Centre for International Governance Innovation

Stanford | Global Digital Policy Incubator

Governance Innovation for a Connected World

Protecting Free Expression, Diversity and Civic Engagement in the Global Digital Ecosystem

SPECIAL REPORT

Governance Innovation for a Connected World

Protecting Free Expression, Diversity and Civic Engagement in the Global Digital Ecosystem

SPECIAL REPORT

Edited by Eileen Donahoe and Fen Osler Hampson

Centre for International Governance Innovation

Stanford | Global Digital Policy Incubator

CIGI Masthead

Executive

President Rohinton P. Medhora
Deputy Director, International Intellectual Property Law and Innovation Bassem Awad
Chief Financial Officer and Director of Operations Shelley Boettger
Director of the Global Economy Program Robert Fay
Director of the International Law Research Program Oonagh Fitzgerald
Director of the Global Security & Politics Program Fen Osler Hampson
Director of Human Resources Laura Kacur
Deputy Director, International Environmental Law Silvia Maciunas
Deputy Director, International Economic Law Hugo Perezcano Díaz
Director, Evaluation and Partnerships Erica Shaw
Managing Director and General Counsel Aaron Shull
Director of Communications and Digital Media Spencer Tripp

Publications

Publisher Carol Bonnett
Senior Publications Editor Jennifer Goyder
Publications Editor Susan Bubak
Publications Editor Patricia Holmes
Publications Editor Nicole Langlois
Publications Editor Lynn Schellenberg
Graphic Designer Melodie Wakefield

For publications enquiries, please contact publications@cigionline.org.

Communications

For media enquiries, please contact communications@cigionline.org.

@cigionline

GDPi Masthead

GDPi Senior Leadership

Executive Director Eileen Donahoe
Principal Investigator Larry Diamond

Research and Program Staff

Associate Director for Research Megan Metzger
Associate Director for Programs Jan Rydzak
Research Associate and Project Leader, Human Rights Roya Pakzad
Program Associate Sarahi Zaldumbide

For media enquiries, please contact Stanford_GDPI@stanford.edu.

@stanford_GDPi

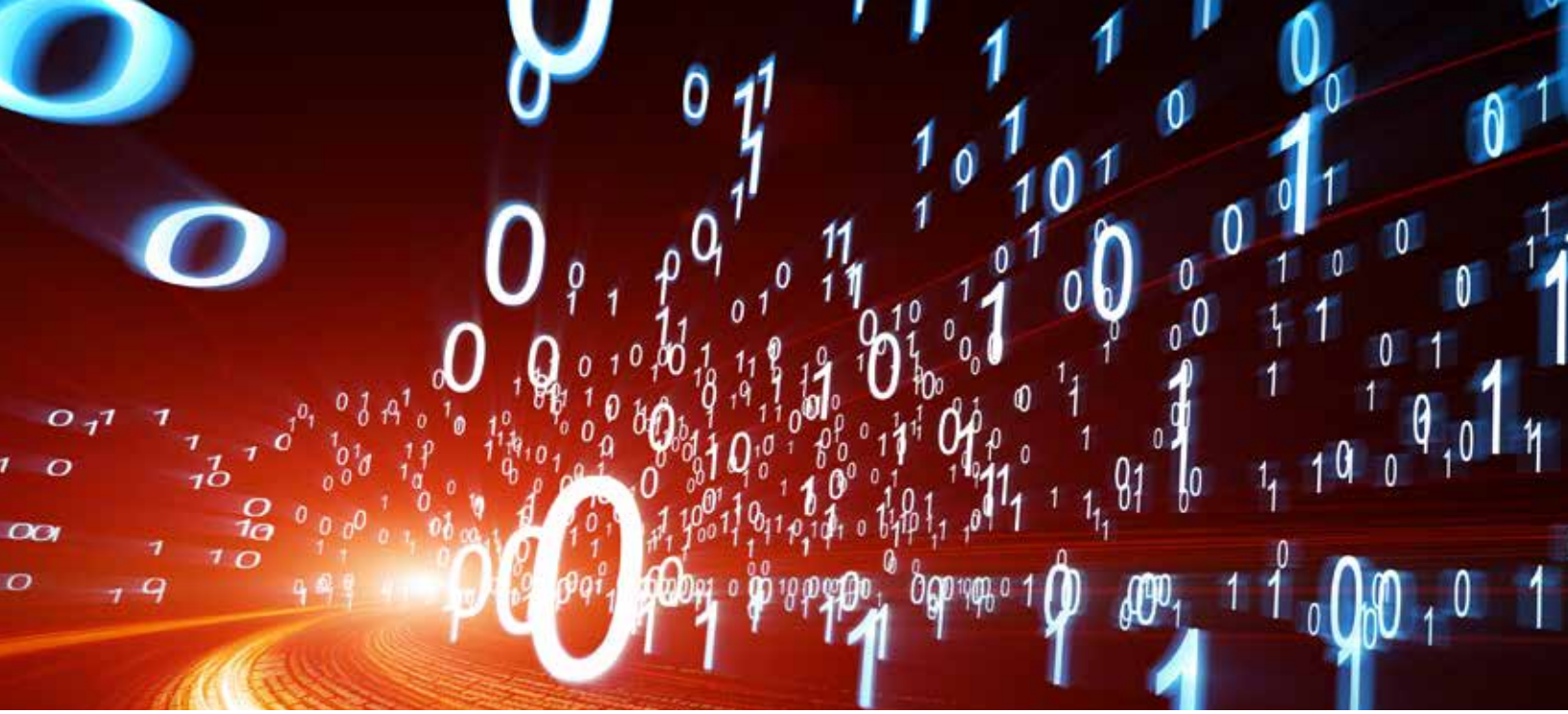
Printed in Canada on paper containing 100% post-consumer fibre and certified by the Forest Stewardship Council® and the Sustainable Forestry Initiative.

Centre for International Governance Innovation and CIGI are registered trademarks.

Centre for International Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Contents



Acronyms and Abbreviations

AfD	Alternative for Germany
API	application program interface
CEO	chief executive officer
CRTC	Canadian Radio-television and Telecommunications Commission
ECHR	European Convention on Human Rights
FCC	Federal Communications Commission
FTC	Federal Trade Commission
GCIG	Global Commission on Internet Governance
GDPi	Global Digital Policy Incubator
GDPR	General Data Protection Regulation
GNI	Global Network Initiative
IANA	Internet Assigned Numbers Authority
ICANN	Internet Corporation for Assigned Names and Numbers
ICCPR	International Covenant on Civil and Political Rights
IETF	Internet Engineering Task Force
IGF	Internet Governance Forum
ISOC	Internet Society
ISP	internet service provider
NetzDG	German Network Enforcement Act
NTIA	National Telecommunications and Information Administration
OECD	Organisation for Economic Co-operation and Development
OSCE	Organization for Security and Co-operation in Europe
RDR	Ranking Digital Rights
TCP/IP	transmission control protocol/ internet protocol
UNESCO	United Nations Educational, Scientific and Cultural Organization
VPN	virtual private network

Introduction

Eileen Donahoe and Fen Osler Hampson

The emergence of a global digital ecosystem has been a boon for global communication and the democratization of the means of distributing information. The internet, and the social media platforms and web applications running on it, have been used to mobilize pro-democracy protests and give members of marginalized communities a chance to share their voices with the world.

However, more recently, we have also seen this technology used to spread propaganda and misinformation, interfere in election campaigns, expose individuals to harassment and abuse, and stir up confusion, animosity and sometimes violence in societies. Even seemingly innocuous digital technologies, such as ranking algorithms on entertainment websites, can have the effect of stifling diversity by failing to reliably promote content from underrepresented groups. At times, it can seem as if technologies that were intended to help people learn and communicate have been irreparably corrupted.

It is easy to say that governments should step in to control this space and prevent further harms, but part of what helped the internet grow and thrive was its lack of heavy regulation, which encouraged openness and innovation. However, the absence of oversight has allowed dysfunction to spread, as malign actors manipulate digital technology for their own ends without fear of the consequences. It has also allowed unprecedented power to be concentrated in the hands of private technology companies, and these giants to act as de facto regulators with little meaningful accountability.

So, who should be in charge of reversing the troubling developments in our global digital spaces? And what, if anything, can be done to let society keep reaping the benefits of these technologies, while protecting it against the risks?

These questions were at the root of an international working meeting organized by the Global Digital Policy Incubator (GDPi) at Stanford University and the Centre for International Governance Innovation (CIGI), in cooperation with the Department of Canadian Heritage. "Governance Innovation for a Connected World: Protecting Free Expression, Diversity and Civic Engagement in the Global Digital Ecosystem" was held at Stanford in March 2018. It brought together representatives from government, academia, technology companies and civil society to discuss potential governance options to deal with these complex global challenges. The participants discussed governance policies rooted in private sector, government and multi-stakeholder approaches.

The essays that comprise this special report helped inform the conversations that took place.

Suzanne Nossel and Viktorya Vilk of PEN America delve further into the topic of freedom of expression in their essay, "Protecting Free Expression, Access to Diverse Information and Democratic Engagement Online." The very nature of the internet — which enables communications to move farther, faster and at greater volumes — has "profound implications for free expression and open discourse," they write. It has opened the door to new forms of state surveillance and censorship, while also giving unprecedented power and influence over how people communicate to a handful of private tech companies. Nossel and Vilk also caution that any stakeholders attempting to address harmful online behaviour through policy need to tread carefully, because interventions may have unintended consequences: "Efforts to address one impairment to free expression, such as the spread of online trolling, can open the door to other forms of infringement, including the policing of speech based on ideology and viewpoint."

Two specific government policies aimed at tackling harmful online content are evaluated through the lens of international human rights law by Evelyn Mary Aswad, in her essay, "Are Recent Governmental Initiatives to Combat Online Hate Speech, Extremism and Fraudulent News Consistent with the International Human Rights Law Regime?" Aswad examines the European Code of Conduct on Countering Illegal Hate Speech Online and the German Network Enforcement Act (known as NetzDG) to see how closely they conform to the provisions of the International Covenant on Civil and Political Rights. Article 19 of the covenant allows states to limit speech only when the limitations are provided by law; use the least intrusive means necessary; and achieve a legitimate objective, such as protecting public order or national security. Aswad finds that both the Code of Conduct and NetzDG leave much to be desired in adhering to the standards set out by article 19. What is needed instead, she writes, is "an open, thorough and ongoing dialogue…among governments, civil society, international organizations and companies to ascertain the nature/scope of the underlying problems that governments are trying to address and to assess properly the range of potential solutions short of broad governmental speech bans enforced by private companies."

The private sector's role in preserving a healthy and diverse online environment is discussed by Rebecca MacKinnon and Roya Pakzad in their essay, "Private Sector Roles and Responsibilities: Protecting Quality of Discourse, Diversity of Content and Civic Engagement on Digital Platforms and Social Media." MacKinnon and Pakzad note that the major tech companies have already taken steps to curb harmful speech and misinformation on their platforms, such as improving content moderation; using automation and machine learning to detect fake accounts and violent content; and partnering with independent fact-checkers. However, a number of gaps remain, including insufficient information about the volume and the nature of the content they remove for violating their terms of service; a lack of transparency around how they use and share information about their users; and inadequate grievance and remedy mechanisms for people who feel their content was unfairly censored. Addressing these gaps would be a good starting place to develop more transparency and accountability for tech companies, MacKinnon and Pakzad write. Such transparency "will in turn increase the chances that stakeholders have enough information — and sufficient basis for trust — to work with companies on solutions that are publicly accountable and do not produce unintended consequences for the human rights of internet users around the world."

The third and final governance model explored in this report is the multi-stakeholder approach, discussed by Larry Strickling and Jonah Force Hill in the fourth essay, "Multi-stakeholder Governance Innovations to Protect Free Expression, Diversity and Civility Online." Strickling and Hill emphasize the commitment to the multi-stakeholder approach as the key, noting that methods, structures and objectives may vary as long as the approach is stakeholder-driven, open, transparent and consensus-based. This approach has several advantages when it comes to internet governance issues, given the rapid pace of technological change and global environment involved. Multi-stakeholder governance may also involve serious challenges, such as ensuring outcomes are seen as legitimate and involve adequate representation from all stakeholders, in particular those who might lack the resources and expertise of more established players. "Yet…when compared to the challenges posed by traditional legislative or regulatory approaches, they produce fewer impediments to effective problem solving," they write.

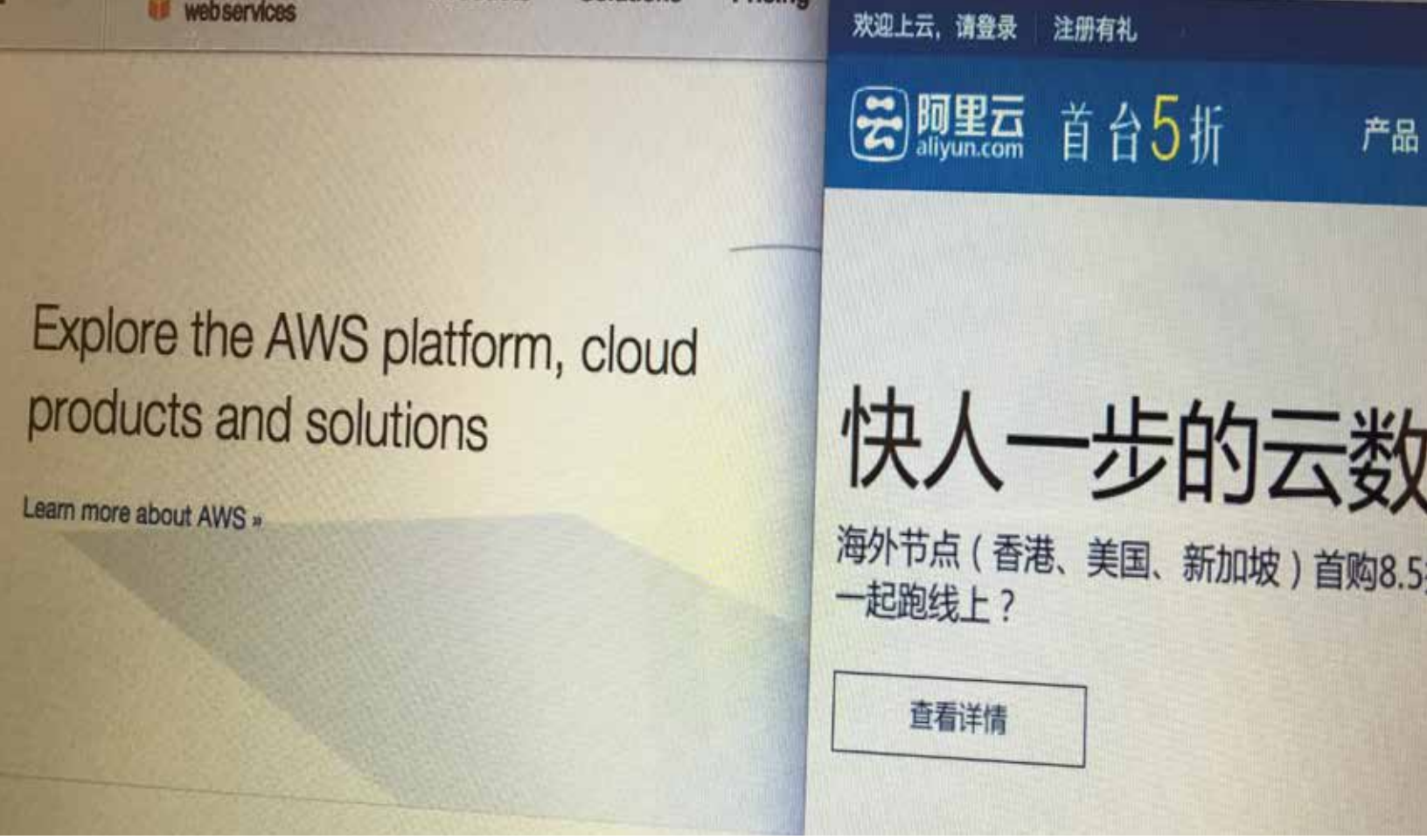
We are far from answering the question of how best to govern the global digital environment. Our shared goal is to enhance free expression, diversity and democracy at the same time as we protect human rights and encourage innovation. The international working meeting, and the contributions from the authors included in this report, offer a starting point for thinking about and discussing the best possible ways to get there.

The organizers of the conference gratefully acknowledge the generous support of the Department of Canadian Heritage.

About the Authors

Eileen Donahoe is executive director of the GDPi at Stanford University and a distinguished fellow at CIGI. She served as US ambassador to the United Nations Human Rights Council in Geneva during the Obama administration. After leaving government, she was director of global affairs at Human Rights Watch. Earlier in her career, she was a technology litigator at Fenwick & West in Silicon Valley. She serves on the National Endowment for Democracy board of directors; Dartmouth College board of trustees; University of Essex Human Rights, Big Data and Technology advisory board; Benetech Human Rights advisory board; and the Freedom Online Coalition advisory network. She is a member of the Council on Foreign Relations and a member of the Transatlantic Commission on Election Integrity.

Fen Osler Hampson is a distinguished fellow and director of CIGI's Global Security & Politics Program, overseeing the research direction of the program and related activities. A fellow of the Royal Society of Canada, he also served as co-director of the GCIG and is the co-director of the D-10 Strategy Forum, jointly managed with the Atlantic Council in Washington, DC. He was director of the Norman Paterson School of International Affairs (2002–2012), and continues to serve as Chancellor's Professor at Carleton University in Ottawa. He is the author or co-author of 13 books, most recently *Look Who's Watching: Surveillance, Treachery and Trust Online* (2016), with Eric Jardine.



Overview of the Challenges Posed by Internet Platforms: Who Should Address Them and How?

Bill Graham and Stephanie MacLellan

Introduction

In March 2018, the Global Digital Policy Incubator (GDPi) at Stanford University and CIGI, in cooperation with the Department of Canadian Heritage, convened an international working meeting to explore governance innovations aimed at protecting free expression, diversity and civic engagement in the global digital ecosystem. The meeting brought together global experts from academia, civil society, several major internet companies and government for the discussions and was held under the Chatham House rule.

There can be no doubt that the internet has created immeasurable benefits for free expression and other social and economic progress, and the plans for the March discussions began with that recognition. Nonetheless, we also recognized the increasing level of concern, among internet users and policy makers alike, about the risks also present in this diverse and evolving digital ecosystem. The following is written in the spirit of contributing to the continuing positive evolution of the internet and the tools it provides.

This overview was prepared for the meeting under the guidance of Eileen Donahoe and Fen Hampson and provided a basis from which the gathered experts could engage. It attempts to explore possible solutions to the negative effects wrought by contemporary digital applications and platforms on free expression, a healthy diversity of views and content, and civic engagement. It is intended to outline some measures currently being implemented or considered that could help maintain and foster robust and functioning democratic engagement and openness to diversity. An addendum recognizes the many significant developments in the space since the March working group meeting.

The Need to Talk about Governance Models

From its simple beginnings, the internet has grown to become the foundation of the world's systems of communication. It has expanded from being a tool for the exchange of scientific data to a vast network of networks essential to the world's commerce and economy, one now enabling a significant portion of the interpersonal and social communication that defines modern societies. Not all of the impacts have been foreseeable or desirable, especially in the realms of society and politics. The pervasive nature of platforms and applications running over the internet means that it is now essential to speak of the global digital ecosystem, rather than generically of the internet.

There is little doubt that some recently revealed uses of digital applications such as social media and content platforms are creating a perception of crisis for the world's democracies (Ferguson 2018). These developments are having detrimental impacts on citizens' ability to exercise freedom of expression, on diversity online and on civic engagement. Examples must include:

- threats to freedom of expression posed, on the one hand, by trolls and bots discouraging speech by attack or simply by swamping the conversation, and, on the other, by increasing censorship or distortion by governments and by the platforms themselves in response to government mandate;
- threats to democracy posed by proven and alleged Russian interventions in democratic elections, accusations of fake news[1] and the resulting threat to a shared understanding of objective reality;
- declining trust in public institutions and traditional media (Chiang and Hoenemeyer 2017);
- increasing isolation of social media users in filter bubbles or echo chambers[2] imposed by the platforms' design, which makes civic engagement across ideological lines difficult, if not impossible; and
- threats to diversity caused by the market dominance of the major platforms, with resulting impacts on the discoverability and economic sustainability of digital cultural expression, including local and linguistically diverse content.

1 When speaking of "fake news," this paper uses the Cambridge Dictionary definition: "Fake news: false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke" (Cambridge Dictionary, n.d.).

2 "These two terms share the same denotation (literal meaning): a phenomenon in which a person is exposed to ideas, people, facts, or news that adhere to or are consistent with a particular political or social ideology" (Lum 2017, para. 2).

Each of the online content challenges that democracies face is real, but their effects and their importance vary, depending upon their national contexts. Some challenges will have more political resonance, depending on their setting; for example, the discussion in the United States may currently be dominated by threats to democracy and political civility, while in Canada and Europe the issue of maintaining cultural and linguistic diversity captures a similar level of political attention. These differences can create the impression that quite separate debates are going on, which compounds the obstacles to reaching a consensus on what, if anything, can be done to correct the situation we find ourselves in.

One of the goals of the international working group meeting, therefore, was to bring different stakeholders with different perspectives together to explore their similarities within a comparative public policy context, and to strive to find principles we might use collectively to guide action.

This essay is intended to review several possible responses to the challenges faced in trying to protect free expression, diversity and civic engagement in the global digital ecosystem. These range from traditional legal and regulatory approaches undertaken by governments, to "softer" approaches to encourage other actors to adopt voluntary, self-protective measures, to efforts to engage social forces more broadly in finding a multi-stakeholder or user-centred way forward.

Traditional Governmental Legislative and Regulatory Approaches

When one thinks about governments' likely reactions to troubling developments, legislation and regulation usually come to mind. To be clear about these terms, legislation refers to statute law passed by the governing authority of a country, establishing a framework of principles within which the government is expected to act in relation to an issue, while regulation is the administrative framework established by a minister or governmental authority to monitor and enforce conditions established by legislation. Legislation is developed to provide conditions that concern rights or prohibitions and are general, therefore not requiring frequent updates. In comparison, regulations tend to be more dynamic; as administrative rules, they may be more readily altered to deal with changing circumstances covered by the legislation.

In most liberal democracies, the global digital ecosystem was permitted to develop within the framework of generally applicable laws; that is to say, something that was permitted or prohibited offline was, by extension, to be treated the same way online, without requiring additional legislation or regulation. This approach was based on a recognition of the rapid

and often unexpected developments enabled by the internet and the realization that legal frameworks would not be rewritten rapidly enough to keep up. Over time, that initial approach, of assuming that online behaviour could be addressed adequately by a legal framework designed for the offline world, has been changing, as the real-world differences between offline and online behaviour have become clearer.

Looking at an example from Canada, the Canadian Radio-television and Telecommunications Commission (CRTC) held public consultations on what they referred to as "new media." In 1999, the regulator issued a public notice outlining its approach to the topic, which included a discussion of how best to deal with offensive and illegal content, and "acknowledg[ing] the views of the majority of parties who argued that Canadian laws of general application, coupled with self-regulatory initiatives, would be more appropriate for dealing with this type of content over the internet than either the Broadcasting Act or Telecommunications Act" (CRTC 1999, para. 121). The notice went on to point to the possibility of giving the Human Rights Commission expanded powers to deal with hate speech, and spoke approvingly of industry efforts at self-regulation; joint government-private sector efforts to combat the problem; and the availability of tools, such as filtering software, to permit end users to control their children's access to undesirable content. But the CRTC declined to take further action itself at that time.

More recently in Canada, "governments are constantly broadening the scope of various laws. The law has come to cover new technological changes, such as electronic meetings, form filing, access to records, and legal authority for using digital or electronic signatures. Also, individual laws dealing with privacy rights, the use of personal information, rights of intellectual property owners, broadcasting over the internet, and other areas often include specific laws to govern internet issues… Courts and lawmakers are starting to develop a body of case law and legislation addressing online rights and obligations" (Legal Line 2013). From the perspective of what is permitted and to be protected on the internet, liberal democracies have viewed the internet as a force capable of reinforcing human rights, including freedom of expression and the promotion of democracy. The governments of many democratic states have recognized that many other countries' governments were increasingly restricting citizens' human rights and fundamental freedoms, and so have joined together to promote the fundamental freedoms on the internet through organizations such as the Freedom Online Coalition.[3]

Along the same lines, the United Nations Human Rights Council in 2012 unanimously passed a resolution entitled "Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development," which "*Affirms* that the same rights that people have offline must also be protected online" (United Nations General Assembly 2012).

