated false information—or even how to define these terms precisely. They represent

classic hard cases, in connection with which platforms, governments, and civil society would benefit from stronger and more thoughtful guidance.

## Terrorist content

Propaganda by terrorist groups long predates the internet and modern online platforms. In the 1970s, organizations ranging from the Irish Republican Army to Italy's Red Brigades issued communiqués taking credit for their bloody exploits. Decades later, Al-Qaeda used cable news channels to disseminate videos of Osama bin Laden urging Muslims to contribute financially to his group and wage war against the US and Israel.<sup>17</sup> As Professor Klaus Schwab notes, "The democratic power of digital media means it can also be used by non-state actors, particularly communities with harmful intentions, to spread propaganda and to mobilize followers in favour of terrorist causes".<sup>18</sup> Today, ISIS spreads its message with unprecedented speed and on an overwhelming scale. Estimates of daily pro-ISIS social media messages—including those generated by members of the group, supporters, and computer programmes—have ranged up to 200,000 a day.<sup>19</sup>

“

No consensus exists for dealing with terrorist content or politically motivated false information—or even how to define these terms precisely. They represent classic hard cases, in connection with which platforms, governments, and civil society would benefit from stronger and more thoughtful guidance.

”

“  
**ISIS has shown an impressive understanding of how to exploit modern media and online communication ... It grasps that winning the information war is crucial to maintaining relevance, especially as it suffers losses on the ground.**  
”

Terrorist content can take many forms, including direct recruitment, incitement to commit violence, displays of terrorist activity, and support or justification of terrorist acts. What qualifies as terrorist content often depends on context. A statement about “spraying for cockroaches” has one connotation if made by a homeowner and a very different one if uttered during the Rwandan genocide by a Hutu radio announcer in reference to Tutsis.

We define terrorist content as that devoted to the advocacy of unlawful violence and intimidation, especially against civilians, in the pursuit of religious or political aims. The relationship to violence can be subtle. A religious preacher may disguise an online call to violence against innocent civilians in an abstract theological discussion such that the hidden message becomes obvious only to already radicalised followers. That would still be terrorist content. But expressions of extremism in ideology or religion that steer clear of exhortations to commit violence would fall outside of our definition.

ISIS has shown an impressive understanding of how to exploit modern media and online communication—far beyond that of any other terrorist organization. It grasps that winning the information war is crucial to maintaining relevance, especially as it suffers losses on the ground.<sup>20</sup> At times, ISIS has sought to terrify people by releasing what have been called online “snuff films”. These have included YouTube videos showing beheadings by English-speaking ISIS member Mohammed “Jihadi John” Emwazi, who himself is believed to have been killed by a US drone strike in Syria.<sup>21</sup> On other occasions, radical preachers use YouTube to reach worldwide audiences of potential terrorists. American imam Ahmad Musa Jibril is one of the most influential spiritual authorities within ISIS foreign-fighter networks, according to the London-based International Centre for the Study of Radicalisation

(ICSR).<sup>22</sup> Jibril doesn’t explicitly endorse violent jihad but “supports individual foreign fighters and justifies the Syrian conflict in highly emotive terms”, the ICSR has reported.<sup>23</sup> One of the three ISIS supporters who carried out a van-and-knife attack in London in June 2017 extensively watched YouTube videos of Jibril.<sup>24</sup>

Al-Qaeda has also used YouTube videos to recruit followers and incite violence. Online sermons by Anwar al-Awlaki, an American-born cleric who joined Al-Qaeda and was later killed in an American drone strike in Yemen, are thought to have influenced numerous terrorists, including Boston Marathon bomber Dzhokhar Tsarnaev and San Bernardino, California, mass shooter Syed Farook. Although YouTube has removed hundreds of Awlaki videos, thousands of the sermons remained on the platform as of August 2017.<sup>25</sup> ISIS, though a bitter rival of Al-Qaeda, regularly draws upon Awlaki’s YouTube image and message to appeal to young English-speaking second- and third-generation Muslims living in the West.<sup>26</sup>

## Politically motivated disinformation

Politically motivated disinformation is content that gives the appearance of being factual and news-like but actually relies on deliberate falsehoods to mislead and/or promote a political agenda. Objective falsehood is a key component of the term. Multiple internet platforms have argued strenuously that they are not editors and cannot be arbiters of truth. Indeed, there are many instances where no objective truth exists and there is room for widely divergent opinions. But there are also demonstrably incorrect statements—the claims, for example, that former US President Barack Obama was born in Kenya and that Pope Francis endorsed

the candidacy of Donald Trump. (We avoid the term “fake news” because it has been used in a range of contexts to confuse the public and serves to undermine legitimate journalism.)

Dissemination of false information is hardly a new problem. In the US presidential election of 1828, pamphlets known as “coffin handbills” accused Andrew Jackson of murder and even cannibalism.<sup>27</sup> In the 1890s, newspaper moguls Joseph Pulitzer and William Randolph Hearst pursued readers and profit with phony “yellow journalism” that helped goad the US into the Spanish-American War. Nation states have long used fake information as a tool, as evidenced by propaganda during World War II and the Cold War. More recently, countries ranging from Iran to Myanmar have engaged in international information operations.

The rise of internet platforms has provided purveyors of disinformation with new avenues of attack, diminishing their chance of detection and improving their ability to quantify impact. At the same time, a range of factors have made large audiences more vulnerable to the spread of false information. Increased political polarization in Western societies has led more people to gravitate towards news sources that reinforce their biases while dismissing the reliability of those expressing different views. And the increasing use of mobile devices means people tend to read more quickly and less carefully, making it less likely that they will critically evaluate the information they consume.

During the 2016 US political season, each of the top 20 spurious news stories about the presidential election received more engagement on Facebook than any of the 20 best-performing election news stories from major media outlets, according to an analysis by *BuzzFeed News*.<sup>28</sup> The profusion of false information had a distinct partisan slant. Seventeen out of the top 20 fraudulent stories were anti-Hillary

Clinton or pro-Donald Trump.<sup>29</sup> Coming at the topic from a different angle, Philip N. Howard, an internet scholar at Oxford University, led a team of researchers who looked at Twitter messages, or tweets, generated by internet robots, or bots, which are software applications that can automatically simulate human activity. Howard and his colleagues found that as election day approached in November 2016, bots were generating five pro-Trump tweets for every pro-Clinton tweet.<sup>30</sup>

pro-Trump messages larded with made-up news and anti-Clinton conspiracy theories. Voluminous bot activity helped drive pro-Trump tweets to the top of Twitter’s “trends” list, which heavily influences mainstream news coverage.<sup>34</sup>

Countries across Europe, from Germany to Ukraine, have faced similar threats to their democratic fabric. Over the course of the 2017 French presidential campaign, multiple “spear-phishing” attacks targeted Emmanuel Macron’s team. Intruders stole

## Digital raid on the 2017 French elections

### Multiple “spear-phishing” attacks targeted Emmanuel Macron's team, as intruders stole thousands of documents and emails, many of which were disseminated online days before voting.

In January 2017, American intelligence agencies released a declassified version of a report concluding that the Russian government had carried out “an influence campaign in 2016 aimed at the presidential election”.<sup>31</sup> The campaign’s goals, the report added, “were to undermine public faith in the US democratic process, denigrate Secretary Clinton, and harm her electability and potential presidency”.<sup>32</sup> The attack included hackers who penetrated the servers of Clinton’s campaign and the Democratic National Committee, stealing and leaking vast amounts of email and other documents. The US intelligence report also cited dubious television broadcasts by RT, an international cable network formerly known as Russia Today; online dispatches by the website Sputnik; and countless social media messages.<sup>33</sup>

The operation exploited the vast reach of leading internet companies. “Trolls” posing as Americans used Twitter to push

thousands of campaign documents and emails in one such assault in April 2017. The pilfered material was interspersed with falsified, seemingly incriminating documents and disseminated online just days before the election. During the 44-hour period prior to voting, French law bars the country’s media and the presidential candidates themselves from publishing news about the election. Consequently, the Macron campaign could not effectively refute the false allegations.<sup>35</sup>

As these examples illustrate, false information perpetuated on internet platforms delegitimizes electoral processes and threatens the human right to participate in politics.

## Freedom of speech

Having established the danger posed by certain forms of harmful content, we turn to a crucial issue to be borne in mind when trying to deal with this danger—namely, the right to free speech.

