



# Elections and Democracy in the Digital Age

A Kofi Annan Foundation Initiative

## PROTECTING ELECTORAL INTEGRITY IN THE DIGITAL AGE

The Report of the Kofi Annan  
Commission on Elections and  
Democracy in the Digital Age

January 2020



# TABLE OF CONTENTS

About the Kofi Annan Commission on Elections and Democracy in the Digital Age .....	1
Members of the Commission .....	5
Foreword .....	9
Executive Summary .....	14
Building Capacity .....	17
Building Norms .....	18
Action by Public Authorities .....	20
Action by Platform .....	22
<b>I. Elections as Focal Points in the Struggle for Democracy .....</b>	<b>25</b>
<b>II. Affective Polarization, Social Media, and Electoral Integrity .....</b>	<b>31</b>
Social Media and Polarization .....	35
Implications for Action .....	38
<b>III. Hate Speech and Electoral Integrity .....</b>	<b>41</b>
Approaches to Regulating Hate Speech .....	44
Implications for Action .....	49
<b>IV. Protecting Electoral Integrity from Disinformation .....</b>	<b>55</b>
Measuring the Prevalence of Disinformation .....	58
The Weaponization of Disinformation .....	61
Implications for Action .....	62
<b>V. Political Advertising in the Digital Age .....</b>	<b>71</b>
Implications for Action .....	76
<b>VI. Protecting Elections from Foreign Interference .....</b>	<b>81</b>
An Emerging Transnational Industry of Election Manipulation .....	83
Protecting Electoral Infrastructure .....	84
Implications for Action .....	86

<b>VII. Summary of Recommendations .....</b>	<b>91</b>
Building Capacity .....	91
Building Norms .....	93
Action by Public Authorities .....	94
Action by Platform .....	96
<b>Acknowledgements .....</b>	<b>97</b>
<b>About the Kofi Annan Foundation .....</b>	<b>99</b>
<b>End Notes .....</b>	<b>101</b>

## LIST OF FIGURES

<b>Box 1</b> Social Media as a Weapon Against Women .....	43
<b>Box 2</b> Internet Shutdowns in Asia and Africa .....	47
<b>Box 3</b> Initiatives for Diversity in Coding .....	52
<b>Box 4</b> Citizen Monitoring of Hate Speech in Kenya, 2012-2013 .....	53
<b>Box 5</b> Mexico's Approach to Counter Disinformation .....	66
<b>Box 6</b> Indonesia's Approach to Counter Disinformation, 2018-2019 .....	68
<b>Box 7</b> Nigeria's Abuja Accord on Election Conduct, 2015 .....	77



# ABOUT THE KOFI ANNAN COMMISSION ON ELECTIONS AND DEMOCRACY IN THE DIGITAL AGE

(KACEDDA)

Kofi Annan was a lifelong advocate for the right of every citizen to have a say in how they are governed, and by whom. He was adamant that democratic governance and citizen empowerment were integral elements to achieving sustainable development, security and lasting peace, and this principle guides much of the work of the Foundation, most notably its Electoral Integrity Initiative.

In 2018, as one of his last major initiatives, Mr Annan convened the Commission on Elections and Democracy in the Digital Age. The Commission includes members from civil society and government, the technology sector, academia and media; over 12 months they examined and reviewed the opportunities and challenges for electoral integrity created by technological innovations.



“Technology does not stand still;  
neither can democracy.”

- Kofi Annan



Assisted by a small secretariat at Stanford University and the Kofi Annan Foundation, the Commission has undertaken extensive consultations to issue recommendations as to how new technologies, social media platforms and communication tools can be harnessed to engage, empower and educate voters, and to strengthen the integrity of elections.

### The Key Questions that Guided the Commission's Deliberations:

- What are the fundamental elements of digital technology which will have a uniquely detrimental, or positive, impact on democracy and electoral processes?
- What is the potential of digital technologies to both strengthen and undermine the integrity of the electoral environment?
- How can the use of technology in elections be made transparent and accountable?
- What opportunities and incentives can digital technology offer voters, especially young people, to engage in democratic processes?
- What role and impact does political finance have in the deployment and use of digitally-based electoral strategies and instruments?

### The Commission's Objectives:

1. To identify and frame the challenges to electoral integrity arising from the global spread of digital technologies and social media platforms;
2. To develop policy measures that address these challenges and which also highlight the opportunities that technological innovation offers for strengthening electoral integrity and political participation;
3. To define and articulate a programme of advocacy to ensure that the key messages emerging from the Commission are widely diffused and debated around the world.

### Statement from the Chair of the Commission

“In this digital age, new technologies and social media platforms are profoundly changing democracies – and democratic processes – all over the world. While these provide the unequalled potential to deliver citizen’s hopes for democratic governance, they also create new challenges and risks for democratic processes and political rights.

Along with the team at Stanford and the Kofi Annan Foundation, my fellow Commissioners and I are determined to honour Mr Annan’s legacy and ensure this Commission plays a leading role in defending and strengthening the electoral processes that are at the heart of democracy.”

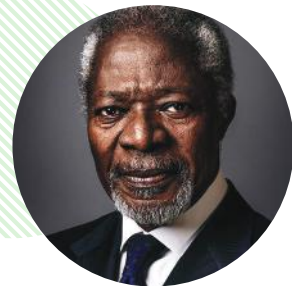
**- Laura Chinchilla**

CHAIR OF THE KOFI ANNAN COMMISSION ON ELECTIONS  
AND DEMOCRACY IN THE DIGITAL AGE



# MEMBERS OF THE COMMISSION

The Kofi Annan Commission on Elections and Democracy in the Digital Age unites some of the most distinguished leaders from the tech sector, academics and of political life to answer one simple question: How can we mitigate the risks of the digital age to our elections while harnessing the opportunities and ultimately strengthen democracy worldwide.



**Kofi Annan †** - Convening Chairman  
*(Ghana)*

Nobel Peace Prize laureate, Secretary-General of the UN from 1997 to 2007, and Founding Chair of the Kofi Annan Foundation



**Laura Chinchilla** - Chair  
*(Costa Rica)*

Vice-President of the Club of Madrid, Former President of Costa Rica



**Yves Leterme** - Vice-Chair  
*(Belgium)*

Former Secretary-General of International IDEA, Former Prime Minister of Belgium



**Stephen Stedman** - Secretary-General of the Commission  
*(United States)*

Senior Fellow at the Freeman Spogli Institute for International Studies and Professor of Political Science, Stanford University



**Noeleen Heyzer**  
*(Singapore)*

Former Executive Secretary of the United Nations Economic and Social Commission for Asia and the Pacific



**Toomas Hendrik Ilves**  
*(Estonia)*

Distinguished Visiting Fellow, Hoover Institution, Former President of Estonia



**Ory Okolloh**  
*(Kenya)*

Managing Director of Africa at Luminate



**Nate Persily**  
*(United States)*

James B. McClatchy Professor of Law at Stanford Law School



**Alex Stamos**  
*(United States)*

Research Professor at Stanford University, Former Chief Security Officer at Facebook



**William Sweeney**  
*(United States)*

Former President and CEO of the International Foundation for Election Systems (IFES)



**Megan Smith**  
*(United States)*

Founder and CEO at Shift7, Former United States Chief Technology Officer



**Ernesto Zedillo**  
*(Mexico)*

Director of the Yale Center for the Study of Globalization, Former President of Mexico

Democratic consolidation around the world currently faces major challenges. Threats to democracy have become more insidious, especially due to the manipulation of legal and constitutional procedures originally designed to guard democracy against arbitrary action and abuse. Free and fair elections, the cornerstone of democratic legitimacy, are under considerable stress from populism and post-truth movements, who abuse new digital communication technologies to confuse and mislead citizens. Today, free and fair elections, the primary expression of democratic will for collective government, are far from guaranteed in many countries around the world. Protecting them will require a new set of policies and actions from technological platforms, governments and citizens.

The vulnerability of electoral integrity worldwide is symptomatic of larger processes of democratic erosion felt in old and new democracies alike: increasing political polarization, declining trust both between fellow citizens and between them and government institutions, systematic attacks on the press and independent media, the decline of political parties as legitimate vehicles to aggregate interests, and an increasing frustration that democratic governments are not satisfying people's basic needs and aspirations.

At the center of these changes is the use of digital communication technologies, often blamed as the source for this democratic erosion. Some claim that social media polarizes public debate, pushing people to political extremes. Others argue that social media creates 'filter bubbles' and 'echo chambers', reducing access to diverse sources of information and perspectives that enable democratic deliberation. Because political campaigns use social

media to target small groups of voters with personalized appeals, others claim social media undermines the public square and the give and take of electoral campaigns. In short, current debates on the causes and effects of democratic erosion would be incomplete without addressing and exploring the factual role that digital communications technologies are playing in that process.

For that reason, Kofi Annan convened the Commission on Elections and Democracy in the Digital Age. Deeply concerned with the effects that information and communication technologies (ICTs) were having on democracy and elections, and based on conversations with experts from around the world, Mr. Annan believed that creating a new commission could shed light on some fundamental issues concerning new ICTs, elections, and democracy. He provided the Commission with a mandate to identify and frame the challenges to electoral integrity arising from the global spread of digital technologies and social media platforms, and to develop policy measures that address these challenges.

Mr. Annan asked the Commission to look globally and to seek to understand these issues as they manifest themselves on different continents, especially among the democracies of the Global South. To accomplish this broader reach, the Commission engaged in consultations with experts and authorities in Brazil, Mexico, Kenya, Cote D'Ivoire, South Africa, and India, and commissioned several research papers from Latin America, Africa, and Asia. The Commission also met with the European Commission and consulted leading actors from the internet and social media industries.

In our report, the Commission puts forward a series of recommendations to strengthen the capacities of electoral integrity authorities, to build norms encompassing shared understandings on the acceptable use of digital technologies in elections, and to encourage action by public authorities and technological companies to enhance electoral integrity. These recommendations originate from one of the main conclusions of the report: all relevant stakeholders – tech and digital platforms, governments, electoral authorities, traditional media and citizens – have a critical role to play in strengthening electoral integrity.

I would like to thank Yves Leterme, the Vice-Chair of the Commission and my fellow commissioners for their contribution and their dedication to this project. I wish to express especially my deep appreciation to the Secretary-General of the Commission, Stephen Stedman, for his work in supervising the research and consultations for the Commission, and for his role in drafting the report. I am also very grateful for the support of the Kofi Annan Foundation, and its president, Alan Doss, under whose auspices the Commission was convened.

Mr. Annan died unexpectedly before the Commission began its work. He cared deeply about issues of electoral integrity, especially motivated by his experience as a mediator after the flawed Kenyan elections of 2007 when thousands of Kenyans lost their lives and hundreds of thousands more were forcibly displaced, bringing the country close to a civil war. For his vision, his defense of democracy and his continuous efforts to developing the rule of law and securing international peace, we dedicate this report to Kofi's memory. Kofi Annan leaves behind a rich legacy as a protector and

proponent of electoral integrity. We hope this report honors that legacy and helps it endure in future debates and conversations, but more importantly, in future actions that strengthen the integrity of elections worldwide.

**- Laura Chinchilla**

CHAIR, KOFI ANNAN COMMISSION ON ELECTIONS  
AND DEMOCRACY IN THE DIGITAL AGE







## EXECUTIVE SUMMARY

New information and communication technologies (ICTs) pose difficult challenges for electoral integrity. In recent years foreign governments have used social media and the Internet to interfere in elections around the globe. Disinformation has been weaponized to discredit democratic institutions, sow societal distrust, and attack political candidates. Social media has proved a useful tool for extremist groups to send messages of hate and to incite violence. Democratic governments strain to respond to a revolution in political advertising brought about by ICTs. Electoral integrity has been at risk from attacks on the electoral process, and on the quality of democratic deliberation.

The relationship between the Internet, social media, elections, and democracy is complex, systemic, and unfolding. Our ability to assess some of the most important claims about social media is constrained by the unwillingness of the major platforms to share data with researchers. Nonetheless, we are confident about several important findings:

- Many of the ills the Internet and social media have been accused of – extreme polarization of democratic politics, decline in trust in governments, traditional media, and fellow citizens, partisan media and the spread of disinformation – predate the rise of social media and the Internet.



- Although social media is not a cause of large-scale political polarization, it exacerbates and intensifies it, and is a tool for anyone who seeks to undermine electoral integrity and healthy democratic deliberation.
- Democracies vary in their vulnerability to disinformation based on pre-existing polarization, distrust, and partisan traditional media, with new and transitional democracies in the Global South being particularly vulnerable.
- For the foreseeable future, elections in the democracies of the Global South will be focal points for networked hate speech, disinformation, external interference, and domestic manipulation.
- The responsibility for social media's abuse as a threat to electoral integrity lies with multiple actors:
  - The large platforms allowed hate speech and disinformation on their platforms to go viral, failed to anticipate how their technologies would be used in transitional democracies with fractured societies and histories of ethnic and religious violence, denied evidence of their products undermining democracy and abetting violence, engaged in smear campaigns against critics and were slow to react in constructive ways;
  - Political candidates and elected leaders have used social media to foment hate, spread disinformation and undermine trust in societal and governmental institutions;

- Some political consultants have sought to manipulate electoral processes to win at all costs and have turned election manipulation into a transnational business that threatens electoral integrity everywhere around the world; and
- Traditional media has often amplified disinformation and propaganda instead of challenging it.

The defense of electoral integrity against the misuse and abuse of social media will depend on the choices and behavior of the major tech companies and platforms, and just as importantly, governments, politicians, traditional media, election management bodies, and citizens. In order to protect electoral integrity in the digital age, we will need to strengthen the capacities of the defenders of electoral integrity, and build shared norms around the acceptable use of digital technologies in elections. Technology platforms and public authorities must act to bolster electoral integrity.

## BUILDING CAPACITY

### Recommendation 1.

Greater attention and resources must be dedicated to promoting election integrity. Public authorities, international organizations, philanthropic foundations, and civil society must invest in tech talent and digital capacity, media efforts, and election management bodies that protect and promote electoral integrity. All relevant stakeholders must cooperate, collaborate and rapidly share information related to threats to election integrity. These efforts should include:

- Building an election vulnerability index that gauges which elections require close monitoring for potential electoral interference, online coordinated inauthentic behavior, and mis-and-disinformation;
- Building the capacity of national partnerships dedicated to defending the integrity of elections against weaponized disinformation and support better evaluation and sharing of practices;
- Funding civil society organizations that counter hate speech, targeted harassment, and the incitement of violence, especially in the lead-up to elections; and
- Helping election management bodies (EMBs) develop expertise in best cybersecurity practice;
- Helping democracies build civic technology programs through the teaching of coding, especially to women and minorities, and by incorporating technical talent into government teams.

### Recommendation 2.

Some EMBs may find themselves in need of short-term technical assistance against threats to electoral integrity by foreign interference in elections, hacking, and hate speech leading to election-related violence. In such cases, international technical assistance to help EMBs defend their election should be quickly available when requested. In order to ensure such assistance is delivered promptly, we recommend the development of standing electoral cybersecurity teams that could be deployed immediately on demand. Such teams could be located in existing international organizations, such as in the United Nations Electoral Assistance Division, or regional organizations, or in a new international institution. Such teams should have the capacity for rotational technical fellow positions for best digital government practice.

## BUILDING NORMS

### Recommendation 3.

We endorse the call by the Transnational Commission on Election Integrity for political candidates, parties, and groups to sign pledges to reject deceptive digital campaign practices. Such practices include the use of stolen data or materials, the use of manipulated imagery such as shallow fakes, deep fakes, and deep nudes, the production, use, or spread of falsified or fabricated materials, and collusion with foreign governments and their agents who seek to manipulate the election.

**Recommendation 4.**

Democratic governments must come together to establish an international convention regarding the role of foreign governments and their agents in other countries' elections. In particular, they should develop international norms that distinguish legitimate cross-border assistance from illicit or unlawful interventions.

**Recommendation 5.**

Democratic governments should consider electronic electoral technologies (EETs) critical infrastructure, and should support the norm endorsed by the G20 that “state[s] should not conduct or knowingly support Information and Communication Technology activity... that intentionally damages critical infrastructure.”

**Recommendation 6.**

Vendors of election equipment and services should commit to a code of conduct to guarantee their products are secure, and their business practices protect the rights, privacy and data of citizens in their client countries, and adhere to honest, transparent practices in procurement. In turn, the international electoral integrity community should pledge that electoral assistance to countries will be conditional on vendors signing and adhering to the code. A multi-stakeholder initiative, involving at a minimum the electoral integrity community, the Global Network of Domestic Election Monitors and international partners should develop such a code of conduct.

**Recommendation 7.**

The electoral integrity community should create norms and standards for transnational political campaign consultants, including public relations and strategic communication firms, and digital marketers. Government regulation should develop procedures for certifying these consultants and prevent any company from continuing to work on elections if it breaks the norms, rules and standards of campaign consulting.

## ACTION BY PUBLIC AUTHORITIES

**Recommendation 8.**

Countries must adapt their political advertising regulations to the online environment. Relevant public authorities should:

- Define in law what is considered to be a political advertisement;
- Compel social media platforms to make public all information involved in the purchase of an ad, including the real identity of the advertiser, the amount spent, targeting criteria, and actual ad creative;
- Specify by law the minimum audience segment size for an ad; and
- Legislate a cooling-off period for digital political ads at least 48 hours before an election.



**Recommendation 9.**

Public authorities must compel major Internet platforms to provide independent parties with meaningful data about the impact social media has on democracy. In particular, platforms must:

- Share secure, privacy-protected data with certified academic institutions to examine issues such as: auditing algorithms for bias towards extremism, understanding the effect of social media on political polarization and information consumption, and disentangling the relationship between online hate speech and offline violence.
- Update transparency reports to provide the public with data about the number of reports of hate speech and abuse online. This should include data about the instances of targeted abuse (against race, gender, sexual orientation, religion) and the frequency with which the abuse targets different communities; and
- Label accounts that use automation. If an account is not correctly labelled as automated (e.g., a bot), platforms should face financial penalties.

**Recommendation 10.**

Public authorities should promote digital and media literacy programs in schools and in public interest programming for the general population.

**ACTION BY PLATFORMS****Recommendation 11.**

Platforms should provide greater transparency surrounding political ads:

- Platforms should require users to choose to opt-out or opt-in to political advertising.
- Platforms should only allow candidates, parties and groups who have pledged to avoid deceptive campaign practices to purchase ads. Such pledges should then become working standards for platforms to decide on whether to accept any given ad.
- To avoid the cloaking of funders behind deceptive organizational labels, platforms should require public disclosure of the identity of human beings funding any political advertisement.