Another early impetus for some governments to act was a fear of US domination of the legal framework governing the global digital ecosystem. In particular, as early as 1998, the European Commission expressed dissatisfaction with what they saw as the US government's de facto imposition of policy authority over the internet (European Commission 1998). The European Commission's response reflected a fundamental difference between EU and US models for internet governance. The European Union had developed a preference for coordinated regulation in the information communications and technology area. Meanwhile, the United States preferred to rely on a private self-regulatory model that had been evolving since the internet's early days, and which drew particular attention during the process of creating the Internet Corporation for Assigned Names and Numbers (ICANN) to manage the domain name system.

As the impact of internet platforms has increased, along with the power of platforms as key influencers in the lives and opinions of users, governments have increasingly faced pressures to do something to counter negative impacts. Prime examples of phenomena drawing governments' attention include the increasing online presence of various forms of hate speech and illegal content, the impacts of algorithms, the effect of filter bubbles on social media and the exploitation of platforms by state and non-state hackers.

Concerns such as these have resulted in a demand for governments to act, and governments have begun to respond. Among the first to take strong action has been the government of Germany.

In 2017, the German Parliament passed the "network enforcement law" (NetzDG), which forces any internet platform having more than two million users to make available ways to report and delete potentially illegal content. Affected platforms include Facebook, Twitter, Google, YouTube, Snapchat and Instagram. Professional networks such as LinkedIn and Xing are expressly excluded, as are messaging services like WhatsApp or Telegram. Under the law, which came into effect on January 1, 2018, content such as threats of violence and slander must be deleted within 24 hours of a complaint being received, or within seven days if cases are more legally complex. Companies are also obliged to produce a yearly report detailing how many posts they deleted and why. Companies can be fined up to €50 million if they fail to meet the deadlines (Knight 2018).

3 See https://freedomonlinecoalition.com/.

NetzDG is widely viewed as a sea change with regard to government regulation of the internet. Many human rights organizations have decried the approach taken by the German law, particularly because it assigns responsibility for enforcing laws and standards of speech to the private sector rather than relying on the legal system (Donahoe 2017; Human Rights Watch 2018). Nonetheless, at the recent launch event for the GDPi at Stanford University, Brittan Heller, director of technology and society for the Anti-Defamation League, said she doubts that it will be possible for platforms to remain entirely unregulated after the German legislation because, as she said, "You can't put the toothpaste back in the tube" (Heller 2017). Others at the meeting agreed, suggesting that the survival of a free, open and unfragmented internet is in a race against time since the passage of NetzDG.

However, internet companies have for years been restricting content at government request. Facebook, Twitter and Google all have policies for blocking content — including social media posts, user accounts and search results — from being seen in countries where it contravenes the law, or where a court has ordered its removal. These companies, and many others, have chosen to provide publicly available transparency reports containing information about the number of requests they receive to remove content and the reasons in general terms. The reports are not standardized, and not always complete, so they do not provide more than a general indication of the number of requests submitted and whether those were accepted or rejected by the company. Further information is available about removal practices, both in the form of databases and independent studies of platforms' transparency (Keller 2015).

Critics contend that governments have manipulated country-specific content removal policies to stifle dissent. For instance, in Turkey, where the law bans online content involving terrorism or defamation, most Twitter accounts that are blocked inside the country express anti-government views or are linked to political opponents (Tanash et al. 2015; Silverman and Singer-Vine 2018). But even when countries with more permissive speech environments, such as Germany, introduce internet content laws, there are concerns that strict terms and large penalties will force companies to err on the side of removal, unnecessarily taking down legal content. According to Daphne Keller of the Center for Internet and Society at the Stanford Law School, "Many of the larger companies make a real effort to identify bad faith or erroneous requests, in order to avoid removing legal user content. But mistakes are inevitable given the sheer volume of requests — and the fact that tech companies simply don't know the context and underlying facts for most real-world disputes that surface as removal requests. And of course, the easiest, cheapest, and most risk-avoidant path for any technical intermediary is simply to process a removal request and not question its validity" (Keller 2015).

So far, early experience of applying the NetzDG legislation has shown that implementation will not be without problems or controversy. According to *Deutsche Welle News*, on January 1, 2018, "a top lawmaker from the anti-immigration Alternative for Germany (AfD) party was blocked from Twitter and Facebook on Monday after slamming the Cologne police for sending a New Year's tweet in Arabic" (Winter 2018); a German satire magazine was blocked from Twitter shortly after, when it parodied the law maker's comments (Reuters 2018a). Others have been puzzled by the exemption provided for messaging systems such as WhatsApp (owned by Facebook) and Telegram. These and similar systems offer the capacity to create private groups of hundreds of members to share strongly encrypted messages, making it difficult for outsiders to know what content is being shared. Of course, the security of group messages is only strong if none of the group's members passes the content outside the group. Still, activists of many persuasions, criminals and terrorists are said to be using such platforms to share illegal content or plan acts of violence without fear of discovery by law enforcement (Hinsliff and Pires 2017; Tan 2017). The problem of spreading misinformation, hoaxes and fake news on WhatsApp is increasingly recognized. Because messages on WhatsApp are often voice recordings, the impact of that medium is more strongly felt in parts of the world with low levels of literacy. As a result of that, the impacts can be quite severe, such as the recently reported violent attacks inspired by WhatsApp messages in Brazil, India, Kenya and elsewhere (Funke 2017; Perera 2017).

Germany may be at the leading edge of countries using a legislative approach in attempting to control undesirable behaviour on internet platforms, but it is not likely to be alone for long. For example, elected or senior government officials from Brazil, France and Great Britain are all on record as considering legislative or regulatory action to require social media platforms to monitor for and take down misleading or fake news during election campaigns (Greenwald 2018; Chrisafis 2018; Lomas 2017). There is also ample documentation showing that countries outside of the Western democratic bloc have laws and regulations that require content or applications to be blocked or censored to various degrees, including, for example, a 2017 law moving through the Russian Duma that has been called "a copy-and-paste of Germany's hate speech law" (Reporters Without Borders 2017). Singapore and the Philippines also regard the German law as a positive example, showing the potential for this approach to spread (Human Rights Watch 2018).

Yet, at the same time, at the international level there is widespread opposition to laws intended to limit free speech in an effort to stop the spread of fake news. On March 3, 2017, the United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Organization for Security and Co-operation in Europe (OSCE), the Organization of American States and the African Commission on Human

and Peoples' Rights issued a "Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda" (OSCE 2017). The declaration states that "general prohibitions on the dissemination of information based on vague and ambiguous ideas, including 'false news' or 'non-objective information', are incompatible with international standards for restrictions on freedom of expression…and should be abolished" (ibid., para. 2 [a]); Rose and Mchangama 2017).

Governments' concerns with fake news and illegal content are not the only challenges that the global digital ecosystem poses for public policy. Another area of concern for many is their desire to ensure that diverse and culturally relevant content is available to citizens. In the offline world, these concerns often have led to developing cultural policies to promote the production and distribution of content reflecting unique cultural or linguistic expression. In some cases, legislation and regulation have been used to ensure the availability of such content on the world stage. Examples include national and subnational programs offering support to publishing and broadcasting, as well as to language preservation, and virtual access to archaeological sites, museums and even traditional medicine, games, food and drink.[4] Naturally, government actions to support diversity increasingly include programs that aim to ensure a presence in the online world, including on key platforms. They also include legislation, such as that governing broadcasting, financial support mechanisms, and joint action in international organizations such as the United Nations Educational, Scientific and Cultural Organization (UNESCO). In this area, such traditional "hard" approaches are becoming less used, as governments rely more on a wide spectrum of alternative approaches. These different actions can operate at various degrees of formality, sometimes being characterized as the "soft power" approach, and may combine several instruments to achieve their objectives.

Private Sector and Joint Public-private Sector Approaches

While some countries have instituted or are considering "hard" legislative or regulatory actions to control and punish the publication of illegal content or to ensure the availability of diverse content, the difficulty of enforcing national law on inherently global enterprises such as internet-based platforms is widely recognized. As a result, other authorities are employing less formal approaches, varying from threats to incentives created to encourage voluntary private sector responses. Some of these softer approaches are time-honoured, while others make use of more innovative "nudge" techniques advocated by behavioural economists such as 2017 Nobel Prize winner Richard Thaler (Chu 2017). At the same time, it is clear that internet platforms are aware of both government and public concerns, and, in response, are taking steps to try to minimize their undesirable impacts.

The OSCE "Joint Declaration on Freedom of Expression and 'Fake News,' Disinformation and Propaganda" offers a good summary of approaches that governments might wish to take (or not take). The document suggests creating an enabling environment for freedom of expression by promoting a free, independent and diverse communications environment; by ensuring that there are strong, independent and adequately resourced public service media; by promoting media and digital literacy in schools and through civil society engagement; and by promoting intercultural understanding and democratic values, with a view to addressing the negative effects of disinformation and propaganda (OSCE 2017).

Governments around the world are making use of these types of approaches. For example, the Italian government, in cooperation with leading digital companies, including Facebook and Google, is reaching out through the education system to train a generation of students steeped in social media how to recognize fake news and conspiracy theories online (Horowitz 2017). Finland is combatting fake news by educating the public and politicians, including through teaching students at a young age how to read news critically; Finland's president also spoke out, encouraging citizens to be skeptical about information found online (Martinelli 2017). Similar programs are in place or being developed in several other countries' schools, many in partnership with concerned private sector and civic organizations, and thus tending in the direction of multi-stakeholder solutions.

The OSCE "Joint Declaration" also offers recommendations that intermediaries, including platforms, should consider (OSCE 2017, para. 4 [a]–[e]). Most of these are aimed at ensuring transparency to protect free speech if companies decide to delete or moderate information posted by third parties. The recommendations include providing the public with readily accessible information on their policies and practices; offering opportunities for redress; and ensuring that automated processes, such as algorithms, operate in keeping with actions to delete or moderate information. The OSCE's final recommendation for intermediaries is that they should support research and development of technological solutions to disinformation and propaganda that users themselves could apply on a voluntary basis. As well, platforms are urged to help make fact-checking services available to users and to review their advertising models to ensure that they do not adversely impact diversity of opinions and ideas.

4 For example, see Republic of Kenya (2009); Department of Canadian Heritage (2017b); Government of Quebec (2016); see also https://ec.europa.eu/culture/policy/culture-policies/cultural-heritage_en and https://en.unesco.org/creativity/monitoring-reporting/periodic-reports.

Most of the leading internet platforms are, of their own accord, taking steps along these lines to show a willingness to deal with illegal content and fake news.

Mark Zuckerberg, Facebook's creator and chief executive officer (CEO), in particular seems to be seized by the need to find ways to avoid the problems that have come to light. Following the revelation of Russian meddling in the 2016 US election, Zuckerberg's public statements began with outright denial, but they have since evolved, through acknowledgement to recent announcements of initiatives intended to keep elections free of influence via his platform and to protect users from fake news (Weiss 2017). The company also announced programs to help users to understand and prevent the spread of fake news, accompanied by projects intended to expand news literacy and improve trust in journalism (for example, The News Literacy Project 2017).

In early 2018, Zuckerberg announced that he was taking on a personal challenge in 2018 to fix Facebook, with a goal of "protecting our community from abuse and hate, defending against interference by nation states, or making sure that time spent on Facebook is time well spent" (Zuckerberg 2018a). One week later, he announced that changes were being introduced to refocus the operations of the platform's "News Feed," to show fewer advertisements and news articles and to favour content posted by a user's friends and family (Zuckerberg 2018b). Next, the CEO announced that, in future, Facebook would "shift the balance of news [users] see towards sources that are determined to be trusted by the community" (Zuckerberg 2018c). The decision about *which* news sources were trusted would be determined by asking some members of the community if there were any news sources they are familiar with and whether they trust those sources. These changes have occasioned alarm from investors fearing a drop in advertising revenues, and from news media companies, concerned that strategies they have developed to profit from a presence on Facebook will no longer work for them (Vanian 2018). It is also questionable whether these changes will help to reduce the spread of fake news or diminish the filter bubble effect — after all, one's friends and family might be as likely to confirm and amplify opinions as to challenge them. Indeed, when Facebook conducted early tests of a similar modification to the News Feed in post-conflict countries, the modified News Feed "surfaced more news stories from friends and family — and fake news increased" (Kosoff 2018).

At the end of January 2018, Facebook announced that it was for the first time posting its privacy principles, increasing users' control over their own information and how it is shared, and would also be launching a series of user-education videos to teach users how to make use of the new tools. Reuters reported: "The announcements on [January 29] by Erin Egan, chief privacy officer at Facebook, are a sign of its efforts to get ready before the European Union's general data protection regulation (GDPR) enters into force on 25 May, marking the biggest overhaul of personal data privacy rules since the birth of the internet" (Reuters 2018b). Egan wrote in a post on Facebook's news blog: "We recognize that people use Facebook to connect, but not everyone wants to share everything with everyone — including with us. It's important that you have choices when it comes to how your data is used" (Egan 2018).

Facebook is also experimenting with a project to safeguard against interference in the political process. The Canadian Election Integrity Initiative will help politicians and parties to protect their accounts from being hacked, provide guidance on how to secure their pages, and work with MediaSmarts, a non-profit group, to educate voters about the dangers of fake news. This initiative is one of several being tried in other countries around the world (Leblanc 2017).

Along with Facebook, Google has been criticized for promoting misinformation through the results of its search algorithms and its YouTube service. In response, the company has announced plans to tighten its policies to ban websites that peddle fake news from using its online advertising service (Wingfield, Isaac and Benner 2016) and to increase human oversight of top-tier YouTube content. (Wakabayashi 2018). It has announced it will block websites from its search results if those sites are masking their country of origin (Wong 2017). As well, in cooperation with other organizations, Google has announced programs to boost media literacy and to help users learn to discriminate between news and fake news or misinformation online. One such initiative is NewsWise, funded by a grant from Google, to be used in Canadian schools (Nanji 2017).

Similarly, Twitter has been taking steps to show it intends to counter fake news. As one might expect, many of its initiatives are specifically directed at reducing the company's impact on elections. It has created an elections task force to prepare for the 2018 US Congressional elections by verifying party candidates' accounts; improving algorithms to weed out bot accounts; monitoring trending topics for fake news; and bringing new transparency to its advertising policies (Ng 2018).

Among social media platforms, Snapchat stands out for its success in keeping fake news off its site. In fact, *Bloomberg Businessweek* says it appears that Snapchat has no fake news at all. The company credits "humans" for this success, saying, "We only work with authoritative and credible media companies, and we unashamedly have a significant team of producers, creators, and journalists" (quoted in Chafkin 2017). As well, "Snapchat doesn't use algorithms to try to keep people clicking on new material; the only posts you see when glancing at the app have either come from your

friends or been vetted for Our Stories. As a result, posts by individuals almost never reach more than a few hundred viewers" (Chafkin 2017).

The public and private sectors alike are also taking steps to address the challenge of promoting diversity of content online. Governments have sought ways to work with leading internet platforms to ensure that their citizens have access to culturally relevant content. At the international level, the parties to UNESCO's 2005 Convention on the Protection and Promotion of the Diversity of Cultural Expression, comprising 145 countries and the European Union, have adopted operational guidelines for implementing the Convention in the digital environment, to provide policy makers with options to protect and preserve cultural expression in the digital age (UNESCO 2017; Stephens 2017). The guidelines address four main areas — creation, production, distribution and access — and recognize the challenges posed by the major internet platforms' increasing dominance of these four areas. The guidelines recommend measures to provide for fair compensation and rights of creators; build digital capacity among small and medium enterprises; encourage collaboration between public authorities and the private sector in encouraging broader distribution and dissemination of national content; encourage the creation of algorithms that help to promote diversity of cultural expression and cultural products; and increase access to diverse linguistic and cultural products. Hugh Stephens, a distinguished fellow of the Asia Pacific Foundation of Canada, says that while the guidelines are not enforceable, they will raise awareness of the issues faced by creators in the digital environment and offer useful suggestions for coordinated policy responses (Stephens 2017).

At the national level, governments are creating and repositioning their programs to promote diverse linguistic and cultural products in the digital ecosystem. A recent example is the Government of Canada's fall 2017 announcement of "Creative Canada," a renewed vision for Canada's cultural and creative industries in a digital world. The policy framework is designed to do three things: to invest in creators and cultural entrepreneurs with the aim of promoting creation and production of diverse cultural works; to promote discovery and distribution of Canadian content at home and globally; and to strengthen public broadcasting and support local news (Department of Canadian Heritage 2017b). One noteworthy aspect of this program was the announcement of an agreement between the government and Netflix. Netflix promised "to invest CDN$500 million in original productions in Canada over five years; to support French-language content through a $25 million market development strategy; to create Netflix Canada — a first of its kind production company for Netflix outside the United States; and to implement measures to ensure Canadians and Netflix subscribers across the world can discover Canadian films and television shows" (Department of Canadian Heritage 2017a). The government also highlighted recent moves by major platforms to increase the visibility of and access to Canadian content in English and French on Amazon's Audible audiobook service, Spotify Canada and YouTube (ibid.). Reactions to the policy framework's announcement were mixed, but the approach can be seen as an example of government engaging with platforms through less formal, more collaborative arrangements to achieve its objectives.

Internet platforms are increasingly aware that governments are concerned about the challenges of maintaining cultural diversity in the global digital ecosystem, and, as with initiatives to deal with fake news and filter bubbles, several are taking steps voluntarily to help. The following examples are not a complete catalogue but an indication that several of the largest platforms in the business of distributing cultural products are attempting to address the problem. On YouTube, countries with high numbers of users have channels dedicated to their content. Facebook offers a number of country-specific pages. Apple offers country- and region-based versions of its iTunes Store to promote domestic music and video content, as does the online music streaming service Spotify. Examples such as these show that the platforms and services recognize there is business to be had by catering to national and regional preferences, but perhaps they also demonstrate that it is better to take voluntary action to avoid the potential of governments resorting to more forceful measures.

Multi-stakeholder or User-centred Approaches

Models other than government regulation or industry self-regulation are possible — models that involve citizens more directly in deciding the best ways to protect free expression, diversity and civic engagement in the global digital ecosystem. One such approach is that of multi-stakeholder governance. For some time, the multi-stakeholder approach has been widely accepted in the field of internet governance. It has proven successful because of its ability to find acceptable, often innovative, ways to deal with complex problems in rapidly changing environments where decisions impact a wide range of people and interests, often across sectors and borders. The approach works best when different types of expertise are needed, and where there may be overlapping rights and responsibilities or where the resulting solutions must be seen as legitimate so that they can successfully be implemented. Those characteristics describe well the attributes that will be needed to overcome the challenges described above. Multi-stakeholder processes do not typically end up suggesting that a new governing body be created. It is more normal for such

a process to end up recommending the application of either a conventional hard or soft power solution, although it is also possible that it will obviate the need for either.

There is no single definition to describe the multi-stakeholder approach. It would be counterproductive to stick to a single cookie-cutter approach; instead, the approach must be adapted to suit the nature of the problem being approached and the constellation of stakeholders to be involved in finding a solution. A recent paper published by the Internet Society (Strickling and Abuhamad 2017) describes a multi-stakeholder approach as being:

- *stakeholder-driven* — stakeholders determine the process and decisions, from agenda setting to workflow, rather than merely fulfill an advisory role;
- *open* — any stakeholder may participate and the process includes and integrates the viewpoints of a diverse range of stakeholders;
- *transparent* — all stakeholders and the public have access to deliberations, creating an environment of trust, legitimacy and accountability; and
- *consensus-based* — outcomes are consensus-based, reached by compromise and a win-win for the greatest number or diversity of stakeholders.

A multi-stakeholder process that hopes to address the challenges confronting our political, social and cultural sectors must involve governments; private sector platforms and content companies of all types; educators; academics; activists; and ordinary citizens that make up or are affected by the global digital ecosystem. Governments cannot solve the problems alone — the transborder nature of the internet and the applications that make use of it simply make that impossible. The private sector almost certainly cannot and will not solve the problem in isolation; companies do not share the incentives to do so and do not have the necessary levers to deal with the impacts of their business models. Civil society also lacks the cohesion, the levers and the experience to deal with the challenges without cooperation from the other players. All of these actors need to come together to develop a shared understanding of the problems and the possible solution space, and then to work in good faith to find the way forward.

Just such an approach was strongly recommended by the Global Commission on Internet Governance (GCIG), convened by the CIGI and Chatham House. The GCIG's final report called for all stakeholders to develop a "social compact [which] must be built on a shared commitment by all stakeholders in developed and less-developed countries to take concrete action in their own jurisdictions to build trust and confidence in the internet. A commitment to the concept of collaborative security and to privacy must replace lengthy and over-politicized negotiations and conferences" (GCIG 2015, 1).

As described above, some governments and private sector players, acting alone or together, have already made efforts to improve the situation. Similarly, some efforts have involved more stakeholders, although few yet are engaging the full range of necessary actors.

In the international arena, the Organisation for Economic Co-operation and Development (OECD) has long been a major player. The OECD is an intergovernmental economic organization with 36 members, primarily from the high-income economies. Although governments are the decision makers, the OECD involves business, academics and civil society organizations in a consultative capacity, and increasingly invites less developed economies to discuss policy issues across a wide variety of economic sectors. The organization is a leading forum for discussion of communications, computational and internet-related policy issues. Recently these have expanded to include discussions about the effects of algorithms, fake news, keeping democracies safe from hackers and the role of internet intermediaries, which include internet platforms of all types (OECD 2010; 2017; Alter 2017; Clarke and Gyimeshi 2017). These discussions have now spread throughout the organization, in recognition of the fundamental role played by applications running on the internet for virtually all sectors of the global economy.

At a more informal level, one current multi-stakeholder example is intended to develop new "trust indicators" to help users better vet the reliability of the publications and journalists behind articles that appear in online news feeds. The indicators were developed by the Trust Project, operating out of Santa Clara University's Markkula Center for Applied Ethics, to boost transparency and media literacy at a time when misinformation is rampant. Twitter, Facebook and Google are engaged in and supporting this project, as are leading publishers, including *The Washington Post*, *The Economist* and *The Globe and Mail*, but so far without governments' involvement (Fiegerman 2017; The Trust Project 2017).

Other examples can be found in the education sector. MediaSmarts has been developing digital and media literacy programs and resources for Canadian homes, schools and communities since 1996. Different levels of government, the private sector (including internet companies, internet service providers, media companies and broadcasters) and civil society have supported this not-for-profit effort. Its goal is to provide adults with information and tools so they can help children and teens develop critical-thinking skills for interacting with media of all types.[5] There are many similar projects in the United States that involve different constellations of partners, both domestic and international. These include the Center for News

5 See http://mediasmarts.ca/about-us/what-we-do.

Literacy in Stony Brook, New York (which has an international outreach program), the News Literacy Project in Chicago and others (Jacobson 2017).

Many suggestions have been put forward that could help to develop the needed multi-stakeholder solutions. One set of principles has usefully been offered by Timothy Garton Ash (2016) in his book *Free Speech: Ten Principles for a Connected World.* In *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, data scientist Cathy O'Neil (2016) introduces the idea of algorithmic auditing as a way to investigate whether algorithms deployed by internet firms and others create harmful outcomes for certain groups of users. In a similar call for transparency, Wael Ghonim and Jacob Rashbass (2017) propose a standardized "public interest API" (application program interface) to be used by platforms that use algorithms to share content; it would show information such as how many people saw a piece of content, how advertisements were targeted and which content was censored. The non-profit Ranking Digital Rights research initiative developed a Corporate Accountability Index, which ranks internet companies' "corporate disclosure of policies and practices that affect users' freedom of expression and privacy" as a means of encouraging accountability through a systematic, regularly updated rating system (Ranking Digital Rights 2017).