In *On Liberty*, John Stuart Mill discussed the importance of freedom of speech to liberalism and democracy. He argued that reliable facts emerge from the clash of ideas: “It is only by the collision of adverse opinions that the remainder of the truth has any chance of being supplied.”<sup>36</sup> Informed by this belief, others have maintained that an ever-contentious “marketplace of ideas” will provide a vibrant forum for political, philosophical, and scientific discourse. To fight bad speech, in the words of US Supreme Court Justice (1916-1939) Louis Brandeis, “the remedy to be applied is more speech.”<sup>37</sup> Referring to Twitter, Karen North, a social media scholar at the University of Southern California, told *The New York Times*: “When the false information is stated, people can jump on false statements and challenge [them].”<sup>38</sup>

The online platforms at times have enabled a marketplace of ideas where more speech has carried the day. But Brandeis himself said his prescription applied only “if there be time to expose through discussion the falsehood and fallacies”. Today, there often is not time. The speed and scale of internet traffic have eroded the more-speech solution. More speech does not suffice when political disinformation and calls to carry out violent attacks can be specifically targeted to susceptible audiences and insidiously designed to exploit pre-conceived biases. Today, online speech can spread so widely and quickly that rebuttal often becomes ineffectual. More speech cannot compete with messages amplified by armies of anonymous bots and paid internet commenters.

While freedom of speech is widely recognized as a core human right, international human rights law acknowledges that carefully drawn

restrictions on expression are appropriate in certain limited circumstances.<sup>39</sup> This is particularly the case when restrictions are instituted to protect other fundamental rights, such as the right to life and security of the person or the right to engage in the political process, which are clearly threatened by terrorist content and politically motivated disinformation. Any restrictions on freedom of speech must be strictly proportionate, however. One legislative approach that could fit these parameters in the US would extend political advertising disclosure rules to the online environment. In October 2017, Republican Senator John McCain joined two Democrats to introduce a bill that would require anyone paying for a digital political ad to be identified, as is already the requirement for radio and television ads. Without endorsing the McCain bill in particular, we applaud the concept of closing what amounts to a digital loophole allowing anonymous political advertising.

But beyond this narrow legislative goal, we believe that government regulation is likely to be too blunt an instrument when seeking to limit terrorist content or politically motivated disinformation. Legislation may impinge on freedom of speech and stifle the very aspects of the internet that have made it so valuable to society. Even if carefully tailored legislation were adopted by countries with strong liberal democratic traditions and protections for free speech, the measures could be cited by authoritarian countries to justify restrictions on intellectual dissent and political opposition online.

Despite the risks, European governments are moving gradually towards greater regulation of the internet. Under a law that went into effect in October 2017, Germany now requires Facebook, Twitter, and other platforms to remove “hate speech” within 24 hours after it is flagged by a user. Companies that fail to comply face fines as high as 50 million euros. Governments, of course, have an interest in ensuring the safety of their citizens online, but laws like this are not the answer.<sup>40</sup>

The German measure puts an enormous burden on the technology companies. It creates an incentive for the platforms to err on the side of taking down excessive amounts of content in an effort to avoid stiff monetary sanctions. When considering censorship of harmful content, governments must weigh the interests of all stakeholders, especially the right of their citizens to free speech.

The undesirability of government intervention does not, however, alter the real harms caused by terrorist content and politically motivated false information, nor the threat such content poses to human rights. If there are serious risks in allowing governments to regulate these issues, then some other party must act. We contend that the primary responsibility falls on the internet platforms. This is not a responsibility to replace the state or act as an organ of government. Nor are we suggesting a legal obligation that would open the companies to new liability in the courts. Rather, it is a responsibility consistent with their stated commitment to uphold human rights. And it is an opportunity to mitigate harms caused by the deliberate manipulation of platforms that have otherwise contributed so much to society.



**The rise of internet platforms has provided purveyors of disinformation with new avenues of attack, diminishing their chance of detection and improving their ability to quantify impact.**





“

To fight bad speech, in the words of Justice (1916-1939) Louis Brandeis, ‘the remedy to be applied is more speech’.

But Brandeis said his prescription applied only ‘if there be time to expose through discussion the falsehood and fallacies’.

Today, there often is not time.

The speed and scale of internet traffic have eroded the more-speech solution.

”

Facebook Takes Action

---

**1 MILLION** accounts are  
taken down every day for a range of reasons.

---



# The Responsibilities of Digital Platforms: Algorithms, Advertisements, and Human Judgement



The internet's social and search platforms don't fit either traditional category ... The platforms are neither old-fashioned editorial publishers nor purely neutral pipes.



Originally a fairly decentralized network in which most publishers operated their own servers, the internet is now increasingly made up of a limited set of massive platforms, each with global scale. These include social platforms, such as Facebook, YouTube, Twitter, Snapchat, and Instagram, which host and distribute users' writings, images, and videos; search engines, like Google and Bing, which steer users to writings, images, and videos; and communication services, such as Messenger, WhatsApp, Skype, Telegram, Signal, and WeChat. The platforms operate at an almost unimaginable scale, processing billions of posts and photos and uploading and streaming millions of hours of video every day.<sup>41</sup> Collectively, the dominant internet companies—especially the social and search platforms—constitute a new beast in the global information ecosystem.

Before the rise of these platforms, through the eras of the newspaper, telephone, radio, broadcast television, cable, and the first decade of the internet, one could divide information discovery and distribution actors into two classes: editorial publishers and neutral pipes. Publishers, acting as gatekeepers, made decisions about which speakers and what speech would reach their audiences. Pipes blindly carried information and communications without reviewing or passing judgement on its content. Accordingly, the law in most countries held publishers responsible for what was said on their pages or broadcasts, while offering immunity from liability to the pipes. If a newspaper reporter committed libel, the publisher of the paper could be held responsible, but if a pair of thieves planned a bank heist by telephone, the phone company could not be prosecuted as an accessory to the crime.

## Building better algorithms

The internet's social and search platforms don't fit either traditional category. For the most part, they engage in selecting and ranking content to present to their users via the operation of complex, constantly changing algorithms, rather than the exercise of human editorial judgement.

Generally speaking, an algorithm is a set of instructions telling a computer how to organize a body of data—in this case, how to choose one type of content and reject another. A user interface algorithm then determines how content is arranged on the screen. These mechanisms of selection and presentation pose a challenge to the long-standing legal dichotomy. The platforms are neither old-fashioned editorial publishers nor purely neutral pipes.

They do not employ people who make individualized determinations about which users see particular pieces of content. Instead, the companies operate systems that automatically amass and organize the material that users will see. Search engines “crawl” the web and present a list of sites corresponding to users' queries, prioritized according to relevance. Social networks, by contrast, tend to play a more active role in shaping an online environment of photos, videos, posts, and outside content that they hope will engage and entertain their users. Both types of platforms make systemic decisions about how to structure their search algorithms and user interfaces. Thus they have editorial control, in a very broad sense, but they generally do not exercise it on a case-by-case basis.

To understand this distinction, it helps to examine two common but misguided assertions about internet companies:

The first, often made by the platforms themselves, is that they are not responsible for the quality or veracity of what appears on their systems. “We, as a company, should not be the arbiter of truth,” Colin Crowell, a vice president at Twitter, wrote on the company’s blog in June 2017. “Twitter’s open and real-time nature is a powerful antidote to the spreading of all types of false information.”<sup>42</sup> On social networks like Facebook and Twitter, users post whatever they want, according to this view, and the platforms just show people what their friends post. Search engines like Google merely fetch lists of websites relevant to users’ queries. The companies remain neutral.

### Common misperception #1 Internet platforms are not responsible for the quality or veracity of what appears on their systems

This claim obscures reality. Consider a Google search for “history of the Holocaust”. Millions of webpages have information on the topic. Google’s algorithms analyse hundreds of factors to present a manageable search result with what the company considers the “best” and “most relevant” information displayed most prominently. According to Google, the factors include the freshness of content, the number of times search terms appear on the site, and whether the site offers “a good user experience”. To assess trustworthiness and authority, the algorithms would favour sites that many other users “seem to value for similar

false information”,<sup>45</sup> Google launched a research-and-reform initiative called Project Owl. In April 2017, the company announced that it had “improved our evaluation methods and made algorithmic updates to surface more authoritative content”. Google noted that before Project Owl, about 0.25 percent of its queries—meaning millions per day—were “returning offensive or clearly misleading content”. With the improvements, incidents “similar to the Holocaust denial results that we saw back in December [2016] are less likely to appear”, the company predicted.<sup>46</sup>

Project Owl appears to have had some effect. If one queried “did the Holocaust happen” in late September 2017, a Holocaust-denial site did not surface until the top of the seventh page of results. And Stormfront was nowhere to be found, as the web hosting firm Network Solutions had revoked its domain name.<sup>47</sup>



Google noted that before Project Owl, about 0.25% of its queries—meaning millions per day—were ‘returning offensive or clearly misleading content’.