**Recommendation 12.**

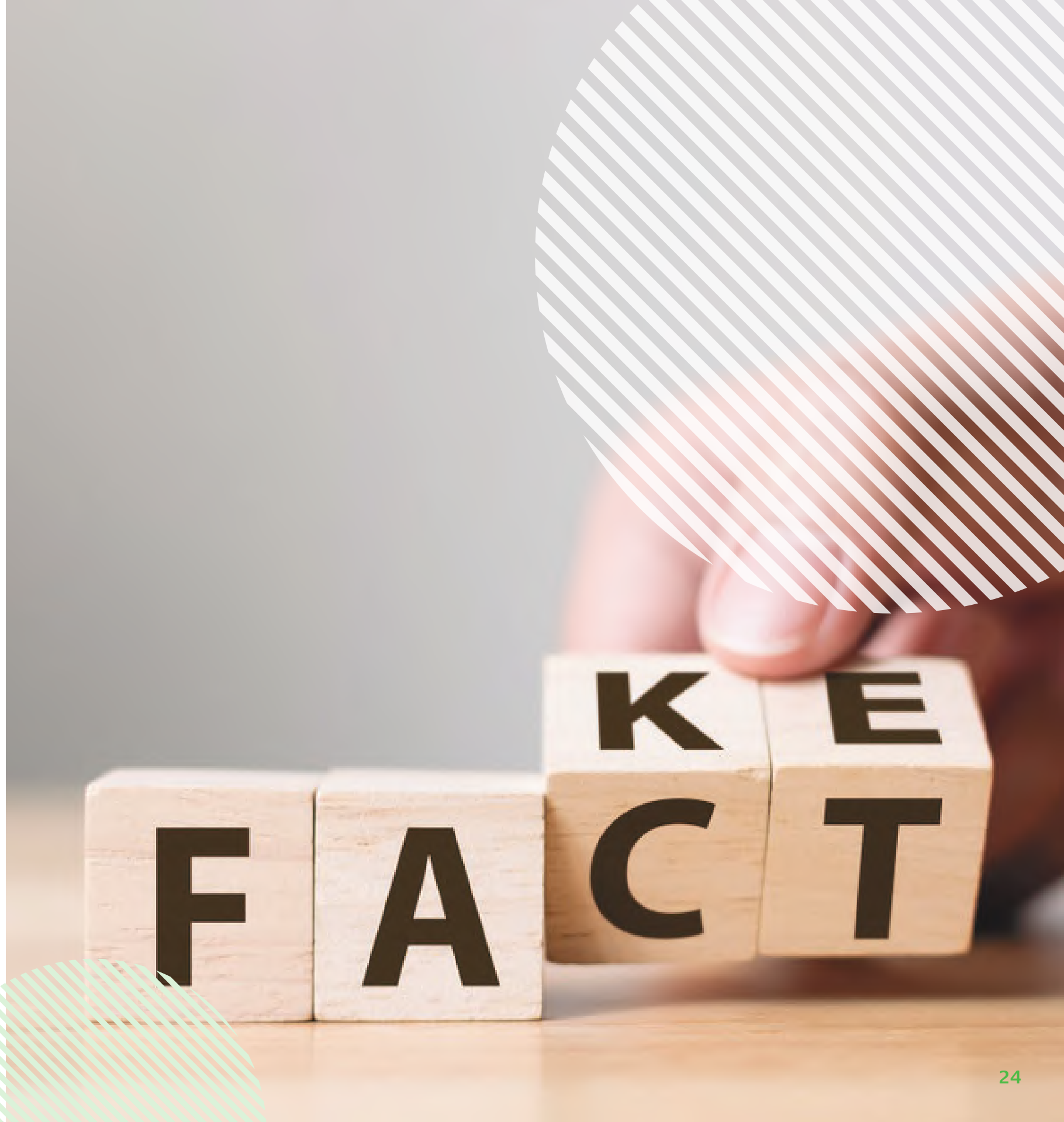
Social media platforms need to develop early warning systems for election-related disinformation, foreign interference, hate crimes, threats to women, violence, and voter suppression:

- Platforms need to employ more experts who speak local languages and have cultural competency where they are operating;

- Because responding once communication achieves virality is too late, early warning systems must initiate human review for accounts and posts that pose a potential threat to elections. Content that achieves a certain level of virality should be subject to human moderation and review.

**Recommendation 13.**

Social media platforms should create a coalition to address digital threats to democracy and election integrity, akin to what they have done collaboratively to address terrorism and child exploitation. Members of the coalitions would meet regularly, and create cross-platform strategies for detecting and limiting the reach of weaponized disinformation and hate speech.



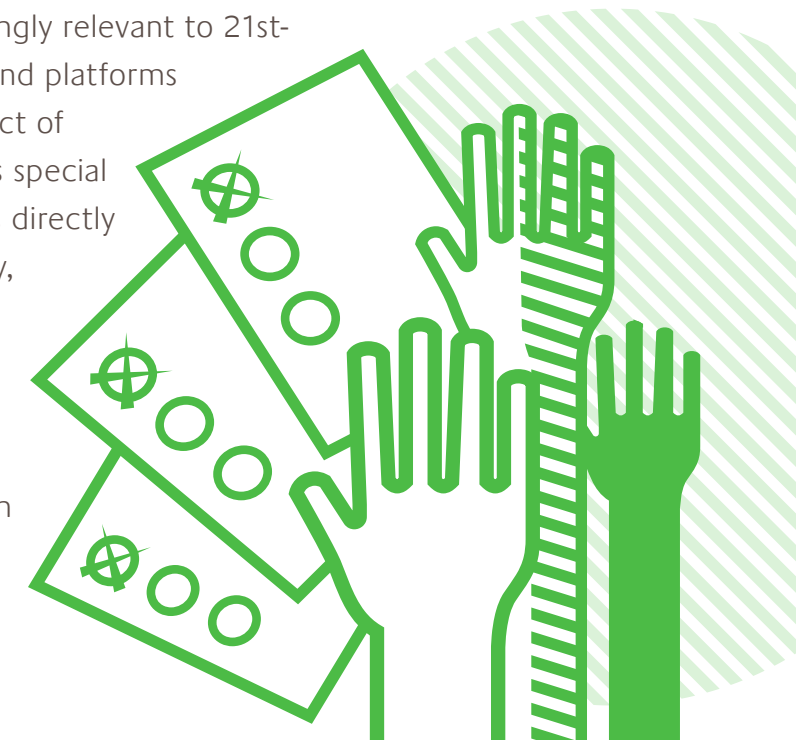
# I. ELECTIONS AS FOCAL POINTS IN THE STRUGGLE FOR DEMOCRACY

Digital communication technologies have enabled those who would exploit the fissures of modern democracy to undermine elections and democratic deliberation. Parties, candidates, campaign consultants, and foreign governments have weaponized social media to spread disinformation, incite hate and violence, and meddle in elections, both domestically and abroad. In a global environment characterized by increasing polarization and growing levels of distrust, elections have become the front line in the struggle for democracy.

The unique properties of the new information and communication technologies (ICTs)—virality, velocity, anonymity, homophily, and transnational reach—create novel challenges for democracy that have reverberated around the globe.<sup>1</sup> Before the public spotlight was cast on the 2016 U.S. Presidential Election—which featured political bots and Russian trolls amplifying disinformation to sow societal distrust and undermine candidates—the same tactics had been developed, refined, and used in Kenya, the Philippines, and Nigeria, among others. Unscrupulous consulting firms like Cambridge Analytica deployed information warfare tactics in elections across Africa, Asia, and Latin America, several years before their campaigning in the United States and the United Kingdom. And populist political parties and candidates have exploited social media to fuel nationalist sentiment, spreading hate and intolerance to push their fringe agendas into mainstream debates.

But research to date suggests that the effects of these technologies are neither uniform nor uniformly negative. Digital technologies can certainly exacerbate political turmoil in countries that suffer from pre-existing polarization and distrust. As our review of social media and democracy in Latin America concludes: the effects of social media “amplify prior patterns rather than create new ones.”<sup>2</sup> Research in Africa and Latin America have also found incremental positive effects on political engagement, such as increased voting, joining social movements, and coordinating political action.<sup>3</sup> In Africa, digital platforms are powerful catalysts for the spread of social movements, and are an important—if not fundamental—medium for political conversation, information sharing, and democratic deliberation.<sup>4</sup>

As social media networks become increasingly relevant to 21st-century politics—both as campaign tools and platforms which enable public deliberation—the effect of technology on electoral integrity demands special consideration. Indeed, digital technologies directly affect how citizens take part in democracy, and elections are critical periods when citizens are particularly attentive to public discourse. Elections are unique tests of democratic legitimacy, and perceptions of their integrity can have outsized effects on democratic stability and performance.



Our mandate is to identify the challenges new ICTs have raised for electoral integrity and develop policy measures that highlight the opportunities for digital technologies to strengthen electoral integrity, while mitigating the harms that have arisen from technological innovation. Electoral integrity is defined as “any election that is based on the democratic principles of universal suffrage and political equality as reflected in international standards and agreements, and is professional, impartial, and transparent in its preparation and administration throughout the electoral cycle.”<sup>5</sup> We focus not just on how social media may affect the conduct of an election, but also on how it may affect *perceptions about the integrity of an election*. As one NGO observes, “If voters and candidates believe that an election is fraudulent or has been poorly administered, they may not accept the outcome. At best, this can breed public dissatisfaction or disinterest; at worst, violence, ineffective governance and long-term instability.”<sup>6</sup>


One component of electoral integrity that we pay particular attention to is the environment in which elections take place, and whether the participants—candidates, parties, and supporters—have confidence in their mutual security.<sup>7</sup> Electoral integrity requires that citizens believe elections are a repeated game and will continue to be held long into the future. In many transitional democracies, political contenders fear that if they lose an election, they will be shut out of power permanently. Candidates, parties, and supporters must be confident that if they lose an election, they will be free to organize and contest future elections and the winners will not use their power to disenfranchise them. They must also be confident that if they win an election, they will earn the ability to form a government and pursue their policy agenda, and the losers will not turn to violence or block their electoral mandate.

ICTs—and social media in particular—pose several potential challenges to electoral integrity, and we address each of these in turn: polarization, hate speech, disinformation, new forms of political advertising, and foreign interference. Although we focus on specific challenges ICTs pose to electoral integrity, two larger concerns about democratic governability were expressed in our various consultations around the world.

First, legislators from Europe to Latin America expressed their belief that the speed and ubiquity of social media create pressures to respond immediately to demands, news, grievances, and accusations, and undermine the mandate of parliaments to deliberate and to set policy agendas. Some parliamentarians felt that the very nature of political representation was breaking down because of the intense mobilization of online opinion and the ability of social media campaigns to put enormous scrutiny on the day to day lives of elected officials. In some countries, the toxicity of personal attacks and threats online have led some Members of Parliament, disproportionately women, to leave political life. Although the effects of social media on the quality of democratic governance are beyond our mandate, we believe that it is an issue that deserves greater scrutiny and research.

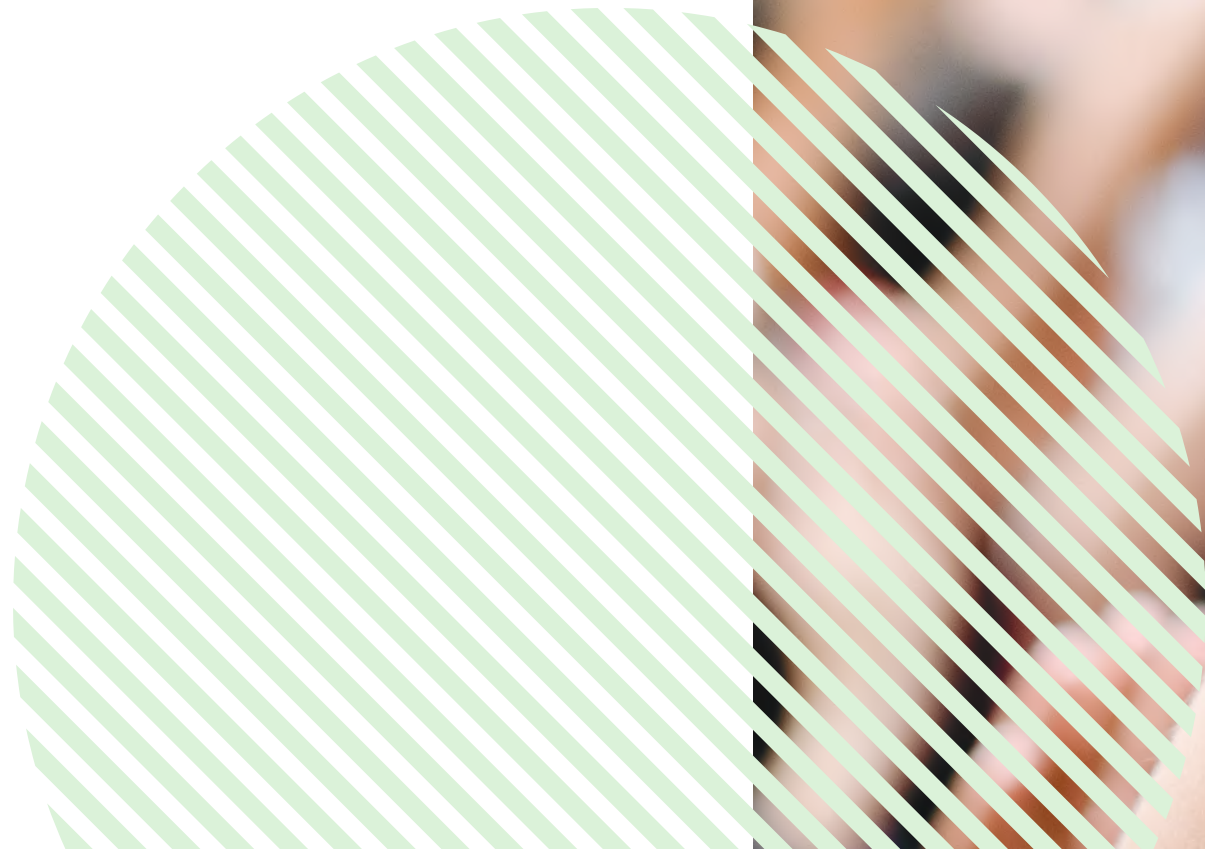
Second, we repeatedly heard anxieties about whether democracy can survive in a world where fake news is rampant and citizens cannot agree on basic facts. Although the quality of democratic deliberation may have declined in some long-standing democracies, it is worth noting that contemporary critiques about the poverty of deliberation in our democracy predate the rise of the Internet.<sup>8</sup> And just as fake news and hate speech have been around for centuries, there has





never been a time when citizens in democracies all shared the same facts or agreed on what constitutes a fact. Democratic citizens often disagree on fundamental facts and certainly do not vote on the basis of shared truths.<sup>9</sup> Democracy is needed precisely because citizens do not agree on fundamental facts. Even in the digital age, democracy provides several advantages that authoritarianism does not: greater protection of rights and liberties, greater access to information, greater opportunity for interaction among citizens, and greater possibility for meaningful debate.

As a normative proposition, we can agree that the quality of democracy is enhanced when citizens agree on facts and what constitutes a fact, and we should strive to improve democratic deliberation.



## II. AFFECTIVE POLARIZATION, SOCIAL MEDIA, AND ELECTORAL INTEGRITY

Polarization is increasingly posing challenges to governability, social cohesion, and democracy. Polarization is multifaceted, but in this report we are most concerned with affective polarization, where partisan animus leads political supporters to hold negative opinions and beliefs about their opponents. When affective polarization becomes severe, “people increasingly perceive and describe politics and society in terms of ‘us’ versus ‘them,’”<sup>10</sup> with potential toxic effects for electoral integrity. As one annual report on democracy states, “Once political elites and their followers no longer believe that political opponents are legitimate and deserve equal respect, or are even acceptable as family and friends, they become less likely to adhere to democratic rules in the struggle for power.”<sup>11</sup>

In the United States, polarization has increased steadily since the 1970s.<sup>12</sup> Before the spread of the Internet, the creation of smartphones, and the rise of social media, political polarization was stoked by partisan traditional media.<sup>13</sup> The opening of cable news in the 1980s, the end of the Fairness Doctrine in 1987, and the emergence of hate radio grew an extreme partisan media ecosystem in the United States, deeply distrustful of traditional media, and vulnerable to conspiracy theories and propaganda.<sup>14</sup> This highly partisan media ecosystem has become a forum for the worst excesses of Internet disinformation and hate-mongering and is a platform for irresponsible politicians, pundits and journalists.<sup>15</sup>

The United States example suggests that countries with pre-existing polarization, distrust in traditional media, and partisan media ecosystems

are much more vulnerable to the manipulations of social media than countries with low polarization and distrust.<sup>16</sup> Where polarization is already high, social media can be easily weaponized to exacerbate and intensify political division and conflict. Where trust in traditional media is low, citizens eschew objective news reporting. Highly partisan media plays to the worst instincts of its readership. This insight counsels caution in generalizing the effects of social media across countries with widely varied institutions, societal divisions, and media ecosystems.

In some ways, the United States is extreme among long-standing democracies in its degree of polarization, partisan media, and low trust.<sup>17</sup> In a recent study that measures affective polarization by the extent that “public debate is respectful, builds on facts, and opponents are open to persuasion by reason,” the United States ranks 98th out of 178 countries and scores closer to India (102nd), Poland (109th), Brazil (117th), and Hungary (127th), than to long-standing democracies such as Norway (1st), Switzerland (2nd), or Denmark (3rd).<sup>18</sup>

Democracies in the rest of the world show a mixed picture. Many democracies in Asia, Latin America and Africa rank among the highest in affective polarization, but some of the largest increases have occurred in Western Europe.





Empirical studies indicate that where we find polarization and distrust in countries globally, they are not new but rather a product of long-term trends predating the rise of social media.<sup>19</sup>

While political polarization has increased in some democracies, trust among citizens towards each other, towards the media, and towards parliaments or legislatures are generally on the decline.<sup>20</sup> In most countries, there has been a steady decline in the amount people trust each other. Moreover, one sees gradual declines over time in the percentage of citizens expressing a “great deal of confidence” in the media throughout Europe, and Central, South, and North America. Africa shows the greatest overall decline and the greatest volatility; Asia and Oceania is the only region that has shown a net increase in confidence in the press over the last three decades.