There have also been suggestions that a "naming and shaming" approach used in areas such as conflict minerals (as explored in Chesterman and Pouligny 2002) could be applied to the digital space, putting pressure on companies that violate users' privacy or allow other harmful behaviours on their platforms.

A recent article in *The New York Times Magazine* describes an even more radical idea: moving beyond sharing power through a multi-stakeholder approach to giving power to individual users through control over their data and usage history (Johnson 2018). The suggestion is that it would be possible, by making use of the highly secure and distributed nature of blockchain technology, to effectively eliminate the virtual monopolies that today's very large platforms have over the data that users create when using their services. The article highlights the considerable advantages that established platforms gain by virtue of the amount of personal data they hold about their users — not only their personal data (often credit card numbers and similarly sensitive material) but also their usage history, their interests, their friends and associates, and much more. If this data resided in a blockchain ledger where the data subject/owner had the power to control who gets to use it, for what purpose and under what terms, much of that advantage and much of the present risk would be removed. This idea may seem utopian, but as blockchain ledgers become more broadly used, and as users' dissatisfaction with how their data is currently managed or mismanaged by today's leading platforms increases, the author believes that a door may open to the possibility of creating entirely new, redesigned platforms, including social media platforms (ibid.).

There is no doubt that solutions must, and will, be found. The effort to find solutions in a multi-stakeholder approach could potentially be originated by any of the stakeholders: by governments; by corporations, as a result of civil society activism; or even by proponents of a new technology such as blockchain. First, though, some basic questions need to be answered. For example, what can be done to encourage broadly based approaches? Are there fora where norms can be discussed and agreed upon, and if not, how and where can such discussions be convened? These questions and others were offered as a basis for discussions during the working group meeting.

Addendum

The six months following the international working meeting at Stanford saw a number of significant developments with implications for digital governance. Indeed, the rapidity of these developments underscores one of the challenges of governing the digital environment: technology and current events can quickly overtake policy. As one participant said at the meeting, there is no guarantee that today's dominant platforms will still be dominant in five years, or that they will function as they do today; accordingly, governance options should be developed with general principles in mind that can adapt to a rapidly changing digital landscape.

What follows is a list of some of the major developments since the Stanford meeting in March 2018. It is highly probable that further discoveries will emerge by the time this report is published, so this list is not intended to be exhaustive but to give some added background on the responses various sectors have taken so far to adapt to these changing circumstances.

Cambridge Analytica

Shortly after the working meeting in March 2018, it was revealed that Cambridge Analytica, a British data analytics company contracted by election campaigns, had acquired data from about 87 million Facebook users, without the consent of a vast majority of those users. The information was harvested through a third-party survey app that researcher Aleksandr Kogan shared on Facebook in 2013. Only about 300,000 people agreed to share their data through the app, but the number of those whose data was shared grew exponentially, because the app gained access to the profile information of the consenting users' friends. This practice was permitted under Facebook's terms of service at the time, but the company reversed

that policy in 2014. Cambridge Analytica acquired the profile information from Kogan and reportedly used it to develop its voter targeting efforts for the 2016 US election — violating Facebook's requirement that developers who collect user data not use it for anything other than its stated purpose. Facebook had been warned about Cambridge Analytica's data hoard in 2015 and asked the firm to delete the data, but it did not follow up to ensure Cambridge Analytica had done so (Cadwalladr and Graham-Harrison 2018; Facebook 2018a).

Facebook's initial reaction was to downplay the events, stating that this was not a data breach because the data had not been acquired illegally. However, as a media and political firestorm ensued, CEO Mark Zuckerberg apologized and the company suspended Cambridge Analytica and Kogan from its services. Cambridge Analytica later filed for bankruptcy. Since then, Facebook has further restricted the user information that developers can access and introduced several changes to help users manage their data privacy, including prompts that show which apps they have allowed to access their data (Schroepfer 2018). Investigations into the matter have been launched by the UK Information Commissioner's Office (Summers and Slawson 2018) and by the US Department of Justice, Federal Bureau of Investigation, US Securities and Exchange Commission and the Federal Trade Commission (Timberg et al. 2018). Parliamentary and congressional committees have also investigated the matter in the United States, the United Kingdom, Canada and the European Union. Financially, Facebook shares took a massive hit in July after the company released its second-quarter results, which showed stagnating user growth and below-expectation revenue; the company's stock dropped by 20 percent after the earnings announcement (Newton 2018).

GDPR

The European Union's GDPR took effect on May 25. This data privacy regulation has had ripples around the world because it applies not only to European companies but to any company that holds data on Europeans — including the Silicon Valley tech giants. Among the key changes: companies must seek users' consent in clear terms to collect, retain or process their personal information; users have the right to access their personal data from a company and to remove it and/or transfer it to another service; and companies must notify users of a data breach within 72 hours. Companies that breach the GDPR can be fined up to four percent of their annual revenue or €20 million, whichever is higher.[6] Governments and observers have been closely watching the GDPR roll out, with some predicting it could set a new global standard for data protection legislation. Brazil and California have already introduced data privacy bills modelled on the GDPR. While it is still too soon to fully evaluate the outcomes of GDPR, some early developments suggest a few issues to watch in the months ahead:

- The day the GDPR took effect, a privacy activist filed complaints against Google, Facebook, Instagram and WhatsApp, claiming they violated the GDPR by forcing users to choose between sharing their data or leaving the platform entirely (Fielder and Busvine 2018).
- Similarly, researchers at the Consumer Council of Norway have found that Facebook and Google, and to a lesser extent Microsoft, use design settings to "nudge" users toward choosing more permissive privacy options (Forbrukerrådet 2018).
- Early numbers suggest the GDPR has drawn digital advertising revenue toward Google and away from its smaller competitors that might be struggling to obtain users' consent for targeted advertising (Kostov and Schechner 2018).
- As of August, more than 1,000 American news websites were still unavailable in Europe, with the media companies choosing to block readers in Europe rather than comply with GDPR standards (South 2018).

Election Advertisements and Political Accounts

This May, Google, Facebook and Twitter all began rolling out stricter conditions for political and election-related advertisements and content on their platforms. Each company introduced new procedures, for verifying the identity of anyone who purchased election ads and for clearly disclosing who paid for the ads, and transparency initiatives, which allow users to view additional information about ad spending and targeting (Walker 2018; Leathern 2018; Gadde and Falck 2018). As might be expected with an initiative of this size, Facebook has run into some early implementation challenges. Its automated, keyword-based ad system flagged several unrelated ads as political content and took them down for not going through the proper procedures. For example, ads for Bush's Beans (a baked bean company) and a bible school in Clinton, Indiana, were flagged because they included politicians' surnames (Frier 2018). Additional scrutiny is also applied to ads related to 20 issues Facebook deems to be political, such as immigration, education, environment, gun control, health and "values." Critics claim that non-political ads — from media outlets sharing political journalism and the US Department of Homeland Security, among others — are getting caught up in the system, something Facebook says it is working to address. (Hamilton 2018). Outside of the United States, Facebook and Google also introduced political advertising restrictions for the May 2018 referendum on abortion in Ireland. Facebook blocked campaign ads that came from outside of Ireland, and Google banned all campaign ads from its platforms (Satariano 2018).

6 See www.eugdpr.org/key-changes.html.

Detection and Removal of Fake Accounts

On July 31, Facebook announced that it had removed 32 pages and accounts that were "involved in coordinated inauthentic behavior" ahead of the 2018 mid-term Congressional elections in the United States. The company did not say who was responsible, but noted that some of the accounts could be linked to accounts from Russia's Internet Research Agency that were active in influence operations around the 2016 elections. Like the earlier accounts that were tied to Russia, these false pages shared events and opinions on opposing sides of American social and cultural debates (Facebook 2018b). Interestingly, Facebook coordinated with the Atlantic Council's Digital Forensic Research Lab to analyze the affected pages and posts, which could be a multi-stakeholder approach worth following (Digital Forensic Research Lab 2018). Earlier in July, Twitter had announced it would remove tens of millions of automated or fake accounts, totalling up to six percent of its user base. The move was aimed at curbing the market for fake followers, which can be used to make Twitter accounts seem more popular or influential than they are, but it could also affect influence operations that rely on bots to amplify their messages (Confessore and Dance 2018).

Works Cited

Alter, Rolf. 2017. "Can we save our democracies from hackers?" In *OECD Yearbook 2017: Bridging Divides*, 58-59. https://issuu.com/oecd.publishing/docs/oecd-yearbook-2017.

Cadwalladr, Carole and Emma Graham-Harrison. 2018. "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach." *The Guardian,* March 17. www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

Cambridge Dictionary. n.d. "Fake news." In *Cambridge Advanced Learner's Dictionary & Thesaurus,* online ed. Cambridge, UK: Cambridge University Press. https://dictionary.cambridge.org/dictionary/english/fake-news.

Chafkin, Max. 2017. "How Snapchat Has Kept Itself Free of Fake News." *Bloomberg Businessweek*, October 26. www.bloomberg.com/news/features/2017-10-26/how-snapchat-has-kept-itself-free-of-fake-news.

Chesterman, Simon and Béatrice Pouligny. 2002. *The Politics of Sanctions*. May. New York, NY: International Peace Academy. www.ipinst.org/wp-content/uploads/publications/politics_of_sanctions.pdf.

Chiang, Lulu and Lauren Hoenemeyer. 2017. "Public trust in institutions sinks to a record low." CBS News, January 15. www.cbsnews.com/news/davos-world-economic-forum-public-trust-in-institutions-at-record-low-media-government-business/.

Chrisafis, Angelique. 2018. "Emmanuel Macron promises ban on fake news during elections." *The Guardian*, January 3. www.theguardian.com/world/2018/jan/03/emmanuel-macron-ban-fake-news-french-president.

Chu, Ben. 2017. "What is 'nudge theory' and why should we care? Explaining Richard Thaler's Nobel economics prize-winning concept." *The Independent*, October 9. www.independent.co.uk/news/business/analysis-and-features/nudge-theory-richard-thaler-meaning-explanation-what-is-it-nobel-economics-prize-winner-2017-a7990461.html.

Clarke, Rory and Balazs Gyimeshi. 2017. "Digging up facts about fake news: The Computational Propaganda Project." In *OECD Yearbook 2017: Bridging Divides*, 76-77. www.oecd.org/governance/digging-up-facts-about-fake-news-the-computational-propaganda-project.htm.

Confessore, Nicholas and Gabriel J. X. Dance. 2018. "Battling Fake Accounts, Twitter to Slash Millions of Followers." *The New York Times*, July 11. www.nytimes.com/2018/07/11/technology/twitter-fake-followers.html.

CRTC. 1999. "Telecom Public Notice CRTC 99-14 and Broadcasting Public Notice CRTC 1999-84." May 17. https://crtc.gc.ca/eng/archive/1999/pt99-14.htm.

Department of Canadian Heritage. 2017a. "Backgrounder: Creative Canada — Changes to Policies, Programs and Legislation." September 28. www.canada.ca/en/canadian-heritage/news/2017/09/creative_canada_changestopoliciesprogramsandlegislation.html.

———. 2017b. "Creative Canada: Policy Framework." CH4-185/2017E-PDF, October 19. www.canada.ca/content/dam/pch/documents/campaigns/creative-canada/CCCadreFramework-EN.pdf.

Digital Forensic Research Lab. 2018. "#TrollTracker: Facebook Uncovers Active Influence Operation." Medium.com, July 31. https://medium.com/dfrlab/trolltracker-facebook-uncovers-active-influence-operation-74bddfb8dc06.

Donahoe, Eileen. 2017. "Protecting Democracy from Online Disinformation Requires Better Algorithms, Not Censorship." *Net Politics* (blog), August 21. New York, NY: Council on Foreign Relations. www.cfr.org/blog/protecting-democracy-online-disinformation-requires-better-algorithms-not-censorship.

Egan, Erin. 2018. "Giving You More Control of Your Privacy on Facebook." Facebook Newsroom, January 29. https://newsroom.fb.com/news/2018/01/control-privacy-principles/.

European Commission. 1998. "European Commission proposes a Draft Reply of the EU and its Member States to the US Green Paper on Internet Governance." European Commission press release IP/98/184, February 25. http://europa.eu/rapid/press-release_IP-98-184_en.htm.

Facebook. 2018a. "Hard Questions: Update on Cambridge Analytica." Facebook Newsroom, March 21. https://newsroom.fb.com/news/2018/03/hard-questions-cambridge-analytica/.

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

Ferguson, Niall. 2018. "Social networks are creating a global crisis of democracy." *The Globe and Mail*, January 20. www.theglobeandmail.com/opinion/niall-ferguson-social-networks-and-the-global-crisis-of-democracy/article37665172/.

Fiegerman, Seth. 2017. "Facebook, Google, Twitter to fight fake news with 'trust indicators.'" CNN, November 16. http://money.cnn.com/2017/11/16/technology/tech-trust-indicators/index.html.

Fielder, Lucy and Douglas Busvine. 2018. "Austrian data privacy activist takes aim at 'forced consent.'" Reuters, May 25. www.reuters.com/article/us-europe-privacy-lawyer/austrian-data-privacy-activist-takes-aim-at-forced-consent-idUSKCN1IQ0ZI.

Forbrukerrådet. 2018. "Deceived by Design: How tech companies use dark patterns to discourage us from exercising our rights to privacy." Forbrukerrådet [Consumer Council of Norway], June 27. https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf.

Frier, Sarah. 2018. "Facebook's Political Rule Blocks Ads for Bush's Beans, Singers Named Clinton." Bloomberg, July 2. www.bloomberg.com/news/articles/2018-07-02/facebook-s-algorithm-blocks-ads-for-bush-s-beans-singers-named-clinton.

Funke, Daniel. 2017. "Here's why fighting fake news is harder on WhatsApp than on Facebook." The Poynter Institute, October 5. www.poynter.org/news/heres-why-fighting-fake-news-harder-whatsapp-facebook.

Gadde, Vijaya and Bruce Falck. 2018. "Increasing Transparency for Political Campaigning Ads on Twitter." *Twitter* (blog), May 24. https://blog.twitter.com/official/en_us/topics/company/2018/Increasing-Transparency-for-Political-Campaigning-Ads-on-Twitter.html.

Garton Ash, Timothy. 2016. *Free Speech: Ten Principles for a Connected World.* New Haven, CT: Yale University Press.

GCIG. 2015. *Toward a Social Compact for Digital Privacy and Security*. Waterloo, ON: CIGI. www.cigionline.org/publications/toward-social-compact-digital-privacy-and-security.

Ghonim, Wael and Jake Rashbass. 2017. "It's time to end the secrecy and opacity of social media." *The Washington Post*, October 31. www.washingtonpost.com/news/democracy-post/wp/2017/10/31/its-time-to-end-the-secrecy-and-opacity-of-social-media/?utm_term=.1a61788f9700.

Government of Quebec. 2016. *Periodic Quadrennial Report on the Implementation of the UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions.* https://en.unesco.org/creativity/sites/creativity/files/periodic_reports/old/quebec_qpr_en.pdf.

Greenwald, Glenn. 2018. "First France, Now Brazil Unveils Plan to Empower the Government to Censor the Internet in the Name of Stopping 'Fake News.'" *The Intercept*, January 10. https://theintercept.com/2018/01/10/first-france-now-brazil-unveils-plans-to-empower-the-government-to-censure-the-internet-in-the-name-of-stopping-fake-news/.

Hamilton, Keegan. 2018. "Facebook Can't Decide if Homeland Security Ads are 'Political Content.'" *Vice News*, July 31. https://news.vice.com/en_us/article/594555/facebook-quietly-deleted-homeland-security-ads-from-political-content-archive.

Heller, Brittan. 2017. "When Freedom of Expression Conflicts with Democracy: Enhancing the Quality of Discourse Necessary to Sustain Democracy." Panel discussion remarks given at launch event for the GDPi, Stanford University. Filmed October 6 in Stanford, CA, 1:39:38. www.youtube.com/watch?v=Y2FsHIqtrUc.

Hinsliff, Gaby and Candice Pires. 2017. "WhatsApp: inside the secret world of group chat." *The Guardian*, November 12. www.theguardian.com/global/2017/nov/12/whatsapp-inside-secret-world-of-group-chat-politics-sexual-harassment.

Horowitz, Jason. 2017. "In Italian Schools, Reading, Writing and Recognizing Fake News." *The New York Times*, October 18. www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html.

Human Rights Watch. 2018. "Germany: Flawed Social Media Law." *Human Rights Watch News*, February 14. www.hrw.org/news/2018/02/14/germany-flawed-social-media-law.

Jacobson, Linda. 2017. "The Smell Test: Educators can counter fake news with information literacy. Here's how." *School Library Journal*, January 1. www.slj.com/2017/01/industry-news/the-smell-test-educators-can-counter-fake-news-with-information-literacy-heres-how/.

Johnson, Steven. 2018. "Beyond the Bitcoin Bubble." *The New York Times Magazine*, January 16. www.nytimes.com/2018/01/16/magazine/beyond-the-bitcoin-bubble.html.

Keller, Daphne. 2015. "Empirical Evidence of 'Over-Removal' by Internet Companies under Intermediary Liability Laws." *The Center for Internet and Society at Stanford Law School* (blog), October 12. http://cyberlaw.stanford.edu/blog/2015/10/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws.

Knight, Ben. 2018. "Germany implements new internet hate speech crackdown." *Deutsche Welle News*, January 1. http://p.dw.com/p/2qBvi.

Kosoff, Maya. 2018. "How Mark Zuckerberg's Media Divorce Could Backfire." *Vanity Fair*, January 15: www.vanityfair.com/news/2018/01/will-facebooks-media-divorce-backfire.

Kostov, Nick and Sam Schechner. 2018. "Google Emerges as Early Winner from Europe's New Data Privacy Law." *The Wall Street Journal,* May 31. www.wsj.com/articles/eus-strict-new-privacy-law-is-sending-more-ad-money-to-google-1527759001?mod=e2tw.

Leathern, Rob. 2018. "Shining a Light on Ads With Political Content." Facebook Newsroom, May 24. https://newsroom.fb.com/news/2018/05/ads-with-political-content/.

Leblanc, Daniel. 2017. "Facebook to launch hotline for hacked Canadian politicians." *The Globe and Mail*, October 19. https://theglobeandmail.com/news/politics/facebook-to-launch-hotline-for-hacked-canadian-politicians/article36657659/.

Legal Line. 2013 "Which laws apply on the Internet?" March 18. www.legalline.ca/legal-answers/which-laws-apply-on-the-internet/.

Lomas, Natasha. 2017. "Social media firms should face fines for hate speech failures, urge UK MPs." *TechCrunch*, May 2. https://techcrunch.com/2017/05/02/social-media-firms-should-face-fines-for-hate-speech-failures-urge-uk-mps/.

Lum, Nick. 2017. "The Surprising Difference Between 'Filter Bubble' and 'Echo Chamber.'" Medium, January 27. https://medium.com/@nicklum/the-surprising-difference-between-filter-bubble-and-echo-chamber-b909ef2542cc.

Martinelli, Marissa. 2017. "Finland Has Figured Out How to Combat Fake News. *Full Frontal* Thinks The U.S. Can Follow Suit." *Slate*, October 12. www.slate.com/blogs/browbeat/2017/10/12/full_frontal_investigates_finland_s_anti_fake_news_efforts_video.html.

Nanji, Sabrina. 2017. "Google bankrolls Canadian school program targeting fake news." *Toronto Star*, September 19. www.thestar.com/news/gta/2017/09/19/google-bankrolls-canadian-school-program-targeting-fake-news.html.

Newton, Casey. 2018. "Facebook's stock price collapses along with user growth." *The Interface*, July 25. www.getrevue.co/profile/caseynewton/issues/facebook-s-stock-price-collapses-along-with-user-growth-125656?utm_campaign=Issue&utm_content=view_in_browser&utm_medium=email&utm_source=The+Interface.

Ng, Alfred. 2018. "How tech giants plan to keep fake news out of 2018 election." CNET News, January 17. www.cnet.com/news/facebook-google-twitter-2018-election-prevent-fake-news-senate/.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Crown.

OECD. 2010. *The economic and social role of Internet intermediaries.* April. www.oecd.org/internet/ieconomy/44949023.pdf.

———. 2017. *Algorithms and collusion: Competition policy in the digital age.* September 14. www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm.

OSCE. 2017. "Joint Declaration on Freedom of Expression and 'Fake News', disinformation and propaganda." FOM.GAL/3/17, March 3. www.osce.org/fom/302796?download=true.

Perera, Ayeshea. 2017. "The people trying to fight fake news in India." BBC News, July 24. www.bbc.com/news/world-asia-india-40657074.

Raymond. 2017. "How Google is stopping fake news from spreading through Google News." Mashable, December 17. http://mashable.com/2017/12/17/google-news-no-hiding-country-origin-stop-fake-news/#yLLj9uJ1GmqW.

Ranking Digital Rights. 2017. "2018 Indicators." July 25. https://rankingdigitalrights.org/2018-indicators/.

Reporters Without Borders. 2017. "Russian bill is copy-and-paste of Germany's hate speech law." Reporters Without Borders, July 19. https://rsf.org/en/news/russian-bill-copy-and-paste-germanys-hate-speech-law.

Republic of Kenya. 2009. *National Policy on Culture and Heritage*. Nairobi, Kenya: Office of the Vice-President, Ministry of State for National Heritage and Culture. https://en.unesco.org/creativity/sites/creativity/files/activities/conv2005_eu_docs_kenya_policy.pdf.

Reuters. 2018a. "German hate speech law tested as Twitter blocks satire account." Reuters, January 3. www.reuters.com/article/us-germany-hatecrime/german-hate-speech-law-tested-as-twitter-blocks-satire-account-idUSKBN1ES1AT.

———. 2018b. "Facebook reveals privacy principles for first time, helps users control access." *The Guardian*, January 29. www.theguardian.com/technology/2018/jan/29/facebook-reveals-privacy-principles-for-first-time-helps-users-control-access.

Rose, Flemming and Jacob Mchangama. 2017. "History proves how dangerous it is to have the government regulate fake news." *The Washington Post*, October 3. www.washingtonpost.com/news/theworldpost/wp/2017/10/03/history-proves-how-dangerous-it-is-to-have-the-government-regulate-fake-news/?utm_term=.169d0f10f89d.

Satariano, Adam. 2018. "Ireland's Abortion Referendum Becomes a Test for Facebook and Google." *The New York Times,* May 25. www.nytimes.com/2018/05/25/technology/ireland-abortion-vote-facebook-google.html?rref=collection/sectioncollection/technology.

Schroepfer, Mike. 2018. "An Update on Our Plans to Restrict Data Access on Facebook." Facebook News, April 4. https://newsroom.fb.com/news/2018/04/restricting-data-access/.

Silverman, Craig and Jeremy Singer-Vine. 2018. "Here's Who's Been Blocked By Twitter's Country-Specific Censorship Program." BuzzFeed News, January 24. www.buzzfeed.com/craigsilverman/country-withheld-twitter-accounts?utm_term=.tuVxx346M#.rq8llJeLM.

South, Jeff. 2018. "More than 1,000 U.S. news sites are still unavailable in Europe, two months after GDPR took effect." NiemanLab, August 7. www.niemanlab.org/2018/08/more-than-1000-u-s-news-sites-are-still-unavailable-in-europe-two-months-after-gdpr-took-effect/.

Stephens, Hugh. 2017. "UNESCO's Guidelines on Contemporary Culture in the Digital Environment: Worth Thinking About." *Hugh Stephens Blog*, September 18. https://hughstephensblog.net/2017/09/18/unescos-guidelines-on-contemporary-culture-in-the-digital-environment-worth-thinking-about/.

Strickling, Lawrence E. and Grace M. Abuhamad. 2017. "The Feasibility of Expanding the Use of Multistakeholder Approaches for Internet Governance: Final Report to the Internet Society." *The Internet Society*, October 26. https://internetsociety.org/wp-content/uploads/2017/10/Feasaibility-Study-Final-Report-Oct-2017.pdf.