### Humans Improving Algorithms

# 10,000

Google “raters” continually evaluate search results with an eye toward improving algorithms.

queries”, Google says. In the case of the Holocaust example, another factor would be whether World War II history websites that are themselves heavily linked to, in turn, linked to a page.<sup>43</sup>

For all their subtlety, algorithms sometimes elevate clearly false information. As of December 2016, if a Google user entered the search “did the Holocaust happen”, the very first result would have been a page from the American neo-Nazi site Stormfront, entitled “Top 10 reasons why the Holocaust didn’t happen”.<sup>44</sup> Alarmed by the Holocaust-denial result and similar incidents where its algorithms served up what it termed “blatantly misleading, low quality, offensive, or downright

Facebook operates differently. Its users do not make Google-like queries. Instead, Facebook’s News Feed algorithm determines which posts, videos, links, shares, or stories to show them. Items are posted by users’ friends, people they follow, and Facebook pages, which are profiles created for celebrities, businesses, and other organizations. Every time a typical user visits the News Feed, a vast number of potential posts await them—everything from wedding photos to political commentary to restaurant reviews to advertisements. The algorithm prioritizes several hundred items, using “signals” it gleans from the user’s past behaviour, including whether they have “liked” similar material; how often they have interacted

with the friend, page, or public figure who posted; and the number of likes, shares, and comments a post has received from their friends and the world at large.<sup>48</sup>

The News Feed algorithm also promotes Facebook's business imperatives. If the company has launched a new photo product, photos may get more priority for a time.

While algorithms select content, user interface programmes determine how content is presented on the screen. Years ago, platforms tended to supply simple vertical lists in reverse chronological order, with the most recent material on top, sometimes organized by topic. Today, posts (or search results) that are highly relevant, authoritative, and/or compelling—all as estimated by proprietary algorithms—are most likely to appear towards the top of the user's screen. Material with low relevancy or authoritativeness scores may appear further down, or not at all.

All this suggests a crucial point: Algorithms are a human construction, incorporating human judgement about the relative value of different types and sources of content. Google reportedly employs more than 10,000 human "raters" who continually evaluate search results with an eye towards improving algorithms.<sup>49</sup> Referring to Facebook's News Feed algorithm, one technology columnist wrote: "Humans decide what data goes into it, what it can do with the data, and what they want to come out at the other end. When the algorithm errs, humans are to blame. When it evolves, it's because a bunch of humans read a bunch of spreadsheets, held a bunch of meetings, ran a bunch of tests, and decided to make it better."<sup>50</sup>

Companies make choices that result in the elevation or suppression of search results and posts—those that include "click bait" headlines ("13 Travel Tips That Will Make You Feel Smart!"), link to known racist or terrorist websites, or traffic in sham news about political candidates.<sup>51</sup> In short, while internet companies do not make editorial judgements about individual search results or posts (unless users take the trouble to flag the items for review), algorithms

incorporate systematic editorial judgement about which categories and types of material ought to be made available to particular users.

## Common misperception #2 Internet platforms should be held responsible for everything they present to users.

A second claim, made by some politicians and other skeptics, is that the internet platforms should be held responsible for everything they present to users. When motivated, the companies can do extraordinary things technologically, this argument goes, so surely they can build a mechanism to detect and block false news and terrorist content.

We think this claim also misses the mark. While internet platforms have the technical ability to promote or demote categories and sources of content, there is no way they can automatically and reliably determine whether a given post constitutes pro-terrorist propaganda or asserts phony political news. The vagaries of human language and the breadth and complexity of human opinion and expression make that task, as yet, infeasible. A phrase such as "the world is flat", for example, is in one sense objectively false, but it may well appear in a post debunking fake theories or one offering a metaphor for modern capitalism, as in columnist Thomas Friedman's book of that title. Platforms can more easily deal with totally blacklisted sources—say, the white supremacist site Daily Stormer, which saw its website-hosting registration cancelled by Google and GoDaddy in the wake of the violence in Charlottesville, Virginia, in August 2017—or specifically recognizable images. But the endless variability of context and the breakneck speed of human creativity mean that automated systems cannot yet reliably determine what is or is not false or terroristic in nature.

The truth lies between the two common but wrongheaded claims: Facebook,

Twitter, and Google can take effective action to counter the advance of false information and terrorist incitement, even if they cannot be expected to develop impregnable defences. We believe they can do so in a manner that also respects the rights to free speech. A handful of promising examples show that the platforms are capable of making progress in this area.

Jigsaw, a research group within Google's parent company, Alphabet, is developing better algorithms designed specifically to identify the handiwork of unsavory online figures, including terrorists and authors of fake political reports. To understand bad actors in this context, Jigsaw employees travelled to Macedonia to meet with purveyors of fake political articles and to Iraq, where they debriefed former ISIS recruits willing to discuss pro-terrorist posts. An early product of this research is a tool known as the Redirect Method. It can detect a Google user's possible extremist sympathies based on search patterns. Once it has identified such a user, the tool redirects them to videos that show the brutality of ISIS in an unflattering light. Over the course of a recent eight-week trial run, some 300,000 people watched videos suggested to them by the Redirect Method for a total of more than half a million minutes.<sup>52</sup> Microsoft's search engine, Bing, announced in April 2017 that it would begin a pilot programme similar to Google's. Bing users looking for terrorism content now see video links to testimonials of former violent extremists, among other "counter-narratives".<sup>53</sup>

YouTube has toughened its stance towards videos that contain inflammatory religious or supremacist content but do not cross the line and violate company policies. Such material now comes with a warning and is not eligible for recommended status, endorsements, or user comments. Borderline videos are harder to find with YouTube's search function and cannot be monetized by selling advertisements next to them.<sup>54</sup>

## Foiling Unsavory Online Actors

---

# 300,000

people watched videos suggested to them by Google's Redirect Method, designed to detect a user's possible extremist sympathies based on search patterns and steer them to videos showing ISIS in an unflattering light.

---

“

In December 2016, Facebook announced the addition of a fact-checking function to its News Feed in some markets.

When third-party fact checkers question a story, Facebook notifies users that it has been 'disputed' and discourages sharing.

”

In December 2016, Facebook announced the addition of a fact-checking function to its News Feed in some markets. Based on user reports and “other signals”, the company sends stories to third-party fact-checking organizations, such as Snopes and Politifact. When fact checkers question a story, Facebook notifies users that it has been “disputed” and discourages sharing.<sup>55</sup> Bing announced more recently that it has added a similar function for major news stories and web pages. Bing alerts users to certain controversial search results for which third-party fact checkers have offered analysis, including verdicts of “true” or “false”. The alerts “allow users to have additional information to judge for themselves what information on the internet is trustworthy”, Microsoft said.<sup>56</sup>

Another approach platforms are reportedly experimenting with involves removing extremist videos by means of a technique used to identify and take down duplicates of copyrighted material. The method, known as “hashing”, entails computers calculating a digital fingerprint that allows them to link original works to copies. YouTube, Facebook, Twitter, and Microsoft are employing the approach to block violent extremist videos that reappear on their sites after having been taken down elsewhere. Commendably, the companies are also collaborating on a shared digital fingerprint database to make the process more efficient.<sup>57</sup> (This set of automated detection techniques and shared resources mirror the way the companies have tackled the problem of online imagery of child sexual abuse.)

In order to find duplicate terrorist material, platforms have to know about the original, of course. And one way to gain that awareness is decidedly low-tech: allowing, and even encouraging, concerned users to flag harmful content. User reports of offensive and dangerous material are a feature of most, if not all, internet companies. Some have shown more creativity than others. YouTube has a “Trusted Flagger” programme that prioritizes company review of the reports

of users who have been accurate more than 90% of time in the past and allows these users to point out problematic videos in bulk, rather than one at a time. Non-governmental organizations that specialize in counterterrorism participate in the programme.

As these illustrations show, the major search engines and social networks are willing and able to take varied steps to counter some, though by no means all, objectionable content. We believe they can do still more, and have an obligation to try. One step companies can take immediately is to launch comprehensive internal risk assessments of the various ways malevolent actors, including governments, are misusing their platforms to spread disinformation. These assessments, which could also address terrorist content, should help each company identify needed operational changes. Beyond fixing their own problems, the leading companies should share their findings with each other and cooperate on industry-wide responses.