Citizen confidence in parliament has declined, but with greater variability. In North America, Europe, Central and Latin America, and Africa, there has been a thirty-year reduction in the percentage of citizens who express a “great deal of confidence” in their legislature. The only outlier region has been Asia and Oceania, which showed a slight overall increase from the early 1980s to the 2010-2014 survey period. The loss of trust over time is less due to generational attitudes, but rather due to citizen judgments about the actions of government and their trustworthiness, suggesting that when democracies deliver, citizens respond positively.<sup>21</sup> Increasing political polarization and declining trust threaten electoral integrity. Political polarization and distrust weaken beliefs in mutual security, and once they begin to erode, organized online disinformation and hate speech can poison electoral environments.

Fear can take hold of voters who believe that if their party loses in the near term, they will lose forever. Elections become solely about winning, with little regard for rules, laws, ethics or democratic norms.

Many transitional democracies in the Global South exhibit high polarization, low trust and partisan media, and thus rank highly vulnerable to online disinformation and hate speech. Elections in these countries are already focal points for violence and destabilization, and social media have already been weaponized to intensify polarization and weaken norms of mutual security necessary for elections with integrity. This trend will not only continue but is likely to escalate.

The finding that countries with high polarization are more vulnerable to the weaponization of social media does not in any way exonerate the platforms from responsibility for the harmful effects of their products. The platforms rushed their products into countries such as Myanmar, Sri Lanka, and Kenya that were extremely vulnerable to disinformation, propaganda and hate speech. The platforms did not consider how their products would be used in extremely polarized countries with repertoires of violence. Once the platforms were aware of the deadly potential of their products, they were way too slow in remedying the problem.

The causes of long-term polarization are complex and multifaceted. There is a growing literature that implicates the steady rise of economic inequality as a cause of polarization<sup>22</sup> and rising support for more extreme political positions.<sup>23</sup>

Other research suggests that electoral systems matter, and that ‘first-past-the-post’ majoritarian democracies are most prone to extreme polarization.<sup>24</sup> Still others point to growing status concerns of rural residents and workers who feel left behind by urban growth and dynamism, and feel threatened by immigration.<sup>25</sup> A number of plausible recommendations follow from the analyses of these causes: implement social and economic policies that protect the middle class and labor,<sup>26</sup> institute political reforms that mitigate zero-sum outcomes and perceptions,<sup>27</sup> and create social opportunities for groups to interact, deliberate together, and build a commitment to larger political identities.<sup>28</sup> It is beyond the scope of this commission to suggest recommendations on how societies can best prevent extreme polarization, but it is clear to us that this is an important first line of defense in building immunity to the distortions of social media.

## **SOCIAL MEDIA AND POLARIZATION**

The importance of the Internet and social media as a platform for news and information consumption has led to growing concerns about the ways in which technology might exacerbate or aggravate polarization. Since the early days of the Internet,

some scholars posited that social media lead to filter bubbles and echo chambers, and segregate citizens into groups who read the same news, communicate only with each other, and therefore think alike, with deleterious effects on how democracies govern.<sup>29</sup> The algorithmic, individual, and social filtering of content affect the kinds of information to which individuals are exposed. Critics suggest that voters who spend time online may not get a representative, balanced, or accurate selection of news and information, nor may the distribution of quality information be equally distributed across a voting population. Similarly, some journalists argue that algorithms radicalize significant portions of the population by feeding content that induces individuals to take more extreme positions. Recent media coverage suggests biases exist in social media algorithms that push users towards extremist content, such as Google Search ranking Holocaust denial websites over legitimate sources of information,<sup>30</sup> or YouTube’s autoplay features recommending increasingly radical content to viewers.<sup>31</sup>

Research on the question, however, has proven inconclusive for two reasons. First, it is difficult to distinguish between users who tend to associate willingly with people and news sources that reinforce their politics and worldview, and users who follow content that they would not choose on their own but is fed algorithmically through newsfeeds and recommendations. Second, the platforms have not shared the appropriate data for researchers to answer the question.

Research suggests that social media actually promotes media diversity and access to a range of political viewpoints and informational sources, especially compared to traditional news



sources.<sup>32</sup> Similarly, platforms offer greater opportunities for cross-cutting interactions by connecting individuals to their family and close friends, as well as relatives, co-workers, or acquaintances—relationships we might consider to be “weak ties”—who are more likely to post or share ideologically diverse content.<sup>33</sup> Through this dynamic, it is possible that social media platforms increase the range of political views to which users are exposed.

At the same time, it is likely that some people do live in echo chambers, and some do not.

Research on Germany, Spain, and the United States shows that Twitter users within heterogeneous networks often develop more politically moderate networks over time, suggesting that those already predisposed to consuming cross-cutting information continue to do so in their online interactions.<sup>34</sup> On the other hand, recent research suggests that partisans actually increase in polarization when exposed to opposing views on social media.<sup>35</sup>

Some people are obviously radicalized through the Internet, hence the substantial resources platforms have devoted to defeating online terrorist recruitment. When we turn to the issue of political extremism, however, the question is whether the algorithms of the platforms are doing the radicalizing. Few rigorous studies of the recommendation algorithm exist, and researchers reach different conclusions: where one study found YouTube had a slight algorithmic bias toward promoting increasingly radical videos,<sup>36</sup> another suggests that the volume of extremist content has simply grown in response to demand, and not as the result of an algorithmic bias that pulls its audience to more radical material.<sup>37</sup>

## IMPLICATIONS FOR ACTION

In order to develop effective policies that limit any harms, policymakers need a strong empirical base to make decisions. The biggest hurdle to understanding the impact social media have on opinion diversity, or the radicalization of viewpoints, is the limited data that the platforms have made available to qualified researchers. Much of the existing research to date is overwhelmingly on the United States and Europe, with little platform data shared on Africa, Asia, and Latin America. Moreover, we have little data on certain platforms that are much more prevalent in the Global South, such as WhatsApp. And on some basic questions, like radicalization through YouTube, the inaccessibility of the recommendation algorithm to outside research prevents us from assessing whether the changes they have made over the last year have led to any difference in outcomes.

In order to study any of the pathologies attributed to the transformation of the digital media ecosystem, social scientists need access to platform-controlled data as to “who” saw or engaged with “what” “when.” That is, scientists need to understand how and when certain people, and the population at large in different countries, interact with new media and what the consequences of those interactions are. Even when the platforms have promised to make available data for independent academic research, those promises have often gone unfulfilled.<sup>38</sup>

*Public authorities must compel major Internet platforms to provide independent parties with meaningful data about the impact social media has on democracy. In particular, platforms must share secure, privacy-protected data with certified academic institutions to examine issues such as: auditing algorithms for bias towards extremism, understanding the effect of social media on political polarization and information consumption, and disentangling the relationship between online hate speech and offline violence.*

Given the large numbers of elections held annually worldwide, and the variable susceptibility of countries to toxic polarization, disinformation, and hate speech, it would be helpful for the international electoral integrity community working in conjunction with the platforms to be able to prioritize which countries will require the greatest attention and resources to protect their electoral integrity.

*The electoral integrity community should invest in building an election vulnerability index that gauges which elections require close monitoring for potential electoral interference, online coordinated inauthentic behavior, and mis-and-disinformation.*





### III. HATE SPEECH AND ELECTORAL INTEGRITY

The Internet facilitates coordination and collective action among groups—including between extremists in geographically scattered places. Gab, 4chan, 8chan, and avowedly racist subreddits have become popular outlets for discussion, the building of shared identity, and mobilization among pariah groups. The anonymity of these networks helps facilitate offensive and hateful discussion while avoiding accountability. The rise of unaccountable speech has raised concern about the relationship between social media and new waves of political extremism, as well as between social media and political violence.

In recent years, the ties between online hate speech and offline abuse have become more apparent. In several Western democracies, white supremacists used social media to publicize mass shootings targeting religious or racial minorities.<sup>39</sup> In India, rumors spread on WhatsApp incited lynch mobs and communal violence, killing dozens.<sup>40</sup> In Sri Lanka, anti-Muslim posts in March 2018 fueled violence against its Muslim minority population, leading to the burning of several hundred houses and businesses.<sup>41</sup> In Myanmar, government operatives flooded Facebook with anti-Muslim rhetoric and called for the ethnic cleansing of the Rohingya minority, contributing to the displacement of 700,000 refugees due to increasing threats, physical attacks, and sexual violence.<sup>42</sup> The risk of social media amplifying hateful discourse is especially consequential for countries in the Global South, where the coalescence of longstanding ethnic or religious tensions and the rapid adoption of new information technologies can intensify political conflict.

Women are targeted, in particular, by online hate. According to a recent poll commissioned by Amnesty International, almost a quarter (23 percent) of women in eight democracies said they had experienced online abuse or harassment at least once.<sup>43</sup> Similarly, the European Parliament found that one-fifth of women in the European Union (EU) had experienced sexual harassment online.<sup>44</sup> Gender-based intimidation online can be a powerful tool of self-censorship that chills freedom of speech and disrupts the democratic participation of targeted groups.<sup>45</sup> According to Amnesty International, most women who experienced online abuse changed the way they used social media—adjusting their privacy settings and changing the content they post<sup>46</sup> [Box 1]. The demobilizing effect of abuse may be particularly acute for women journalists and political candidates.



## BOX 1

## Social Media as a Weapon Against Women

Social media is increasingly weaponized to target female journalists, activists, and politicians with threats of rape and violence. Moreover, disinformation is used to undermine the credibility and capacity of prominent female voices on the frontlines of democratic politics. For example, when Maria Ressa—a prominent Filipino journalist—started covering President Rodrigo Duterte’s use of propaganda in the 2016 Philippine election campaign, his keyboard army used online threats, targeted harassment, and allegations of corruption to intimidate her.<sup>47</sup> After Rappler, a popular news outlet, published the transcript of a telephone conversation between President Trump and President Duterte, a coordinated network of bots and fake accounts flooded social media with the hashtag #ArrestMariaRessa, leading her to receive a consistent flow of hate messages and threats, including a call for her to be “raped repeatedly until she died.”<sup>48</sup>

Ressa’s experiences with state-sponsored trolling mirrors several campaigns waged against prominent female figures around the world. When Ukrainian MP Svitlana Zalishchuk spoke at the United Nations about the effect of the War in Donbass on women in Ukraine, a Russian-backed disinformation campaign began sharing a fake tweet of hers claiming she would “run naked through the streets of Kiev if the Ukrainian army lost a key battle.”<sup>49</sup> Doctored nude images of her circulated on social media to further discredit and shame her.

These targeted attacks seek to suppress the voice and political participation of women through shame and intimidation. Unlike other forms of trolling, attacks on women often last for much longer periods of time and focus on demeaning, sexualized insults, which makes them more vicious and enduring. Gender-based cyber intimidation tactics can be a powerful tool to chill freedom of speech and disrupt women’s online and offline political participation.<sup>50</sup>

Hate speech is a concern for electoral integrity because it undermines the mutual security necessary for peaceful contestation. It is also a weapon for those candidates and parties who stoke violence in order to suppress voting by their opponents. This is most worrisome in transitional democracies in the Global South that already have a history of electoral violence and weak rule of law to hold offenders accountable.

## APPROACHES TO REGULATING HATE SPEECH

A critical challenge for moderating hate speech online is that it is nearly impossible to define and separate from other instances of offensive language. Even when platforms create a working definition, the complexity and variation within natural language constructs makes the task of automatic hate speech detection imprecise and inconsistent. Although industry leaders and policymakers generally agree that more should be done to combat hate speech, there is substantial disagreement about what should be done, given these challenges to conceptualization and implementation. It is particularly difficult to separate hate speech from legitimate political speech when countries’ leaders themselves engage in the precise kinds of wordplay and sometimes outright racism that represents the basis for the kinds of censorship critics ask of platforms.



Debates about balancing freedom of expression with protecting individuals from discrimination are longstanding. Moderating hate speech in the digital era, however, is complicated by the extra-jurisdictional reach of platforms. Multiple regional models have emerged in response to this challenge, each making an inherent trade-off between protecting free speech and regulating speech that could justify or incite group-based hatred and violence.

The self-regulatory model is most prominent in the United States, where hate speech is legally protected under the First Amendment,<sup>51</sup> and social media platforms are immunized from legal liability for almost all except intellectual property and federal crimes claims.<sup>52</sup> However, tech companies have taken a number of steps to limit the spread of hateful content, including banning accounts, deleting and demoting certain content, and diluting the reach of extremist ideas by countering information with alternative resources, using both artificial intelligence and human moderation. The self-regulatory approach has limitations: moderators based in Silicon Valley do not necessarily have the deep cultural, political, or religious knowledge necessary to review content being shared on the other side of the globe<sup>53</sup> and there is little transparency or accountability for how technical and policy decisions are implemented.<sup>54</sup>

A more quasi-regulatory model has developed in the European Union. Under pre-existing EU law, platforms already had an obligation to take down any illegal content, including illegal hate speech, upon notification. In 2016, the European Commission created a Hate Speech Code of Conduct, under which Facebook, Microsoft, Twitter and YouTube agreed to incorporate hate speech prohibitions into their Community Guidelines. The Code of Conduct thus relies on tech platforms' Community Guidelines

as formal enforcement mechanisms to take down hate speech as defined by European law, following specific timelines for responding to notices. By 2018, Instagram, Google+, and Snapchat also signed on this EU Code of Conduct.<sup>55</sup> This EU initiative is relatively hands-off, insofar that it did not create an actual regulator in the role of content removal enforcer. European civil liberties organizations have criticized this framework as an unaccountable system, leaving too much discretion at the hands of private companies.<sup>56</sup> Although the Commission's Code of Conduct hoped to develop networks of information and trust among the tech industry, civil society, and government, it has received mixed evaluations.

In Germany, there has been a turn toward more direct government regulation under the 2017 NetzDG law. NetzDG protects German social media users against hate speech and harassment by making social media companies accountable to addressing user complaints in a timely fashion,<sup>57</sup> threatening platforms with fines of up to 50 million Euros for failure to remove 'clearly' unlawful content within 24 hours' notice.<sup>58</sup> However, the short removal time frame makes it infeasible for platforms to consider flagged content in political and cultural context, creating incentives for blanket over-removal of free speech. In 2017, the United Nations Special Rapporteur for the Protection of Freedom of Opinion criticized the law for creating incentives for "precautionary censorship" which "would interfere with the right to seek, receive and impart information of all kinds on the Internet."<sup>59</sup> Moreover, if authoritarian governments implement laws similar to NetzDG in mandate, they could legitimize further controls on information<sup>60</sup> and undermine international human rights standards.<sup>61</sup> Indeed, copycat laws have already been passed in Russia and Venezuela, among other authoritarian states.<sup>62</sup> In more extreme cases, authoritarian governments use Internet shutdowns to exert control over information **[Box 2]**.

## BOX 2

### Internet Shutdowns in Asia and Africa

Authoritarian governments continue to rely on internet shutdowns to quell political protests in Asia and Africa. In 2015, international legal experts from the UN, OSCE, OAS and ACHPR (African Commission on Human and Peoples' Rights) condemned shutdowns for aggressively limiting speech and restricting access to information and emergency services in times of unrest. Nevertheless, some politicians are perpetuating an unfree internet culture by removing access to social media and online forums in response to crises.<sup>63</sup>

Approximately half of global network shutdowns, including intentional disruptions on social media, cell phone service, and/or internet access, have occurred in India. Although the Indian government has employed these shutdowns under the guise of quelling violence during politically contentious periods over the past few decades, shutdowns have been associated with disruptions to both violent and non-violent collective action by internet users.<sup>64</sup>

In the wake of the 2019 Easter bombings in Sri Lanka, the government blocked access to social media sites claiming similar objectives. Although some media outlets praised the government for preventing the spread of hate speech, later analyses criticized the blunt decision. In a country with relatively weak and unfree media, lost internet communications added to the chaos and confusion of the day. Furthermore, malicious actors intent on spreading fake news and violent rhetoric were still able to spread vitriol around the region using virtual private networks.<sup>65</sup>

Internet shutdowns in Africa have increased in recent years. More than a dozen African countries have disrupted internet access during elections or periods of political dissent, such as Egypt,<sup>66</sup> Togo, and Ethiopia. A study of ten sub-Saharan African countries found that blackouts have not only been dangerous but expensive; between 2015-2017 blackouts led to an estimated loss of \$235 million in those countries.<sup>67</sup>

Strikingly in Uganda and Benin, shutdowns have accompanied unequivocal state repression. Just prior to Uganda's 2016 election, the state's Electoral Commission disrupted access to Facebook, Twitter, and similar forums. During the shutdown, the state held two presidential candidates under house arrest and did not reinstate full internet access until voting had concluded. Critics argued that this allowed the government to shield itself from criticism for interference in the electoral proceedings.<sup>68</sup> In famously stable and democratic Benin, the electoral authorities similarly ruled candidates from five opposition parties to President Talon ineligible to compete in the 2019 parliamentary elections. The government cracked down on protests and public outcry, shut down certain social media sites, and finally blocked all internet access on election day.<sup>69</sup>

These examples and more have threatened access to information and electoral freedoms across Asia and Africa. Although social media and other online communication forums pose challenges for regulators, increasing authoritarian reliance on shutdowns reinforces the value of these technologies for democratic activists.

Notably, none of these models offer a clear policy framework for grappling with the fact that hate speech has ‘gone mainstream’ in many countries and is increasingly integrated into normal political conversation and the discourse of numerous prominent political elites. Governments have yet to find a way for platforms to address this trend without interfering in electoral campaigns or creating partisan rulings about what political speech should be censored.

## IMPLICATIONS FOR ACTION

Under different regulatory regimes, distinct principles guide how, when, and why hate speech is removed from social media. However, across all regulatory models, tech platforms must do more to scale their practices to the size of their market. There are still several limitations to AI-based solutions, especially when applied to the Global South and regions where English is not the primary language. At the same time, platforms have a responsibility to enforce the same content moderation standards globally—and must be especially responsive in communities that are vulnerable to ethnic conflict, riots, and hate crimes. In these contexts, platforms need to invest especially heavily in developing early warning systems, which would flag content that can pose a potential threat to elections for human review before it can reach a certain level of virality. Tech companies need to allocate a greater share of resources toward automated content moderation, text translation, cultural competency in human moderation, and other similar tools to ensure that their users have equal opportunity to participate in online political discourse. They also need to provide more data about the extent to which hate speech

is a problem on their platforms. While transparency reports provide general statistics on the number of takedowns, greater insight into the kinds of content being flagged and removed will not only help improve early warning systems, but will help inform evidence-based policymaking around the protection of communities more vulnerable to hate speech, harassment and abuse.