Summers, Hannah and Nicola Slawson. 2018. "Investigators complete seven-hour Cambridge Analytica HQ search." *The Guardian,* March 24. www.theguardian.com/news/2018/mar/23/judge-grants-search-warrant-for-cambridge-analyticas-offices.

Tan, Rebecca. 2017. "Terrorists' love for Telegram, explained." *Vox*, June 30. www.vox.com/world/2017/6/30/15886506/terrorism-isis-telegram-social-media-russia-pavel-durov-twitter.

Tanash, Rima S., Zhouhan Chen, Tanmay Thakur, Dan S. Wallach and Devika Subramanian. 2015. "Known Unknowns: An Analysis of Twitter Censorship in Turkey." In *WPES '15: Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society*, 11–20. Association for Computing Machinery Conference on Computer and Communications Security, Workshop on Privacy in the Electronic Society, Denver, Colorado, October 12–16. https://pdfs.semanticscholar.org/1cea/7416ec8a7d862f570f759a69421dd4a70f34.pdf.

The News Literacy Project. 2017. "Welcome news from Facebook on the fake news front." NLP Updates, April 6. https://newslit.org/updates/welcome-news-from-facebook-on-the-fake-news-front/.

The Trust Project. 2017. "News with integrity." https://thetrustproject.org/.

Timberg, Craig, Elizabeth Dwoskin, Matt Zapotosky and Devlin Barrett. 2018. "Facebook's disclosures under scrutiny as federal agencies join probe of tech giant's role in sharing data with Cambridge Analytica." *The Washington Post*, July 2. www.washingtonpost.com/technology/2018/07/02/federal-investigators-broaden-focus-facebooks-role-sharing-data-with-cambridge-analytica-examining-statements-tech-giant/?utm_term=.7d1fc8158e62.

UNESCO. 2017. "Draft operational guidelines on the implementation of the Convention in the digital environment." DCE/17/6.CP/11, June 15. https://en.unesco.org/creativity/sites/creativity/files/sessions/6cp_11_do_numerique_en.pdf.

United Nations General Assembly. 2012. *Human Rights Council, Twentieth Session, Agenda item 3: Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development.* A/HRC/20/L.13, June 29. http://ap.ohchr.org/documents/E/HRC/d_res_dec/A_HRC_20_L13.doc.

Vanian, Jonathan. 2018. "Everything to Know About Facebook's Big News Feed Change." *Fortune*, January 12. http://fortune.com/2018/01/12/facebook-news-feed-change/.

Wakabayashi, Daisuke. 2018. "YouTube Adds More Scrutiny to Top-Tier Videos." *The New York Times*, January 16. www.nytimes.com/2018/01/16/technology/youtube-ads-scrutiny.html?_r=0.

Walker, Kent. 2018. "Supporting election integrity through greater advertising transparency." *Public Policy* (blog), May 4. www.blog.google/outreach-initiatives/public-policy/supporting-election-integrity-through-greater-advertising-transparency/.

Weiss, Brennan. 2017. "From 'crazy' to 'regret' — here's how Facebook's positions on Russian interference evolved over time." *Business Insider*, November 1. www.businessinsider.com/facebook-changing-statements-russian-meddling-2016-election-2017-11/#november-10-2016-mark-zuckerberg-dismisses-russias-influence-1.

Wingfield, Nick, Mike Isaac and Katie Benner. 2016. "Google and Facebook Take Aim at Fake News Sites." *The New York Times*, November 14. www.nytimes.com/2016/11/15/technology/google-will-ban-websites-that-host-fake-news-from-using-its-ad-service.html.

Winter, Chase. 2018. "AfD politician 'censored' under new German hate speech law for anti-Muslim tweet." *Deutsche Welle News*, January 2. http://p.dw.com/p/2qCDH.Wong.

Zuckerberg, Mark. 2018a. "Every year I take on a personal challenge to learn something new. I've visited every US state, run 365 miles, built an AI for my home, read 25 books..." Facebook status update, January 4. www.facebook.com/zuck/posts/10104380170714571.

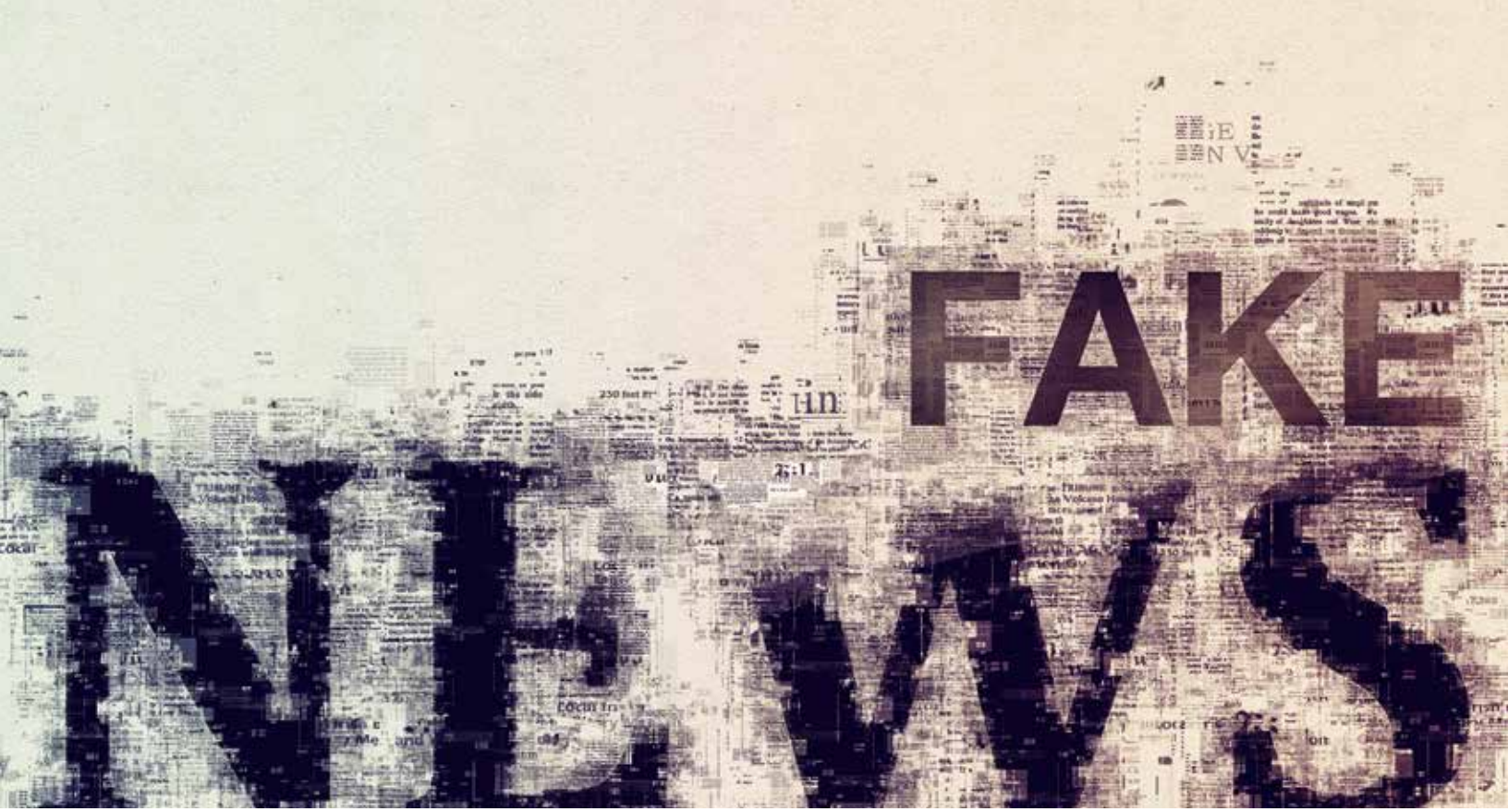
———. 2018b. "One of our big focus areas for 2018 is making sure the time we all spend on Facebook is time well spent." Facebook status update, January 11. www.facebook.com/zuck/posts/10104413015393571.

———. 2018c. "Continuing our focus for 2018 to make sure the time we all spend on Facebook is time well spent..." Facebook status update, January 19. www.facebook.com/zuck/posts/10104445245963251.

About the Authors

Bill Graham is a senior fellow with CIGI, contributing to the Global Security & Politics Program research on internet governance. Most recently, Bill was a contributing author of the GCIG report *One Internet.* Bill served on the board of ICANN from 2011 to 2014. From 2007 to 2011, he was a senior executive with the Internet Society (ISOC), responsible for expanding its engagement in international organizations involved in internet policy and technical issues, including the United Nations, the OECD, the World Intellectual Property Organization and the International Telecommunication Union. Bill was a founding member of the Internet Governance Forum's Multistakeholder Advisory Group from 2006 to 2012. Prior to joining ISOC, Bill was director of international telecommunications policy in the Government of Canada, heading Canada's delegation to the UN World Summits on the Information Society and leading Canada's participation in a range of bilateral, regional and international telecommunication policy organizations. He holds a master's degree in public administration and a B.A. in Pacific studies from the University of Victoria.

Stephanie MacLellan is a senior research associate with CIGI. She joined CIGI's Global Security & Politics Program in July 2016 and specializes in cyber security, digital rights and internet governance. She spent more than a decade working as an editor and reporter for newspapers such as the *Toronto Star*, *The Hamilton Spectator* and *The Slovak Spectator*, an English-language weekly based in Bratislava, Slovakia. Her work has been nominated for three National Newspaper Awards. She holds a bachelor of journalism degree from Carleton University and a master's degree in global affairs from the Munk School of Global Affairs.



Protecting Free Expression, Access to Diverse Information and Democratic Engagement Online: Conceptual and Practical Challenges

Suzanne Nossel and Viktorya Vilk

Introduction

In its early years, the internet was celebrated for ushering in an era of unprecedented freedom of expression and offering the world's largest and most open marketplace of ideas. Anyone with a networked device could express themselves publicly and reach audiences anywhere, at least in theory, and the global digital ecosystem could accommodate a virtually infinite quantity of content. As University of North Carolina sociologist Zeynep Tufecki (2018, para. 12) points out: "In the 21st century, the capacity to spread ideas and reach an audience is no longer limited by access to expensive, centralized broadcasting infrastructure. It's limited instead by one's ability to garner and distribute attention." Communication and information flows have shifted online, and the volume of content, the speed at which it travels, its permanence, its broad accessibility and the potential anonymity of its creators all have profound implications for free expression and open discourse.

The global character of the internet, coupled with the absence of a centralized gatekeeper, has fuelled its growth, innovation and influence, but also greatly complicates its regulation. The digital realm has been described alternatively as a new public square, a private domain and a privately owned public space (Jørgensen 2018). A great deal of digital content, and the majority of the technology companies that publish or host it, are increasingly crossing national borders. Yet, a growing number of states, led by China, are advancing the concept of cyber sovereignty — sweeping control over the digital space within their national borders. By blocking access to international digital platforms and dominating domestic digital platforms, authoritarian regimes have been able to integrate their powers of intelligence gathering, surveillance and censorship into

the digital realm. Even democratic societies with relatively strong protections for free speech are developing laws and policies to restrict content and hold international technology companies liable for content published by users on their platforms.

At the same time, a handful of national and international technology companies are consolidating power, acting as potent gatekeepers and brokers. As states delegate responsibility to assess and restrict content to companies, which are under no legal obligation to protect free expression, censorship is privatized. Meanwhile, the business model underpinning many of these private companies — the collection, analysis and sale of user data for micro-targeted advertising — is proving ripe for mass abuse. Big data can be, and has been, appropriated for state surveillance and used to manipulate civil discourse and the political process within and across national borders.

Acts of free expression online can also, paradoxically, impinge upon the expression of others. The deliberate dissemination of fraudulent information and propaganda online undermines the fundamental human right to receive and impart information. Online harassment, threats and hateful speech interfere with the right to self-expression; silencing voices that are often already marginalized and fostering self-censorship. If the ideas that spread the farthest and fastest are those best able to attract attention, then speech that drowns out other speech or shuts down other speakers poses a threat to free expression, civil discourse and the diversity of viewpoints online.

Determining which content in the digital realm is harmful and dangerous, how it should be restricted and by whom is exceedingly complex. Nearly every attempt to restrict content, no matter how necessary or just, represents a constraint on free expression and has the potential to backfire.

Threats to freedom of expression online are varied and multi-vectored. Some of the most serious concerns are outlined briefly in the following sections.

Cyber Sovereignty and Centralizing Internet Governance

The internet's open and decentralized system of governance laid the foundation for rapid expansion and staggering innovation, but that system is being challenged by an alternative concept: cyber sovereignty. Mirroring the concept of territorial sovereignty, cyber sovereignty asserts the right of each state to control the internet within its own borders. The term originated in a white paper, "The Internet in China," published by the Chinese government in 2010 (Government of the People's Republic of China 2010), which laid the groundwork for the country's restrictive cyber laws and policies (Blomquist 2017).

A leader in advocating for cyber sovereignty on a global scale, China provides an illustrative case study. As of 2017, China became the world's largest internet market with more than 720 million internet users (United Nations Educational, Scientific and Cultural Organization [UNESCO] 2016) and several of the world's top 10 internet companies (Mozur 2017). The state has cultivated a robust, lively, multi-dimensional social media universe that masks tightly policed parameters. The state tightly controls what content can be accessed online through a combination of regulatory and technological filters referred to collectively as the Great Firewall (Barmé and Ye 1997). Many international digital platforms are banned, and China has developed its own digital giants instead.[1] Domestic digital platforms enjoy a symbiotic relationship with the state, with surveillance and censorship fully integrated into the system. To break into this new market, international companies are required to comply with national regulation and must, unavoidably, become complicit in, or directly enforce, state-mandated censorship. A Chinese cyber security law, passed in 2017, requires international companies to store their data within the country, cooperate with any security or criminal investigations and undergo security spot-checks. The new law also forces individual users to register their real names to use messaging services (Wee 2017).

Having laid the foundation domestically, China is now advocating for the concept of cyber sovereignty on the international stage, including through its annual World Internet Conference, a forum at which Chinese concepts of internet regulation are presented to international officials and executives. In the void left by the current US government, which has pulled back from its prior role of defending an open internet, China is stepping in to propose an alternative model that holds appeal for other authoritarian regimes.

Disruption and Blocking of Digital Platforms

As digital platforms become the primary means of communication and information dissemination, states have an unprecedented ability to take down the whole system. Internet-wide blackouts are becoming a favourite strategy for suppressing opposition and protest, used by 19 out of the 65 countries tracked by Freedom House for its *Freedom of the Net* report in 2017. That number has more than doubled since 2015 and includes countries across Africa and the Middle East. Interfering with mobile connectivity and throttling bandwidth at a local level, particularly in regions populated by opposition parties or ethnic or

1 These include: Tencent, Alibaba and Baidu (Google), as well as Youku Tudou (YouTube), weibo.com (Twitter), Renren and WeChat (Facebook) (PEN America 2018, 13).

religious minorities, can effectively seal off resistance and paralyze society (Kelly et al. 2017). Within their own borders, states can refuse technology companies access to their markets, as China has done with Google, Twitter, Facebook and countless others. More targeted strategies include temporarily blocking specific messaging apps, such as Facebook Messenger and WhatsApp[2] and outlawing virtual private networks (VPNs) and TORs, which protect user privacy and enable the circumnavigation of content filters.[3]

State Regulation of Content on Digital Platforms

States are increasingly holding technology companies — internet service providers (ISPs), search engines, hosting services and social media platforms — legally responsible for the content hosted on their sites. In June 2017, the German parliament passed the Network Enforcement Act, a law requiring digital media companies to remove content that violates the country's strict defamation and hate speech laws within 24 hours to one week, or face fines as high as 50 million euros (Center for Democracy & Technology [CDT] 2017). The United Kingdom (Walker 2018) and the Czech Republic (Noack 2017) have established units to tackle fraudulent news. French President Emmanuel Macron announced plans to introduce a law that requires technology companies to take down fraudulent news during election periods (McAuley 2018). In April 2018, Malaysia passed the Anti-Fake News Act, which sets large fines and up to six years in jail for individuals and online service providers for creating, publishing or circulating "fake" news (Beech 2018). The United States has, to date, taken a highly permissive approach, granting technology companies near-total immunity from legal prosecution for content hosted by their sites.[4] However, the passage of two new bills in 2018, both intended to combat sexual exploitation, has opened the door for digital publishers to be held responsible for hosting certain kinds of content on their platforms (Romano 2018). Fears around the dissemination of misinformation and election interference have spurred a debate about more far-reaching US legislation to regulate the digital realm, content included.[5]

Holding digital platforms responsible for content posted by users poses fundamental conceptual challenges for free expression. There are no universally agreed upon definitions of what constitutes hateful speech or fraudulent news, and methods of online harassment and deception constantly evolve. Legislation that bans content without clearly defining it can lead to overbroad enforcement by companies that may opt to simply remove content that skates anywhere near ill-defined legal lines.

Privatization of Content Restriction

In recent decades, technology companies and the wider telecommunications industry have dramatically consolidated. A handful of private companies reaching nearly every corner of the globe dominate the digital ecosystem. These companies are under no legal obligation to protect free expression. In fact, they are increasingly confronted with government regulations and pressure from civil society to police content defined (often loosely) as hateful speech, fraudulent news, an incitement to violence or a threat to national security or public safety.

To comply with regulations and respond to consumer and advertiser demands, technology companies are rapidly developing and deploying a host of tools and strategies to prioritize and restrict content. The most fundamental and ubiquitous of these tools are the terms of service, which outline the rules and responsibilities of the platform and its users in relation to one another. Within these terms, technology companies can include broad restrictions of their own design and must incorporate restrictions mandated by each jurisdiction. To address content that violates terms of service, digital platforms can suspend or shut down individual accounts, take down websites and flag or remove targeted content. To comb through virtually infinite quantities of content, digital platforms are hiring human content monitors and developing automated content monitors that filter and flag objectionable material, augmenting reliance on individual complaints. They are also refining algorithms to de-prioritize certain kinds of content. In its search engine, Google will now "surface more authoritative pages and demote low-quality content," and Facebook is pushing "low-quality" content to the bottom of its news feed (PEN America 2017a, 32-33, 42). Finally, digital platforms are revamping their ad management systems to vet advertisers more thoroughly in an effort

2 For example, since 2016, Turkey has shut down the internet in parts of the country and temporarily blocked Facebook, Twitter, YouTube and WhatsApp as part of a wider crackdown on dissent (Bulman 2016).

3 In 2017, six countries — including Belarus, China, Egypt, Russia, Turkey and the United Arab Emirates — stepped up efforts to control VPNs, by either passing legislation or blocking associated websites or network traffic (Kelly et al. 2017).

4 Section 230 of the US Communications Decency Act states: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider" (Zara 2017, para. 3).

5 In California, a bill was proposed to make "false or deceptive statements designed to influence the vote" illegal (Kravets 2017, para. 2).

to financially starve purveyors of fraudulent news and hateful speech and block foreign interference in domestic politics.[6]

Given the scale and ubiquity of many of these technology companies, the new tools and strategies they deploy can have potent and unpredictable consequences on free expression. The internal policies established to police and remove content and vet advertisers are largely opaque. The algorithms that prioritize content can be unpredictable and their retooling can lead to unintended and hard-to-detect costs. Human monitors, and the human engineers who build automated monitors, are fallible and susceptible to bias.[7] Creative self-expression — in particular work that is intended to be controversial, hyperbolic, satirical or ironic — is especially susceptible to censorship by human or automated monitors seeking to target offensive content (Chase 2018). When content is taken down, recourse to appeal can be cumbersome, slow and ineffective.[8]

Surveillance and Manipulation of Big Data

Many of the world's most successful digital platforms make a profit by collecting, analyzing and selling user data. For states, this represents a surveillance goldmine. By prioritizing domestic technology companies and forcing international platforms to adhere to restrictive and invasive national laws, authoritarian states are striving to secure unfettered access to private user data and to integrate censorship and intelligence. Through targeted surveillance, states can focus their attention on specific individuals and groups in order to shut down accounts, censor content and gather material for prosecution. While VPNs, TORs and encryption can offer avenues for anonymity and circumnavigation, states have caught on and are making many of these tools illegal.[9] States, including a handful of democracies, are also exerting legal and financial pressure on technology companies to store user data within national borders, hand over encryption keys[10] and reveal private information.[11]

In contrast to the direct approach of targeted surveillance, dragnet surveillance pre-emptively collects the metadata and the content of millions of online communications, which can then be mined for intelligence or information. The documents leaked by Edward Snowden in 2013, and subsequently bolstered by further disclosures, revealed the vast scale of surveillance conducted by the US National Security Agency in cooperation with private American technology companies.[12] A survey conducted in 2013 by PEN America revealed dragnet surveillance revelations had a chilling effect on American writers, academics and thinkers. The respondents were not only more worried about government surveillance than the rest of the population, but many actively engaged in self-censorship as a result: 28 percent reported curtailing social media use, 24 percent reported avoiding discussion of certain topics online or on the telephone and 16 percent avoided writing or speaking about certain topics (PEN America 2013, 3). Pen America conducted a follow-up survey in 2014 and revealed similar results for international writers and found that writers working in free countries with expansive surveillance regimes felt nearly as alarmed by intrusions into their privacy as those living in authoritarian states (PEN America 2015).

Attack and Prosecution of Content Producers

Increasingly, individuals are finding themselves in the crosshairs of the state as a result of some form of expression *online*. Nearly half of PEN America's caseload now centres on writers and artists who have been persecuted and imprisoned for their work in

6 For a more in-depth discussion of tools and strategies used by technology companies to prioritize and restrict content, see PEN America (2017a, 29–49).

7 According to a Pro-Publica report, content monitors operating under Facebook's guidelines for identifying and removing hateful speech deleted the statement "All white people are racist. Start from this reference point, or you've already failed," made by Didi Delgado, a poet and Black Lives Matter activist, but did not remove a post from Clay Higgins, a US Republican congressman from Louisiana, in which he referred to "radicalized Muslims" and said, "hunt them, identify them, and kill them" (Angwin and Grassegger 2017, paras 1–4).

8 For a more in-depth discussion of potential pitfalls of privatized content regulation, see PEN America (2017a, 29–49).

9 In 2017, 14 countries restricted VPNs in some form and another six countries introduced new restrictions (Kelly et al. 2017).

10 At least five countries — China, Hungary, Thailand, the United Kingdom and Vietnam — recently passed or implemented laws that may require companies or individuals to break encryption (Kelly et al. 2017). In 2018, Russia blocked Telegram, among the most popular messaging services across Eastern Europe, for refusing to hand over access to encrypted user communications (Stubbs and Ostroukh 2018).

11 In 2017, web hosting company DreamHost was required to hand over the data of users who visited disruptj20.org, a website that helped coordinate protests on the inauguration of Donald Trump (Hsu 2017).

12 Since 2007, the US National Security Agency has collected internet communications from nearly every major technology company (including Google, Facebook, Yahoo, YouTube, Microsoft, Skype and Apple) through a program code named PRISM (Greenwald and MacAskill 2013).

the digital sphere (Olukotun 2013).[13] In 2017, Freedom House tracked physical attacks in retaliation for online activities in 30 countries, in which there was a 50 percent increase compared to the previous year; in eight of the 30 countries, people were murdered for online activities. Bloggers tackling religious subjects, citizen journalists filming unrest and reporters investigating politics, corruption and crime have proven particularly vulnerable (Kelly et al. 2017, 26). Egypt, China and multiple other authoritarian regimes routinely hand out lengthy prison sentences to content producers and publishers (ibid., 5); in Russia, bloggers with more than 3,000 daily visitors are required to register with the state and comply with mass media laws (ibid., 13).

While attacking, prosecuting and jailing people for self-expression is a tried and true strategy for crushing dissent, the digital realm has exacerbated existing threats and generated new ones. Although digital platforms can offer a measure of anonymity, they also provide difficult-to-evade digital fingerprints that allow determined governments to find virtually anything, and anyone, who they wish to target. The reach of the global digital ecosystem means that national borders do not necessarily offer protection.