That terrorists and architects of disinformation will strive to get around any defences companies put in place is no justification for surrender. This is a continuing battle but one worth fighting.

## Running better advertising businesses

Advertising is the lifeblood of the internet. Digital ads provide the revenue that allows Google, Facebook, and Twitter to offer their services to users for free. It would be misleading to discuss the responsibilities of the major platform companies without acknowledging the overwhelmingly dominant source of their revenue. Maximizing advertising dollars shapes most major decisions in the internet industry. As it happens, there is much to be learned from how the companies do and do not protect their ad-sales businesses from violent terrorism content and politically driven disinformation.

What is known as programmatic advertising now dominates the internet. The buying and selling of ads is automated, targeted to specific audiences, and occurs very rapidly. Programmatic advertising offers businesses large and small access to global audiences, with the ability to zero in on highly specific groups of consumers based on masses of demographic, behavioural, and interest-related user data compiled and segmented by the internet companies. A car company can pick out consumers whose online profiles indicate that they drive a lot and admire German engineering. A member of parliament can seek out female, politically active conservatives or liberals in her district. And until recently, Facebook's algorithms would accommodate advertisers targeting individuals who described themselves as "Jew haters".<sup>58</sup> Programmatic ad systems allow for the promotion of commerce, politics, and also repugnant views. With no human involvement at the point of sale, it is no wonder that problems occur.

The Facebook/"Jew haters" episode did not include content that falls neatly into the two categories examined in this paper, but it is still worth pausing over because it underscores how algorithms operating alone can fail. In September 2017, the non-profit journalism organization *ProPublica* reported that someone who wanted to sell Nazi memorabilia or recruit marchers for a right-wing rally could use Facebook's automated ad-buying service to find people who had expressed interest in such topics as "Jew haters", "how to burn Jews", and "history of 'why Jews ruin the world'".<sup>59</sup> It does not take much imagination to figure that an automated ad-sales system that recognized "Jew haters" might identify other objectionable categories, including, potentially, advocates of terrorism. In an earlier investigation, *ProPublica* had also found that Facebook advertisers were allowed to exclude certain "ethnic affinity" groups.<sup>60</sup>

Facebook no longer allows either the ethnic exclusions or the anti-Semitic categories. After the "Jew haters" affair, a contrite Sheryl Sandberg, the company's chief operating officer, publicly expressed regret and promised "more human review and oversight" of Facebook's automated ad programme.<sup>61</sup> Her vow followed a similar one in May 2017, when Mark Zuckerberg, chief executive of the company, said it would add 3,000 more people to the 4,500 it then had screening for harmful videos and other posts reported by users.<sup>62</sup> That the number of reviewers is now slated to rise above 7,500 is a hopeful sign—but we cannot know yet if this will be anywhere near sufficient for a platform with 2 billion users. Facebook's intertwined social and advertising systems clearly need additional human involvement where people can apply the common sense that algorithms are not always capable of exercising.

Only days before the anti-Semitic advertising revelations, Facebook made a separate disclosure that became headline news around the world. The company said it had shut down 470 phony accounts that it believes were created by the Internet Research Agency, a mysterious Russian troll farm. During the 2016 presidential election campaign, the fake Facebook accounts—with names such as "Blacktivist", "Secured Borders", and "United Muslims of America"—spent about \$100,000 to buy more than 3,000 ads on divisive social and political issues ranging from race to immigration to the influence of Islam.<sup>63</sup> The ads reached some 10 million users, but Facebook later told Congress that unpaid Russian posts may have been seen by 126 million people.<sup>64</sup> Google, for its part, acknowledged to lawmakers that Russian operatives seeking to interfere with the election posted more than 1,100 videos on YouTube.<sup>65</sup> And Twitter identified more than 2,700 accounts controlled by Russian agents and more than 36,000 bots that tweeted 1.4 million times during the election.<sup>66</sup>

“

**If it takes complaints from powerful corporate advertisers to jump-start the process, so be it. But the platforms should not merely react to crises; they need to take the initiative and move proactively.**

”

Facebook and Twitter have said that they are stepping up efforts to purge fake accounts.<sup>67</sup> On a weekly basis, Facebook said it already reviews "millions of ads around the world", using "a mix of automated and manual processes".<sup>68</sup> Each day, the social network said, it takes down an astounding 1 million accounts for a range of reasons.<sup>69</sup> Focusing specifically on the prospect of foreign interference in American elections, Zuckerberg acknowledged in a 21 September 2017 video statement: "We can make it harder. We can make it much harder, and that's what we're going to focus on doing."<sup>70</sup>

Zuckerberg said his company will make political advertising on the platform more transparent by requiring ads to identify which Facebook page has paid for them. In addition, interested people will be able to identify in a list all of the stories and ads a given page paid to promote across the site. The company will enhance oversight, he said, by adding another 250 in-house reviewers (a figure the company later raised to 1,000). Facebook also will

share more information with other technology companies and digital security firms. “It’s a new challenge for internet communities to have to deal with nation states attempting to subvert elections,” Zuckerberg added. “But if that’s what we must do, then we are committed to rising to the occasion.”

The promises from Zuckerberg and Sandberg are encouraging, as is Twitter’s late-October 2017 announcement of similar new transparency policies for political ads. At the same time, all of these assurances represent an acknowledgement that until now the companies have not done enough to screen out questionable content. “We can do better,” Zuckerberg said. Now we are urging his company and the other major platforms to embrace fully their responsibilities to protect not only users, but the democratic institutions upon which we all rely.

Another glimpse of the online ad world shows how catering more diligently to nervous corporate advertisers—real ones, in this case—could help efforts to control terrorism content and false information. In early 2017, *The Times* of London revealed that ads for major companies were turning up next to YouTube videos promoting the likes of Combat 18, a violent pro-Nazi group, and ISIS.<sup>71</sup> Over the next several months, such brands as AT&T, Johnson & Johnson, Pepsi, and Walmart suspended millions of dollars in advertising from YouTube. “The content with which we are being associated is appalling and completely against our company values,” Walmart said in a written statement. Making the whole situation even more alarming, the ad dollars in question had been flowing not only to YouTube, but also, in part, to the creators of some of the hateful videos.<sup>72</sup>

YouTube apologized for its automated ad programme juxtaposing the companies’ sales pitches with some of the most offensive offerings posted to the service. The company vowed to hire more

people to review videos and develop more sophisticated algorithms to keep mainstream ads separate from controversial content. Perhaps most important, YouTube said it intends to block more offensive videos from being posted in the first place.<sup>73</sup>

Once again, the implication relevant to this white paper is that internet platforms can—and should—do more to rid their sites of terrorist content and politically motivated false information. If it takes complaints from powerful corporate advertisers to jump-start the process, so be it. But the platforms should not merely react to crises; they need to take the initiative and move proactively.

Some players in the digital marketplace are raising questions about whether, at least in some cases, more traditional human judgement should supplant programmatic advertising. AppNexus, an ad broker that brings together buyers and sellers, has created “blacklists” of online publishers whose sites contain incitements to violence. In November 2016, AppNexus blacklisted *Breitbart News* because it breached the company’s hate speech policy.<sup>74</sup> “We did a human audit of Breitbart and determined there were enough articles and headlines that cross that line, using either coded or overt language,” an AppNexus executive said.<sup>75</sup>

AppNexus also offers clients a “whitelist” of sites deemed suitable for advertising, an approach that appeals to some large companies. “When it comes to making a judgement about a site or channel focused on fake news or hate speech, what’s needed is a human judgement, an editorial assessment,” a spokesman for the British telecommunications company Vodafone told *The Guardian*. “It’s not possible to rely on algorithms alone.”<sup>76</sup>

Human-curated lists—white or black—are not a panacea. Given how quickly new sites appear online, blacklists become out of date almost by the time they are completed. Whitelists

can endanger media diversity if they overlook smaller publishers. Still, for all their imperfections, the lists are reminders of the shortcomings of programmatic advertising and the need for direct human involvement as the platform companies seek to exclude some of the internet’s worst content.