*Social media platforms need to develop early warning systems for election-related disinformation, foreign interference, hate crimes, threats to women, violence, and voter suppression.*

*Platforms need to employ more experts who speak local languages and have cultural competency where they are operating.*

*Because responding once communication achieves virality is too late, early warning systems must initiate human review for accounts and posts that pose a potential threat to elections. Content that achieves a certain level of virality should be subject to human moderation and review.*

*Governments must compel social media platforms to update transparency reports to provide the public with data about the number of reports of hate speech and abuse online. This should include data about the instances of targeted abuse (against race, gender, sexual orientation, religion) and the frequency with which the abuse targets different communities.*

One major challenge for scaling moderation efforts concerns the diversity of the experts designing content review technologies. Algorithms, AI, and other technologies are not neutral tools; they embed the values and biases of their creators and users. A growing body of empirical work has pointed to how these biases inherent in

company technologies and practices can lead to inconsistent—and often discriminatory—automated decisions.<sup>70</sup> Yet, most of the engineers imagining the next generation of technologies are young white men based in Silicon Valley. Part of the long-term solution will involve ensuring engineers and industry leaders have diverse backgrounds and experiences. This will require long-term investments in diversifying the fields of computer science, engineering, and data science [Box 3]. This long-term investment will strengthen electoral integrity—broadly speaking—by granting decision-making power over platform design to a group of experts that more closely represents the demographics and cultural characteristics of social media consumers.

*Public authorities, international organizations, philanthropic foundations, and civil society should help democracies build civic technology programs through the teaching of coding, especially to women and minorities, and by incorporating technical talent into government teams.*

Citizens, civil society, and government can also do more to promote a healthy online environment—contributing to a greater share of high quality, respectful political dialogue. The efforts by civil society organizations in Kenya during the 2013 elections to reduce the amount of hate speech online were impressive. By creating an environment that empowered citizens to counter, discredit, and shame those who shared violent speech, civil society organizations were ultimately able to reduce the proportion of speech that was hateful and create an electoral environment that focused on nonviolent dialogue [Box 4].

*The electoral integrity community should fund civil society organizations that counter hate speech, targeted harassment, and the incitement of violence, especially in the lead-up to elections.*

## BOX 3

### Initiatives for Diversity in Coding

Electoral integrity necessitates equal access to political participation among marginalized social groups. However, online abuse makes vulnerable populations feel unsafe, depresses their participation in online political debate, and—in extreme cases—can make their political or journalistic careers unsustainable. Current algorithmic design and content moderation tools fall short in addressing concerns about hate speech and harassment. One long-term solution is to substantially increase diversity among the engineers and tech leaders who develop algorithms and make decisions about content.

In recent years, a number of non-profit organizations have taken aim at creating diverse pipelines leading to tech careers; these initiatives will help extend decision-making power to the marginalized communities most affected by algorithmic biases and weak content moderation systems. These include high-profile organizations that focus on increasing the tech presence of women—such as Girls Who Code, the Grace Hopper Program, and Women Techmakers—as well as programs aimed at minorities—including Black Girls Code, #YesWeCode, and the Algorithmic Justice League.

One high-profile model of success is Girls Who Code—a nonprofit organization that works to increase the number of women in computer science and engineering careers by offering a range of targeted coding programs to young girls. The organization runs after school programs during the academic year to teach computing skills including programming, robotics, web design, and app development.<sup>71</sup> According to their 2018 Annual Report, program alumni major in CS-related fields at nearly fifteen times the national rate.<sup>72</sup> These pipeline initiatives are gaining the support of major technology companies; Google, Twitter, and GE are among the sponsors of Girls Who Code.<sup>73</sup>



## BOX 4

### Citizen Monitoring of Hate Speech in Kenya, 2012-2013

The 2007 Kenyan presidential elections triggered two months of widespread violence. Over 1,200 people were killed and over 600,000 displaced from their homes.<sup>74</sup> Anecdotal evidence suggests that prolific online hate speech was influential in mobilizing violent behavior.<sup>75</sup>

In the months leading up to the 2013 elections, more than 477 people were killed in inter-communal violence amidst inflammatory rhetoric surrounding land disputes. The government's post-2007 National Cohesion and Integration Commission, tasked with prosecuting proponents of hate speech, struggled to define or enforce their own policies.<sup>76</sup> However, violence faded during the elections due to the collective action of civil society, who took innovative approaches to countering hate speech.

Colloquially dubbed 'peace propaganda,' civil society organizations issued widespread pleas for peace and unity on social media, billboards, and through advertising campaigns.

A nonprofit SMS-texting initiative strategically messaged 65,000 Kenyans with peaceful implorations. The popular television show Vioja Mahakamani aired four special episodes about nonviolent dialogue. Research later found that these episodes made viewers significantly more sceptical of hate rhetoric.<sup>77</sup>

A handful of groups piloted hate speech monitoring programs that empowered individuals to act against hate speech in

online forums. One initiative—called the Umati project—ran from September 2012 to May 2013. Scouring the most popular sites in the Kenyan web space, a team of monitors reported nearly 6,000 instances of hate speech. It found that one fourth of these instances involved 'very dangerous' calls to violence and that more than 80% of hate speech occurred on Facebook.<sup>78</sup>

This research spawned the Nipe Ukweli project, which distributed instructions for citizen monitoring of hate speech across social networks, traditional media, and community forums. Fliers described characteristics of hate speech and encouraged citizens to report abusive users to a text hotline and internet database. Produced in English and Swahili, the project emphasized citizens' choice and power to counter, discredit, or simply disengage with violent speech.<sup>79</sup>

## IV. PROTECTING ELECTORAL INTEGRITY FROM DISINFORMATION

Disinformation—defined as the intentional dissemination of false or misleading information—has become a critical threat to electoral integrity. In recent years, a wide range of politically and economically motivated actors have exploited social media to spread and amplify disinformation and propaganda to potential voters in the lead up to elections around the globe, exacerbating long-standing ethnic, religious, and social divides, and sowing distrust in the media and in democratic institutions.

From a normative standpoint, we want informed voter decision-making. When voters are misinformed, they may choose a candidate who does not actually meet their preferences. Voters should understand the consequences of their decisions and be able to hold their representatives accountable. Disinformation can therefore remove accountability from elections. But beyond normative concerns, disinformation can directly diminish electoral integrity by undermining mutual security. Disinformation can also undermine trust in free and fair elections by sowing doubt about the integrity of the ballot box and the professional, impartial behavior by election management bodies (EMBs), and spreading rumors that call into question the legitimacy and accuracy of an election.

Disinformation is not a new phenomenon nor is it unique to contemporary information communication technologies.<sup>80</sup> Arguments that frame disinformation as purely a “social media” challenge detract attention from the role of traditional media and politicians in creating an information ecosystem where disinformation and hate speech thrive. In countries where there is heavily partisan traditional media, citizens

find some of the worst online rumors and disinformation amplified and legitimized in the press and on television and radio. Even where traditional media is largely responsible and objective, it can be susceptible to online disinformation or amplify narratives that serve to undermine trust and electoral integrity.

Nonetheless, social media is often the place where users first discover false or misleading stories.<sup>81</sup> This is, in part, because social media has fundamentally changed the way in which news and information are both produced and consumed. It has become a truism to say that in the digital age, anyone with a keyboard can be a publisher. Microblogging and citizen journalism have provided the average person with greater opportunity to reach broad audiences, in real-time, with news and content. And traditional media and broadcasting organizations—who provide the guardrails for political communication—no longer hold a monopoly over information dissemination.<sup>82</sup> Although social media empowers users as information producers, the decline of traditional ‘gatekeeping’ intermediaries has meant that the norms governing legacy media are not always extended to the panoply of user-generated content. Thus, disinformation—and other forms of low quality, hyper-partisan, or conspiratorial content—can readily find a home on social media.



Social media algorithms—and the policies and practices that govern their use—also contribute to the ways in which disinformation is spread online. Today’s digital information environment is ‘mutually shaped’ by algorithms that sort, rank, prioritize, and deliver content, and users who influence the kinds of recommendations algorithms make through the data generated by online interactions.<sup>83</sup> Algorithms are not neutral technologies.<sup>84</sup> Rather, they are infrastructures of advertising and persuasion, designed to maximize user attention, and subsequently, advertising revenue.<sup>85</sup> Because people are drawn to content that is emotive, vivid, and compelling, both the attention-maximizing algorithms and human preference tend to favor tantalizing fiction over tedious fact. And disinformation—which is purposefully designed to elicit strong and emotional responses—can generate much more engagement than other forms of news and information.<sup>86</sup>

Newfound concerns over the relationship between social media and disinformation rose to prominence following the June 2016 Brexit Referendum in the United Kingdom and the November 2016 Presidential Election in the United States. So-called fake news stories pushing outrageous headlines—such as Hillary Clinton’s alleged involvement in a pedophile ring in the basement of a DC Pizzeria—occupied a prominent position in some users’ social media feeds. Many of these ‘fake news’ stories outperformed professional news,<sup>87</sup> distracting from other important public conversations about the election. Although “fake news” stories like ‘#Pizzagate’ or ‘the Pope’s endorsement of Donald Trump for President’ contained clearly falsifiable information, most disinformation blends truth and falsities in such a way that make it difficult to clearly define what is disinformation and what is not. Not all disinformation is completely false: factually correct information can be used to “disparage

opposing viewpoints” or misrepresent facts.<sup>88</sup> Compounded with this problem is the use of satire, parody, and exaggeration, which may be interpreted as humorous by some, or as news by others.<sup>89</sup> Should disinformation include mistakes in reporting, political satire, misstatements by politicians, or only outright fabrications? Attempts to find a middle ground and flag potential false news as “disputed” have been shown to have unintended consequences, such as making untagged stories seem more accurate and verified.<sup>90</sup>

## MEASURING THE PREVALENCE OF DISINFORMATION

How one defines the problem of disinformation significantly affects its estimated prevalence. But democracies are not—as sometimes claimed—drowning in disinformation. There remains limited empirical research on the actual prevalence of disinformation, and where studies do exist, they tend to conceptualize disinformation differently or focus mainly on the United States, making broad generalizations about disinformation difficult. However, a growing corpus of evidence reminds us that the scale of disinformation differs around the world, across platforms, and between communities of users. Using a narrow definition of “fake news,” Allcott and Gentzkow estimated that in the period leading up to the 2016 U.S. Presidential Election, the average American adult saw and remembered one fake news article on social media.<sup>91</sup> Another study by Guess, Nyhan and Reifler estimated that approximately 27 percent—or 65 million—Americans were exposed to at least one fake news article during a similar timeframe.<sup>92</sup> Taking a



broader definitional approach, Howard et al. found that Twitter users in the United States shared as much “junk news”—or content that was conspiratorial, hyper-partisan, and lacked professional journalism standards—as professionally produced news in the two weeks leading up to the 2016 Presidential Election.<sup>93</sup>

However, the prevalence of “fake” or “junk news” on social media is globally varied. In the United Kingdom, France, and Germany, Twitter users shared significantly more professional news (49%, 46% and 40% respectively) than junk news (10%, 4% and 9% respectively).<sup>94</sup> In Sweden, one in every three URLs shared on Twitter was classified as junk news.<sup>95</sup> And in Brazil, only 1.2 percent of all Twitter content shared about the 2018 elections was junk news.<sup>96</sup> Here, however, a concurrent study of political communication on WhatsApp painted a different picture: On WhatsApp—where 90 percent of Brazilian Internet users are online—disinformation took the form of visual content including memes, images and links to YouTube videos.<sup>97</sup> Of the top-50 images shared in public WhatsApp groups, only four contained factual information.<sup>98</sup> Qualitative analysis showed that disinformation on WhatsApp tapped into and exacerbated political divisions and amplified anti-feminist and anti-LGBTQ sentiment.<sup>99</sup> Similar patterns of political communication were found in India, where WhatsApp has more than 200 million users; a third of all images shared by the Bharatiya Janata Party (BJP) and a quarter of the images shared by the Indian National Congress (INC) were classified as divisive and conspiratorial.<sup>100</sup>

Beyond differences around the world and across platforms, there are also differences among audiences who share and consume disinformation. Although younger users sometimes have trouble





judging the credibility of information online, the sharing and consumption of disinformation tends to be generational and partisan. In the United States, users who identified as conservative were more likely to share fake news stories than those who identified as liberal or moderate.<sup>101</sup> However, age also played an important role, with individuals over 65 sharing the most ‘fake news’ regardless of political affiliation or ideology.<sup>102</sup> This is reflected in other countries, like Nigeria, where fake news stories were shared more widely on WhatsApp by older users.<sup>103</sup>

## THE WEAPONIZATION OF DISINFORMATION

Although the current body of research is unclear about the extent to which disinformation spreads on social media, bad actors have increasingly weaponized disinformation for political or economic purposes. By exploiting the virality and anonymity afforded by social media platforms, coordinated networks of fake accounts have used disinformation to pollute the digital public sphere and push ideas at the fringe into the newsfeeds of moderate users. This is often done organically, by generating inauthentic engagement through automated and coordinated networks of fake accounts that like, share, retweet and forward messages in order to generate virality. In many democracies, we have seen the use of ‘computational propaganda’ to amplify political memes, videos, and disinformation in order to spread fear, anger, and outrage, which, in some cases,

has even led to political unrest and violence. As innovations in technology—including artificial intelligence and big data analytics—continue, the tools and techniques of disinformation will also evolve. We have already seen examples of ‘shallow fakes’—crudely manipulated audio and video—spread online, such as the viral video of United States Speaker of the House Nancy Pelosi that was edited to give the impression she was intoxicated. ‘Deepfake’ technologies—which use artificial intelligence to simulate facial expression, body movement and voice modulation—might further change the digital information landscape, especially when social media companies already struggle to stop harmful content from spreading on their platforms.

## IMPLICATIONS FOR ACTION

Growing concerns over the prevalence, effect, and weaponization of disinformation have put tremendous pressure on policymakers to “do something” about it. Since 2016, more than 40 governments have proposed or implemented new laws designed to address “fake news” on social media.<sup>104</sup> However, this pressure—and rapid response to a complex and multifaceted issue—presents a serious challenge because research has only scratched the surface of what we know about disinformation. There remains little empirically-grounded research that provides an understanding as to who shares disinformation, why they share it, and if they know it is fake.<sup>105</sup> There is even less research that explains the effects of exposure to disinformation and whether it drives polarization, partisanship, extremism, or violence. In a hybrid media

ecosystem, it is also hard to disentangle the independent effect of social media on democracy.<sup>106</sup>

When considering the role of government and social media companies in regulating disinformation—and ultimately speech—online, the global context in which the platforms operate must also be considered. In some countries, where institutions are weak, human rights are repressed, and democratic norms are regularly broken, social media platforms can be powerful normative entrepreneurs that champion democratic values. There is value in social media as a space—especially for those in closed or repressive regimes—to speak, organize, protest, and share information. But there are also significant harms in empowering companies to be the “arbiters of truth,” and so the benefits of reducing exposure to disinformation must be considerable to give that kind of power over speech to profit-maximizing companies.<sup>107</sup>

Ensuring citizens have access to high-quality and fact-checked information is important for electoral integrity. While bad actors have attempted to flood the digital ecosystem with disinformation, there have been a number of positive examples of civil society organizations working to combat this. The cooperative efforts in elections made by civil society organizations, traditional media, and the tech platforms to debunk fake news and halt its spread in countries such as Mexico and Indonesia have been impressive [see Boxes 5 and 6]. These efforts require enormous time and effort, but evidence suggests that national partnerships dedicated to defending electoral integrity can be effective in making themselves accessible to citizens, building trust in their judgment, and successfully discrediting egregious examples of disinformation. These efforts work best when traditional media values objectivity, and where EMBs adequately warn the public of the threat of fake news.

*The electoral integrity community should help build the capacity of national partnerships dedicated to defending the integrity of elections against weaponized disinformation and support better evaluation and sharing of practices.*

It is also important to consider implementing policy responses that limit the ability of bad actors to weaponize social media platforms and generate inauthentic reach. Because of the anonymous and pseudo-anonymous nature of social media, bad actors have been able to create networks of fake accounts that can enhance the speed and scale at which disinformation spreads. The Commission views anonymity as essential to democracy and electoral integrity, since it provides a valuable shield in countries that lack the human right protections fundamental for democratic participation. However, when combined with automation, anonymity can pose great risks to the digital public sphere. Instead of reducing the amount of anonymity afforded by the Internet and social media platforms, companies should provide greater transparency around accounts that use automation. Greater transparency efforts to label automated accounts will help the public evaluate the popularity of information sources, and whether certain content has been artificially inflated, while preserving the anonymity essential to democratic participation.

*Governments should compel platforms to label accounts that use automation. If an account is not correctly labelled as automated (e.g., a bot), platforms should face financial penalties mandated by public authorities.*

Weaponized disinformation affects the entire media ecosystem. Although a campaign might start on a single platform, the same images, memes, videos, or URLs might be shared across others. Platforms need to do more

to coordinate their responses to weaponized disinformation campaigns. This will require greater collaboration across companies to identify dubious content and the networks of accounts sharing them. Responsibility for combatting coordinated inauthentic behavior should not end at each of the platform's doorsteps. Better information sharing and cross-platform strategies for detecting and limiting the reach of disinformation and hate speech is necessary for enhancing election integrity.

*Social media platforms should create a coalition to address digital threats to democracy and election integrity, akin to what they have done collaboratively to address terrorism and child exploitation.*

*Members of the coalitions would meet regularly, and create cross-platform strategies for detecting and limiting the reach of weaponized disinformation and hate speech.*

Over the long-term democratic societies must work to inoculate themselves from weaponized disinformation. Citizens in the digital age will need to become savvy about information, propaganda, and sources on the Internet.<sup>108</sup> They will need to know how to recognize falsehoods and identify conspiracy theories. The ability to find out who is behind a particular web page and what they stand for is a critical part of digital education. But as important as it is to train citizens about where a piece of information may come from, it is equally important to train them on identifying their own potential sources of bias. We want voters to ask, “where did this piece of information come from and whose interest does it promote?” But as importantly, we want voters to ask, “Why am I predisposed to believe or dismiss this piece of information?”