The Proliferation of Misinformation and Propaganda

The use of propaganda and deliberate misinformation to influence public opinion is hardly new, but the digital sphere has enabled state and non-state actors to deploy these methods with unprecedented efficiency. Within and across national borders, states have been using intelligence analysts, mercenary trolls and automated bots to flood online platforms with pro-government messages and attacks against opponents, interfering with civil discourse and influencing elections. Content creators take advantage of algorithms that prioritize shares and clicks to pedal sensationalized content and tailor misinformation to specific audiences through micro-targeted advertising. By mimicking the visual language of vetted articles, content creators disguise political ads and fraudulent stories as articles from established news media.[14] Hackers have the capacity to hijack social media accounts and put words into the mouths of public figures, including politicians, intellectuals and other leaders (Kelly et al. 2017). New technologies now make it possible to produce fake videos that can be virtually impossible to detect as inauthentic (Roose 2018).

All of these tools and strategies are increasingly lumped together and labelled "fake news," a term used cavalierly by so many different actors and interest groups that it has nearly lost all meaning.[15] PEN America, in its report *Faking News* (2017a, 23), proposed an alternate term, "fraudulent news," which was defined as "demonstrably false information that is being presented as a factual news report with the intention to deceive the public." Within that definition, the report identified several distinct categories: state-sponsored misinformation, fraudulent news for political motives, fraudulent news for profit and conspiracy theories (ibid.).

International human rights law, as outlined by article 19 of the International Covenant on Civil and Political Rights, specifies the right "to seek, receive and impart information and ideas through any media and regardless of frontiers" (United Nations 1966). The virulence with which fraudulent news spreads online interferes with the fundamental human right to receive information. Fraudulent news sows distrust not only in the media but in sources of legitimization (including fact-checkers and investigations). The inability to distinguish fact from fiction and the erosion of the concept of truth poses an existential threat to civic engagement and to democracy itself (Nossel 2017).

Online Harassment, Threats and Hateful Speech

PEN America defines online harassment as "the repeated or severe targeting of an individual or group in an online setting through harmful behaviours," (PEN America 2017b) an umbrella that includes: cyber stalking; filling comment feeds with violent threats and inflammatory content (trolling); impersonating people to sully their reputations; flooding email inboxes with spam; disseminating sexually explicit images without consent (revenge porn); and myriad other tactics (ibid.).

Because of the anonymity, volume, virality and permanence of content in the digital realm, online harassment can have serious consequences on free expression. Journalists, writers, artists and other creative and media professionals who express themselves publicly online — in particular women, people of colour and people identifying with other marginalized groups — are more likely to be targeted (Duggan 2017). In a 2017 PEN America study on the impact of online harassment on writers and journalists, two-thirds of survey respondents reported experiencing a severe reaction to online harassment,

13 To give just one example, Ye Haiyan, a Chinese feminist, activist and artist, was attacked in her own home, jailed and later evicted after images of one of her public performances went viral online (PEN America 2018, 46–47).

14 For a more in-depth discussion, see PEN America (2017a).

15 Many states and politicians now use the term to discredit any unfavourable or critical media coverage. In his first 100 days in office (between January 20 and April 28, 2017), US President Donald Trump used "fake news" in tweets more than 30 times (Lanktree 2017).

including refraining from publishing their work and/or permanently deleting their social media accounts. Severe online harassment can disrupt communication channels, making it difficult to use email, cell phones and social media accounts. The psychological and emotional trauma it causes can lead to self-censorship. More than one-third of survey respondents reported avoiding certain topics in their writing due to online harassment (Macomber 2018, para. 7). When victims are doxed with the public posting of private information, such as home addresses and social security numbers, online harassment can shift offline and threaten physical safety (Hess 2014; Buni and Chemaly 2014).

Not limited to individuals and interest groups, online harassment is utilized by states and mercenaries to terrorize activists and journalists.[16] Harassment is also crossing national borders as troll farms target individuals and organizations for sport, money or political purposes (Griffith 2018).

Conclusion

Threats to digital expression derive from governments, from powerful digital media platforms and from users engaging in modes of speech that can suppress and silence others. These risks intertwine, and they are evolving as quickly as the digital technologies themselves. Efforts to address one impairment to free expression, such as the spread of online trolling, can open the door to other forms of infringement, including the policing of speech based on ideology and viewpoint. While the bedrock principles of free speech protection enshrined in national laws and international instruments offer a lodestar in terms of the values that should govern in the digital realm, their practical application, the legal precedents that interpret them and the premises and analogies on which they rest are all now being tested.

16 The 2017 *Freedom of the Net* report has documented multiple cases of state-sponsored or sanctioned cyber harassment of activists and journalists, including in Venezuela and Turkey (Kelly et al. 2017).

Works Cited

Angwin, Julia and Hannes Grassegger. 2017. "Facebook's Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children." *ProPublica*, June 28. www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms.

Barmé, Geremie R. and Sang Ye. 1997. "The Great Firewall of China." *Wired*, June 1. www.wired.com/1997/06/china-3/.

Beech, Hannah. 2018. "As Malaysia Moves to Ban 'Fake News,' Worries About Who Decides the Truth." *The New York Times,* April 2. www.nytimes.com/2018/04/02/world/asia/malaysia-fake-news-law.html.

Blomquist, Kayla. 2017. "China's Push for Cyber Sovereignty." *FAO Global*, November 8. www.faoglobal.com/article-chinas-push-for-cyber-sovereignty/#_edn3.

Bulman, May. 2016. "Facebook, Twitter and WhatsApp blocked in Turkey after arrest of opposition leaders." *The Independent*, November 4. www.independent.co.uk/news/world/asia/facebook-twitter-whatsapp-turkey-erdogan-blocked-opposition-leaders-arrested-a7396831.html.

Buni, Catherine and Soraya Chemaly. 2014. "The Unsafety Net: How Social Media Turned Against Women." *The Atlantic*, October 9. www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/.

CDT 2017. "Overview of the NetzDG Network Enforcement Law." July 17. https://cdt.org/insight/overview-of-the-netzdg-network-enforcement-law/.

Chase, Jefferson. 2018. "Facebook slammed for censoring German street artist." *Deutsche Welle*, January 15. www.dw.com/en/facebook-slammed-for-censoring-german-street-artist/a-42155218.

Duggan, Maeve. 2017. "Online Harassment 2017." Pew Research Center, July 11. www.pewinternet.org/2017/07/11/online-harassment-2017/.

Google. 2017. "Google searches for 'fake news' have skyrocketed in recent years." http://trends.google.com/trends/explore?date=today%205-y&geo=US&q=Fake%20News.

Government of the People's Republic of China. 2010. "White paper on the Internet in China." Information Office of the State Council. August 6. www.chinadaily.com.cn/china/2010-06/08/content_9950198.htm.

Greenwald, Glenn and Ewen MacAskill. 2013. "NSA Prism program taps in to user data of Apple, Google and others." *The Guardian*, June 7. www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data.

Griffith, Erin. 2018. "Pro-Gun Russian Bots Flood Twitter after Parkland Shooting." *Wired*, February 15. www.wired.com/story/pro-gun-russian-bots-flood-twitter-after-parkland-shooting/.

Hess, Amanda. 2014. "Why Women Aren't Welcome on the Internet." *Pacific Standard*, January 6. https://psmag.com/social-justice/women-arent-welcome-internet-72170#.147kfrhlj.

Hsu, Tiffany. 2017. "DreamHost Ordered to Release Some Trump Protest Website Data to U.S." *The New York Times*, August 25. www.nytimes.com/2017/08/25/business/dreamhost-trump-doj-privacy-ddos-dailystormer.html.

Jørgensen, Rikke Frank. 2018. "The Privatised Public Sphere." Alexander von Humboldt Institute for Internet and Society, February 18. www.hiig.de/en/the-privatised-public-sphere/.

Kelly, Sanja, Mai Truong, Adrian Shahbaz, Madeline Earp and Jessica White. 2017. *Freedom on the Net 2017*. Freedom House, November. https://freedomhouse.org/report/freedom-net/freedom-net-2017.

Kravets, David. 2017. "Alternative facts alert: Proposed legislation bans 'fake news'." *Ars Technica,* March 28. https://arstechnica.com/tech-policy/2017/03/alternative-facts-alert-proposed-legislation-bans-fake-news/.

Lanktree, Graham. 2017. "Trump's First 100 Days Shows a President Obsessed with 'Fake News' — and Twitter." *Newsweek*, April 28. www.newsweek.com/twitter-trumps-first-100-days-show-president-obsessed-fake-news-591595.

Macomber, Laura. 2018. "Writers and Online Harassment: Evidence of a Chilling Effect." PEN America, April 20. https://pen.org/writers-and-online-harassment-evidence-of-a-chilling-effect/.

McAuley, James. 2018. "France weighs a law to rein in 'fake news,' raising fears for freedom of speech." *The Washington Post*, January 10. www.washingtonpost.com/world/europe/france-weighs-a-law-to-rein-in-fake-news-raising-fears-for-freedom-of-speech/2018/01/10/78256962-f558-11e7-9af7-a50bc3300042_story.html?utm_term=.dcaf52c9750f.

Mozur, Paul. 2017. "The World's Biggest Tech Companies Are No Longer Just American." *The New York Times*, August 17. www.nytimes.com/2017/08/17/business/dealbook/alibaba-sales-revenue-first-quarter-profit.html.

Noack, Rick. 2017. "Czech elections show how difficult it is to fix the fake news problem." *The Washington Post*, October 20. www.washingtonpost.com/news/worldviews/wp/2017/10/20/czech-elections-show-how-difficult-it-is-to-fix-the-fake-news-problem/?utm_term=.aceab7539c2c.

Nossel, Suzanne. 2017. "Introductory Essay: Fraudulent News and the Threat to Free Expression." In *Faking News: Fraudulent News and the Fight for Truth*, PEN America, 8–22. https://pen.org/wp-content/uploads/2017/11/2017-Faking-News-11.2.pdf.

Olukotun, Deji. 2013. "Pen Joins All-Star Coalition to Protect Digital Freedoms." PEN America, October 8. https://pen.org/pen-joins-all-star-coalition-to-protect-digital-freedoms/.

PEN America. 2013. *Chilling Effects: NSA Surveillance Drives U.S. Writers to Self-Censor*. https://pen.org/sites/default/files/2014-08-01_Full Report_Chilling Effects w Color cover-UPDATED.pdf.

———. 2015. *Global Chilling: The Impact of Mass Surveillance on International Writers. Results from PEN's International Survey of Writers*. https://pen.org/sites/default/files/globalchilling_2015.pdf.

———. 2017a. *Faking News: Fraudulent News and the Fight for Truth*. https://pen.org/wp-content/uploads/2017/11/2017-Faking-News-11.2.pdf.

———. 2017b. "Online Harassment: Key Findings from Survey of Writers and Journalists." https://pen.org/online-harassment-survey-key-findings/.

———. 2018. *Forbidden Feeds: Government Controls on Social Media in China*, https://pen.org/wp-content/uploads/2018/03/PENAmerica_Forbidden-Feeds-3.13-3.pdf.

Romano, Aja. 2018. "A new law intended to curb sex trafficking threatens the future of the internet as we know it." *Vox*, April 18. www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom.

Roose, Kevin. 2018. "Here Come the Fake Videos, Too." *The New York Times*, March 4. www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html.

Stubbs, Jack and Andrey Ostroukh. 2018. Russia to ban Telegram messenger over encryption dispute." *Reuters*, April 13. www.reuters.com/article/us-russia-telegram-block/russia-to-ban-telegram-messenger-over-encryption-dispute-idUSKBN1HK10B.

Tufecki, Zeynep. 2018. "It's the (Democracy-Poisoning) Golden Age of Free Speech." *Wired,* January 6. www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/.

UNESCO. 2016. "China, India now world's largest Internet markets." www.unesco.org/new/en/media-services/single-view/news/china_india_now_worlds_largest_internet_markets/.

United Nations. 1966. *International Covenant on Civil and Political Rights*. General Assembly resolution 2200A (XXI) (entered into force 23 March, 1976). www.ohchr.org/Documents/ProfessionalInterest/ccpr.pdf.

Walker, Peter. 2018. "New national security unit set up to tackle fake news in UK." *Guardian*, January 23. www.theguardian.com/politics/2018/jan/23/new-national-security-unit-will-tackle-spread-of-fake-news-in-uk.

Wee, Sui-Lee. 2017. "China's New Cybersecurity Law Leaves Foreign Firms Guessing." *The New York Times*, May 31. www.nytimes.com/2017/05/31/business/china-cybersecurity-law.html.

Zara, Christopher. 2017. "The Most Important Law in Tech Has a Problem." *Wired*, January 3. www.wired.com/2017/01/the-most-important-law-in-tech-has-a-problem/.

About the Authors

Suzanne Nossel is a leading voice on free expression issues in the United States and globally. She is chief executive officer of PEN America, the leading human rights and free expression organization, and was previously chief operating officer of Human Rights Watch and executive director of Amnesty International USA. She is a veteran of both the Obama and Clinton administrations, most recently serving as deputy assistant secretary of state for international organizations under President Obama. Nossel coined the term "Smart Power," which was the title of a 2004 article she published in *Foreign Affairs Magazine* and later became the theme of Secretary of State Hillary Clinton's tenure in office. Nossel serves on the board of directors of the Tides Foundation and was a former senior fellow at the Century Foundation, the Center for American Progress and the Council on Foreign Relations. She is a *magna cum laude* graduate of both Harvard College and Harvard Law School.

Viktorya Vilk is chief of staff and manager of special projects at PEN America, where she works closely with leadership on organizational governance and development and manages initiatives around a range of free expression issues, including online harassment of writers and journalists, media transparency and information deserts. She has nearly a decade of experience working in museums and arts nonprofits expanding access to the arts and defending freedom of expression. She graduated *summa cum laude* with a B.A. from Boston University and completed graduate degrees as a Marshall Scholar at the Courtauld Institute of Art, University of London.

Are Recent Governmental Initiatives to Combat Online Hate Speech, Extremism and Fraudulent News Consistent with the International Human Rights Law Regime?

Evelyn Mary Aswad

Executive Summary

There have been a variety of high-profile European governmental and inter-governmental norm-setting initiatives involving freedom of expression online. Indeed, the UN Special Rapporteur on Freedom of Opinion and Expression has expressed concern about a "wave" of European content restrictions (Kaye 2017; Amnesty International 2017, 37–44; Keller 2017).[1] This essay focuses on Europe's Code of Conduct on Countering Illegal Hate Speech Online (the Code) and Germany's Network Enforcement Act (NetzDG law), but the legal analysis is applicable to numerous similar initiatives. The introductory section of this essay provides a summary of the two initiatives. The remainder examines whether these measures are consistent with the international human rights law framework and concludes that they pose serious human rights issues.

Summary of Initiatives

On May 31, 2016, the European Commission, together with several large US information and communication technology (ICT) companies, issued the Code, which requires removal of hate speech.[2] Although there is no universally accepted definition of hate speech (Nossel 2016), the Code's approach is linked to the 2008 European Framework Decision 2008/913/JHA's (the Framework) definition: "conduct publicly inciting to

1 The Amnesty International document describes free expression concerns in countering violent extremism. Keller (2017) discusses the free expression problems with the European Commission's 2017 Communication on Tackling Illegal Content Online.

2 For a more in-depth analysis of the hate speech code of conduct, see Aswad (2016).

violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin" (European Commission 2016, 1). The Framework further defines hate speech more broadly, including certain speech involving denials of historic atrocities (Council of the European Union 2008, art. 1.1 [c]). The breadth of "hate speech" is emphasized by a provision allowing member states to "choose to punish *only* conduct which is either carried out in a manner likely to disturb public order or which is threatening, abusive or *insulting*" (ibid., art. 1.2 [emphasis added]). This provision implies that "hate speech" covers speech that is not likely to affect the peace and mere insults. Criminal sanctions are contemplated for perpetrators (ibid., art. 3). ICT companies pledged to review removal requests against their own guidelines and, as necessary, with national laws that implement the Framework; the companies are to review the majority of notifications within 24 hours and remove illegal content (European Commission 2016, 2). Civil society groups criticized the Code as endangering freedom of expression and lamented their exclusion from its drafting process (Llansó 2016; Access Now 2016). The European Commission's most recent evaluation of implementation noted "IT companies removed on average 70% of illegal hate speech notified to them" and "companies are now increasingly fulfilling their commitment to remove the majority of illegal hate speech within 24 hours" (European Commission 2018, paras 1, 2).

Germany's 2017 NetzDG Law, which took full effect at the beginning of 2018, requires large social media companies (i.e., those with more than two million registered users in Germany) to develop procedures to review complaints and remove illegal speech within certain brief time periods (BBC News 2018; Germany 2017). Noncompliance with the law can result in hefty fines of up to 50 million euros. The law requires removal of (or blocking of access to) "manifestly unlawful" content within 24 hours of receiving complaints (Germany 2017, sec. 3 [2]). It requires removal of (or blocking of access to) "unlawful" content within seven days of receiving complaints (with extra time in certain situations) (ibid., sec. 3 [3]). Illegal speech is defined by reference to existing provisions in the German criminal code, including bans on defamation of religions, denial of historic atrocities, depictions of violence and insults (ibid., sec. 1 [3]; Article 19 2017, 14). Civil society has criticized the law for incentivizing corporate removal of content, as well as not providing for meaningful judicial adjudication of rights or remedies (ibid, 2; Human Rights Watch 2018).[3] There were some widely reported inappropriate removals of speech since the law went into full effect, and several large political parties in Germany are against the law (Human Rights Watch 2018).

3 The Global Network Initiative (GNI) (2017) also criticized an earlier version of the law.

Relevant International Human Rights Law Framework

Summary of the Framework

The key international human rights treaty for purposes of freedom of expression is the International Covenant on Civil and Political Rights (ICCPR), which has 172 state parties, including all 47 members of the Council of Europe, Canada and the United States (United Nations 1966; 2018).[4] Article 19 provides for a broad right to seek and receive information of all kinds, regardless of frontiers and through any media. It permits states to limit speech when a three-pronged test is met. To be valid, speech restrictions must be:

- "provided by law" (i.e., properly promulgated and not vague);
- "necessary" (i.e., the speech restriction must, *inter alia*, be the least intrusive means of achieving governmental purposes); and
- imposed for an enumerated legitimate government objective (i.e., protection of the rights or reputations of others, national security, public order, public health or morals) (United Nations 1966, art. 19 [3]).[5]

The burden is on the government to prove that any limitation on speech, including hate speech, extremist speech and fraudulent news,[6] meets article 19's tripartite test (United Nations 2011, para. 35).

The ICCPR also contains article 20 (2), which mandates bans on speech for "[a]ny advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence." The scope of this provision remains the subject of much controversy. For example, a 2006 report by the UN High Commissioner for Human Rights found that there was no consensus among nation states about the meaning of three key terms in this article: "incitement," "hatred" and "hostility" (UN Human Rights Council 2006, paras 3, 5). The United Nations subsequently convened experts in four regional workshops to propose a way forward on the scope of article 20, which culminated in the Rabat Plan of Action, but it was not endorsed by states (United Nations 2013a, para. 1). Although the scope of article 20 remains under discussion, the UN Human Rights Committee has made clear that any restrictions

4 Some of these countries have reservations, understandings and/or declarations with respect to their ICCPR obligations.

5 The interpretations of the tripartite test come from the UN Human Rights Committee, the body charged with monitoring implementation of the ICCPR (United Nations 2011, paras. 25–34).

6 This essay uses the phrase "fraudulent news" as defined by PEN America: "demonstrably false information that is being presented as a factual news report with the intention to deceive the public" (Pen America 2017, 23).

under article 20 must still meet the tripartite test of article 19 (United Nations 2011, paras 50–52).[7]

Application of the Framework

Whether the Code and NetzDG law are imposed under ICCPR articles 19 or 20, the tripartite test must be met. This analysis assumes (without taking a position) that the third prong of the ICCPR's tripartite test is met, i.e., that legitimate governmental reasons exist for the initiatives under review in this essay.[8] The analysis, therefore, focuses on whether the initiatives are unduly vague and if they constitute the least intrusive means to achieve legitimate governmental objectives.

Do the initiatives contain vague language?

- *Europe's Hate Speech Code of Conduct:* The Code requires ICT companies to remove "hate speech," which encompasses speech that, among other things, is merely insulting and speech that is not likely to affect the public order (Council of the European Union 2008, art. 1.2). Such concepts are quite vague and likely will not meet the first prong of the ICCPR's article 19 (3) tripartite test. Even the UN Special Rapporteur on Freedom of Opinion and Expression highlighted in 2016 that European human rights law fails to "define hate speech adequately" (United Nations 2016, para. 25).
- *Germany's NetzDG Law*: This law is set up to implement removal of speech that violates a variety of provisions in Germany's criminal code. Some of these provisions involve amorphous concepts. For example, one provision criminalizes the "defamation of religions." Although the concept of blasphemy is an inherently ambiguous one (one person's blasphemy is another's truth), Germany's particular ban was recently ranked as one of the most vague blasphemy laws in the world (Fiss and Kestenbaum 2017, 24). Banning speech that "defames" *religions* (as opposed to *individuals*) has also been rejected by the UN Human Rights Committee (United Nations 2011, para. 48).[9]

Do the speech restrictions reflect the "least intrusive means" to achieving legitimate governmental ends?

A number of serious questions exist about whether European governments have met their burden[10] of proving that the various bans on speech constitute the "least intrusive means" to achieve legitimate governmental objectives. Some of these questions include:

- *Incentives to Censor Speech:* Can these initiatives reflect the "least intrusive means" of achieving governmental goals when they appear to be based on foundations that encourage censoring speech? For example, they are set up in ways that incentivize the implementers of these initiatives (i.e., social media companies) to be overly assertive in banning speech or face serious repercussions. In addition to short time frames for removals, there are serious repercussions for under-implementation (for example, the threat of EU regulation in the case of the Code or the fines under the German law), but there are few repercussions for overly active censorship. In addition, the initiatives seem to assume social media companies have automated technology that can accurately, quickly and independently spot "illegal content" in all contexts when, in reality, as discussed in a recent report by the Center for Democracy & Technology, it is wrong to "assume that automated technology can accomplish on a large scale the kind of nuanced analysis that humans can accomplish on a small scale" (Duarte, Llansó and Loup 2017, 3). As set up, these initiatives risk removals of broad swaths of speech rather than the least amount to achieve governmental objectives.
- *Criminal Sanctions*: Both initiatives enlist companies to help governments enforce criminal — rather than civil — sanctions on speech violations. The UN's human rights machinery has been skeptical of the appropriateness of criminal sanctions for speech violations, and it is unlikely that criminal sanctions, for some of the speech violations contemplated in these initiatives, would be viewed as the least intrusive means of achieving governmental objectives (United Nations 2011, para. 47).
- *Outsourcing Speech Decisions to Private Companies*: There are also serious concerns with the initiatives' privatization of law enforcement, the limited judicial role in making determinations about the legality of speech, the lack of meaningful remedies or appeals for improper take downs and the lack

7 The UN Convention on the Elimination of Racial Discrimination requires that states parties, with "due regard" to other human rights, including freedom of expression, prohibit racist hate speech (United Nations 1965, art. 4). The UN Committee charged with monitoring implementation of this treaty has stated that such restrictions must also pass ICCPR article 19's tripartite test (United Nations 2013b, paras. 8, 12, 19).

8 When asserting a legitimate reason for restricting speech, governments must properly assess the scope of the problem they are seeking to solve. For example, when trying to combat fraudulent news, it is important to assess properly the nature/scope of this issue. See, for example, Carey (2018).

9 The German law's proviso that speech that defames religion is actionable when it is likely to breach the peace does not sufficiently narrow its scope to be consistent with the requirements of human rights law for incitement, which does not allow for a heckler's veto.

10 The UN Human Rights Committee has stated that: "[w]hen a State party invokes a legitimate ground for restriction of freedom of expression, it must demonstrate in specific and individualized fashion the precise nature of the threat, and the necessity and proportionality of the specific action taken, in particular by establishing a direct and immediate connection between the expression and the threat" (United Nations 2011, para. 35).

of transparency in the standards applied (Human Rights Watch 2018).[11] Can initiatives with such levels of "outsourcing" represent the least intrusive means for dealing with governmental concerns about hate speech, extremism and fraudulent news?