“

**When it comes to making a judgement about a site or channel focused on fake news or hate speech, what’s needed is a human judgement, an editorial assessment. It’s not possible to rely on algorithms alone.**

— Vodafone spokesperson

”



“

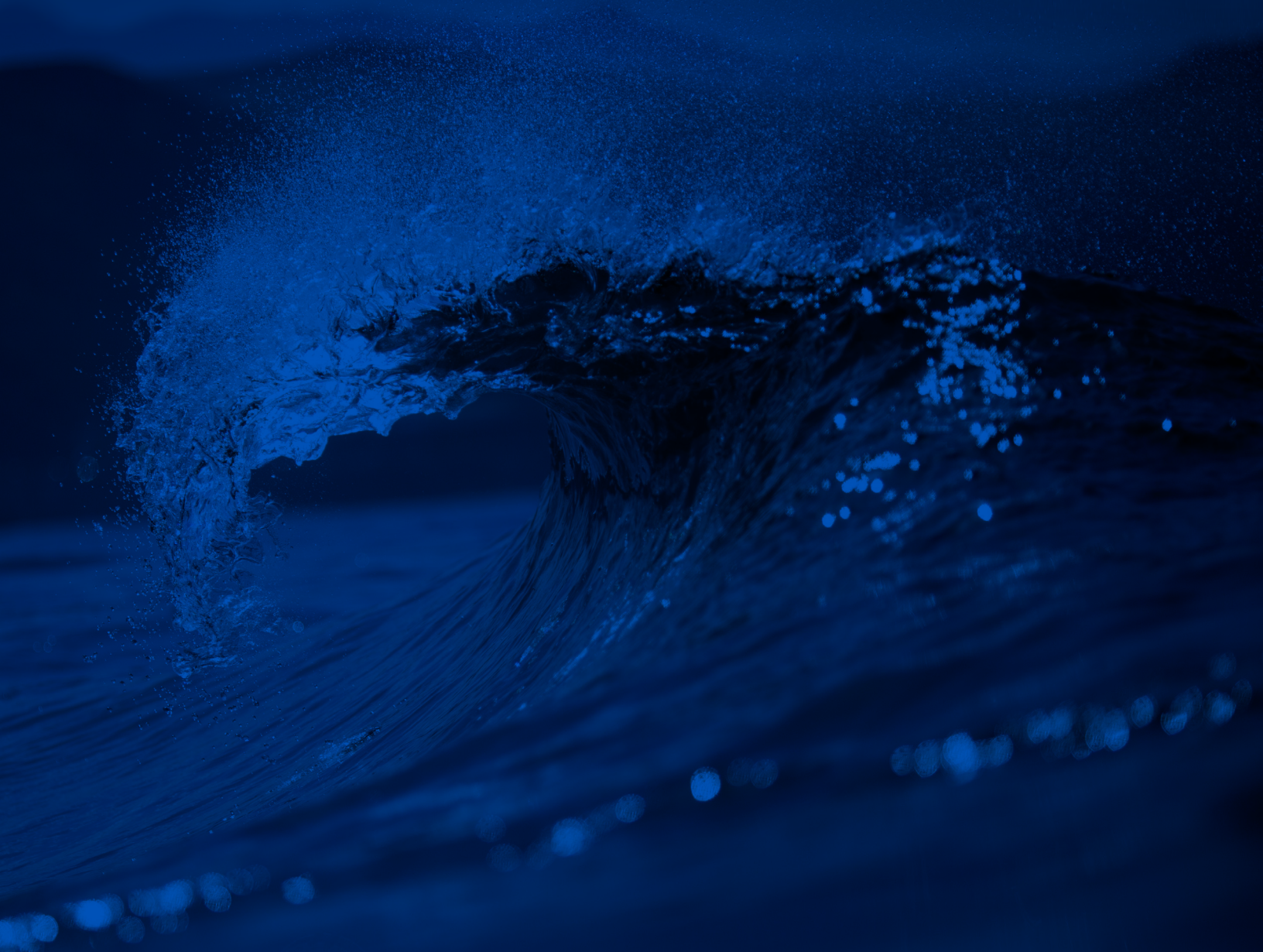
That terrorists and architects  
of disinformation will strive  
to get around any defences  
companies put in place is  
no justification for surrender.  
This is a continuing battle  
but one worth fighting.

”

---

The internet contains a vast and roiling sea of harmful content, posing a danger for those who seek to use the web for constructive purposes.

---



# Conclusions and Recommendations

“  
**Internet platforms should abandon the tired debate over whether they serve as content editors or merely as passive hosts. Their role lies somewhere in-between, difficult as it may be to define with precision.**  
”

The internet is a major driver of the Fourth Industrial Revolution. The largest online platforms are developing innovative technologies that allow users to gain information more easily, communicate with one another more efficiently, and advocate for causes more effectively. The internet also contains a vast and roiling sea of harmful content, posing a danger for those who seek to use the web for constructive purposes.

In this white paper, we have looked at the challenges created by terrorist content and politically motivated false information. We have asked how more can be done to control these categories of harmful content while still preserving the right to free speech. It is a complex and strenuous task; no one solution addresses all situations. Internet companies have begun to grapple with the challenges, but more needs to be done to preserve our democratic institutions and values.

In the broadest terms, the internet platforms must move from a reactive to a proactive approach to the problems raised here. The first step is public acknowledgement of their accountability. The companies (and their critics) should abandon the tired debate over whether they serve as content editors or merely as passive hosts. Their role lies somewhere in-between, difficult as it may be to define with precision. To understand the scope of their role, each company should undertake an internal assessment of the risks posed by the rhetoric of violent extremist

groups and politically motivated disinformation. Similar to the analysis companies undertake when entering new markets, these assessments should explore the steps available to reduce risks unilaterally, as well as possibilities for industry collaboration.

We believe that a combination of strategies is necessary for success. These include improvements in corporate governance, alterations to platform algorithms, and increased human resources dedicated to monitoring and evaluation. To make any of these steps work companies must continually experiment, conducting self-assessments of both the positive and unintended consequences.

Honest assessments—made at both the company and industry-wide levels—will help shape future action and contribute to the crucial dialogue about how companies can best meet user demands and at the same time address broader societal needs.

“  
We believe that a combination of strategies is necessary for success, including improvements in corporate governance, alterations to platform algorithms, and increased human resources dedicated to monitoring and evaluation.  
”

### 1. Enhance company governance

Given the sheer size of the largest technology companies and the number of products and services they offer, piecemeal reforms will prove inadequate. **We urge the companies to conduct across-the-board internal assessments of the threats posed by terrorist content and political disinformation. This risk analysis should call on engineering, product, sales, and public policy groups to identify problematic content, as well as the algorithmic and social pathways by which it is distributed.** The assessments, in turn, should lead to meaningful changes across each company that better protect users and society at large from the effects of the harmful content. As a further step, the fruits of this analysis should be shared with industry colleagues, much as the platforms already cooperate to eliminate child sexual abuse imagery.

### 2. Refine the algorithms

The unique algorithms employed by each internet platform take into account numberless signals and indicators to accomplish their assigned tasks. Because of this complexity, **the platform companies must continually refine these programmes to account for changing circumstances, such as the rise and proliferation of bot accounts injecting fake information into election campaigns.** Identifying new and more precise indicators of the credibility of content could significantly reduce political disinformation and terrorist content. Concurrently, companies should continue testing machine learning, or the development of algorithms that can peruse vast amounts of data and use the information to learn for themselves, without relying entirely on rules-based programming. Machine learning offers the platforms promising new capabilities to keep up with and better control harmful material online. All this should come naturally to the platforms, which are famously data-driven and in the habit of continually improving their products.

### 3. Introduce more “friction”

Companies can adjust their user interfaces to include warnings, notifications, and other forms of friction between suspicious content and individuals. A Facebook user attempting to share an article that third-party fact checkers have challenged gets a pop-up warning and is asked whether they wish to continue posting the piece. **This kind of mild but informative friction discourages a user from sharing content or from clicking on fraudulent material.** Conversely, platforms can mark verified material with a symbol similar to the account-verification check mark used to authenticate Twitter and Facebook accounts. Finally, companies can strip harmful content of engagement tools, as YouTube does when it allows borderline videos to remain on the site but prevents the content from being shared or appearing in the results of a search.

### 4. Increase human oversight

Algorithms and machines alone will never be able to identify accurately all harmful content. The vagaries of context and disguise are too many. **That means internet platforms will need to devote a significant number of people to monitoring and evaluating content.** Obviously Google cannot check every search result, but human audits help determine when algorithms have failed. (Humans are also needed to review complaints from users that content has been blocked or taken down precipitously.) The flip side of more in-house monitoring is giving users additional real-time tools to report pages, videos, and advertisements that appear to be inauthentic or harmful. Data from human alerts—raised by employees, contractors, or users—should then be used to enhance algorithms to limit similar content from being displayed in the future. YouTube’s Trusted Flagger programme provides a model by encouraging expert users to make bulk reports that are then prioritized for internal review.