*Public authorities should promote digital and media literacy programs in schools and in public interest programming for the general population.*

## BOX 5

### Mexico's Approach to Counter Disinformation

Mexico's federal National Electoral Institute (INE) was well prepared to counter fake news that might sway voter opinion, erode public trust in electoral processes, or increase polarization and fragmentation during the 2018 Mexican elections.<sup>109</sup> Over the course of the election cycle, INE launched a successful tripartite strategy to develop an 'alliance' with social networks and other media companies, support the civil society initiative *Verificado 2018* and institute a fact-checking computer system *Certeza 2018* to correct false information online.<sup>110</sup>

In a unique coordinated effort, INE secured formal cooperation agreements with Facebook, Twitter, and Google. This facilitated the first-ever live streaming of the Mexican presidential debates and INE election announcements, watched by millions of voters around the country.<sup>111</sup> INE also worked with Facebook to implement interactive 'buttons' that allowed users to access the official election authority website and spread get-out-the-vote messages, as well as engage users in debate topic selection.<sup>112</sup> Google similarly implemented a button to reroute users to INE web content, and hosted a Google Maps application to provide voters with polling location information.<sup>113</sup> Twitter and INE jointly established presidential debate discussions centered around trackable hashtags, created a forum for real-time journalistic commentary, and utilized an automated response function to provide real-time election results.<sup>114</sup>

INE also supported the fact-checking initiative *Verificado 2018* led by Animal Politico, AJ+ Español, Newsweek Español, and Pop-Up Newsroom that brought together more than 60 other civil society organizations. Verificado created a WhatsApp channel and a host of additional social media forums for users to inquire about the veracity of political claims and receive trustworthy and timely replies. The project operators successfully responded to 400 inquiries, produced 50 informative videos, and attracted millions of visitors to its official website.<sup>115</sup>

INE also established a technological mechanism called *Certeza 2018*. Using both human and computer-monitoring systems, field operators, and an assessment team, *Certeza* addressed misinformation in five stages: (1) monitoring for false information using keywords, (2) assessing flagged cases, (3) verifying appropriate courses of action, (4) gathering evidence, and (5) disseminating media notifications following case assessment.<sup>116</sup> *Certeza* benefited from cooperation agreements with media platforms as well as *Verificado*'s established outreach capacity to spread its verifications. After monitoring millions of social media posts, the system identified 217 cases of politically weaponized disinformation on election day while simultaneously addressing user requests for specific election-related information.<sup>117</sup>

In addition to initiating and supporting these avenues to combat disinformation, INE protected itself from cyberattacks through 2,000 informatics system audits and 24/7 real-time security monitoring.<sup>118</sup> Altogether, INE's approach during the 2018 elections offer a ground-breaking and effective model for other nations to emulate during future elections.

## BOX 6

### Indonesia's Approach to Counter Disinformation, 2018-2019

Indonesia experienced an increase in the prevalence of fake news during its 2018-19 elections.<sup>119</sup> The country's electoral bodies responded with a number of efforts in collaboration with civil society organizations, relevant government agencies and social media platforms. Focused on actively combatting disinformation and hate speech and safeguarding public trust in elections, these partnerships included awareness-building activities, enforcement of legal and regulatory provisions, and joint fact-checking initiatives.

The two election management bodies - General Election Commission (KPU) and Election Supervisory Agency (Bawaslu) – along with civil society organizations including Indonesian Anti-Slander Community (MAFINDO) and Centre for the Study of Religion and Democracy (PUSAD) Paramadina developed strategies to counter disinformation in the Indonesia context. Their work drew insights from best practices shared by The International Foundation for Electoral Systems (IFES), based on its experience in this area most recently in Kenya.<sup>120</sup> These included engaging diverse stakeholders, raising awareness, collecting data, developing counter-messaging, and adjudicating cases fairly.<sup>121</sup> The KPU and Bawaslu coordinated with the Ministry of Communication and Information Technology to establish a 'war room' that continuously monitored social media activity. The National Police were tasked with enforcing Indonesia's existing disinformation and discrimination laws, while the President's Office held dialogues with partner organizations.<sup>122</sup> Bawaslu also initiated a declaration to "Reject and Counter Vote Buying, Insults, Incitements, and Divisive Conflict in the 2018 Pilkada and 2019 General Elections," that secured signatures from 102 civil society groups along with relevant platforms including Google, Facebook, and Twitter.<sup>123</sup>



Nonprofit-led initiatives included MAFINDO's operation of a fact-checking 'Hoax Crisis Center' to debunk social media hoaxes during the 2018 local executive elections. In 2019, MAFINDO and 24 news organizations similarly operated CekFakta.com with funding from Google News to correct hoaxes and false candidate statements.<sup>124</sup> These partners also developed fact-checking tools for public use, including a phone application that gave provincial government offices the opportunity to correct false stories in real-time, as users reported them. Other projects included the Center for the Study of Religion and Democracy's counter-disinformation trainings for civil society organizations, MAFINDO- and Bawaslu-generated public service announcements, and IFES's workshops for election management bodies on fake news and identity-based hate speech.<sup>125</sup>

On a national scale, these coordinated efforts fact-checked 821 instances of political disinformation—more than half of which pertained to the elections at hand.<sup>126</sup> Sub-nationally, a joint 'Hoax Crisis Center' and other locally-targeted projects contributed to maintaining peace in conflict-prone areas such as West Kalimantan.<sup>127</sup> Indonesian public, private, and civil society efforts provide a model for collective action that could defend democratic processes challenged by viral digital falsehoods and violent rhetoric.



## V. POLITICAL ADVERTISING IN THE DIGITAL AGE

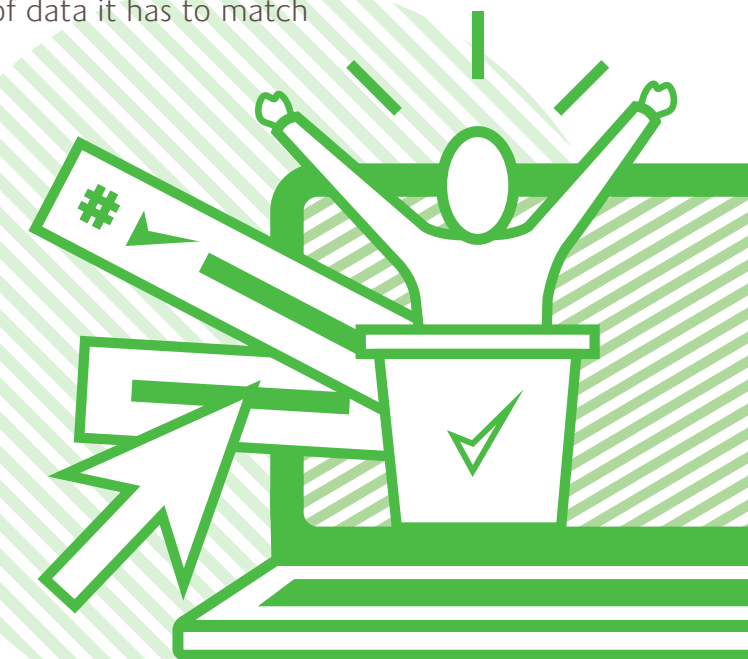
No area of Internet regulation has received more attention in the last three years than the reform of online political advertising. Given the connection of online advertisement to polarization, disinformation, and foreign election interference, increased public attention is quite appropriate. Moreover, through their acceptance and cultivation of political advertising, the major Internet platforms profit from, amplify, and microtarget messages that endanger healthy democratic deliberation and electoral integrity. In many respects, online advertising is a lens through which one can view each of the threats and benefits of the Internet for democracy.

Online political advertising, like the Internet and digital communication technologies in general, has significant benefits for democracy. Digital ads are often significantly less expensive than television or radio ads, allowing poorly financed candidates to get their message out, especially at the local level.<sup>128</sup> Moreover, despite all the criticism surrounding it, microtargeting allows campaigns and interest groups to efficiently deliver messages to the people they want to reach, as marketers have done for years with direct mail, phone calls, and later—email. Finally, the Internet has proven particularly beneficial for small-donor fundraising, as candidates and groups have used it to raise significant amounts of money from large numbers of donors, thereby “democratizing” political finance.<sup>129</sup>

But online political advertising has also proven fertile ground for all of the democracy-related pathologies attributed to digital communication technologies. Just as microtargeting can assist in mobilization and fundraising efforts, it enables campaigns to send

selective messages of demobilization or polarizing propaganda to persuadable voters, whose psychological and political affinities can be identified with increasing accuracy due to the data harvesting enabled by Internet platforms and other firms. Content testing and machine learning also allow advertisers to perfect the messages they wish to send to increase engagement with their content, often leading campaigns to create hundreds of variations on a single ad.<sup>130</sup> As a result, different (and even contradictory) messages can be sent to different people without others knowing about it. In this way, the dark side of microtargeting represents the artificial creation of (as opposed to self-selection into) online echo chambers: Internet platforms enable advertisers to send selective messages to narrow groups of voters based on demographics, interests, geography, and other political variables.<sup>131</sup> In fact, they will also allow targeting of “custom audiences” derived from lists of individual email addresses provided by campaigns, and “lookalike audiences”, which the platform derives from the immense amount of data it has to match the group provided by a campaign.

The lack of regulation of online political advertising has opened up the door for non-transparent activity from domestic and foreign actors, alike, to push polarizing and misleading messages during an election campaign. Russian activity in the 2016 U.S. Presidential election campaign exposed the vulnerabilities of the political advertising



system even to foreign-sponsored ads pushing polarizing messages intended both to affect electoral outcomes, as well as prey on the existing social division among the polity.<sup>132</sup> With over \$100,000 of Facebook ads, some of which was purchased in Russian rubles, the Russian Internet Research Agency delivered targeted messages that sometimes mentioned the presidential candidates, but more often sent polarizing propaganda surrounding divisive social issues such as immigration, gun rights, and race-based police brutality.<sup>133</sup> At times, different ads would propagandize on both sides of an issue, seeking to foment division and polarization, as opposed to persuading targets of the merits of one side of the argument. Tending toward sensationalism and masking the identity of the true spender, these ads then allowed the Russian Internet Research Agency (IRA) to develop audiences that could be targeted with organic messages and forwarded on to the user's broader network.

The controversy surrounding Russian ads in the 2016 election did not lead to legislative or regulatory action in the United States, but it has forced the major Internet platforms to implement new measures surrounding political ads. In particular, in the two years following the election, Facebook, Google, and Twitter adopted transparency rules for political ads, providing ad archives where users and watchdogs can view all the ads as well as information about the spender and some limited information about targeting and exposure. Furthermore, to prevent foreign and anonymous purchasing of ads, the platforms have adopted verification systems in which all political advertisers receive and must return postcards from their stated address to ensure that an identifiable domestic actor is the actual purchaser of an ad. While social media platforms have implemented countermeasures to prevent foreign entities from purchasing political

advertisements in the United States, these steps have not been effective at preventing foreign-sponsored advertisements from targeting other nations.<sup>134</sup>

Legislators around the world have offered proposals to further rein in online political advertising, but they often confront very difficult challenges. The main challenge concerns the definition of “political advertising”: how can the law specify what types of issues are sufficiently “political” such that they should be captured by political advertising regulations? This is all the more difficult when companies blend political messages with consumer ones, as when an ad for razors referenced the cause of fighting sexual harassment, an ad for sneakers referenced American football players fighting against race-based police brutality, or a beer ad sent a positive message about immigration by referencing the story of its founding. Moreover, if any purchased presentation of political topics should be treated as a political ad, then journalists and media organizations will be captured each time they pay (as they frequently do) to boost their reporting on social media to reach a larger audience. Because so much of the political advertising often deemed problematic (either because of foreign influence, polarizing messages, or disinformation) concerns issue advocacy, advertising reforms that do not deal with this dimension will not be dealing with “the problem.”

Each of the platforms has dealt with the issue advocacy problem differently. Twitter recently announced that it will no longer run ads that mention candidates, referenda, or pending legislation, but will allow “cause related advertising.” Google does not include issue advertisements in its political ad archives, but recently announced that it will no longer allow advertisers to target on the basis of



political affiliation and that it will prevent political advertisers from making fraudulent claims in ads, just as it does for consumer fraud. Facebook provides the most expansive archive of political ads, which includes not only candidate ads but also advertisements that mention a range of issues the company has identified by reference to an academic project called the “Comparative Agendas Project.” However, its “no censorship” policy stance on political advertisements, which exempts politicians from third-party fact-checking in order to preserve free speech<sup>135</sup> received significant criticism, even from internal employees who penned a letter to the New York Times arguing that “free speech and paid speech are not the same thing”<sup>136</sup>

In addition, the companies and regulators are considering other types of reforms to deal with concerns about manipulation by way of political advertising. Some platforms have considered limiting the ability to target with custom audiences or limiting the minimum audience size for a political ad. Others would require fact-checking of disputed ads to ensure claims made therein are not false. Still others would adapt existing television regulations to prevent, for example, political advertising in the period immediately before the election when it would be most difficult for campaigns adequately to confront disinformation advertisements in time to make a difference. In many of these areas, the platforms themselves have asked for legislative direction, having finally recognized that these decisions are simply too important and too central to democracy for profit-maximizing multinational corporations to be writing the rules.

## IMPLICATIONS FOR ACTION

Responsible digital political advertising requires action from political parties, social media platforms, and relevant public authorities. First, politicians, political parties, and candidates must take responsibility

for using digital advertising in ways that uphold electoral integrity. How they use new digital advertising technologies to campaign can set the tone for an entire election—providing citizens with a high degree of confidence, trust and information, or by undermining election integrity by engaging in deceptive campaign practices that emphasize rumor, conspiracy, disinformation, and manipulated media. There are already examples of politicians adhering to codes of conduct and best practices to support a healthy environment for political campaigning [Box 7]. Electoral integrity in the digital era requires higher ethical standards for how politicians, parties and candidates use social media and digital advertising.

*We endorse the call by the Transatlantic Commission on Election Integrity for political candidates, parties, and groups to sign pledges to reject deceptive digital campaign practices. Such practices include the use of stolen data or materials, the use of manipulated imagery such as shallow fakes, deep fakes, and deep nudes, the production, use, or spread of falsified or fabricated materials, and collusion with foreign governments and their agents who seek to manipulate the election.*



## BOX 7

## Nigeria's Abuja Accord on Election Conduct, 2015

In the lead up to the 2015 Nigerian general election, many Nigerians were concerned about a repeat of the 2011 election, which was marked by large-scale pre-election violence that left more than 800 dead and thousands internally displaced.<sup>137</sup>

Alarmed by the potential for an electoral catastrophe, Kofi Annan visited Nigeria and urged the presidential candidates to sign a peace agreement committing them to running clean, issue-based campaigns. In January 2015 in Abuja, with Annan as a witness, the candidates pledged to refrain from negative campaigning that could incite violence along religious, tribal, or ethnic lines.<sup>138</sup> The agreement was put in place by numerous domestic organizations, including both Christian and Muslim religious leaders, and assisted by several NGOs.<sup>139</sup> The Accord called for the creation of the National Peace Committee, which was charged with monitoring adherence to the agreement.<sup>140</sup>

The Abuja Accord was a clear success. In March 2015, just two days before the presidential election, the Accord was renewed by the two main presidential contenders, as a symbol of national unity, stability, and security ahead of election day.<sup>141</sup> The Accord was rapidly adopted as model for other elections, including in the lead up to Mozambique's October 2019 election.<sup>142</sup> Variants on the Accord were also signed by most state representatives in Nigeria.<sup>143</sup> During the 2019 general election cycle, Nigerian party leaders signed another agreement in the same spirit.<sup>144</sup> The legacy of the Abuja Accord demonstrates the potential efficacy of campaign pledges on non-violence and civil political conduct.

Second, all of the major social media companies have taken measures to address some of the new digital challenges to political advertising. However, we believe more can be done to support election integrity. In particular, all platforms could take further steps to improve the transparency of political advertising on their platform—including publishing more data about microtargeting and disclosing the identity of advertisement purchasers—as well as give users more control over the kinds of ads they are targeted with. Furthermore, platforms could help reinforce positive norms of political campaigning by requiring politicians, parties, and candidates purchasing advertisements to sign a pledge to avoid deceptive campaign practices, and holding their political advertisers to their commitments. All of these steps could help bolster election integrity.

*Platforms should provide greater transparency surrounding political ads.*

- Platforms should require users to choose to opt-out or opt-in to political advertising.
- Platforms should only allow candidates, parties and groups who have pledged to avoid deceptive campaign practices to purchase ads. Such pledges should then become working standards for platforms to decide on whether to accept any given ad.
- To avoid the cloaking of funders behind deceptive organizational labels, platforms should require public disclosure of the identity of human beings funding any political advertisement.

Finally, relevant public authorities must accept their responsibility to protect electoral integrity. Most importantly, laws and regulations for political campaigning and advertising must be adapted to the digital age. In particular, defining what constitutes a political advertisement should be a matter of law, and not left up to profit-maximizing companies to determine. Relevant public authorities should also specify the minimum audience size for microtargeting political advertisements, and consider legislating a cooling-off period for digital political advertisements, similar to some countries' current broadcasting laws. While the platforms have made significant strides at improving the transparency of political advertisements, not enough has been done to promote an environment that enhances electoral integrity. Governments should take steps to compel social media companies to make political advertising data public, including requiring platforms to publish information about the identity of the advertiser, targeting criteria, the amount spent and the actual ad creative.

*Countries must adapt their political advertising regulations to the online environment. Relevant public authorities should:*

- Define in law what is considered to be a political advertisement;
- Compel social media platforms to make public all information involved in the purchase of an ad, including the real identity of the advertiser, the amount spent, targeting criteria, and actual ad creative;
- Specify by law the minimum audience segment size for an ad; and
- Legislate a cooling-off period for digital political ads at least 48 hours before an election.



# VI. PROTECTING ELECTIONS FROM FOREIGN INTERFERENCE

Although the Internet and social media have many positive effects on democracy and elections—including the promotion of free speech, opportunities for political mobilization, and the democratization of information—their potential misuse are most apparent when seen through the lens of foreign interference operations. Over the past decade, state and non-state actors have used the Internet to pursue their political, economic, and military agendas, strategically combining traditional military operations with cyberattacks and online propaganda campaigns. By exploiting the open, anonymous, and borderless nature of digital technologies, social media have provided novel opportunities for bad actors to meddle transnationally. Electoral integrity depends on the sovereignty of elections, and outside actors should not be able to determine the outcome of an election.