→ *Other Means of Achieving Governmental Objectives*: These concerns tend to fall into two baskets: do these types of governmental speech bans work to create societies that are tolerant,[12] resistant to radicalization and savvy about political disinformation; and whether there are alternative (equally or even more effective) means of achieving such objectives short of such broad speech bans. With regard to the first basket, serious questions that have been raised include whether bans cause dangerous ideas to fester underground (or in an online world on alternative smaller platforms) (Plucinska 2018; Schulberg, Liebelson and Craggs 2017) and, if so, are they counterproductive? Do speech bans create other unintended consequences, including with respect to free speech? (Price 2017; Berger and Morgan 2015, 54–58).[13] Might speech bans raise the profile of ideas and create opinion/speech martyrs (Oltermann 2018)?[14]

Even if broad speech bans are deemed to work, are they the least intrusive means of accomplishing the governmental goals? The international community recently wrestled with this second basket of concerns, (i.e., identifying means of achieving governmental objectives short of speech bans), in the context of religious intolerance and hate. Long-standing annual UN resolutions by the Organization of Islamic Cooperation that called for bans on the "defamation of religions" were replaced in 2011 by Human Rights Council Resolution 16/18, which called for implementation of a proactive and time-proven governmental toolkit to promote religious tolerance that rejected broad bans on blasphemy/defamation of religion (UN Human Rights Council 2011).[15] Serious questions are being raised about whether new toolkits could provide a better means for achieving the ends contemplated in the Code and NetzDG law. For example, in a recent report, PEN America thoughtfully developed a robust toolkit for combatting fraudulent news without broad speech bans (PEN America 2017).

Given such serious concerns with the Code and NetzDG law, it is unlikely that governments have met their burden of showing these initiatives meet the "least intrusive means" test.

Differentiating between Regional and International Law

It is important to note that the level of protection under international human rights treaties and under regional human rights treaties is not always the same, although the language of the treaties may be quite similar. European countries may reasonably believe many of their actions could be upheld in their own regional human rights system, but this does not release them from their international human rights obligations under the ICCPR. The European Convention on Human Rights (ECHR) article 10 has similar language regarding freedom of expression as the ICCPR article 19 (Council of Europe 1950, art. 10), but the ECHR article 10 has been interpreted by the European Court on Human Rights in ways that depart significantly from ICCPR interpretations. For example, the European Court has upheld criminal sanctions for Holocaust denial without engaging in a serious analysis of the tripartite test because it deemed the offensive speech unworthy of Convention protections (*Garaudy v. France* 2003). The UN Human Rights Committee, on the other hand, has stated that the ICCPR does not condone general prohibitions on denials of historic facts (United Nations 2011, para. 49).

Similarly, the European Court and the UN Human Rights Committee have approached blasphemy differently. For example, the European Court upheld Austria's decision to engage in prior censorship of a film that dealt with Christian beliefs in a highly offensive manner because it was "disparaging religious doctrines" (*Otto-Preminger-Institut v. Austria* 1994, para. 11). The court held that protecting citizens from having their religious feelings insulted was a legitimate government purpose (ibid., para. 48) and banning the film was necessary as it could have offended the

11 The United Nations and regional freedom of expression experts have jointly taken the position that "[i]ntermediaries should never be liable for any third party content relating to those services unless they specifically intervene in that content or refuse to obey an order adopted in accordance with due process guarantees by an independent, impartial, authoritative oversight body (such as a court) to remove it and they have the technical capacity to do that" (Organization for Security and Co-operation in Europe [OSCE] 2017, para. 1[d]).

12 Such issues were prominently raised by President Barack Obama in 2012 at the UN General Assembly when the United States was criticized for not banning the *Innocence of Muslims* video. President Obama discussed the dangers of empowering governments to ban speech, as well as the potential futility of speech bans in a digital age (The White House 2012).

13 J.M. Berger and Jonathon Morgan (2015) explain that account suspensions could result in potential loss of key information for law enforcement and terror networks could turn insular, reducing de-radicalizing influences.

14 Oltermann (2018) reports Germany's largest newspaper called for the repeal of the NetzDG law and noted concerns it creates "opinion martyrs."

15 This UN resolution calls on states to promote inter-religious dialogues, educational initiatives, government outreach to minorities, implementation of discrimination and hate crime laws and other actions rather than broad speech bans to promote religious tolerance.

majority Catholic population and thus disturbed the peace (although it cited to no evidence in reaching this conclusion) (ibid., para. 42).[16] The UN Human Rights Committee, on the other hand, has stated that prohibitions on lack of respect for religions or beliefs are generally incompatible with the ICCPR unless they meet the high standard in article 20 (2) and other treaty provisions such as article 19's tripartite test (United Nations 2011, para. 48). In addition, the European Court applies the "margin of appreciation" (i.e., a measure of deference to local authorities, particularly where there is no European consensus on an issue) in freedom of expression cases; the UN Human Rights Committee does not (ibid., para. 36).

Conclusion

The norm-setting initiatives reviewed in this essay raise serious concerns in terms of ICCPR article 19's tripartite test. They contain language that is unduly vague, and the governments, also, do not appear to have met their burden of showing that the speech bans constitute the least intrusive means of achieving their objectives. An open, thorough and ongoing dialogue is needed among governments, civil society, international organizations and companies to ascertain the nature/scope of the underlying problems (for example, fraudulent news, intolerance and extremism) that governments are trying to address and to assess properly the range of potential solutions short of broad governmental speech bans enforced through private companies.

16 Although this is a case from 1994, the court's existing overview of its religious freedom jurisprudence continues to reference this case as good law (European Court of Human Rights 2013).

Works Cited

Access Now. 2016. EDRi and Access Now Withdraw from EU Commission Discussions. May 31. www.accessnow.org/edri-access-now-withdraw-eu-commission-forum-discussions/.

Amnesty International. 2017. *Dangerously Disproportionate: The Ever-expanding National Security State in Europe.* EUR 01/5342/2017. www.amnesty.org/download/Documents/EUR0153422017ENGLISH.PDF.

Article 19. 2017. *Germany: The Act to Improve Enforcement of the Law in Social Networks*. August. www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf.

Aswad, Evelyn. 2016. "The Role of U.S. Technology Companies as Enforcers of Europe's New Internet Hate Speech Ban." *Columbia Human Rights Law Review* 1:1. http://dx.doi.org/10.2139/ssrn.2829175.

BBC News. 2018. "Germany Starts Enforcing Hate Speech Law." BBC News, January 1. www.bbc.com/news/technology-42510868.

Berger, J.M. and Jonathon Morgan. 2015. "The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter." The Brookings Project on U.S. Relations with the Islamic World, Analysis Paper No. 20, March. www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf.

Carey, Benedict. 2018. "'Fake News': Wide Reach but Little Impact, Study Suggests." *The New York Times*, January 2. www.nytimes.com/2018/01/02/health/fake-news-conservative-liberal.html.

Council of Europe. 1950. Convention for the Protection of Human Rights and Fundamental Freedoms, (European Treaty Series No. 5), UN Treaty Series 213: 221, November 4. Entered into force 3 September, 1953. https://treaties.un.org/doc/Publication/UNTS/Volume%20213/volume-213-I-2889-English.pdf.

Council of the European Union. 2008. Council Framework Decision 2008/913/JHA, Combatting Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law. *Official Journal of the European Union* 328/55, December 6.

Duarte, Natasha, E. Llansó and Anna Loup. 2017. "Mixed Messages? The Limits of Automated Social Media Content Analysis." Center for Democracy & Technology. November. https://cdt.org/insight/mixed-messages-the-limits-of-automated-social-media-content-analysis/.

European Commission. 2016. "Code of Conduct on Countering Illegal Hate Speech Online." May 31. http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300.

———. 2018. "Countering Illegal Hate Speech Online: Commission Initiative Shows Continued Improvement, Further Platforms Join." European Commission Press Release IP/18/261, January 19. http://europa.eu/rapid/press-release_IP-18-261_en.htm.

European Court of Human Rights. 2013. "Overview of the Court's Case-law on Freedom of Religion." Council of Europe, October 31. www.echr.coe.int/documents/research_report_religion_eng.pdf.

Fiss, J. and J. G. Kestenbaum. 2017. "Respecting Rights? Measuring the World's Blasphemy Laws." U.S. Commission on International Religious Freedom, July.

Garaudy v. France. 2003. App. No. 65831/01 European Court of Human Rights, IX 369, June 24. http://hudoc.echr.coe.int/eng?i=002-4830.

Germany. 2017. *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* [*Act to Improve Enforcement of the Law in Social Networks* (Network Enforcement Act)]. Bundesministerium der Justiz und für Verbraucherschutz [Federal Ministry of Justice and Consumer Protection], July 12. www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2 (in English).

GNI. 2017. "Proposed German Legislation Threatens Free Expression Around the World." GNI, April 19. https://globalnetworkinitiative.org/proposed-german-legislation-threatens-free-expression-around-the-world/.

Human Rights Watch. 2018. "Germany: Flawed Social Media Law: NetzDG is Wrong Response to Online Abuse." Human Rights Watch, February 14. www.hrw.org/news/2018/02/14/germany-flawed-social-media-law.

Kaye, David. 2017. "How Europe's New Internet Laws Threaten Freedom of Expression: Recent Regulation Risk Censoring Legitimate Content." *Foreign Affairs*, December 18. www.foreignaffairs.com/articles/europe/2017-12-18/how-europes-new-internet-laws-threaten-freedom-expression.

Keller, Daphne. 2017. "The European Commission, For One, Welcomes Our New Robot Overlords." *Stanford Center for Internet and Society* (blog), October 9. http://cyberlaw.stanford.edu/blog/2017/10/european-commission-one-welcomes-our-new-robot-overlords.

Llansó, Emma J. 2016. "Letter to European Commission on Code of Conduct for 'Illegal' Hate Speech Online." Center for Democracy & Technology, June 3. https://cdt.org/insight/letter-to-european-commissioner-on-code-of-conduct-for-illegal-hate-speech-online/.

Nossel, Suzanne. 2016. "To Fight 'Hate Speech', Stop Talking About It." *The Washington Post*, June 3. www.washingtonpost.com/posteverything/wp/2016/06/03/we-dont-need-laws-banning-hate-speech-because-it-doesnt-exist/?utm_term=.b78feb15ebdf.

Oltermann, Philip. 2018. "Tough New German Law Puts Tech Firms and Free Speech in Spotlight." *The Guardian*, January 5. www.theguardian.com/world/2018/jan/05/tough-new-german-law-puts-tech-firms-and-free-speech-in-spotlight.

OSCE. 2017. "Joint Declaration on Freedom of Expression and 'Fake News', Disinformation and Propaganda." The UN Special Rapporteur on Freedom of Opinion and Expression, in cooperation with: the OSCE Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access to Information. FOM.GAL/3/17. March 3. www.osce.org/fom/302796?download=true.

Otto-Preminger-Institut v. Austria. 1994. 11/1993/406/485. Council of Europe: European Court of Human Rights, 295 (ser A), August 23. http://hudoc.echr.coe.int/eng?i=001-57897.

Pen America. 2017. *Faking News: Fraudulent News and the Fight for Truth*, October 12. https://pen.org/wp-content/uploads/2017/11/2017-Faking-News-11.2.pdf.

Plucinska, Joanna. 2018. "Hate Speech Thrives Underground." *Politico*, February 7. www.politico.eu/article/hate-speech-and-terrorist-content-proliferate-on-web-beyond-eu-reach-experts/.

Price, Rob. 2017. "YouTube's Crackdown on Extremist Content and ISIS is also Hurting Researchers and Journalists." *Business Insider*, August 14. www.businessinsider.com/youtube-crackdown-terrorist-extremist-isis-content-hurting-journalists-researchers-2017-8.

Schulberg, Jessica, Dana Liebelson and Tommy Craggs. 2017. "The NeoNazis are Back Online." *Huffington Post*, October 3. www.huffingtonpost.ca/entry/nazis-are-back-online_us_59d40719e4b06226e3f46941.

The White House. 2012. "Remarks by the President to the UN General Assembly." The White House, President Barack Obama Archives, September 25. https://obamawhitehouse.archives.gov/the-press-office/2012/09/25/remarks-president-un-general-assembly.

United Nations. 1965. *International Convention on the Elimination of All Forms of Racial Discrimination*. UN Treaty Series 660: 195, December 21. Entered into force 4 January, 1969. https://treaties.un.org/doc/Publication/UNTS/Volume%20660/v660.pdf.

———. 1966. *ICCPR*. UN Treaty Series 999: 171. December 19. Entered into force 23 March, 1976. https://treaties.un.org/doc/Publication/UNTS/Volume%20999/v999.pdf.

———. 2011. *ICCPR General Comment No. 34*. UN Human Rights Committee. UN Document CCPR/C/GC/34, September 12. www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.

———. 2013a. *Report of the United Nations High Commissioner for Human Rights on the Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred*. UN Document A/HRC/22/17/Add.4, January 11. www.ohchr.org/Documents/Issues/Opinion/SeminarRabat/Rabat_draft_outcome.pdf.

———. 2013b. *UN Committee on the Elimination of Racial Discrimination, General Recommendation No. 35*. UN Document CERD/C/GC/35, September 26. https://tbinternet.ohchr.org/_layouts/treatybodyexternal/TBSearch.aspx?Lang=en&TreatyID=6&DocTypeID=11.

———. 2016. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression.* UN Document A/71/373, September 6. www.un.org/ga/search/view_doc.asp?symbol=A/71/373.

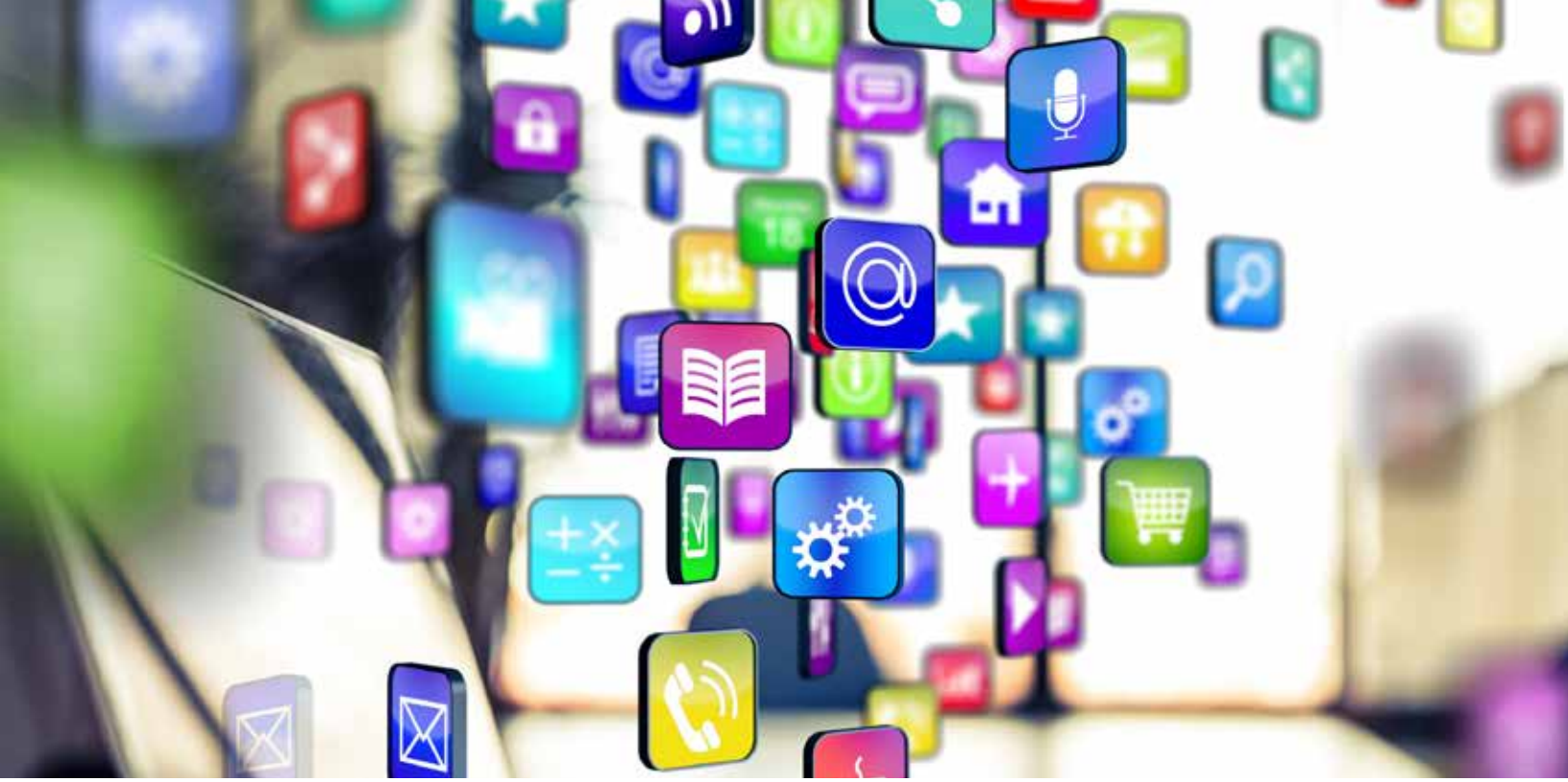
———. 2018. "4. ICCPR" https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-4&chapter=4&clang=_en.

UN Human Rights Council. 2006. *Incitement to Racial and Religious Hatred and the Promotion of Tolerance. Report of the High Commissioner for Human Rights.* UN Document A/HRC/2/6. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G06/139/97/PDF/G0613997.pdf?OpenElement.

———. 2011. *Combating intolerance, negative stereotyping and stigmatization of, and discrimination, incitement to violence and violence against, persons based on religion or belief*. UN Document A/HRC/RES/16/18, April 12. www2.ohchr.org/english/bodies/hrcouncil/docs/16session/A.HRC.RES.16.18_en.pdf.

About the Author

Professor Evelyn Mary Aswad is the Herman G. Kaiser Chair in International Law and director of the Center for International Business & Human Rights at the University of Oklahoma College of Law. Previously, she served as the director of the human rights law office at the US State Department. Prior to joining the State Department, she was an attorney with the law firm of Arnold & Porter in Washington, DC and clerked for the Honourable Arthur Gajarsa on the US Court of Appeals for the Federal Circuit. She has also taught international courses at Georgetown University's School of Foreign Service, as well as its Law Center.



Private Sector Roles and Responsibilities: Protecting Quality of Discourse, Diversity of Content and Civic Engagement on Digital Platforms and Social Media

Rebecca MacKinnon and Roya Pakzad

Amid rising concerns about the spread of disinformation across social media platforms — in particular after the 2016 US presidential elections — a range of stakeholders from around the world have expressed concern that internet platforms wield too much power over the global public discourse and with too little accountability to the public interest. As documented in detail in "Digital Deceit," a recent report by New America's Dipayan Ghosh and Ben Scott (2018), the same advertising technology used by digital marketers to promote products and causes is being abused by actors whose objective is to skew and manipulate public discourse with propaganda and disinformation. Platforms that were once celebrated for giving voice to freedom fighters in the most oppressive countries are now being condemned as the purveyors of cheap weapons of information warfare to the enemies of democracy, human rights and accountable governance.

How should we — how can we — define and operationalize corporate responsibility for democracy's survival?

If we succeed in doing so, it will not be the first time a crisis has brought a range of stakeholders together to forge new and innovative solutions. More than a decade ago, a reputational crisis for global technology companies, combined with the threat of regulation, led to the creation of a corporate responsibility standard for how companies should address government censorship and surveillance demands.

In 2006, Yahoo, Microsoft, Google and Cisco were called to the carpet in the US Congress for their role in aiding Chinese censorship and surveillance (Zeller 2006). Hearings were followed by regulatory proposals that, if implemented, would have resulted in blunt and counter-productive US government interference in corporate operations. Partially in response to the threat of regulation, in 2008, Google, Microsoft and Yahoo joined with civil society, academics and responsible investors to launch the Global Network Initiative (GNI) and, at that time, made public commitments to respect and protect their users' human rights in the face of government surveillance and censorship demands. Facebook and several European telecommunication and equipment companies more recently joined them (GNI 2018).

Even today, platforms' handling of government censorship demands in authoritarian countries is far from problem-free, as the recent censorship of Russia's most popular opposition politician underscores (Washington Post 2018). At least there is a broadly accepted framework, which is grounded in international human rights law and was developed through a contentious multi-stakeholder process over several years, for how companies should address the human rights risks posed by government demands to censor speech, block information flows or hand over user data. Companies have a policy toolbox that includes due diligence and impact assessment, stakeholder engagement, transparency reporting and a commitment to a narrow interpretation of government requests and to push back against (or not comply with) legally invalid demands (GNI 2017). Projects such as New America's Ranking Digital Rights (RDR) and the Electronic Frontier Foundation's Who Has Your Back have emerged to measure and benchmark how well and how comprehensively companies are using such tools (RDR 2018; Reitman 2017).

Now the world's most powerful social media platforms are under fire in Congress and across the democratic world for enabling social media disinformation campaigns whose goal is to destroy and discredit democratic discourse and governance. Many of the regulatory solutions now being proposed, if not implemented — in particular those that increase companies' liability for content appearing on their platforms — are overly blunt, not fit for purpose and fraught with collateral damage for human rights (Human Rights Watch 2018).

Today, many of the key stakeholders in media, civil society, academia and business share an incentive to work together and innovate, lest politicians and companies come under pressure to "do something" quickly and take actions that not only fail to solve the problem but make things worse in the long run.

Other papers in this special report focus on policy responses. This paper focuses on the responses of companies, both unilaterally and in collaboration with other stakeholders. After a broad overview of the actions taken to date, this paper will point to some of the gaps that have been observed by research, including the RDR project, and conclude with recommendations.

Company Responses

Although major companies have chosen different responses, depending on the nature of their audience and their media format, their actions can be divided into the following general categories (note that the examples listed are by no means comprehensive).

Implementing and adhering to trust and safety policies: In the past year, several leading internet companies have demonstrated a commitment to policing false and abusive content by improving trust and safety policies. Both Facebook and Google have announced plans to hire thousands more content moderators tasked with the job of monitoring user-created content that violates their terms of service and community guidelines (Wojcicki 2017; Zuckerberg 2017). The goal of these efforts is to accelerate the response process for reporting and addressing abusive content. In addition to expanding their workforces, companies have also been developing increasingly sophisticated machine-learning techniques to detect fake accounts and violent content. Some companies are trying to engage stakeholders in the improvement of these policies. Twitter, for example, has established a Trust and Safety Council comprised of more than 40 civil society groups and academics from 13 regions (selected by the company itself) to develop best practices to enable their users to express themselves freely and safely on Twitter (Cartes 2016).

Refining algorithms and addressing automation: Updates to "news feed" algorithms have been another important method by which companies have sought to minimize users' exposure to exaggerated headlines, bots and violent or hateful content. Facebook, in particular, has repeatedly tweaked its news feed algorithms in the past two years. Most recently, in a much-heralded rollout of a new formula in January 2018, Facebook claimed that their news feed had been re-designed to ensure that users see more content from friends and family, with the aim of inspiring a greater sense of community and sociability among users (Mosseri 2018). Twitter also developed "safe search" filters and provided an option to users to exclude sensitive content in their search results (Twitter, n.d.(b)). Further steps are being taken to curb automated content posting by bots. For example, Twitter recently announced changes to Tweetdeck and its application protocol interface to prevent the simultaneous posting of the same content from multiple accounts (Roth 2018).

Improving ad management systems: In general, social media and search products have seemingly begun to reverse course away from their emphasis on optimizing feeds primarily on the basis of ad placement and targeting, at least in theory giving users more options over their news feed and search settings. However, the economic model of internet companies remains heavily dependent on advertising. The advancement of algorithmic technologies, fuelled by large quantities of behavioural data, has helped brands track user preferences, profile them in remarkable detail and, some would argue, manipulate them. Because the same ad-targeting methods are used by political propaganda campaigns led by both state and non-state actors, internet companies have begun taking much-needed steps toward verifying ads and confirming the legitimacy and lawfulness of the advertisers who use their services.