## 5. Reform advertising models

Our recommendations about governance, technology, and human oversight all apply to advertising, the main commercial engine of the internet. **The Russian fake-ads scandal has already led to promises of change. Facebook vows to disclose who is paying for political ads.** The company also says it will reveal each of the ads a Facebook page is running. And the social network is expanding its advertising-review teams. Other internet platforms should follow suit. YouTube’s advertiser protest—sparked by major brands discovering their ads had been displayed next to violent extremist content—prompted the video site to promise it would improve algorithms to keep its paying customers’ ads at a distance from harmful content. Going a step further, YouTube vowed it would improve its overall advertising environment by weeding out more offensive videos altogether. This is an example of how bottom-line interests can prompt improvements that serve users and the larger society.

## 6. Advance industry cooperation

To maximize the benefits of combating problematic content, **digital platforms should share their knowledge with one another. They should also interact, where appropriate, with civil society groups, advertisers, content producers, and users.** Multistakeholder initiatives have the potential to boost accountability and build user trust. Such initiatives allow for the consideration of diverse views—surely a plus when the problem to be addressed is so tricky. Multistakeholder approaches have a long history in internet governance. ICANN, the organization that coordinates the assignment of domain names and internet protocol addresses, benefits from a hybrid open- and closed-committee system, allowing it to solicit the views of experts, governments, and users. Other examples include the Global Internet Forum to Counter Terrorism, a collaboration including major technology companies, civil society groups, and government organizations; the Global Network Initiative, which focuses on freedom of expression and privacy; and the PhotoDNA initiative, which deals with child pornography. As noted earlier, Google, Facebook, Twitter, and Microsoft are already cooperating on a common database of digital fingerprints identifying violent extremist videos. That worthy effort should be expanded. The World Economic Forum could act as an impartial platform to bring stakeholders together to address online disinformation and amplify news literacy initiatives around the world.

## 7. Identify government’s role

We have discussed our opposition in general to government regulation as a remedy for distasteful and even dangerous internet content. One exception, though, is currently under consideration in the US Congress, where lawmakers are debating whether to impose the same transparency requirements on digital political ads that already exist for television and radio advertising. **A narrow law requiring disclosure of who has paid for online political advertisements might deter foreign interference without any damage to free speech rights.** Broader attempts to outlaw harmful content, however, could encourage censorship and provide cover for authoritarian regimes. In the main, government should focus on mitigating the damage caused by malignant content. Sweden and Italy have changed their school curricula to include instruction on how to spot false news stories and critically evaluate sources. Media literacy seems like one possible mission government could take on. Counter-extremism programmes promoting voices of moderation online are another. Government need not be passive in the face of internet dangers.

“

**Broader attempts to outlaw harmful content could encourage censorship and provide cover for authoritarian regimes. In the main, government should focus on mitigating the damage caused by malignant content.**

”