To date, the most prolific example of foreign meddling has been Russia's interference in the 2016 US Presidential Election. By combining traditional hacking techniques with digitally coordinated campaigns across a range of old and new media, Russian operatives attempted to influence the American public and sway the outcome of the vote. Stolen data was strategically leaked to undermine Hilary Clinton's presidential nomination.<sup>145</sup> Russian-controlled media outlets—such as Russia Today (RT) and Sputnik—broadcast conspiracy tales and fueled anti-Clinton narratives about corruption on television and YouTube channels. Fake and automated accounts operated by the IRA—amplified these messages on Facebook, Twitter, and Instagram. Political advertisements—designed to polarize audiences around highly sensitive political debates—targeted communities of voters in order to foster division and distrust among the American

population. And posts, pages, and groups generated significant virality for free, reaching more than 126 million Americans in the lead-up to the 2016 vote.<sup>146</sup>

Although Russia's interference during the US 2016 election was one of the most prolific—especially given the sheer scale and sophistication of the campaign—the US has not been the only country targeted by Russian meddling. Traces of Russian-sponsored disinformation campaigns have been found in many parts of the world, including in Ukraine, the United Kingdom, and throughout Africa.<sup>147</sup> More troubling, the Russian model for election interference, which centers on a nation-state acting to destabilize an adversary, has been adapted by other countries looking to exert geopolitical power via social media. One only has to look to China's recent disinformation campaign against Hong Kong protestors, which paints political activists as violent and unpopular,<sup>148</sup> or Iranian influence operations, which promote anti-Saudi and anti-Israeli narratives, and urge support for US policies favorable to Iran.<sup>149</sup>

One of the challenges for combatting foreign influence operations is that it is increasingly difficult to distinguish between normal campaign activity by official arms of domestic political actors, and anti-democratic information operations by foreign governments, dubious commercial entities, or national groups. Populist politicians and parties have used the same tools and strategies as foreign agents to drive ultra-nationalist and anti-immigrant



rhetoric into mainstream political debates. Interest groups have used social media to target citizens in foreign countries with partisan messages. This was the case when pro-life groups in the United States targeted Irish citizens with political advertisements in the lead-up to Ireland's 2018 referendum on abortion.<sup>150</sup> Often, these efforts and activities overlap, making it increasingly difficult to draw some of the more traditional lines between foreign and domestic political activity, government and non-governmental organizations, and information operations and permissible campaign activity.

## AN EMERGING TRANSNATIONAL INDUSTRY OF ELECTION MANIPULATION

The interference playbook has also been monetized by private actors and strategic communication firms, who sell the various commodities of election interference to those interested in these services. The scandal involving Cambridge Analytica—which gained notoriety for misusing Facebook data to target voters with propaganda during the US 2016 election—is one of the most egregious examples highlighting professionalization of election manipulation.<sup>151</sup> By exploiting private social media data and its infrastructure, Cambridge Analytica—and its parent company Strategic Communications Laboratories (SLC Group)—crafted, targeted, and tailored messages of persuasion and demobilization to try and affect election outcomes in countries around the world, including Nigeria, Sri Lanka, Kenya, the Philippines, Trinidad and Tobago, and the United Kingdom.<sup>152</sup> Although the effectiveness

of Cambridge Analytica's data mining and 'psychographic profiling' techniques have been largely overstated;<sup>153</sup> this case highlights the more general phenomenon of the professionalization of election manipulation.

Today, there are a range of companies, consultancies, political communications agencies, and digital marketing firms that use the tools of the marketing industry to sway voters.<sup>154</sup> These companies exist around the world and have worked with politicians and governments to spread disinformation and propaganda, and target voters with messages aimed at suppressing their vote.<sup>155</sup> In some cases, these firms work internationally in order to conceal the true identity of the individual or organization behind the influence campaign,<sup>156</sup> as well as to take advantage of cheap digital labor in countries such as India<sup>157</sup> or the Philippines,<sup>158</sup> where a lucrative 'troll farm' industry has emerged.

## PROTECTING ELECTORAL INFRASTRUCTURE

All citizens have the right to have their vote counted equally and accurately, and the importance of citizen confidence in the vote tabulation is imperative for electoral integrity. But, much of electoral integrity is a black box maintained by voters' faith that they are legally registered, their vote was counted, and the results announced by officials are accurate. In addition to the soft



and subtle effects that propaganda and disinformation can have on elections, it is important to recognize the hard cybersecurity concerns around electronic electoral technologies (EETs).

Protecting the computer-based hardware and software for voter registration and vote casting is essential for election integrity. Beyond EETs themselves, there is a vast and decentralized ecosystem of technologies that support elections, including online voter registration systems, voter tabulation systems, and auditing systems. All of these technologies are susceptible to digital attacks as well as internal errors, both of which can erode voter confidence and impact the integrity of elections.

Hacking into electoral hardware and software can be done to alter results, manipulate voter lists, or simply undermine citizen confidence in their elections. Motives behind such hacking can be as simple as a foreign government or domestic actor seeking to ensure a favorable outcome. Hacking is only one possible tool of EET manipulation. For example, in Mozambique’s 2014 election, the government suppressed biometric voter registration in opposition-held areas “by sending inadequate equipment and undertrained teams.”<sup>159</sup> Beyond interference for a particular candidate, foreign governments may have a general interest in sowing domestic trouble, creating chaos, undermining legitimacy and creating distrust, and thus weakening the targeted country. When doubts are raised about the security of EETs it is all too easy for disgruntled political actors to blame losses at the polls to the manipulation of voting hardware and software and further undermine citizen trust in the process and outcome.

Moreover, EET security is not only a technical issue, and election officials themselves might also compromise EETs—wittingly or not.<sup>160</sup> Because humans are “intimately involved in all electoral operations, human vulnerabilities can certainly be exploited.”<sup>161</sup> The resiliency of the electoral infrastructure will depend on protecting technical systems as well as ensuring that human beings involved in all parts of the EET ecosystem are trained in cybersecurity best practices. It is worth reemphasizing the essential role of professional, capable, independent EMBs in protecting electoral integrity in the digital age.

## IMPLICATIONS FOR ACTION

When state actors interfere in foreign elections, there are a number of legal remedies set out under international law. Article 2 of the United Nations Charter guarantees the territorial and political integrity of states, and foreign influence campaigns by state actors offend the later guarantee. This can be bolstered by governments declaring that election hardware and software are critical infrastructure, and negotiating an international norm against cyber-attacks against critical infrastructure.

*Democratic governments should consider EETs critical infrastructure, and should support the norm endorsed by the G20 that “state[s] should not conduct or knowingly support Information and Communication Technology activity... that intentionally damages critical infrastructure.”*

A larger challenge, however, is how to distinguish and protect legitimate foreign assistance for the promotion of democracy and electoral integrity from illegitimate foreign interference in elections. All too often defenders of recent foreign interference in democratic elections claim that what the Russians did is no different from what Western democracies do when they support the building of political parties in Africa, or promote civil society organizations striving to hold authoritarian governments accountable. The best way to counter arguments based on false equivalence is for democracies to spell out what is and what is not legitimate transnational support for democracy.

*Democratic governments must come together to establish an international convention regarding the role of foreign governments and their agents in other countries' elections. In particular, they should develop international norms that distinguish legitimate cross-border assistance from illicit or unlawful interventions.*

Beyond foreign interference by state actors, there are also no rules or regulations on the emerging industry of election manipulation. Despite all of the scandals and unethical practices, companies like Cambridge Analytica were able to rebrand and continue their work. There is a need for government regulation, codes of conduct, and best practices for political consultants and strategic communication firms. Two voluntary professional associations, the International Association of Political Consultants and the American Association of Political Consultants, have both developed codes of conduct and could serve as a forum for dialogue between the election consultant industry, government, and the larger election integrity

community in creating and enforcing mandatory, transnational norms of ethical campaign consulting.

*The electoral integrity community should create norms and standards for transnational political campaign consultants, including public relations and strategic communication firms, and digital marketers. Government regulation should develop procedures for certifying these consultants and prevent any company from continuing to work on elections if it breaks the norms, rules and standards of campaign consulting.*

Electronic election technologies play central roles in almost every aspect of the election process, and because of this, robust cybersecurity is an essential element for ensuring elections with integrity. However, there are a number of challenges for securing EETs. One major barrier to safeguarding democratic elections from hacking is the lack of transparency and cooperation of the major vendors of EETs, who have been very slow to adapt to the digital threats to electoral integrity. Election officials might not have the technical knowledge to vouchsafe for the systems that they oversee. Sometimes EMBs are more worried about voter confidence waning because of warnings that their systems can be hacked, rather than the actual threat of outside interference.<sup>162</sup> In some countries, the procurement of EETs is corrupted by vendors willing to provide kickbacks to officials who purchase expensive equipment that is neither secure nor appropriate for the level of development of the country.<sup>163</sup>

*The electoral integrity community should help EMBs develop expertise in best cybersecurity practice.*

*Some EMBs may find themselves in need of short-term technical assistance against threats to electoral integrity by foreign interference in elections, hacking, and hate speech leading to election-related violence. In such cases, international technical assistance to help EMBs defend their election should be quickly available when requested. In order to ensure such assistance is delivered promptly, we recommend the development of standing electoral cybersecurity teams that could be deployed immediately on demand. Such teams could be located in existing international organizations, such as in the United Nations Electoral Assistance Division, or regional organizations, or in a new international institution. Such teams should have the capacity for rotational technical fellow positions for best digital government practice.*

Even though it is central to electoral integrity and public confidence in the results of elections, the EET industry is unregulated globally. Already several governments and private technology firms with close ties to their governments have become vendors of electoral equipment and election support services. There is no guarantee that such vendors will not become tools of foreign policy rather than independent purveyors of election services. Authoritarian governments are increasingly marketing dual-use technologies that can provide voter registration and identification, but with the potential for government surveillance of citizens and opponents.

The global election technology industry has an obligation to work with global standard-setting efforts to protect digital information, vote counting; and results transmission as well as the hardware and

software equipment from domestic or foreign intrusion, hacking, manipulation, or interference. This would not only benefit democracies around the world, but would also benefit election technology companies that show leadership in upholding election integrity.

*Vendors of election equipment and services should commit to a code of conduct to guarantee their products are secure, and their business practices protect the rights, privacy and data of citizens in their client countries, and adhere to honest, transparent practices in procurement. In turn, the international electoral integrity community should pledge that electoral assistance to countries will be conditional on vendors signing and adhering to the code. A multi-stakeholder initiative, involving the electoral integrity community, the Global Network of Domestic Election Monitors, and international partners should develop such a code of conduct.*

## VII. SUMMARY OF RECOMMENDATIONS

The defense of electoral integrity against the misuse and abuse of social media will depend on the choices and behavior of the major tech companies and platforms, and just as importantly, governments, politicians, traditional media, election management bodies, and citizens. In order to protect electoral integrity in the digital age, we will need to strengthen the capacities of the defenders of electoral integrity, and build shared norms around the acceptable use of digital technologies in elections. Technology platforms and public authorities must act to bolster electoral integrity.

### BUILDING CAPACITY

#### Recommendation 1.

Greater attention and resources must be dedicated to promoting election integrity. Public authorities, international organizations, philanthropic foundations, and civil society must invest in tech talent and digital capacity, media efforts, and election management bodies that protect and promote electoral integrity. All relevant stakeholders must cooperate, collaborate and rapidly share information related to threats to election integrity. These efforts should include:

- Building an election vulnerability index that gauges which elections require close monitoring for potential electoral interference, online coordinated inauthentic behavior, and mis-and-disinformation;

- Building the capacity of national partnerships dedicated to defending the integrity of elections against weaponized disinformation and support better evaluation and sharing of practices;
- Funding civil society organizations that counter hate speech, targeted harassment, and the incitement of violence, especially in the lead-up to elections; and
- Helping EMBs develop expertise in best cybersecurity practice;
- Helping democracies build civic technology programs through the teaching of coding, especially to women and minorities, and by incorporating technical talent into government teams.

#### Recommendation 2.

Some EMBs may find themselves in need of short-term technical assistance against threats to electoral integrity by foreign interference in elections, hacking, and hate speech leading to election-related violence. In such cases, international technical assistance to help EMBs defend their election should be quickly available when requested. In order to ensure such assistance is delivered promptly, we recommend the development of standing electoral cybersecurity teams that could be deployed immediately on demand. Such teams could be located in existing international organizations, such as in the United Nations Electoral Assistance Division, or regional organizations, or in a new international institution. Such teams should have the capacity for rotational technical fellow positions for best digital government practice.



## BUILDING NORMS

### Recommendation 3.

We endorse the call by the Transnational Commission on Election Integrity for political candidates, parties, and groups to sign pledges to reject deceptive digital campaign practices. Such practices include the use of stolen data or materials, the use of manipulated imagery such as shallow fakes, deep fakes, and deep nudes, the production, use, or spread of falsified or fabricated materials, and collusion with foreign governments and their agents who seek to manipulate the election.

### Recommendation 4.

Democratic governments must come together to establish an international convention regarding the role of foreign governments and their agents in other countries' elections. In particular, they should develop international norms that distinguish legitimate cross-border assistance from illicit or unlawful interventions.

### Recommendation 5.

Democratic governments should consider EETs critical infrastructure, and should support the norm endorsed by the G20 that “state[s] should not conduct or knowingly support Information and Communication Technology activity... that intentionally damages critical infrastructure.”

### Recommendation 6.

Vendors of election equipment and services should commit to a code of conduct to guarantee their products are secure, and their business practices protect the rights, privacy and data of citizens in their client

countries, and adhere to honest, transparent practices in procurement. In turn, the international electoral integrity community should pledge that electoral assistance to countries will be conditional on vendors signing and adhering to the code. A multi-stake holder initiative, involving at a minimum the electoral integrity community, the Global Network of Domestic Election Monitors and international partners should develop such a code of conduct.

### Recommendation 7.

The electoral integrity community should create norms and standards for transnational political campaign consultants, including public relations and strategic communication firms, and digital marketers. Government regulation should develop procedures for certifying these consultants and prevent any company from continuing to work on elections if it breaks the norms, rules and standards of campaign consulting.

## ACTION BY PUBLIC AUTHORITIES

### Recommendation 8.

Countries must adapt their political advertising regulations to the online environment. Relevant public authorities should:

- Define in law what is considered to be a political advertisement;
- Compel social media platforms to make public all information involved in the purchase of an ad, including the real identity of advertiser, amount spent, targeting criteria, and actual ad creative;

- Specify by law the minimum audience segment size for an ad; and
- Legislate a cooling-off period for digital political ads at least 48 hours before an election.

### **Recommendation 9.**

Public authorities must compel major Internet platforms to provide independent parties with meaningful data about the impact social media has on democracy. In particular, platforms must:

- Share secure, privacy-protected data with certified academic institutions to examine issues such as: auditing algorithms for bias towards extremism, understanding the effect of social media on political polarization and information consumption, and disentangling the relationship between online hate speech and offline violence.
- Update transparency reports to provide the public with data about the number of reports of hate speech and abuse online. This should include data about the instances of targeted abuse (against race, gender, sexual orientation, religion) and the frequency with which the abuse targets different communities; and
- Label accounts that use automation. If an account is not correctly labelled as automated (e.g., a bot), platforms should face financial penalties.

### **Recommendation 10.**

Public authorities should promote digital and media literacy programs in schools and in public interest programming for the general population.

## **ACTION BY PLATFORMS**

### **Recommendation 11.**

Platforms should provide greater transparency surrounding political ads.

- Platforms should require users to choose to opt out or opt in to political advertising.
- Platforms should only allow candidates, parties and groups who have pledged to avoid deceptive campaign practices to purchase ads. Such pledges should then become working standards for platforms to decide on whether to accept any given ad.
- To avoid the cloaking of funders behind deceptive organizational labels, platforms should require public disclosure of the identity of human beings funding any political advertisement.

### **Recommendation 12.**

Social media platforms need to develop early warning systems for election-related disinformation, foreign interference, hate crimes, threats to women, violence, and voter suppression.

- Platforms need to employ more experts who speak local languages and have cultural competency where they are operating;

- Because responding once communication achieves virality is too late, early warning systems must initiate human review for accounts and posts that pose a potential threat to elections. Content that achieves a certain level of virality should be subject to human moderation and review.

### **Recommendation 13.**

Social media platforms should create a coalition to address digital threats to democracy and election integrity, akin to what they have done collaboratively to address terrorism and child exploitation. Members of the coalitions would meet regularly, and create cross-platform strategies for detecting and limiting the reach of weaponized disinformation and hate speech.

## **ACKNOWLEDGEMENTS**

We would like to thank the individuals and institutions that contributed to the Commission's work over the course of our deliberations.

Laura Jakli, a doctoral student at the University of California, Berkeley led a small research team at Stanford University's Center on Democracy, Development, and Rule of Law that supported Stephen Stedman: Sylvie Ashford, Carolyn Chun, Yingjie Fan, and Whitney McIntosh. Laura Jakli and Samantha Bradshaw, a DPhil student at the Oxford Internet Institute contributed greatly to the drafting

and revising of the report. Young Lee and Eloise Duvillier provided budgetary and administrative support at Stanford.

Declan O'Brien, Sebastian Brack, Li Ling Low, and Stephanie Lewis of the Kofi Annan Foundation provided organizational support for the running of the commission. Bijan Farnoudi and Genna Ingold were responsible for communications and outreach strategy.

Several scholars and practitioners offered their insights to the commission: Tanja Bosch, Pablo Boczkowski, Michelle Brown, Chipo Dendere, Katherine Ellena, Jonas Kaiser, Daphne Keller, Admire Mare, Mora Matassi, Pat Merloe, Eugenia Mitchelstein, Vasu Mohan, Joyojeet Pal, and Erica Shein.

The Core Group of the Kofi Annan Foundation's Electoral Integrity Initiative served as a sounding board for ideas and recommendations.

The Kofi Annan Foundation would also like to acknowledge the balanced and diverse group of funding partners mobilised to support the work of the Commission; the Federal Chancellery of Austria, the Open Society Foundation, the United Nations Foundation, William H. Draper III, Facebook, and Twitter.