As an example, Twitter launched a separate category for political (electioneering) ads. According to Twitter's transparency centre, "to make it clear when you are seeing or engaging with an electioneering ad, we will now require that electioneering advertisers identify their campaigns as such" (Falck 2017, para. 5). Twitter also decided to incorporate this distinction in its product design, adding that "we will also change the look and feel of these ads and include a visual political ad indicator" (ibid.). More recently, in addition to a new safe browsing feature, Google has enabled a built-in ad blocker for Chrome to filter ads that violate the Coalition for Better Ads standard (Roy-Chowdhury 2018).[1]

Educating the public: Internet companies have been working steadily on the task of educating their users about ways to be more cautious about malicious and fake content. One goal of these efforts is to enlist users themselves in the attempt to more rapidly detect and flag this type of content. From developing easy-to-understand community guidelines to warning users about "free followers" apps, public policy departments at companies such as Twitter, LinkedIn and Google have sought to develop new methods for empowering and educating users (Twitter, n.d.(a)). However, progress on this front is still in its early stages and has been subject to some criticism, notably around Facebook's fake news "warning flags," which were discontinued by the company in December 2017 (Lyons 2017).

Partnering with third-party fact checkers: In addition to developing in-house solutions, Facebook partnered with third-party organizations, such as Factcheck.org, to create "warning flags" attached to potentially malicious or untruthful content. Although Facebook has since modified this feature and removed the original format of the warnings, independent fact-checking organizations continue to offer an important public service. In recent months, signatories to the non-partisan International Fact-Checking Network Code of Principles have agreed upon a set of guiding principles to direct their efforts to mitigate the spread of disinformation while preserving users' right to freedom of expression (Poynter Institute 2018).

Supporting independent and local journalism: In January 2017, Facebook initiated the Facebook Journalism Project to better collaborate with the news industry. The company intends to promote new storytelling formats including Instant Article to better engage and inform Facebook users. According to Facebook, other objectives include supporting local news, promoting independent media and improving public news literacy (Simo 2017).

Immediate Gaps to Be Addressed

In assessing the disclosed policies of 22 companies, including Facebook, Twitter, Google and Microsoft, the RDR 2018 Corporate Accountability Index found that companies disclose inadequate information about policies and practices that shape the flow of content on platforms, as well as policies and practices that determine who has access to user information under what circumstances (RDR 2018; Ullman, Reed and MacKinnon 2017).

Lack of transparency around policing content: Through their transparency reports and other public documents, the largest US-based internet platforms disclose a range of information (albeit with varying levels of thoroughness) about their policies for handling government requests to remove or restrict content, as well as the volume and nature of the requests and the percentage of compliance. However, there is little information available about the volume and nature of content removed to enforce companies' private terms of service. The process for enforcing these rules is also opaque. Such opacity has contributed to a lack of understanding on the part of users, policy makers and civil society about how content-policing processes and mechanisms work. This, in turn, not only makes it easier for users to be manipulated but also makes it more difficult for policy makers and other stakeholders to work with companies on effective and constructive solutions.

While RDR does not currently evaluate companies' transparency about their processes for validating advertisers, this is an area where more transparency is badly needed. While companies try to facilitate their advertising services for legitimate brands (such as Facebook Ads Manager or Twitter Promote Mode), they are clearly not transparent enough in their process for

1 Leading international trade associations and companies involved in online media formed the Coalition for Better Ads to improve consumers' experience with online advertising. See Coalition for Better Ads (2018).

validating advertisers. For example, in November 2017, Facebook admitted that the Internet Research Agency, a troll farm in Russia, had bought 3,000 ads for political purposes from June 2015 to May 2017 (Stamos 2017).

Lack of transparency and accountability around collection and sharing of user information: RDR research data shows that companies fail to disclose adequate information to users about how their information is handled: what is collected, how it is used, by whom and for what purposes. In an environment where user information is being collected for the purpose of sharing with advertisers, some of whom we now know use the information to craft disinformation campaigns targeted at people with very specific attributes, the public-interest implications of this lack of transparency have become quite clear.

In addition to transparency about collection, use and sharing of users' information, there is the question of whether company practices should be reined in through regulation if companies do not voluntarily start to give users greater control over what is collected about them and how it is used. Some experts, such as former Organisation for Economic Co-operation and Development Ambassador Karen Kornbluh, are hopeful that Europe's new General Data Protection Regulation (GDPR) will help curb the exploitation of user information to spread political disinformation. She suggests that companies should consider extending many of their GDPR-compliant practices to users in the United States (Kornbluh 2018).

Corporate governance gaps: GNI member companies undergo independent third-party assessment to verify whether they are implementing their commitment to respect user rights when governments come knocking. This includes: corporate oversight over how the company's business operations are affecting users' freedom of expression and privacy; company-wide training on dealing with government demands; and a commitment to carry out human rights impact assessments on business operations and products that are affected by government censorship and surveillance demands around the world. However, the GNI does not require member companies to carry out due diligence or engage with stakeholders on the human rights implications of commercial policies and practices not involving direct government demands. Indeed, RDR's 2018 Index identified only two companies (Oath and Microsoft) that disclose any sort of risk assessments related to terms-of-service enforcement.

RDR does not currently measure whether companies also carry out human rights impact assessments on how technologies and practices related to targeted advertising affect human rights. Nor does it measure whether companies assess the broader social or political and economic impact of their products and services. As of early 2018, researchers found no evidence of such efforts taking place among any companies that were evaluated with the exception of Microsoft's human rights impact assessment process, which is focused on the use of artificial intelligence (Microsoft, n.d.). That could change. On April 10, 2018 Facebook recently announced the launch of a "new initiative to help provide independent, credible research about the role of social media in elections, as well as democracy more generally" (Schrage and Ginsberg 2018). It remains unclear whether or how that research would become part of a systematic and regular impact assessment process, but it has the potential to inform the scope and methodology of effective impact assessments in this area.

Impact assessment will also need to include communities and civil rights groups in new ways. In a Senate hearing on April 10, 2018, New Jersey Senator Cory Booker suggested to Facebook's Chief Executive Officer Mark Zuckerberg that the company work with civil liberties groups to conduct a "civil rights audit" (Booker 2018), echoing a demand by Muslim Advocates and Color of Change to conduct "an aggressive and thorough independent audit of its privacy and security policies and of the civil rights impact of those policies. The audit should explain how Facebook's privacy controls have been implemented, whether risk assessments were thoroughly conducted, and if those assessments included a focus on civil rights violations" (Simpson 2018).

Traditionally, human rights impact assessments have been carried out by experts in corporate social responsibility, public policy and human rights. However, with the proliferation of machine-learning techniques in advertising and content management — and unresolved challenges with regard to black-box models and interoperability — it is also important to engage engineering and product-development groups in evaluating ethical implications around certain products.

Inadequate grievance and remedy mechanisms: More than one billion hours of YouTube videos are watched daily by more than one billion users from more than 88 countries who speak more than 76 languages (YouTube n.d.). Issues, such as the circulation of hateful content against the Rohingya community in Myanmar, misinformation passed between Syrian and Afghan refugees and the large number of bots active on Twitter during the most recent Iranian protests, have proved that the threat of online disinformation and hate is a global issue with global consequences.

Meanwhile, as companies come under growing pressure from governments and other stakeholders to delete hate speech and illegal content, the collateral damage is also growing. For example, several Rohingya activists have reported that their Facebook accounts have been suspended and/or content has been repeatedly deleted. This content ranged from news about military atrocities, military action in the Rakhine state and even a poem about refugees fleeing

military violence (Woodruff 2017). Users in these and other similar situations have repeatedly reported that their attempts to appeal via mechanisms provided by the company were unsuccessful. Media attention or assistance from civil society groups with contacts at the companies appear to be the main way that such cases are addressed. As companies continue to hire more people to evaluate content and experiment with more automated solutions to detect content that violates their rules, in the absence of adequate grievance and remedy mechanisms, the collateral damage for human rights activists and journalists is likely to grow.

What Next?

Addressing all the issues listed above will not solve the entire problem of online disinformation and manipulation. However, tackling clear and pressing gaps in company policy and practice will help bring about more transparency and accountability. That will, in turn, increase the chances that stakeholders have enough information — and sufficient basis for trust — to work with companies on solutions that are publicly accountable and do not produce unintended consequences for the human rights of internet users around the world.

Platforms have grown too powerful, leading some critics to argue that they have become new kinds of monopolies and need to be broken up. These critics believe antitrust law in the United States should be upgraded for the networked age so that it can be used effectively to moderate the platforms' power (Khan 2017; Manjoo 2017). Others suggest that antitrust measures might not be necessary if people are given the mechanisms not just to control the use of their data by companies (as Europe's GDPR aims to do) but to "own" their personal data, so that they can switch more easily from one platform to another and also share in the profits that companies accumulate through the use of their data (Powles 2017; The Economist 2018).

Legal innovation: Legal scholars and practitioners have begun to suggest that legal frameworks around intermediary liability are no longer fit for purpose and that legal innovation is necessary to avoid untenable situations. In Germany, companies are facing impossible regulatory expectations to eliminate all illegal content proactively without inflicting collateral damage on journalism and activism, resulting in what Danielle Citron (2018) calls "censorship creep." Tarleton Gillespie (2018, 19) has suggested new approaches to safe harbour from liability might be one way forward, extending platform responsibility "to second order consequences from the proper working of these systems, not just their misuse." A company could obtain safe harbour from first-order liability for what individuals choose to upload onto its service if it accepts its broader responsibility for how user content is managed and policed. Such responsibility would be demonstrated through concrete measures including transparency reporting, frameworks for shared best practices and standards for moderation, public ombudspeople, expert advisory councils and independent audit processes (Gillespie 2018). Such innovative ideas deserve further exploration. Related recommendations have been made by the UN Special Rapporteur on freedom of expression David Kaye, who also emphasizes the importance of aligning corporate content moderation policies to international human rights standards and implementing grievance and remedy in alignment with the UN Guiding Principles on Business and Human Rights (Freedex 2018).

Business model innovation: As Gosh and Scott (2018) and others (Bradshaw and Howard 2018; Wood and Ravel 2018; Morgan 2018) have pointed out, company measures to address disinformation are merely tweaks on the margins as long as their core business model relies on targeted advertising technology. As the authors of a recent report by the Data & Society Research Institute put it: "The problems associated with "fake news" appear moored to platform corporations' business models" (Caplan, Hanson and Donovan 2018, 27).

For over a generation we have been aware of how society's over-dependence on fossil fuels has contributed to climate change, and still we struggle to make the investments and take the regulatory measures necessary to reduce such dependence. Our society has only just begun to grapple with the question of if and how our information ecosystem's over-dependence on advertising – and especially advertising technology – is affecting democracy. The question is if democracy will still exist within the span of a generation if we do not quickly come to terms with how advertising in the digital information ecosystem affects democratic discourse and take the necessary action. Development of alternatives to fossil fuels has taken large amounts of political and financial risk, substantial investment and hard work on the part of public and private sectors. Similar levels of commitment from all concerned stakeholders will be necessary to design an ecosystem of digital platforms and services that can sustain democracy.

Corporate commitment to democracy: The UN Guiding Principles on Business and Human Rights expects companies to make a public commitment to respect human rights and embed that commitment in their business operations (Office of the United Nations High Commissioner for Human Rights [OHCHR] 2011). If companies are to operate in a manner that supports democracy, perhaps the time has come for them to commit to a related set of guiding principles. Companies that commit to design and operate their products and services in a manner that is supportive of democratic discourse and politics would be expected to carry out impact assessments to better understand

the political effects of their design and business choices. They would work with stakeholders in government, journalism and civil society to ensure that these choices are not inadvertently empowering deliberate attacks against democratic discourse and institutions. Perhaps such a commitment might be related to the safe harbour innovations discussed above.

Ultimately, however, the problem of disinformation will be difficult to address unless public and private institutions and civil society work together more creatively. Healthy, sustainable democracies need robust public interest news and information ecosystems at local, national and global levels.

Authors' Note

Laura Reed, senior research analyst and coordinator for RDR, provided editorial and research support for this paper, which also draws from Ullman, Reed and MacKinnon (2017). See also Rebecca MacKinnon. 2018. "RDR Submission to 2018 Human Rights Council Study on Content Regulation in the Digital Age." *RDR Blog* (blog), January 3. https://rankingdigitalrights.org/2018/01/31/2018-human-rights-council-study-content-regulation.

Works Cited

Booker, Cory. 2018. "Booker Urges Facebook to Enact Broader Reforms." Cory Booker: US Senator for New Jersey press release, May 14. www.booker.senate.gov/?p=press_release&id=792.

Bradshaw, S. and Philip N. Howard. 2018. "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation." Computational Propaganda Research Project, Oxford Internet Institute and University of Oxford. July 20.

Caplan, Robyn, Lauren Hanson and Joan Donovan. 2018. "Dead Reckoning: Navigating Content Moderation After Fake News." Data & Society Research Institute, February. https://datasociety.net/pubs/oh/DataAndSociety_Dead_Reckoning_2018.pdf.

Cartes, Patricia. 2016. "Announcing the Twitter Trust & Safety Council." *Twitter Blog* (blog), February 9. https://blog.twitter.com/official/en_us/a/2016/announcing-the-twitter-trust-safety-council.html.

Citron, Danielle Keats. 2018. "Extremist Speech, Compelled Conformity, and Censorship Creep." *Notre Dame Law Review* 93 (3): 1035; U of Maryland Legal Studies Research Paper No. 2017-12. https://ssrn.com/abstract=2941880.

Coalition for Better Ads. 2018. "The Initial Better Ads Standards." www.betterads.org/standards.

The Economist. 2018. "The techlash against Amazon, Facebook and Google — and what they can do." *The Economist,* January 20. www.economist.com/briefing/2018/01/20/the-techlash-against-amazon-facebook-and-google-and-what-they-can-do.

Falck, Bruce. 2017. "New Transparency for Ads on Twitter." *Twitter Blog* (blog), October 24. https://blog.twitter.com/official/en_us/topics/product/2017/New-Transparency-For-Ads-on-Twitter.html.

Freedex. 2018. "A Human Rights Approach to Platform Content Regulation." https://freedex.org/a-human-rights-approach-to-platform-content-regulation/.

Ghosh, Dipayan and Ben Scott. 2018. "Digital Deceit: The Technologies Behind Precision Propaganda on the Internet." *New America*, January 23. www.newamerica.org/public-interest-technology/policy-papers/digitaldeceit/.

Gillespie, T. 2018. "Platforms are not intermediaries." *Georgetown Law Technology Review* 2: 198. www.georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Gilespie-pp-198-216.pdf.

GNI. 2017. "GNI Principles on Freedom of Expression and Privacy." GNI, April. https://globalnetworkinitiative.org/gin_tnetnoc/uploads/2018/04/GNI-Principles-on-Freedom-of-Expression-and-Privacy.pdf.

———. 2018. "Our Members." GNI. https://globalnetworkinitiative.org/.

Human Rights Watch. 2018. "Germany: Flawed Social Media Law." Human Rights Watch, February 14. www.hrw.org/news/2018/02/14/germany-flawed-social-media-law.

Khan, Lina M. 2017. "Amazon's Antitrust Paradox." *The Yale Law Journal* 126 (3): 710–805.

Kornbluh, Karen. 2018. "Could Europe's New Data Protection Regulation Curb Online Disinformation?" Council on Foreign Relations, February 20. www.cfr.org/blog/could-europes-new-data-protection-regulation-curb-online-disinformation.

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

Manjoo, Farhad. 2017. "Can Washington Stop Big Tech Companies? Don't Bet on It." *The New York Times*, October 25. www.nytimes.com/2017/10/25/technology/regulating-tech-companies.html?mtrref=www.google.com.

Microsoft. n.d. "Microsoft Salient Human Rights Issues: Report - FY17." Microsoft Corporation. http://download.microsoft.com/download/6/9/2/692766EB-D542-49A2-AF27-CC8F9E6D3D54/Microsoft_Salient_Human_Rights_Issues_Report-FY17.pdf.

Morgan, Susan. 2018. "Fake news, disinformation, manipulation and online tactics to undermine democracy." *Journal of Cyber Policy* 3 (1) 39–43. https://doi.org/10.1080/23738871.2018.1462395.

Mosseri, Adam. 2018. "Bringing People Closer Together." Facebook Newsroom, January 11. https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/.

OHCHR. 2011. "Guiding Principles on Business and Human Rights." United Nations. www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

Powles, Julia. 2017. "The EU is right to take on Facebook, but mere fines don't protect us from tech giants." *The Guardian*, May 21. www.theguardian.com/commentisfree/2017/may/20/eu-right-to-take-on-facebook-fines-dont-protect-us-from-tech-giants.

Poynter Institute. 2018. "Commit to transparency – sign up for the International Fact-Checking Network's code of principles." Poynter, July 13. https://ifcncodeofprinciples.poynter.org/.

RDR. 2018. "2018 Corporate Accountability Index." RDR, April. https://rankingdigitalrights.org/index2018/assets/static/download/RDRindex2018report.pdf.

Reitman, Rainey. 2017. "Who Has Your Back? Government Data Requests 2017." Electronic Frontier Foundation, July 10. www.eff.org/who-has-your-back-2017.

Roth, Yoel. 2018. "Automation and the use of multiple accounts." *Twitter Blog* (blog), February 21. https://blog.twitter.com/developer/en_us/topics/tips/2018/automation-and-the-use-of-multiple-accounts.html.

Roy-Chowdhury, Rahul. 2018. "The browser for a web worth protecting." *Google Blog* (blog), February 13. www.blog.google/products/chrome/browser-web-worth-protecting/.

Schrage, Elliot and David Ginsberg. 2018. "Facebook Launches New Initiative to Help Scholars Assess Social Media's Impact on Elections." Facebook Newsroom, April 9. https://newsroom.fb.com/news/2018/04/new-elections-initiative/https://newsroom.fb.com/news/2018/04/new-elections-initiative/.

Simo, Fidji. 2017. "Introducing: The Facebook Journalism Project." *Facebook Blog* (blog), January 11. https://media.fb.com/2017/01/11/facebook-journalism-project/.

Simpson, Scott. 2018. "Muslim Advocates and Color Of Change Demand Independent Civil Rights Audit of Facebook." *Muslim Advocates*, April 13. www.muslimadvocates.org/muslim-advocates-and-color-of-change-demand-independent-civil-rights-audit-of-facebook/.

Stamos, Alex. 2017. "An Update on Information Operations on Facebook." Facebook Newsroom, September 6. https://newsroom.fb.com/news/2017/09/information-operations-update/.

The Washington Post. 2018. "Tech Giants should resist Russia's iron grip of censorship." *The Washington Post*, February 20. http://wapo.st/2EWuBLt?tid=ss_tw&utm_term=.a8ddbc8ecde8.

Twitter. n.d.(a) "About 'free followers' apps." Twitter Help Center. https://help.twitter.com/en/safety-and-security/free-twitter-followers-policy.

———. n.d.(b) "How to use Twitter search." Twitter Help Center. https://help.twitter.com/en/using-twitter/twitter-search.

Ullman, Ilana, Laura Reed and Rebecca MacKinnon. 2017. "Submission to UN Special Rapporteur for Freedom of Expression and Opinion David Kaye: Content Regulation in the Digital Age." RDR, December 15. https://rankingdigitalrights.org/wp-content/uploads/2018/01/RDR-2018-David-Kaye-Submission.pdf.

Wojcicki, Susan. 2017. "Expanding our work against abuse of our platform." *YouTube Official Blog* (blog), December 4. https://youtube.googleblog.com/2017/12/expanding-our-work-against-abuse-of-our.html.

Wood, Abby K. and Ann M. Ravel. 2018. "Fool Me Once: Regulating 'Fake News' and Other Online Advertising." *Southern California Law Review* 91: 1227; USC Legal Studies Research Papers Series No. 261; University of California, Berkeley Public Law Research Paper. https://ssrn.com/abstract=3137311.

Woodruff, Betsy. 2017. "Exclusive: Facebook Silences Rohingya Reports of Ethnic Cleansing." *Daily Beast*, September 18. www.thedailybeast.com/exclusive-rohingya-activists-say-facebook-silences-them?ref=scroll.

YouTube. n.d. "YouTube for Press." www.youtube.com/yt/about/press/.

Zeller Jr., Tom. 2006. "Web Firms Are Grilled on Dealings in China." *The New York Times*, February 16. www.nytimes.com/2006/02/16/technology/web-firms-are-grilled-on-dealings-in-china.html.

Zuckerberg, Mark. 2017. "Over the last few weeks..." May 3. www.facebook.com/zuck/posts/10103695315624661?notif_t=notify_me&notif_id=1493820261300939.

About the Authors

Rebecca MacKinnon directs the RDR project at New America, evaluating internet, mobile and telecommunications companies on their respect for users' privacy, security and freedom of expression. She is co-founder of the citizen media network Global Voices and author of *Consent of the Networked: The Worldwide Struggle For Internet Freedom.* She is on the board of directors of the Committee to Protect Journalists and was a founding board member of the GNI. MacKinnon was CNN's bureau chief and correspondent in China and Japan between 1998 and 2004. She has taught at the University of Hong Kong and the University of Pennsylvania Law School and held fellowships at Harvard University's Shorenstein and Berkman centers, the Open Society Foundations and Princeton University's Center for Information Technology Policy.

Roya Pakzad serves as a research associate and project leader in Technology and Human Rights at Stanford University's Global Digital Policy Incubator. She has worked on initiatives relating to everything from the role of technology in the Syrian refugee crisis to the human rights implications of artificial intelligence. Roya also researches the human rights policy and practices of technology companies and explores the role of technology in empowering marginalized groups. After three years working at Advanced Micro Devices as an electrical engineer, she decided to converge her interests in technology and human rights. She holds degrees from Shahid Beheshti University (B.Sc. in electrical engineering), the University of Southern California (M.Sc. in electrical engineering) and Columbia University (M.A. in human rights studies). Roya was born and raised in Tehran, Iran and currently lives in Santa Cruz, California. You can follow her work on her website and Twitter.

Multi-stakeholder Governance Innovations to Protect Free Expression, Diversity and Civility Online

Lawrence E. Strickling and Jonah Force Hill

Reflecting on the issues, strategies and approaches discussed in the previous essays, this final essay will consider how the multi-stakeholder approach can help achieve solutions, offer examples of cross-sector collaboration in this space and generate further ideas of how multi-stakeholder solutions might be pursued.

What is the Multi-stakeholder Approach?

A threshold question for this discussion is to try to arrive at a shared understanding of what is the multi-stakeholder approach. There is no one single concept of what is appropriately viewed to be a multi-stakeholder approach. There are, instead, numerous models currently in use today, each with its own unique contours. Few, if any, of the models currently in use are static; rather, they are constantly evolving to meet new and yet uncharted governance challenges.

Multi-stakeholder approaches are just that, approaches. They encompass a range of procedures, formats, resolution mechanisms and outcomes. In the same way that democratic governments may follow a parliamentary or a presidential system of governance, so too do multi-stakeholder approaches vary and adapt to fit the particular governance question at hand. Some models lead to decisions while others are merely consultative. Some have established membership rules and criteria, while others allow anyone to participate. Some models are intended to last decades, while others are one-off processes designed to address a specific challenge of the day.

As an illustration of this diversity, compare the Internet Governance Forum (IGF) to the Internet

Engineering Task Force (IETF). The IGF is an annual UN conference that brings together stakeholders from government, civil society, academia, the private sector and the technical community to debate global internet governance issues in a public and transparent fashion. The IGF does not produce decisions. Instead, it serves as a once-a-year forum for discussion — planned through regular consultations throughout the year — at which various communities can set the future internet governance agenda and air concerns publicly in the presence of other stakeholder groups.