# Endnotes

- 1 Klaus Schwab, *The Fourth Industrial Revolution* (Geneva: World Economic Forum, 2016), 1.
- 2 Ibid., 90.
- 3 U.N. Human Rights Council Res. 20/8, U.N. Doc. A/HRC/RES/20/8 (July 5, 2012). In 2011, the UN Human Rights Council adopted the UN Guiding Principles on Business and Human Rights, which assert that companies have a “responsibility to respect” human rights. The challenge now is to develop industry-specific substantive standards and metrics that are consistent with the Universal Declaration of Human Rights and subsequent treaties.
- 4 K.G. Coffman and A.M. Odlyzko, “The Size and Growth Rate of the Internet,” *First Monday* 3, no. 10 (October 1998), accessed September 7, 2017, <http://firstmonday.org/ojs/index.php/fm/article/view/620/541>.
- 5 “Press Release: ITU Releases 2016 ICT Figures,” *International Telecommunication Union*, July 22, 2016, <http://www.itu.int/en/mediacentre/Pages/2016-PR30.aspx>.
- 6 Cisco has projected that 5.5 billion people will use the Internet by 2025, and the United Nations has forecasted that the world’s population will reach 8.2 billion by the same year. See “The Evolving Internet: Driving Forces, Uncertainties, and Four Scenarios to 2025,” Cisco, last modified 2010, [https://newsroom.cisco.com/dlls/2010/ekits/Evolving\\_Internet\\_GBN\\_Cisco\\_2010\\_Aug.pdf](https://newsroom.cisco.com/dlls/2010/ekits/Evolving_Internet_GBN_Cisco_2010_Aug.pdf); “World Population to Increase by One Billion by 2025,” *United Nations Population Fund*, June 13, 2013, <http://www.unfpa.org/news/world-population-increase-one-billion-2025>.
- 7 “Google Search Statistics,” *Internet Live Stats*, accessed September 8, 2017, <http://www.internetlivestats.com/google-search-statistics/>.
- 8 Sarah Perez, “YouTube Reaches 4 Billion Views Per Day,” *TechCrunch*, January 23, 2012, <https://techcrunch.com/2012/01/23/youtube-reaches-4-billion-views-per-day/>.
- 9 Josh Constone, “Facebook Now Has 2 Billion Monthly Users... And Responsibility,” *TechCrunch*, June 27, 2017, <https://techcrunch.com/2017/06/27/facebook-2-billion-users/>.
- 10 Alex D’Angelo, “Twitter Reports \$116-Million Loss and Flat User Growth; Its Stock Drops 14%,” *Los Angeles Times*, July 27, 2017, <http://www.latimes.com/business/technology/la-fi-tn-twitter-earnings-20170727-story.html>.
- 11 Barb Darrow, “LinkedIn Claims Half a Billion Users,” *Fortune*, April 24, 2017, <http://fortune.com/2017/04/24/linkedin-users/>.
- 12 As of September 11, 2017 market close. See “Technology Companies,” *Nasdaq*, <http://www.nasdaq.com/screening/companies-by-industry.aspx?industry=Technology&sort-name=marketcap&sorttype=1>. Google trades under the name of its parent company, Alphabet.
- 13 Jonathan Landay, “Many Foreign Fighters Likely to Stay in Syria, Iraq: US Official,” *Reuters*, July 21, 2017, <https://www.reuters.com/article/us-mideast-crisis-usa-baghdadi/many-foreign-fighters-likely-to-stay-in-syria-iraq-u-s-official-idUSKBN1A61ZJ>
- 14 Patrick Wintour, “Islamic State fighters returning to UK ‘pose huge challenge,’” *The Guardian*, March 9, 2017, <https://www.theguardian.com/uk-news/2017/mar/09/islamic-state-fighters-returning-to-uk-pose-huge-challenge>.
- 15 European Parliament Resolution on EU Strategic Communication to Counteract Propaganda Against It by Third Parties, November 23, 2016, accessed September 28, 2017, <http://www.refworld.org/docid/584abdf24.html>.
- 16 By urging that internet platforms take steps to screen out certain harmful content, we are not arguing that the companies should be exposed to new forms of legal liability. In the US, under Section 230 of the Communications Decency Act of 1996, the companies enjoy immunity from civil liability for most of the content they publish.
- 17 Andrew Sanders, *Inside the IRA: Dissident Republicans and the War for Legitimacy* (Edinburgh: Edinburgh University Press, 2011), 120-122; Associated Press, “Al-Qaida Tapes Often Come Through Al-Jazeera,” *NBC News*, January 20, 2006, [http://www.nbcnews.com/id/10948626/ns/world\\_news-terrorism/t/al-qaida-tapes-often-come-through-al-jazeera/#.WbJD-dOGPGI](http://www.nbcnews.com/id/10948626/ns/world_news-terrorism/t/al-qaida-tapes-often-come-through-al-jazeera/#.WbJD-dOGPGI); “Video Shows Bin Laden Urging Muslims To Prepare For Fighting,” *CNN*, June 21, 2001, <http://www.cnn.com/2001/WORLD/europe/06/21/video.binladen>.
- 18 Schwab, *The Fourth Industrial Revolution*, 95.
- 19 Jon Greenberg, “Does the Islamic State Post 90,000 Social Media Messages Each Day?” *Politifact*, February 19, 2015, <http://www.politifact.com/punditfact/statements/2015/feb/19/hillary-mann-leverett/cnn-expert-islamic-state-posts-90000-social-media/>.
- 20 Charlie Winter, “ISIS Is Using the Media Against Itself,” *The Atlantic*, March 23, 2016, <https://www.theatlantic.com/international/archive/2016/03/isis-propaganda-brussels/475002/>.
- 21 Missy Ryan, Griff Witte, and Adam Goldman, “US Strike Believed to Have Killed ‘Jihadi John,’” Islamic State Executioner,” *The Washington Post*, November 13, 2015.
- 22 Joseph A. Carter, Shiraz Maher, and Peter Neumann, “#Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks,” *The International Centre for the Study of Radicalization and Political Violence (ICSR)*, April 2014, <http://icsr.info/wp-content/uploads/2014/04/ICSR-Report-Greenbirds-Measuring-Importance-and-Influence-in-Syrian-Foreign-Fighter-Networks.pdf>.
- 23 Ibid.: 2, 23.
- 24 Robert Booth, Ian Cobain, Vikram Dodd, Matthew Taylor and Lisa O’Carroll, “London Bridge Attacker Named As Khuram Butt,” *The Guardian*, June 5, 2017, <https://www.theguardian.com/uk-news/2017/jun/05/london-bridge-attacker-named-as-khuram-butt>.
- 25 “Anwar al-Awlaki — Part III: Anwar al-Awlaki Online,” *Counter Extremism Project*, August 2017, [https://www.counterextremism.com/sites/default/themes/bricktheme/pdfs/Anwar\\_al-Awlaki\\_on\\_YouTube.pdf](https://www.counterextremism.com/sites/default/themes/bricktheme/pdfs/Anwar_al-Awlaki_on_YouTube.pdf).
- 26 Harold J. Ingram and Craig Whiteside, “The Yemen Raid and the Ghost of Anwar al-Awlaki,” *The Atlantic*, February 9, 2017, <https://www.theatlantic.com/international/archive/2017/02/yemen-raid-trump-awlaki-al-qaeda-isis/516180/>.
- 27 Rebecca Onion, “The ‘Coffin Handbill’ Andrew Jackson’s Enemies Used to Circulate Word of His ‘Bloody Deeds,’” *Slate*, March 5, 2014, [http://www.slate.com/blogs/the\\_vault/2014/03/05/andrew\\_jackson\\_the\\_coffin\\_handbill\\_distributed\\_by\\_opponents\\_in\\_the\\_1828.html](http://www.slate.com/blogs/the_vault/2014/03/05/andrew_jackson_the_coffin_handbill_distributed_by_opponents_in_the_1828.html).
- 28 Craig Silverman, “This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook,” *BuzzFeed News*, November 16, 2016, [https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm\\_term=.ykd68mGqb#.bg0DG6L58](https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.ykd68mGqb#.bg0DG6L58).
- 29 Ibid.
- 30 Bence Kollanyi, Philip N. Howard, and Samuel C. Woolley, “Bots and Automation Over Twitter During the US Election,” *Data Memo 2016.4* (Oxford, UK: Project on Computational Propaganda), accessed September 8, 2017, <http://comprop.oi.ox.ac.uk/wp-content/uploads/sites/89/2016/11/Data-Memo-US-Election.pdf>.
- 31 US Office of the Director of National Intelligence, *Assessing Russian Activities and Intentions in Recent US Elections*, January 6, 2017, accessed July 18, 2017, [https://www.dni.gov/files/documents/ICA\\_2017\\_01.pdf](https://www.dni.gov/files/documents/ICA_2017_01.pdf).
- 32 Ibid.
- 33 Jim Rutenberg, “RT, Sputnik and Russia’s New Theory of War,” *The New York Times Magazine*, September 13, 2017, [https://www.nytimes.com/2017/09/13/magazine/rt-sputnik-and-russias-new-theory-of-war.html?\\_r=0](https://www.nytimes.com/2017/09/13/magazine/rt-sputnik-and-russias-new-theory-of-war.html?_r=0).
- 34 Evan Osnos, David Rennick, and Joshua Yaffa, “Trump, Putin, and the New Cold War,” *The New Yorker*, March 6, 2017, <https://www.newyorker.com/magazine/2017/03/06/trump-putin-and-the-new-cold-war>.
- 35 Eric Auchard and Bate Felix, “French Candidate Macron Claims Massive Hack As Emails Leaked,” *Reuters*, May 5, 2017, <http://www.reuters.com/article/us-france-election-macron-leaks/french-candidate-macron-claims-massive-hack-as-emails-leaked-idUSKB-N1812AZ>. The New York-based cyber intelligence firm Flashpoint said its review indicated that APT 28, a hacking group linked to the Russian military intelligence directorate, was behind the French episode. But Guillaume Poupard, the director general of the French cyber-security agency, has said that no conclusive evidence has been found tracing the attack back to APT 28. See “The Latest: France Says No Trace of Russian Hacking Macron,” *Associated Press*, June 1, 2017, <https://www.apnews.com/fc570e4b400f4c7db3b0d739e9dc5d4d>.
- 36 John Stuart Mill, *On Liberty*, 2nd ed., (London: John W. Parker and Son, 1857), 95.
- 37 *Whitney v. California*, 274 US 357 (1927), accessed September 8, 2017, <https://www.law.cornell.edu/supremecourt/text/274/357>.
- 38 Daisuke Wakabayashi and Scott Shane, “Twitter Seen as Key Battlefield In Russian Influence Campaign,” *The New York Times*, September 27, 2017, <https://www.nytimes.com/2017/09/27/technology/twitter-russia-election.html>.
- 39 See, for instance, Article 19 of the International Convention on Civil and Political Rights, Article 10 of the European Convention on Human Rights, and Article 13 of the American Convention on Human Rights.
- 40 Tara Wadhwa and Gabriel Ng, “Tech Companies Policing the Web Will Do More Harm Than Good,” *Wired*, July 31, 2017, <https://www.wired.com/story/tech-companies-policing-the-web-will-do-more-harm-than-good/>.
- 41 Jeff Desjardins, “Here’s How Much Activity Happens In Just One Minute On the Internet,” *Business Insider*, August 3, 2017, <http://www.businessinsider.com/how-much-activity-happens-in-just-one-minute-on-the-internet-2017-8>.
- 42 Colin Crowell, “Our Approach to Bots & Misinformation,” *Twitter* (blog), June 14, 2017, [https://blog.twitter.com/official/en\\_us/topics/company/2017/Our-Approach-Bots-Misinformation.html](https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html). See also Mathew Ingram, “Facebook Changes Its News Feed Algorithm (And Its Control Over Publishers),” *Fortune*, April 22, 2015, <http://fortune.com/2015/04/22/facebook-newsfeed-algorithm-publishers/>.
- 43 “How Search Works,” *Google*, accessed September 28, 2017, <https://www.google.com/search/howsearchworks/>.
- 44 Jeff John Roberts, “A Top Google Result for the Holocaust Is Now a White Supremacist Site,” *Fortune*, December 12, 2016, <http://fortune.com/2016/12/12/google-holocaust/>.