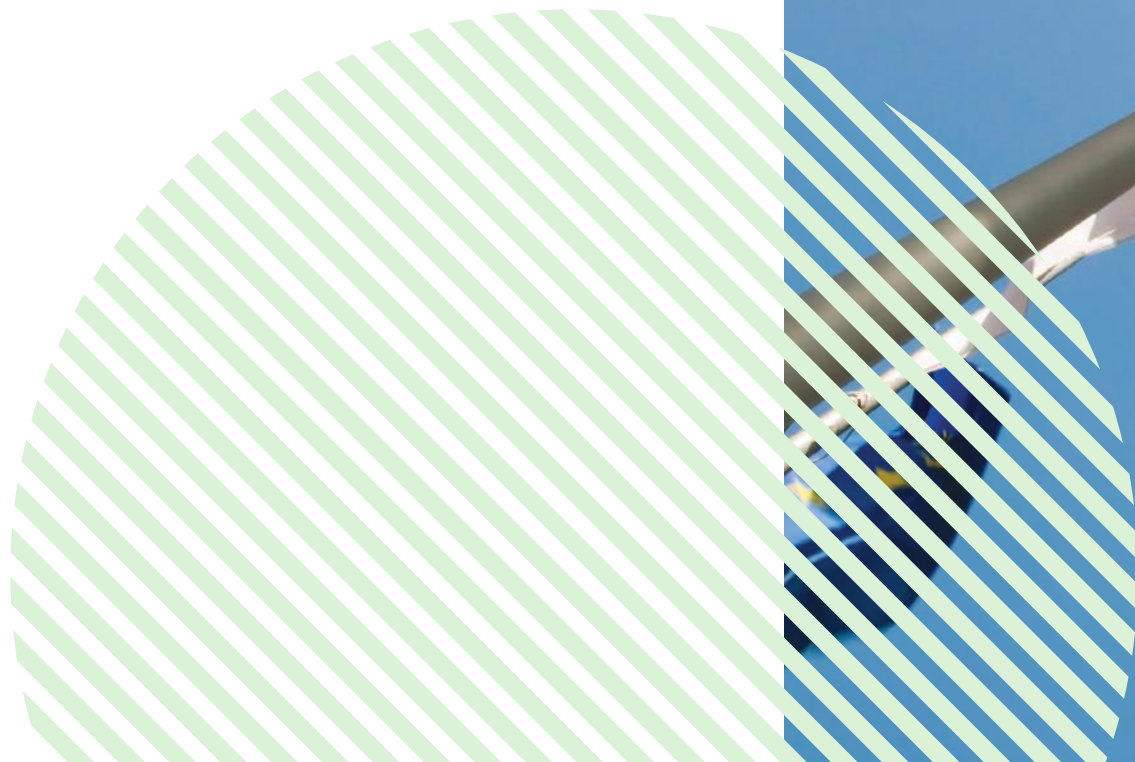
This mix of industry actors, governments, Foundations and private individuals encouraged a constructive engagement without compromising the independence of the Commission.



# ABOUT THE KOFI ANNAN FOUNDATION

The Kofi Annan Foundation is an independent, not-for profit organization that works to promote better global governance and strengthen the capacities of people and countries to achieve a fairer, more peaceful world.

The Kofi Annan Foundation mobilises political will to overcome threats to peace, development and human rights. In most cases the expertise and evidence needed to solve pressing problems such as poverty, armed conflict and poor governance already exist. What holds us back is lack of leadership or political will to identify and deliver solutions. The Foundation rallies those who are in a position to influence and bring leadership to the world's most pressing problems.



# END NOTES

**1** Nathaniel Persily, *The Internet's Challenge to Democracy: Framing the Problem and Assessing Reforms*, report for the Kofi Annan Commission on Elections and Democracy in the Digital Age, 2019. **2** Pablo J. Boczkowski, Eugenia Mitchelstein, and Mora Matassi, *Social Media and Democracy in Latin America*, report for the Kofi Annan Commission on Elections and Democracy in the Digital Age, 2019. **3** Boczkowski, Mitchelstein, and Matassi, *Social Media and Democracy in Latin America*. See also Tanja Bosch, Chipso Dendere, and Admire Mare, *The Effect of Social Media on Democracy and Elections in Africa*, report for the Kofi Annan Commission on Elections and Democracy in the Digital Age, 2019. **4** Bosch, Dendere, and Mare, *The Effect of Social Media on Democracy and Elections in Africa*. **5** Global Commission on Elections, Democracy and Security, *Deepening Democracy: A Strategy for Improving the Integrity of Elections Worldwide* (Geneva: Kofi Annan Foundation, 2012). **6** "Electoral Integrity Assessments," *International Foundation for Electoral Systems*, [www.ifes.org/issues/electoral-integrity-assessments](http://www.ifes.org/issues/electoral-integrity-assessments). **7** Global Commission on Elections, Democracy and Security, *Deepening Democracy*, 24-26. Mutual security as a precondition for democracy comes from Robert Dahl, *Polyarchy: Participation and Opposition* (New Haven: Yale University Press, 1971). **8** James S. Fishkin, *Democracy and Deliberation: New Directions for Democratic Reform* (New Haven: Yale University Press, 1991). **9** Christopher H. Achen and Larry M. Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government* (Princeton: Princeton University Press, 2016). **10** Murat Somer and Jennifer McCoy, "Déjà vu? Polarization and Endangered Democracies in the 21st Century," *American Behavioral Scientist* 62, no. 1 (2018): 3-15. **11** Anna Lührmann et al., "State of the World 2018: Democracy Facing Global Challenges," *Democratization* 26, no. 6 (2019): 895-915. **12** Nolan McCarty, Keith T Poole, and Howard Rosenthal, *Polarized America: the Dance of Ideology and Unequal Riches* (Cambridge: MIT Press, 2006); Jacob Jensen et al., "Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech," *Brookings Papers on Economic Activity* 43, no. 2 (2012): 1-81; Matthew Gentzkow, "Polarization in 2016," Whitepaper, *Toulouse Network for Information Technology*, 2016. **13** Yochai Benkler, Rob Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (New York: Oxford University Press, 2018). **14** Benkler, Faris, and Roberts, *Network Propaganda*. **15** Benkler, Faris, and Roberts, *Network Propaganda*. **16** Edda Humprecht, "Where 'Fake News' Flourishes: A Comparison Across Four Western Democracies," *Information, Communication, and Society* 22, no. 13 (2019): 1973-88. **17** Humprecht, "Where 'Fake News' Flourishes," 1973-88. **18** Lührmann et al., "Democracy Facing Global Challenges," 895-915. **19** Noam Gidron, James Adams, and Will Horne, "Toward a Comparative Research Agenda on Affective Polarization in Mass Publics," *APSA Comparative Politics Newsletter* 29 (2019): 30-36; Jennifer McCoy and Murat Somer, "Toward a Theory of Pernicious Polarization and How It Harms Democracies: Comparative Evidence and Possible Remedies," *The ANNALS of the American Academy of Political and Social Science* 681, no. 1 (2019): 234-271. **20** Nic Newman et al., "Reuters Institute Digital News Report 2019," *Reuters Institute for the Study of Journalism*, 2019; Ronald Inglehart et al., "World Values Survey: All Rounds-Country-Pooled Datafile 1981-2014," (Madrid: JD Systems Institute, 2014). **21** Pippa Norris, *Why Electoral Integrity Matters* (Cambridge: Cambridge University Press, 2014).

**22** McCarty et al., *Polarized America*, 78-81. **23** Herman Winkler, "The Effect of Income Inequality on Political Polarization: Evidence from European Regions, 2002-2014," *Economics and Politics* 31, no. 2 (2019): 137-162. **24** Benjamin Reilly, *Democracy in Divided Societies: Electoral Engineering for Conflict Management* (Cambridge: Cambridge University Press, 2001); Lee Drutman, "The Case for Proportional Voting," *National Affairs* 34 (2017): 50-63. **25** Ryan D. Enos, *The Space between Us: Social Geography and Politics* (Cambridge: Cambridge University Press, 2017); Dante J. Scala and Kenneth M. Johnson, "Political Polarization along the Rural-Urban Continuum? The Geography of the Presidential Vote, 2000-2016," *The ANNALS of the American Academy of Political and Social Science* 672, no. 1 (2017): 162-184. **26** Jonas Pontusson and David Rueda, "Inequality as a Source of Political Polarization: A Comparative Analysis of Twelve OECD Countries," in *Democracy, Inequality, and Representation*, ed. Pablo Beramendi and Christopher J. Anderson (New York: Russell Sage Foundation, 2008), 312-353. **27** Drutman, "The Case for Proportional Voting." **28** McCoy and Somer, "Toward a Theory of Pernicious Polarization," 234-271. **29** Cass R. Sunstein, *Republic.Com* (Princeton: Princeton University Press, 2001). **30** Carole Cadwalladr, "Google Is Not 'Just' a Platform. It Frames, Shapes and Distorts How We See the World," *The Guardian*, December 11, 2016, sec. Opinion, <https://www.theguardian.com/commentisfree/2016/dec/11/google-frames-shapes-and-distorts-how-we-see-world>. **31** Kelly Weill, "How YouTube Built a Radicalization Machine for the Far-Right," *The Daily Beast*, December 17, 2018, <https://www.thedailybeast.com/how-youtube-pulled-these-men-down-a-vortex-of-far-right-hate>. See also Kevin Roose, "The Making of a YouTube Radical," *The New York Times*, June 8, 2019, <https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html>. **32** Richard Fletcher and Rasmus Kleis Nielsen, "Are People Incidentally Exposed to News on Social Media? A Comparative Analysis," *New Media & Society* 20, no. 7 (2018): 2450-2468. **33** Matthew Barnidge, "Exposure to Political Disagreement in Social Media versus Face-to-Face and Anonymous Online Settings," *Political Communications* 34, no. 2 (2017): 302-321. **34** Pablo Barbera, "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S." *Working paper*, 2014. **35** Christopher A. Bail et al., "Exposure to Opposing Views on Social Media can Increase Political Polarization," *Proceedings of the National Academy of Sciences* 115, no. 37 (2018): 9216-9221. **36** Manoel Horta Ribeiro et al., "Auditing Radicalization Pathways on YouTube," 2019, <http://arxiv.org/abs/1908.08313>. **37** Kevin Munger and Joseph Phillips, "A Supply and Demand Framework for YouTube Politics," *Working Paper*, 2019. **38** "Public Statement from the Co-Chairs and European Advisory Committee of Social Science One," December 11, 2019, <https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one>. **39** Jenni Marsh and Tara Mulholland, "How the Christchurch Terrorist Attack Was Made for Social Media," *CNN*, March 16, 2019, <https://www.cnn.com/2019/03/15/tech/christchurch-internet-radicalization-intl/index.html>; Kevin Roose, "On Gab, an Extremist-Friendly Site, Pittsburgh Shooting Suspect Aired His Hatred in Full," *The New York Times*, October 28, 2018, <https://www.nytimes.com/2018/10/28/us/gab-robert-bowers-pittsburgh-synagogue-shootings.html>. **40** Timothy McLaughlin, "How WhatsApp Fuels Fake News and Violence in India," *Wired*, December 12, 2018, <https://www.wired.com/story/how-whatsapp-fuels-fake-news-and-violence-in-india/>.



41 Amalini De Sayrah, "Facebook Helped Foment Anti-Muslim Violence in Sri Lanka. What Now?," *The Guardian*, May 5, 2018, <https://www.theguardian.com/commentisfree/2018/may/05/facebook-anti-muslim-violence-sri-lanka>. 42 Paul Mozur, "A Genocide Incited on Facebook, With Posts From Myanmar's Military," *The New York Times*, October 18, 2018, sec. Technology, <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>. 43 "Amnesty Reveals Alarming Impact of Online Abuse against Women," *Amnesty International*, November 20, 2017, <https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/>. 44 Adriane Van Der Wilk, "Cyber Violence and Hate Speech Online against Women," *European Parliament Committee on Women's Rights and Gender Equality*, September 2018, [http://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL\\_STU\(2018\)604979\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2018/604979/IPOL_STU(2018)604979_EN.pdf) 45 David Kaye, "UN Experts Urge States and Companies to Address Online Gender-Based Abuse but Warn against Censorship," *OHCHR*, March 8, 2017, [www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21317&LangID=E](http://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21317&LangID=E). 46 "Toxic Twitter - The Silencing Effect." *Amnesty International*, March 2018, [www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-5/](http://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-5/). 47 Lauren Etter, "Rodrigo Duterte Turned Facebook Into a Weapon, With a Little Help from Facebook," *Bloomberg News*, December 7, 2017, <https://www.bloomberg.com/news/features/2017-12-07/how-rodrido-duterte-turned-facebook-into-a-weapon-with-a-little-help-from-facebook> 48 Carly Nyst and Nick Monaco, "State-Sponsored Trolling: How Governments are Deploying Disinformation as Part of Broader Digital Harassment Campaigns," *Institute for the Future*, 2018, [http://www.iftf.org/fileadmin/user\\_upload/images/DigIntel/IFTF\\_State\\_sponsored\\_trolling\\_report.pdf](http://www.iftf.org/fileadmin/user_upload/images/DigIntel/IFTF_State_sponsored_trolling_report.pdf) 49 Nina Jankowicz, "How Disinformation Became a New Threat to Women," *World Policy*, December 20, 2017, <https://worldpolicy.org/2017/12/20/how-disinformation-became-a-new-threat-to-women/> 50 Amnesty International, "Toxic Twitter—The Silencing Effect." 51 The most recent Supreme Court ruling against regulating hate speech was under a unanimous June 2017 decision on *Matal v. Tam*, 52 U.S. (2017). [https://www.supremecourt.gov/opinions/16pdf/15-1293\\_1o13.pdf](https://www.supremecourt.gov/opinions/16pdf/15-1293_1o13.pdf) 52 Daphne Keller and Paddy Leerssen, "Facts and Where to Find Them: Empirical Research on Internet Platforms and Online Speech," in *Social Media and Democracy: The State of the Field*, ed. Nathaniel Persily and Joshua Tucker (New York: Cambridge University Press, forthcoming). 53 Steve Stecklow, "Why Facebook Is Losing the War on Hate Speech in Myanmar," *Reuters*, August 15, 2018, <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>. 54 Tarelton Gillespie, *Custodians of the Internet: Platforms, Content Moderation and the Hidden Decisions That Shape Social Media* (New Haven: Yale University Press, 2018). 55 "The EU Code of Conduct," *European Commission*, February 4, 2019, [https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online\\_en](https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/countering-illegal-hate-speech-online_en) 56 Keller and Leerssen, "Facts and Where to Find Them." 57 "Germany to Enforce Hate Speech Law," *BBC News*, sec. Technology, January 1, 2018, <https://www.bbc.com/news/technology-42510868>. 58 Daphne Keller, "Internet Platforms: Observations on Speech, Danger, and Money," Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 1807 (2018): 2. 59 David Kaye, "Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression," (June 2017): 4. <https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf> 60 Samantha Bradshaw, Lisa-Maria Neudert, and Philip Howard, "Government Responses to the Malicious Use of Social Media," *NATO Strategic Communications Centre of Excellence*, November 2018, [https://](https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Nato-Report.pdf)

[comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Nato-Report.pdf](https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Nato-Report.pdf) 61 David Kaye and Fionnuala Ni Aolain, "Mandates of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression; and the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms While Countering Terrorism," 2019, <https://freedex.org/wp-content/blogs.dir/2015/files/2019/04/OL-AUS-04.04.19-5.2019-2.pdf>. 62 Jacob Mchangama and Joelle Fiss, "Germany's Online Crackdowns Inspire the World's Dictators," *Foreign Policy*, November 6, 2019, [foreignpolicy.com/2019/11/06/germany-online-crackdowns-inspired-the-worlds-dictators-russia-venezuela-india/](http://foreignpolicy.com/2019/11/06/germany-online-crackdowns-inspired-the-worlds-dictators-russia-venezuela-india/). 63 Access Now, "Joint Letter on Internet Shutdown in Uganda," *CIPESA*, February 24, 2016, <https://cipesa.org/2016/02/joint-letter-on-internet-shutdown-in-uganda/>. 64 Jan Rydzak, "Of Blackouts and Bandhs: The Strategy and Structure of Disconnected Protest in India," February 7, 2019, <https://ssrn.com/abstract=3330413>. 65 Louise Matsakis and Issie Lapowsky, "Don't Praise the Sri Lankan Government for Blocking Facebook," *Wired*, April 23, 2019, <https://www.wired.com/story/sri-lanka-bombings-social-media-shutdown/>. 66 Daniel Arnaudo et al., "Political and Economic Implications of Authoritarian Control of the Internet," in Jonathan Butts and Sujeet Sheno, *Critical Infrastructure Protection VII*. (Berlin, Heidelberg: Springer Berlin Heidelberg, 2013), 3–19. 67 Abdi Latif Dahir, "Internet Shutdowns Are Costing African Governments More than We Thought," *Quartz Africa*, September 28, 2017, <https://qz.com/africa/1089749/internet-shutdowns-are-increasingly-taking-a-toll-on-african-economies/>. 68 Access Now, "Joint Letter on Internet Shutdown in Uganda." 69 Roderick Fanou et al., "OONI - Open Observatory of Network Interference," April 30, 2019, <https://ooni.org/>. 70 Ribeiro et al., "Auditing Radicalization Pathways on YouTube." 71 Christopher Ross, "Reshma Saujani's Ambitious Plan for Technology," *The Wall Street Journal*, November 6, 2014, [www.wsj.com/articles/reshma-saujani-ambitious-plan-for-technology-1415237831](http://www.wsj.com/articles/reshma-saujani-ambitious-plan-for-technology-1415237831). 72 "Girls Who Code: Annual Report 2018," [girlswhocode.com/2018report/](http://girlswhocode.com/2018report/). 73 Ross, "Reshma Saujani's Ambitious Plan for Technology." 74 Kimiko de Freytas-Tamura, "Kenyan Election Official Is Killed on Eve of Vote," *The New York Times*, July 31, 2017, sec. World, <https://www.nytimes.com/2017/07/31/world/africa/chris-musando-kenya-election-official-dead.html>. 75 Nanjira Sambuli, "The Importance of Monitoring Online Hate Speech," *Deutsche Welle*, October 3, 2016, <https://www.dw.com/en/the-importance-of-monitoring-online-hate-speech/a-19104789>. 76 Ibid. 77 Susan Benesch, "Countering Dangerous Speech to Prevent Mass Violence during Kenya's 2013 Elections," *Berkman Center for Internet and Society*, February 9, 2014, <https://www.ushmm.org/m/pdfs/20140212-benesch-kenya.pdf>. 78 Kagonya Awori, "Umati Final Report," *iHub Research*, June 2013, <https://preventviolentextremism.info/sites/default/files/Umati%20Final%20Report.pdf>. 79 Ibid. 80 Luciano Floridi, "Fake News and a 400-Year-Old Problem: We Need to Resolve the 'Post-Truth' Crisis," *The Guardian*, November 29, 2016, [www.theguardian.com/technology/2016/nov/29/fake-news-echo-chamber-ethics-infosphere-internet-digital](http://www.theguardian.com/technology/2016/nov/29/fake-news-echo-chamber-ethics-infosphere-internet-digital). 81 Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* 31, no. 2 (2017): 211-36. 82 Stephen Coleman, "The Digital Difference: Media Technology and the Theory of Communication Effects," *Journal of Communication* 67, no. 6 (2017): E7-E8. 83 Ralph Schroeder, "Does Google Shape What we Know?" *Prometheus* 32, no. 2 (2014): 145-160. 84 Tarelton Gillespie, "The Relevance of Algorithms," in: Tarelton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, eds. *Media Technologies: Essays on Communication, Materiality, and Society* (Cambridge: MIT Press, 2014), 167-194. 85 Tim Wu, *The Attention Merchants: The Epic Scramble to Get Inside Our Heads* (New