The IETF, by contrast, does function as a sort of rule-making entity; it is responsible for developing and updating many of the internet's core technical standards, including, by way of example, Transmission Control Protocol/Internet Protocol (TCP/IP). However, unlike conventional technical standards bodies, such as the International Standards Organization, any interested person may participate in the IETF's standard-setting work. All participants in the IETF are volunteers, and there is no official membership criteria. All IETF work products and communications among the participants are available to the public and are listed on the organization's website and mailing lists. IETF standards are not mandatory; instead, the internet community adopts its standards because they are based on the consensus and combined engineering judgment of the internet's technical experts and upon those experts' real-world experience in implementing and deploying technical specifications. The IGF and the IETF are both unmistakably multi-stakeholder, but they represent different models of multi-stakeholder processes in action.

The diversity of multi-stakeholder approaches is revealed, perhaps most comprehensively, in a study by the Global Network of Internet and Society Research Centers and the Berkman Center for Internet and Society at Harvard University (Budish, Gasser and Myers West 2015). The report analyzes 12 geographically and topically diverse case studies of internet, as well as non-internet, multi-stakeholder governance processes, ranging from water resource management in the Volta River Basin in West Africa, to arbitration of disputes within the Bitcoin community. The research was conducted with an eye toward describing the formation, operation and critical success factors for multi-stakeholder governance. The study's authors unambiguously conclude from their research that: "[t]here is no single best-fit model for multi-stakeholder governance groups that can be applied in all instances" (ibid., 2). Other scholarly studies have reached similar conclusions.

Yet, despite efforts to highlight the range of multi-stakeholder approaches (or even perhaps because of them), there is no agreed upon definition of "multi-stakeholder governance." Scholars have attempted to identify its core features, however, and to provide heuristics for differentiating among models (Drake 2005; Drake and Wilson 2008). Notably, DeNardis and Raymond (2013) have constructed a taxonomy of multi-stakeholder processes in which different approaches are defined by the types of actors involved and the nature of the authority relations between those actors. They argue that in order for a process or organization to qualify as multi-stakeholder, at least two broad categories of actors, such as states, civil society, firms and intergovernmental organizations must be involved. Hemmati (2002, 2), in her book on multi-stakeholder processes, which focuses narrowly on climate and environmental governance, posits "multi-stakeholder processes are processes which aim to bring together all major stakeholders in a new form of communication, decision-finding (and possibly decision-making) on a particular issue."

The term multi-stakeholder is often used casually, even haphazardly. It has become a bit of a buzzword in governance circles. Actors often mistakenly — indeed, sometimes manipulatively — attach the multi-stakeholder label to what are, in practice, multilateral and top-down processes. To fully assess the prospects of expanding the use of multi-stakeholder processes, it is appropriate to fashion an outline of a definition that serves both to reinforce the internet's core values and to protect the term multi-stakeholder from becoming little more than a marketing meme for governance schemes.

It is proposed that an "authentic" multi-stakeholder process display the following attributes:

- *stakeholder-driven*: stakeholders determine the process and decisions, from agenda setting to workflow, rather than simply fulfilling an advisory role;
- *open*: any stakeholder can participate, and the process includes and integrates the viewpoints of a diverse range of stakeholders, including the viewpoints of those stakeholders who hold specialized expertise applicable to the governance challenge at hand;
- *transparent*: all stakeholders and the public have access to deliberations, creating an environment of trust, legitimacy and accountability; and
- *consensus-based*: outcomes (when outcomes are desirable) are consensus-based, arrived at by compromise and are a win-win for the greatest number or diversity of stakeholders.

How Do Multi-stakeholder Approaches Work in Comparison to, and in Conjunction with, More Traditional Legislative and Regulatory Actions of States, Including Multilateral Treaty Negotiations?

Multi-stakeholder approaches have repeatedly proven themselves to be exceptionally well-suited to rapidly changing technologies and business practices and to the global environment in which the internet exists. Moreover, these processes match up well when compared to more traditional regulatory and legislative models. Complex policy issues often take years to make their way over the regulatory and legislative hurdles found in Washington, Brussels or Geneva. Many efforts at policy making end in indecision. Many that do reach a conclusion commonly end up solving a problem that no longer exists, while the regulatory or legislative debate has been overtaken by newer, unanticipated issues that need urgently to be addressed. Consider if the internet's technological challenges had been handed off to a typical regulatory or legislative process. More than likely, the world would still be waiting for a resolution. Worse, the result might have been technical protocols that were hopelessly out of date, hamstringing technologists and users from creating the robust, evolving internet that exists today.

Traditional communications regulatory approaches, such as those employed by the US Federal Communications Commission (FCC), are often not suited to current policy challenges. The time it takes for the FCC to publish a final rule from the date at which rule drafting begins can take upwards of several years, a time frame that is not realistic in the context of the rapidly evolving internet. Moreover, these processes can fall prey to rigid regulatory procedures, bureaucratic inertia, self-serving interest-group lobbying, judicial appeals and legislative roadblocks. Many ultimately end in stalemate. Often, they do not adequately incorporate the views of all stakeholders in decision making. Granted, in the "notice-and-comment" process, stakeholders are given an opportunity to provide input and recommendations on a proposed rule as part of a formal consultation period, but final decision-making authority still ultimately resides with the regulator and not with the community of stakeholders who are developing and using the technology. The process is not designed to lead to consensus and, as a result, the publication of new rules often triggers extensive legal challenges, which themselves can delay, or even discredit, a new rule.

The US government has made support for multi-stakeholder governance a cornerstone of its global internet policy and has been a determined advocate for the approach. This advocacy goes beyond simply rhetoric. The US government, through the National Telecommunications and Information Administration (NTIA), has utilized the multi-stakeholder approach to address a range of internet governance questions and to demonstrate and better understand the strengths and opportunities of the approach.

Perhaps nowhere was this commitment clearer than in NTIA's support for the transition of its stewardship of the key functions of the Domain Name System to the global internet multi-stakeholder community. From 1998 until 2016, the Internet Corporation for Assigned Names and Numbers (ICANN) performed several of these important coordination functions, known as the Internet Assigned Numbers Authority (IANA) functions, pursuant to a contract between ICANN and NTIA. NTIA's role in administering the IANA function contract over those 18 years was largely procedural. NTIA had had no operational role, but simply verified that ICANN had followed policies, procedures and contractual obligations in processing domain name change requests.

In 1998, the US government declared that NTIA's "stewardship" role over the IANA functions would be temporary and that the multi-stakeholder community would eventually assume responsibility over the administration of functions. Toward that end, NTIA announced in March 2014 that the US government would transition stewardship over the IANA functions contract to ICANN and the global internet community.

The history of the IANA transition is well documented in other NTIA reports and speeches (NTIA 2016a). Looking back on the two-year effort, it is clear that only a multi-stakeholder process could have brought together the views and ideas of so many people in such a short period of time to find a consensus solution to such complicated and important issues. The transition was an audacious experiment in global governance. The multi-stakeholder approach, while previously employed in ICANN's technical and policy processes, had never been tested in such a large, complex or as geopolitically contentious governance challenge as the IANA transition. Accordingly, the success of the transition, as well as the multi-stakeholder approach utilized to plan it, served as a validation of the theoretical premise. It demonstrated that the multi-stakeholder model was both flexible and adaptive enough to address even the most difficult internet governance challenges.

The IANA transition is an example where a multi-stakeholder approach was utilized as an alternative to direct government action. However, the approach can also work in tandem with legislation and regulation to fill in the details of a more general legislative or regulatory declaration. For example, NTIA has organized multi-stakeholder processes to develop codes of conduct or best practices that specify how the White House's "Consumer Privacy Bill of Rights" (White

House 2012) applies in specific business contexts. The Consumer Privacy Bill of Rights was one of the four components of the White House's Privacy Blueprint (US Government 2012), a multi-stage, multi-component commercial privacy plan, which included baseline privacy legislation, codes of conduct to establish specific industry practices and an expansion of the Federal Trade Commission's (FTC's) enforcement expertise and authority.

The plan proposed that Congress enact the Bill of Rights in baseline privacy legislation. That legislation was never passed. Nonetheless, NTIA convened multi-stakeholder discussions on certain elements of the Bill of Rights with the goal of having stakeholders develop codes of conduct that would provide specific guidance and flexibility as to how the Bill of Rights could apply to issues such as mobile app transparency, unmanned drones and facial recognition.

The processes encountered some challenges, of course. The Code of Conduct on mobile app transparency (NTIA 2013) took a year of meetings to draft, in part because participants, who were accustomed to engaging in adversarial legislative and regulatory proceedings, found that new skills were required to succeed in the consensus-based discussions of a multi-stakeholder process. In the end, the process enabled industry and civil society groups to reach consensus on a number of key ideas to improve privacy in an area that traditional policy making and regulatory approaches had been unable to address effectively.

The unmanned drone privacy process successfully produced a voluntary privacy best practices guide for the use of commercial drones (NTIA 2016b), one that was driven and drafted entirely by stakeholders from the drone and aviation industry and a range of civil society and consumer rights groups. The best practices guide continues to be supported and promoted, and it is a centrepiece of industry and civil society conversation about drone privacy protections. The facial recognition process, likewise, produced a best practices guide for commercial facial recognition privacy; however, the final product of that effort was weakened by the privacy stakeholder community's early rejection of the process and eventual refusal to sign off on the product that was developed by the remaining participants.

While participating stakeholders have generally lauded the outcomes of the NTIA-led initiatives and the processes by which they were derived, those initiatives were not without their skeptics. Two papers in particular, one by Kaminski (2016) and another by Rubinstein (2016), argue that the kind of self-regulatory frameworks developed through multi-stakeholder processes, and convened by NTIA, are often rendered ineffective without a strong enforcement backstop. Enforcement by agencies like the FTC, Rubinstein (2016, 5-6) asserts, is "necessary to deter bad actors and outliers and ensure the widest possible participation in any self-regulatory schemes." This was a sentiment shared by many of the privacy advocates who walked out of the facial recognition forum.

The concern that enforcement may be necessary to guarantee that actors adhere to best practices is valid. However, it is not a reason to forego a multi-stakeholder process, particularly when there is no likely alternative that might produce a more enforceable outcome. The simple act of bringing parties together to discuss and reach consensus on a particular technology policy question, especially those surrounding nascent and emerging technologies, can ultimately lead to a more thoughtful and less reactionary regulatory environment, one more responsive to consumers and industry alike. Multi-stakeholder processes, even when they only achieve the most baseline level of consensus, can provide the blueprint for a path forward, if not the actual brick and mortar building.

What Barriers Need to Be Overcome to Operationalize Greater Use of Multi-stakeholder Processes?

A number of challenges must be addressed to allow the expansion and enhancement of the use of multi-stakeholder processes. These challenges are not intractable and can be overcome. However, for multi-stakeholder processes to be truly effective, the global internet community must work diligently, thoughtfully and collaboratively to come up with ways to address these difficulties.

Legitimacy

First and foremost is the question of legitimacy. A multi-stakeholder approach must provide participants with confidence that the world at large will accept and recognize the outcome of the process as authoritative (UK Internet Governance Forum 2016). Where does legitimacy come from? In many cases, legitimacy comes from a government or some other "official" entity that convenes — but does not control — the process. In the United States, for instance, NTIA's domestic multi-stakeholder initiatives have been accepted as legitimate by virtue of the fact that they were convened by NTIA, by statute and by the principal adviser to the president on telecommunications and information policy. Of course, there are other sources of legitimacy. The IETF, for example, has gained legitimacy over its 30-plus years by producing voluntary standards of the highest quality; criteria that have become the gold standard for the internet since the body's inception.

More than a well-regarded convening agency, more even than a history of sound practices, the legitimacy of any process derives from its openness to any participant, its conscious inclusion of a diversity of

stakeholders and its commitment to reaching decisions by consensus. Also, to maximize the possibility of success, participants must be the ones who make the final decision on a particular issue, not the convening body. This feature is one of the fundamental differences between a multi-stakeholder process and consultation. If participants are not empowered to make a final decision, then a process is merely consultative. By contrast, multi-stakeholder processes that place responsibility for final decision making on the participants themselves are generally viewed as more legitimate. They also tend to be more successful because the prospect of fashioning policy, and not just offering commentary, frequently induces the participants to put in the extra effort needed to reach a consensus. Further, entrusting the participants with the power to make decisions also reduces the possibility of non-participants mounting a collateral challenge of the outcome by appealing to others who did not choose to participate.

This issue can create a particular challenge for governments that might seek to convene multi-stakeholder discussions in conjunction with regulatory and legislative proceedings. Notwithstanding the desire of government officials to allow a group of stakeholders to reach a consensus decision, the laws of the government, such as the Administrative Procedures Act in the United States, may prohibit giving the decision-making power to a group of stakeholders and require the agency to conduct subsequent notice and comment on the rule-making processes, thus diminishing the incentive of stakeholders to work together to reach consensus in the multi-stakeholder discussions.

Consensus

One of the key attributes of the "ideal" multi-stakeholder process is that decisions are reached by consensus. Consensus decision making requires parties to persuade one another of the merits of their position. In consensus decision making, participants must compromise if they are to accomplish anything; they must ultimately either persuade, or be persuaded by, the other participants, at least insofar as it is necessary to achieve the required consensus. What is consensus? A standard of unanimity is nearly impossible to achieve. If the standard is not unanimity, how should it be defined and who sets the standard? There is no one standard that works for every situation, but many convenings have found that a standard of "can you live with it?" works well. Perhaps the best solution is to leave the definition of consensus to the participating stakeholders in the process.

In order to achieve some form of consensus, conveners and stakeholders must set the tone and the culture of a process early and anticipate the complex social dynamics of the group. All parties need to be dedicated to reaching a consensus outcome and must be willing to compromise to achieve that goal. The strategy of some participants in normal legislative and regulatory proceedings is sometimes to make maximalist demands or simply to preserve the status quo — those strategies will not work in a multi-stakeholder process. Everyone needs to come to the table with open minds, committed to collaborating fully in the process.

Representation

Multi-stakeholder processes are generally quite resource-intensive, both in terms of time and money. A single initiative focusing on a specific policy issue can take months from start to completion. Many multi-stakeholder organizations hold multiple meetings a year, often in far-flung places across the globe. For stakeholders with limited resources, in-person attendance can be prohibitively expensive. While most venues try to provide remote participation opportunities for stakeholders who are unable to travel, there is a sense that stakeholders who participate in person can have more impact on the group decision than those who engage remotely.

The inclusion of underrepresented groups is critical to the success of multi-stakeholder governance and, thus, future convenings need to address this disparity between well and poorly resourced stakeholders. Companies and organizations may want to try to pool resources with other like-minded organizations to lighten the load. Conveners may want to consider funding programs to subsidize travel to meetings for those stakeholders who are eager to participate, but cannot afford to pay their own way. ICANN and the Internet Society, for instance, have created fellowship programs to provide travel assistance to individuals from underrepresented communities to attend ICANN and IETF meetings (Internet Society Fellowship 2011).

The multi-stakeholder approach also poses a subtle, yet inescapable, problem for new entrants. Start-up companies, governments of developing nations and new civil society groups all have difficulty establishing themselves in multi-stakeholder processes. It is in the nature of negotiations that the most persuasive stakeholders, and thus the most effective and influential participants, are those who possess expertise in both the subject matter (for example, the technology or policy issue in question) and the politics and institutional history of the multi-stakeholder process or entity in which they are operating. New entrants often lack these competence and, as a result, their views are less likely to be incorporated into the group's decision making. This handicap alongside resource constraints are among the primary reasons why stakeholders from the developing world are so often frustrated by the approach.

None of these challenges are trivial. Yet — at least in the context of internet governance — when compared to the challenges posed by traditional legislative or regulatory approaches, they produce fewer

impediments to effective problem solving. Multi-stakeholder processes can be resource intensive, but they are still generally less financially burdensome than traditional regulatory proceedings or litigation. Reaching multi-stakeholder consensus can be difficult and time-consuming, but compare the time it takes to achieve consensus to the time it takes the US Congress to enact legislation. New entrants may have a strategic disadvantage in multi-stakeholder settings, but they at least have a seat at the table and a say in the outcome. Traditional government and multilateral rulemaking settings afford them no such right.

One Proposal for a Way Forward

In February 2018, the Internet Society launched a new program, the Collaborative Governance Project, to expand the global knowledge and use of collaborative governance processes to solve problems and develop norms. (Internet Society 2018) The key activities of the project are as follows:

- convening stakeholders to solve concrete problems and develop norms on a consensus basis;
- training stakeholders on how to be effective in collaborative governance discussions; and
- building and promoting academic research and writing on collaborative governance approaches.

The project is based on the findings from more than 150 interviews of global stakeholders and is designed to address the types of barriers identified in this paper (Internet Society 2017). Of course, the success of this effort will depend on the voluntary commitment and participation of global stakeholders to come together in discussions to reach consensus on solutions and then to implement them. The project will emphasize the need for convenings to develop concrete and actionable outcomes that will be implemented by the parties to the discussions. With expert facilitation, preparation and the careful curation of issues to be discussed, the Internet Society is hopeful that the project will successfully deliver concrete, positive outcomes and will create capacity around the world for stakeholders to make greater utilization of collaborative, multi-stakeholder approaches.

Works Cited

Budish, R., U. Gasser and S. Myers West. 2015. "Multi-stakeholder as Governance Groups: Observations from Case Studies." January 15. Berkman Klein Center for Internet and Society at Harvard University. Berkman Center Research Publication No. 2015-1. https://cyber.harvard.edu/publications/2014/internet_governance.

DeNardis, L. and M. Raymond. 2013. "Thinking Clearly About Multi-Stakeholder Internet Governance." November 14. Global Internet Governance Academic Network, Annual Symposium.

Drake, W. 2005. *Reforming Internet Governance: Perspectives from the Working Group on Internet Governance*. New York: The UN Information and Communication Technologies Task Force. www.wgig.org/docs/book/WGIG_book.pdf.

Drake, W. J. and E. J. Wilson. 2008. *Governing Global Electronic Networks: International Perspectives on Policy*. Cambridge, MA: MIT Press.

Hemmati, M. 2002. *Multi-stakeholder Processes for Governance and Sustainability: Beyond Deadlock and Conflict*. London, UK: Earthscan Publications.

Internet Society. 2017. *The Feasibility of Expanding the Use of Multi-stakeholder Approaches for Internet Governance, Final Report to the Internet Society*. https://cdn.prod.internetsociety.org/wp-content/uploads/2018/01/Feasibility-Study-Final-Report-Oct-2017.pdf.

———. 2018. "Collaborative Governance Project." www.internetsociety.org/collaborativegovernance/.

Internet Society Fellowship. 2011. "Internet Society Fellowship Awards Build Technical Leadership in Developing Regions." September 8. www.internetsociety.org/what-we-do/ education-and-leadership-programmes/ietf-and-ois-programmes/internet-society-fellowship.

Kaminski, M. E. 2016. "When the Default is No Penalty: Negotiating Privacy at the NTIA." *Denver University Law Review* 93 (4); Ohio State Public Law Working Paper 366. https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2835434.

NTIA. 2013. *Short Form Notice of Code of Conduct to Promote Transparency in Mobile App Practices*. July 25. www.ntia.doc.gov/files/ntia/publications/july_25_code_draft.pdf.

———. 2016a. "Remarks of Lawrence E. Strickling, Assistant Secretary of Commerce for Communications and Information, IGF, Opening Session Guadalajara, Mexico, December 6, 2016." www.ntia.doc.gov/speechtestimony/2016/remarks-assistant-secretary-strickling-internet-governance-forum-opening.

———. 2016b. *Voluntary Best Practices for UAS Privacy, Transparency, and Accountability: Consensus, Stakeholder-Drafted Best Practices Created in the NTIA-Convened Multi-Stakeholder Process*. March 16. www.ftc.gov/system/files/documents/public_comments/2016/10/00008-129242.pdf.

Rubinstein, I. 2016. "The Future of Self-Regulation is Co-Regulation." In *The Cambridge Handbook of Consumer Privacy*, edited by Evan Selinger, Jules Polonetsky and Omer Tene, 503–23. Cambridge, UK: Cambridge University Press.

UK Internet Governance Forum. 2016. "Remarks of Assistant Secretary Strickling at the UK Internet Governance Forum." November 17. https://ukigf.org.uk/events/forthcoming-events/.

US Government. 2012. "Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy." *Journal of Privacy and Confidentiality* 4 (2): 95–142.

The White House. 2012. "Fact Sheet: Plan to Protect Privacy in the Internet Age by Adopting a Consumer Privacy Bill of Rights." February 23. https://obamawhitehouse.archives.gov/the-press-office/2012/02/23/fact-sheet-plan-protect-privacy-internet-age-adopting-consumer-privacy-b.

About the Authors

Lawrence E. Strickling is the executive director of the Collaborative Governance Project of the Internet Society, a project whose goal is to expand the knowledge and use of multi-stakeholder, collaborative processes to solve problems and develop norms. From 2009 to 2017, Strickling served as assistant secretary for Communications and Information at the US Department of Commerce. In this role, Strickling served as administrator of the National Telecommunications and Information Administration, where he led the US government's engagement with ICANN and directed the US role in transitioning its stewardship of the Domain Name System to the global multi-stakeholder community from 2014 to 2016. Strickling also expanded the use of multi-stakeholder convenings as an alternative to regulation/legislation in areas of privacy, cybersecurity and intellectual property.

Jonah Force Hill is a fellow in New America's Cybersecurity Initiative. Formerly, he served as a policy specialist in the National Telecommunications and Information Administration of the US Department of Commerce, where he worked on a range of domestic and global internet and technology policy matters. He represented the United States at government-to-government engagements and in venues such as the Organisation for Economic Co-operation and Development, the Internet Corporation for Assigned Names and Numbers, and the Asia-Pacific Economic Cooperation forum. He has been published in a number of academic journals, including the *Harvard National Security Law Journal*, the *Georgetown Journal of International Affairs* and the *Lawfare Research Paper Series*.

About CIGI

We are the Centre for International Governance Innovation: an independent, non-partisan think tank with an objective and uniquely global perspective. Our research, opinions and public voice make a difference in today's world by bringing clarity and innovative thinking to global policy making. By working across disciplines and in partnership with the best peers and experts, we are the benchmark for influential research and trusted analysis.

Our research programs focus on governance of the global economy, global security and politics, and international law in collaboration with a range of strategic partners and support from the Government of Canada, the Government of Ontario, as well as founder Jim Balsillie.

À propos du CIGI

Au Centre pour l'innovation dans la gouvernance internationale (CIGI), nous formons un groupe de réflexion indépendant et non partisan doté d'un point de vue objectif et unique de portée mondiale. Nos recherches, nos avis et nos interventions publiques ont des effets réels sur le monde d'aujourd'hui car ils apportent de la clarté et une réflexion novatrice pour l'élaboration des politiques à l'échelle internationale. En raison des travaux accomplis en collaboration et en partenariat avec des pairs et des spécialistes interdisciplinaires des plus compétents, nous sommes devenus une référence grâce à l'influence de nos recherches et à la fiabilité de nos analyses.

Nos programmes de recherche ont trait à la gouvernance dans les domaines suivants : l'économie mondiale, la sécurité et les politiques mondiales, et le droit international, et nous les exécutons avec la collaboration de nombreux partenaires stratégiques et le soutien des gouvernements du Canada et de l'Ontario ainsi que du fondateur du CIGI, Jim Balsillie.

About GDPi

The mission of the Global Digital Policy Incubator (GDPi) at Stanford's Center for Democracy Development and the Rule of Law is to inspire policy and governance innovations that reinforce democratic values, universal human rights, and the rule of law in the digital realm. Its purpose is to serve as a collaboration hub for the development of norms, guidelines, and laws that enhance freedom, security, and trust in the global digital ecosystem. GDPi provides a vehicle for global multi-stakeholder collaboration between technologists, governments, private sector companies, diplomats, international organizations, academics, and civil society in a shared purpose: to develop norms and policies that enhance security, promote economic development, and reinforce respect for human rights in or our global trans-border digital ecosystem. We aim to reinforce existing frameworks of international human rights law and international humanitarian law but lead in articulation of how to apply these norms and values in a global digital context. Our goal is to facilitate the development of operational policies and processes that meet the societal challenges that arise from digitization and technological innovation.

Centre for International
Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

@cigionline