- 45 Ben Gomes, "Our Latest Quality Improvements for Search," *Google* (blog), April 25, 2017, <https://blog.google/products/search/our-latest-quality-improvements-search/>;
- 46 Ibid. Daisuke Wakabayashi, "Google as Traffic Cop," *The New York Times*, September 26, 2017, <https://www.nytimes.com/2017/09/26/technology/google-search-bias-claims.html>.
- 47 Associated Press, "World's Oldest Neo-Nazi Website Stormfront Shut Down," *The Telegraph*, August 29, 2017, <http://www.telegraph.co.uk/technology/2017/08/29/worlds-oldest-neo-nazi-website-stormfront-shut/>.
- 48 Lars Backstrom, "News Feed FYI: A Window Into News Feed," *Facebook Business*, August 6, 2013, <https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed>; Will Oremus, "Who Controls Your Facebook Feed," *Slate*, January 3, 2016, [http://www.slate.com/articles/technology/cover\\_story/2016/01/how\\_facebook\\_s\\_news\\_feed\\_algorithm\\_works.html](http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html).
- 49 Wakabayashi, "Google as Traffic Cop."
- 50 Oremus, "Who Controls Your Facebook Feed."
- 51 Steve Rayson, "We Analyzed 100 Million Headlines. Here's What We Learned (New Research)," *Buzzsumo*, June 26, 2017, accessed September 27, 2017, <http://buzzsumo.com/blog/most-shared-headlines-study/>.
- 52 Emily Dreyfuss, "Hacking Online Hate Means Talking to the Humans Behind It," *Wired*, June 7, 2017, <https://www.wired.com/2017/06/hacking-online-hate-means-talking-humans-behind/>.
- 53 "Microsoft Partners with Institute for Strategic Dialogue and NGOs to Discourage Online Radicalization to Violence," (blog), <https://blogs.microsoft.com/on-the-issues/2017/04/18/microsoft-partners-institute-strategic-dialogue-ngos-discourage-online-radicalization-violence/>
- 54 Kent Walker, "Four Ways Google Will Help to Tackle Extremism," *Financial Times*, June 18, 2017, <https://www.ft.com/content/ac7ef18c-52bb-11e7-a1f2-dc19572361bb>.
- 55 Adam Mosseri, "News Feed FYI: Addressing Hoaxes and Fake News," *Facebook Newsroom* (blog), December 15, 2016, <https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news>. In October 2017, Facebook announced it was testing a new "I" button on articles posted on its News Feed that allows users to jump conveniently to Wikipedia and access other information sources about publishers. See Andrew Anker, Sara Su, and Jeff Smith, "News Feed FYI: New Test to Provide Context About Articles," *Facebook Newsroom* (blog), October 5, 2017, <https://newsroom.fb.com/news/2017/10/news-feed-fyi-new-test-to-provide-context-about-articles/>.
- 56 "Bing Adds Fact Check Label in SERP to Support the ClaimReview Markup," *Bing* (blog), September 14, 2017, <https://blogs.bing.com/Webmaster-Blog/September-2017/Bing-adds-Fact-Check-label-in-SERP-to-support-the-ClaimReview-markup>.
- 57 Olivia Solon, "Facebook, Twitter, Google, and Microsoft Team Up to Tackle Extremist Content," *The Guardian*, December 5, 2016, <https://www.theguardian.com/technology/2016/dec/05/facebook-twitter-google-microsoft-terrorist-extremist-content>. Targeting objectionable content can produce unintended consequences. Activists have said that, as YouTube reins in certain violent imagery, it has taken down evidence of human rights violations—for example, amateur video recordings of the civil war in Syria. Working with YouTube, activists have had some of the videos of Syria restored, but the danger of overly broad takedowns persists. This example also illustrates the importance of platforms having formalized appeal processes by which users can object to the removal of content. See Sara Ashley O'Brien, "YouTube and Syria: Tech's Role as Archivist," *CNN*, August 24, 2017.
- 58 Julia Angwin, Madeleine Varner, and Ariana Tobin, "Facebook Enabled Advertisers to Reach 'Jew Haters,'" *ProPublica*, September 14, 2017, <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>.
- 59 Ibid.
- 60 Julia Angwin and Terry Parris Jr., "Facebook Lets Advertisers Exclude Users by Race," *ProPublica*, October 28, 2016, <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>.
- 61 Sapna Maheshwari and Mike Isaac, "Facebook, After 'Fail,' Over Ads Targeting Racists, Makes Changes," *The New York Times*, September 20, 2017, [https://www.nytimes.com/2017/09/20/business/media/facebook-racist-ads.html?\\_r=0](https://www.nytimes.com/2017/09/20/business/media/facebook-racist-ads.html?_r=0).
- 62 Ingrid Lunden, "Facebook to Add 3,000 to Team Reviewing Posts with Hate Speech, Crimes, and Other Harming Posts," *TechCrunch*, May 3, 2017, <https://techcrunch.com/2017/05/03/facebook-to-hire-3000-to-review-posts-with-hate-speech-crimes-and-other-harming-posts/>.
- 63 Alex Stamos, "An Update on Information Operations on Facebook," *Facebook Newsroom* (blog), September 6, 2017, <https://newsroom.fb.com/news/2017/09/information-operations-update/>; Nicholas Confessore and Daisuke Wakabayashi, "How Russia Harvested American Rage to Reshape US Politics," *The New York Times*, October 9, 2017, <https://www.nytimes.com/2017/10/09/technology/russia-election-facebook-ads-rage.html>.
- 64 Craig Timberg and Elizabeth Dwoskin, "Russian Content on Facebook, Google, and Twitter Reached Far More Users than Companies First Disclosed, Congressional Testimony Says," *The Washington Post*, October 30, 2017, [https://www.washingtonpost.com/business/technology/2017/10/30/4509587e-bd84-11e7-97d9-bdab5a0ab381\\_story.html?utm\\_term=.f527987ff094](https://www.washingtonpost.com/business/technology/2017/10/30/4509587e-bd84-11e7-97d9-bdab5a0ab381_story.html?utm_term=.f527987ff094).
- 65 Ibid.
- 66 Ibid.
- 67 Cecilia Kang, Nicholas Fandos, and Mike Isaac, "Internet Giants Pique Senators With Restraint," *The New York Times*, November 1, 2017, [https://www.nytimes.com/2017/10/31/us/politics/facebook-twitter-google-hearings-congress.html?\\_r=0](https://www.nytimes.com/2017/10/31/us/politics/facebook-twitter-google-hearings-congress.html?_r=0)
- 68 Kevin Roose, "We Asked Facebook 12 Questions About the Election, and Got 5 Answers," *The New York Times*, October 11, 2017, <https://www.nytimes.com/2017/10/11/technology/facebook-election.html>.
- 69 John Shinal, "Facebook Shuts Down 1 Million Accounts Per Day But Can't Stop All 'Threat Actors,' Security Chief Says," *CNBC*, August 24, 2017, <https://www.cnn.com/2017/08/24/facebook-removes-1-million-accounts-every-day-security-chief-says.html>.
- 70 "Facebook CEO Mark Zuckerberg Discusses 'Next Steps In Protecting Election Integrity,'" YouTube video, 7:23, posted by *ABC News*, September 21, 2017, <https://www.youtube.com/watch?v=dAay3FribnE>.
- 71 Alexi Mostrous, "Google Faces Questions Over Videos on YouTube," *The Times*, February 9, 2017, <https://www.thetimes.co.uk/article/google-faces-questions-over-videos-on-youtube-3km257v8d>.
- 72 Olivia Solon, "Google's Bad Week: YouTube Loses Millions as Advertising Row Reaches US," *The Guardian*, March 25, 2017, <https://www.theguardian.com/technology/2017/mar/25/google-youtube-advertising-extremist-content-att-verizon>.
- 73 Michael Liedtke, "Starbucks, Pepsi, Walmart Pull YouTube Ads After They Were Placed on Racist Videos," *Chicago Tribune*, March 24, 2017, <http://www.chicagotribune.com/bluesky/technology/ct-google-youtube-ad-boycott-20170324-story.html>.
- 74 Nick Statt, "Advertising Company AppNexus Bans Breitbart Over Hate Speech," *The Verge*, November 22, 2016, <https://www.theverge.com/2016/11/22/13719510/appnexus-breitbart-news-ad-ban-hate-speech-steve-bannon>.
- 75 Mark Bergen, "Major Advertising Technology Company Bars Breitbart News for Hate Speech," *Bloomberg*, November 22, 2016, <https://www.bloomberg.com/news/articles/2016-11-22/major-advertising-technology-company-bars-breitbart-news-for-hate-speech>. Technology companies Moat and Storyful have joined forces to create a database of suspect web domains and video URLs that will allow advertisers to avoid juxtaposing their brands with "fake or extremist" content. <https://newscorp.com/2017/05/02/storyful-and-moat-launch-initiative-to-combat-fake-news/>.
- 76 Mark Sweney, "Vodafone to Stop Its Ads Appearing on Fake News and Hate Speech Sites," *The Guardian*, June 6, 2017, <https://www.theguardian.com/business/2017/jun/06/vodafone-ads-fake-news-hate-speech-google-facebook-advertising>.

## Acknowledgements

The NYU Stern Center for Business and Human Rights acknowledges the diligent work on this report of the following staff members and fellows of the Center: Paul M. Barrett, Tara Wadhwa, Gabriel Ng, and Zoe Rubin.

NYU Stern Center for Business and Human Rights  
Leonard N. Stern School of Business  
44 West 4th Street, Suite 800  
New York, NY 10012  
+1 212-998-0722  
[bhr@stern.nyu.edu](mailto:bhr@stern.nyu.edu)  
[bhr.stern.nyu.edu](http://bhr.stern.nyu.edu)



Center for Business  
and Human Rights