York: Knopf, 2016); James Williams, *Stand Out of Our Light: Freedom and Resistance in the Attention Economy* (Cambridge: Cambridge University Press, 2018). **86** Benkler, Faris, and Roberts, *Network Propaganda*; Marchal et al., “Junk News during the EU Parliamentary Elections,” *Oxford Internet Institute*, Data Memo 2019.3. **87** Craig Silverman, “This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook,” *BuzzFeed News*, November 16, 2016, [www.buzzfeednews.com/article/craig-silverman/viral-fake-election-news-outperformed-real-news-on-facebook](http://www.buzzfeednews.com/article/craig-silverman/viral-fake-election-news-outperformed-real-news-on-facebook). **88** Joshua Tucker et al., *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature* (William and Flora Hewlett Foundation, 2018). **89** Claire Wardle and Hossein Derakhshan, “Information Disorder: Toward and Interdisciplinary Framework for Research and Policy Making,” Council of Europe Report DGI(2017)09. *Council of Europe*. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> **90** Gordon Pennycook and David Rand, “Assessing the Effect of ‘Disputed’ Warnings and Source Salience on Perceptions of Fake News Accuracy,” *Working Paper*, 2017, 10.2139/ssrn.3035384. **91** Allcott and Gentzkow, “Social Media and Fake News in the 2016 Election” **92** Andrew Guess, Brendan Nyhan, and Jason Reifler, “Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 US Presidential Campaign,” *European Research Council 9* (2018). **93** Samantha Bradshaw et al., “Sourcing and Automation of Political News and Information over Social Media in the United States, 2016-2018,” *Political Communication* (2019): 1-21. **94** Lisa-Maria Neudert, Philip Howard, and Bence Kollanyi, “Sourcing and Automation of Political News and Information During Three European Elections,” *Social Media+ Society* 5, no. 3 (2019): 2056305119863147. **95** Freja Hedman et al., “News and Political Information Consumption in Sweden: Mapping the 2018 Swedish General Election on Twitter,” *Oxford Internet Institute*, Data Memo 2018.3. **96** Caio Machado et al., “News and Political Information Consumption in Brazil: Mapping the First Round of the 2018 Brazilian Presidential Election on Twitter,” *Oxford Internet Institute*, Data Memo 2018.4. **97** Caio Machado, “WhatsApp’s Influence in the Brazilian Election and How It Helped Jair Bolsonaro Win.” *Council on Foreign Relations*, November 13, 2018, [www.cfr.org/blog/whatsapps-influence-brazilian-election-and-how-it-helped-jair-bolsonaro-win](http://www.cfr.org/blog/whatsapps-influence-brazilian-election-and-how-it-helped-jair-bolsonaro-win). **98** Chico Mares and Clara Becker, “Só 4 Das 50 Imagens Mais Compartilhadas Por 347 Grupos de WhatsApp São Verdadeiras,” <https://piaui.folha.uol.com.br/lupa/wp-content/uploads/2018/10/Relat%C3%B3rio-WhatsApp-1- turno-Lupa-2F-USP-2F-UFGM.pdf> **99** Rafael Evangelista and Fernanda Bruno, “WhatsApp and Political Instability in Brazil: Targeted Messages and Political Radicalization,” *Data-Driven Elections: Workshop Papers, Surveillance Studies Centre*, 2019. [https://www.sscqueens.org/sites/sscqueens.org/files/evangelista\\_bruno-2019-04.pdf](https://www.sscqueens.org/sites/sscqueens.org/files/evangelista_bruno-2019-04.pdf) **100** Vidya Narayanan et al., “Polarization, Partisanship and Junk News Consumption over Social Media in the US,” *Oxford Internet Institute*, Data Memo 2018.1. **101** Andrew Guess, Jonathan Nagler, and Joshua Tucker, “Less than you Think: Prevalence and Predictors of Fake News Dissemination on Facebook,” *Science Advances* 5, no. 1 (2019): eaau4586. **102** Guess, Nagler, and Tucker, “Less than You Think.” **103** Jamie Hitchen et al., “WhatsApp and Nigeria’s 2019 Elections: Mobilizing the People, Protecting the Vote.” Abuja: Centre for Democracy and Development, <https://www.cddwestafrica.org/whatsapp-nigeria-2019-elections/> **104** Bradshaw, Neudert, and Howard, “Government Responses to the Malicious Use of Social Media.” **105** Nathaniel Persily and Joshua Tucker, “Introduction,” in *Social Media and Democracy: The State of the Field*, ed. Nathaniel Persily and Joshua Tucker (New York: Cambridge University Press, forthcoming). **106** Andrew Chadwick, *The Hybrid Media System: Politics and Power* (Oxford: Oxford University Press, 2013). **107** Persily and

Tucker, *Social Media and Democracy*. **108** Sam Wineburg, *Why Learn History (When It’s Already on Your Phone)* (Chicago: University of Chicago Press, 2018), 139-159. **109** “New Challenges for Democracy: Elections in Times of Disinformation,” *Instituto Nacional Electoral*, June 2019, 4-5. **110** Ibid., 7. **111** Ibid., 7. **112** Ibid., 8. **113** Ibid., 9. **114** Ibid., 12. **115** Ibid., 9-11. **116** Ibid., 12-13. **117** Ibid., 14-15. **118** Ibid., 16. **119** Vasu Mohan, Maya Jacobs, Tana Azuaje, Kyle Lemargie, and Carla Chianese, “The Race Against SARA and Hoaxes in Indonesian Elections,” Working paper, *International Foundation for Electoral Systems*, August 20, 2019, 3. **120** Vasu Mohan and Catherine Barnes, “Countering Hate Speech in Elections: Strategies for Electoral Management Bodies,” White Paper, *International Foundation for Electoral Systems*, January 2018, ivi. [https://www.ifes.org/sites/default/files/2017\\_ifes\\_countering\\_hate\\_speech\\_white\\_paper\\_final.pdf](https://www.ifes.org/sites/default/files/2017_ifes_countering_hate_speech_white_paper_final.pdf) **121** Ibid., 20-34. **122** Mohan et al., “The Race Against SARA1 and Hoaxes2 in Indonesian Elections,” 4. **123** Ibid., 6. **124** Fanny Potkin and Agustinus Beo de Costa, “Fact-checkers vs. hoax peddlers: a fake news battle ahead of Indonesia’s election,” *Reuters*, August 10, 2019, <https://www.reuters.com/article/us-indonesia-election-fake-news-insight/fact-checkers-vs-hoax-peddlers-a-fake-news-battle-ahead-of-indonesias-election-idUSKCN1RM2ZE> **125** “Mohan et al., “The Race Against SARA1 and Hoaxes2 in Indonesian Elections,” 6-7. **126** Ibid., 8. **127** “Update on Local Election Results in West Kalimantan And Papua,” Report No. 50, *Institute for Policy Analysis of Conflict*, August 2018, 6. [http://file.understandingconflict.org/file/2018/08/IPAC\\_Report\\_50\\_Update.pdf](http://file.understandingconflict.org/file/2018/08/IPAC_Report_50_Update.pdf) **128** Erika Franklin Fowler et al., “Political Advertising Online and Offline,” *Working Paper*, [https://web.stanford.edu/~gjmartin/papers/Ads\\_Online\\_and\\_Offline\\_Working.pdf](https://web.stanford.edu/~gjmartin/papers/Ads_Online_and_Offline_Working.pdf) **129** Jessica Baldwin-Philippi, “The Myths of Data-Driven Campaigning,” *Political Communication* 34, no. 4 (2017): 627-633. **130** Jeff Chester and Kathryn Montgomery, “Follow the Tech: Emerging Digital Practices in the 2020 Election,” *Data-Driven Elections: Workshop Papers, Surveillance Studies Centre*, 2019. [https://www.sscqueens.org/sites/sscqueens.org/files/chester\\_montgomery-2019-04.pdf](https://www.sscqueens.org/sites/sscqueens.org/files/chester_montgomery-2019-04.pdf) **131** Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. (London: Penguin, 2016). **132** Robert S. Mueller, “Report on the Investigation into Russian Interference in the 2016 Presidential Election (Vol. I of II)” (Washington, DC: US Department of Justice, 2019). **133** Philip N. Howard, Bhath Ganesh, and Dimitra Liotsiu, “The IRA, Social Media and Political Polarization in the United States, 2012-2018,” *Computational Propaganda Research Project*, University of Oxford, 2019, <https://assets.documentcloud.org/documents/5632779/IRA-Report-2018.pdf>; Renee DiResta et al., “The Tactics & Tropes of the Internet Research Agency.” *New Knowledge Whitepaper*, 2018. **134** Shannon McGregor, Bridget Barrett, and Daniel Kreiss, “Barely Legal: Digital Politics and Foreign Propaganda,” *Working Paper*, 2019. [https://digitalpoliticsethics.weebly.com/uploads/5/0/9/9/50994643/mcgregorbarrettkreissapsa19\\_submit.pdf](https://digitalpoliticsethics.weebly.com/uploads/5/0/9/9/50994643/mcgregorbarrettkreissapsa19_submit.pdf) **135** Nick Clegg, “Facebook, Elections and Political Speech.” Facebook Newsroom, September 24, 2019, [about.fb.com/news/2019/09/elections-and-political-speech/](http://about.fb.com/news/2019/09/elections-and-political-speech/). **136** “Read the Letter Facebook Employees Sent to Mark Zuckerberg About Political Ads,” *The New York Times*, October 28, 2019, [www.nytimes.com/2019/10/28/technology/facebook-mark-zuckerberg-letter.html](http://www.nytimes.com/2019/10/28/technology/facebook-mark-zuckerberg-letter.html). **137** “Nigeria National Elections: 2015 Nigeria Election Observation Report,” *International Republication Institute*, March 28, 2015, 5. <http://www.iri.org/2015%20Nigeria%20Election%20Observation%20Report/1/assets/basic-html/index.html#III> **138** “Abuja Accord On the Prevention of Violence and Acceptance of Elections Results by Presidential Candidates and Chairpersons of Political Parties contesting the 2015 General Elections,” *Institute for Democracy and Electoral Assistance*, 2015, <https://www.idea.int/sites/default/files/codesofconduct/Abuja%20Accord%20January%202015.pdf>



139 International Republication Institute, “Nigeria National Elections,” 5. 140 Institute for Democracy and Electoral Assistance, “Abuja Accord,” 2. 141 International Republication Institute, “Nigeria National Elections,” 6. 142 The Associated Press, “Mozambique Peace Accord Is Signed, Paving Way for Elections,” *The New York Times*, August 6, 2019, [www.nytimes.com/2019/08/06/world/africa/mozambique-peace-accord-signed-paves-way-for-elections.html](http://www.nytimes.com/2019/08/06/world/africa/mozambique-peace-accord-signed-paves-way-for-elections.html). 143 International Republication Institute, “Nigeria National Elections,” 5. 144 Fidelis Mbah, “Nigeria Elections: Presidential Candidates Sign ‘Peace Deal,’” *Al Jazeera*, February 13, 2019, [www.aljazeera.com/news/2019/02/nigeria-elections-presidential-candidates-sign-peace-deal-190213154706618.html](http://www.aljazeera.com/news/2019/02/nigeria-elections-presidential-candidates-sign-peace-deal-190213154706618.html). 145 Michael McFaul and Bronte Kass, “Understanding Putin’s Intentions and Actions in the 2016 US Presidential Election,” In: Michael McFaul, *Securing American Elections: Prescriptions for Enhancing the Integrity and Independence of the 2020 U.S. Presidential Election and Beyond*, 2019, 1-16. 146 “HPSCI Minority Open Hearing Exhibits,” U.S House of Representatives Permanent Select Committee on Intelligence (HPSCI), Minority Report, Nov. 1, 2017. 147 Shelby Grossman, Daniel Bush, and Renee DiResta, “Evidence of Russia-Linked Influence Operations in Africa,” *Stanford Internet Observatory*, October 29, 2019, [https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/29oct2019\\_sio\\_-\\_russia\\_linked\\_influence\\_operations\\_in\\_africa.final\\_.pdf](https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/29oct2019_sio_-_russia_linked_influence_operations_in_africa.final_.pdf) 148 Steven Lee Myers and Paul Mozur, “China Is Waging a Disinformation War Against Hong Kong Protesters,” *The New York Times*, August 14, 2019, <https://www.nytimes.com/2019/08/13/world/asia/hong-kong-protests-china.html> 149 FireEye Intelligence, “Suspected Iranian Influence Operation Leverages Network of Inauthentic News Sites & Social Media Targeting Audiences in U.S., UK, Latin America, Middle East,” *FireEye Threat Research*, August 21, 2018, <https://www.fireeye.com/blog/threat-research/2018/08/suspected-iranian-influence-operation.html> 150 Adam Satariano, “Ireland’s Abortion Referendum Becomes a Test for Facebook and Google,” *The New York Times*, May 25, 2018, <https://www.nytimes.com/2018/05/25/technology/ireland-abortion-vote-facebook-google.html> 151 Hannes Grassegger and Mikael Krogerus, “The Data That Turned the World Upside Down,” *Vice News*, January 28, 2017, [www.vice.com/en\\_us/article/mg9vvn/how-our-likes-helped-trump-win](http://www.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win). 152 Colin Bennett and Smith Oduro-Marfo, “Privacy, Voter Surveillance and Democratic Engagement: Challenges for Data Protection Authorities,” UK Office of the Information Commissioner paper presented at the 2019 International Conference of Data Protection and Privacy Commissioners (ICDPPC), [https://privacyconference2019.info/wp-content/uploads/2019/11/Privacy-and-International-Democratic-Engagement\\_finalv2.pdf](https://privacyconference2019.info/wp-content/uploads/2019/11/Privacy-and-International-Democratic-Engagement_finalv2.pdf) 153 Baldwin-Philippi, “The Myths of Data-Driven Campaigning”; Daniel Kreiss, “Micro-Targeting, the Quantified Persuasion,” *Internet Policy Review* 6, no. 4 (2017): 1-14. 154 Tactical Tech Data and Politics Team, “Personal Data: Political Persuasion. Inside the Influence Industry,” *Tactical Tech*, <https://ourdataourselves.tacticaltech.org/posts/inside-the-influence-industry> 155 Samantha Bradshaw and Philip Howard, “The Global Disinformation Disorder: 2019 Global Inventory of Social Media Manipulation,” Oxford Internet Institute Working Paper 2019.3. Oxford, UK: Project on Computational Propaganda. 156 Jane Lytvynenko and Logan McDonald, “Hundreds Of Propaganda Accounts Targeting Iran And Qatar Have Been Removed From Facebook,” *BuzzFeed News*, October 7, 2019, [www.buzzfeednews.com/article/janelytvynenko/uae-propaganda](http://www.buzzfeednews.com/article/janelytvynenko/uae-propaganda). 157 Michael Riley, Lauren Etter, and Bibhudatta Pradhan, “A Global Guide to State-Sponsored Trolling,” *Bloomberg News*, July 19, 2018, <https://www.bloomberg.com/features/2018-government-sponsored-cyber-militia-cookbook/> 158 Jonathan Corpus Ong and Jason Vincent A. Cabanes, “Architects of Networked Disinformation,” *Newton Tech4Dev Network*, 2018, <http://>

[newtontechfordev.com/wp-content/uploads/2018/02/Architects-of-Networked-Disinformation-Executive-Summary-Final.pdf](http://newtontechfordev.com/wp-content/uploads/2018/02/Architects-of-Networked-Disinformation-Executive-Summary-Final.pdf) 159 Nic Cheeseman, Gabrielle Lynch, and Justin Willis, “Digital Dilemmas: The Unintended Consequences of Election Technology,” *Democratization* 25 no. 8 (2018): 1397-1418. 160 Herbert Lin, Alex Stamos, Nathaniel Persily and Andrew Grotto, “Increasing the Security of the US Election Infrastructure,” In: Michael McFaul, *Securing American Elections: Prescriptions for Enhancing the Integrity and Independence of the 2020 U.S. Presidential Election and Beyond*, 2019, 17-26. 161 Ibid. 162 Peter Wolf, “Cybersecurity and Elections: An International IDEA Round-Table,” *Institute for Democracy and Electoral Assistance*, July 8 2017, [www.idea.int/news-media/news/cybersecurity-and-elections-international-idea-round-table-summary](http://www.idea.int/news-media/news/cybersecurity-and-elections-international-idea-round-table-summary). 163 Cheeseman, Lynch, and Willis, “Digital Dilemmas,” 1397-1418.

In 2018, Kofi Annan, Chair of the Kofi Annan Foundation, convened the Annan Commission on Elections and Democracy in the Digital Age. The Commission consists of senior figures from the public and private sectors and civil society, with high-level experience in government, the technology sector, academia and the media who Mr. Annan tasked with exploring the opportunities and challenges to electoral integrity created by technological innovations.

This report captures the main findings of the research and consultations conducted by the Commission, along with their recommendations to ensure that new technologies, social media platforms and communications tools can be used to realize, not inhibit, citizen's aspirations for democratic governance.

Published by

---



**Kofi Annan**  
FOUNDATION

Towards a fairer, more peaceful world