Social Media, Disinformation and Electoral Integrity: *IFES Working Paper*

August 2019

INTERNATIONAL FOUNDATION FOR ELECTORAL SYSTEMS
IFES
30 Years

Social Media, Disinformation and Electoral Integrity

IFES Working Paper

August 2019

Dr. Beata Martin-Rozumiłowicz and Rasťo Kužel

The authors wish to thanks Ms. Lisa Reppell for her review and input on this paper.

Social Media, Disinformation and Electoral Integrity: IFES Working Paper

International Foundation for Electoral Systems
2011 Crystal Drive, 10th Floor
Arlington, VA 22202
Email: editor@ifes.org
Fax: 202.350.6701

Table of Contents

Executive Summary

Since the 2016 United States (U.S.) presidential election, the issue of social media and disinformation has gained increasing attention as a fundamental threat to the integrity of elections worldwide. Whether by domestic actors, such as candidates and campaigns, or through foreign influence campaigns, the ability of voters to make informed choices based on fair and balanced information has been significantly skewed. This working paper attempts to examine the challenges that this issue poses to electoral integrity and what responses election management bodies (EMBs) and international nongovernmental organizations (INGOs) such as the International Foundation for Electoral Systems (IFES) can take to attempt to mitigate the negative consequences. The solutions presented in this paper aim to assist key stakeholders to meet this emergent and mutable threat.

The paper starts with an introduction of the subject and presents some background on its development, including the important distinction between system and information aspects of technological impact upon electoral processes. This is developed within the literature review that presents an overview of this newly emerging field. While many aspects of traditional media and elections literature are pertinent to the topic of disinformation, there are many fundamental differences that make thinking in this field unique and potentially require an altered set of analytical tools to help theoreticians and practitioners more accurately navigate this space.

The paper then incorporates IFES' key concepts and working definitions on disinformation, which draw on definitions and standards around which academic and practitioner communities have begun to coalesce. The paper also undertakes a review of key international standards that pertain, both from the traditional media angle, as well as new emerging good practices that bring important distinctions to our understanding of this field. This review also analyzes the arguments for and against regulation versus self-regulation, an important aspect of the discourse in this emerging field. This is particularly relevant to EMBs, who have already begun to question whether oversight and sanction is appropriate for social media, or whether this is better left to the traditional legal tools of hate speech legislation and defamation or libel laws that govern free speech in a traditional media setting, although this dynamic is already starting to change.

The paper then goes on to present a comprehensive overview of the constitutional provisions and legal frameworks pertaining to this field. It examines the responses of different states globally as they seek to grapple with this problem and analyzes the responses in legal terms regarding the efficacy of such solutions to the fundamental issue of how it may impact overall election integrity, and thus public trust and confidence.

It also examines some of the issues that EMBs face in this area at a more fundamental level. The paper details the variety of ways that EMBs have responded in trying to mitigate the risks emerging from disinformation in elections and highlights a number of emerging good practices that might be considered by others. Nevertheless, given the newness of this field, the level of information on this front is still developing and expected to widen rapidly in the future.

The final section of the paper examines strategies for countering disinformation in social media in terms of possible future innovations and new areas of potential research. It examines what EMBs, as well as INGOs, might do, both internal-facing and external-facing, to present more truthful information as a possible antidote to false information being spread. It also covers both shorter- and longer-term solutions on how the goal of countering disinformation might be achieved.

The paper is supplemented by two annexes. One lists the extant international obligations and good practices in this area. The other presents an extensive bibliography of recent work in this specific subfield that addresses a wide array of related considerations to this question.

In all, this working paper presents a way forward in terms of thinking on this complicated issue. With a better definition of problems and terms and a deeper understanding of the fundamental challenge that social media disinformation poses to electoral integrity, both those administering elections and those providing technical assistance can move forward on a more solid basis and in a more informed manner.

Introduction and Background

In recent years, the prevalence of disinformation, particularly through social media, and its fundamental impact on electoral integrity, has become an issue of global concern. While IFES has developed and implemented programs in traditional areas of media's impact on elections, this new aspect presents important challenges. Is social media fundamentally different from traditional media and in what way? Does the state have a role in regulating this sector, and if so, what level of regulation is appropriate to ensure respect for fundamental freedoms? How can citizens be better informed to ensure that they are making choices based on accurate information and facts, rather than disinformation, hate speech and other types of divisive influence campaigns as there is a conceptual link between these two different but often interrelated phenomena of disinformation and hate speech?[1] This working paper seeks to tease out these questions and to present possible avenues of programming and assistance for both INGOs, such as IFES, and EMBs to engage in.

Mis/disinformation, particularly through social media, has become an increasing problem to electoral integrity and citizens' trust in their democratic institutions. The Oxford Internet Institute's recent report shows that "formally organized social media campaigns" were taking place in 48 countries – a quarter of countries recognized by the United Nations (UN) – up steeply from 28 in 2017.[2] As well, there is the issue of the sheer volume of information, which is becoming an increasing problem as voters struggle to make sense of all the disparate sources, regardless of levels of expertise.

---

1 For further detail about the distinctions between the two areas, please see the two IFES publications: Vasu Mohan (2018), *Countering Hate Speech in Elections: Strategies for Electoral Management Bodies*, http://www.ifes.org/sites/default/files/2017_ifes_countering_hate_speech_white_paper_final.pdf and Lisa Reppell and Erica Shein (2019), *Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions*, https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf

2 See Bradshaw and Howard (2018), pg.3

Thus, it is increasingly clear that all citizens – and especially EMBs – should be exposed to concepts that not too long ago were familiar only to strategic communication experts: primary vs. secondary source attribution, source validation, definition of reliable sources, source systems, asymmetric threats, information operations and related terminology that are defined and explained further below in the paper.

Another area of concern is deepfake videos, which use artificial intelligence (AI) to produce falsified videos that are almost undetectable by swapping out someone's face and voice with that of an imposter. The danger is that deepfake videos could be used during elections to undermine the reputation of candidates, especially women, who are targeted by deepfake attacks depicting them in pornographic or other sexually degrading contexts.[3] Some experts predict that such videos could be shared and spread even faster than fake news as we know it today. Public confidence in electoral integrity is a cornerstone of democratic resilience and, as such, these emerging threats need to be addressed appropriately.

At the same time, the question emerges regarding the niche that INGOs such as IFES can fill, given its global reputation, relationship with EMBs worldwide and engagement with civil society organizations (CSOs) to promote better electoral processes and, thus, to increase voters' trust. There is a clear and increasing need for EMBs to consolidate their transparency and, thus, credibility so that citizens can come to regard them as primary and reliable sources of truthful information.

This often cuts against the grain of traditional perceptions, both internal and external, of the appropriate EMB role vis-à-vis the public. EMBs have traditionally tended to step aside when it comes to taking responsibility for addressing disinformation threats. INGOs could, therefore, play an increasing role in supporting and encouraging EMBs to implement appropriate measures and expand cooperation with other entities – state and international institutions, social media companies, etc. – in order to counteract and possibly prevent such disinformation operations.

The area of inquiry is vast and cross-cutting, involving a variety of electoral fields, from legislative framework issues to ones of technology, voter education, gender and inclusion. Solutions to the problems posed must have an interdisciplinary focus. As well, there may not be a panacea for this multifaceted issue, but many experts currently espouse a manifold solution composed of various correlated measures that have a common thread: digital literacy and greater civic education of voters, journalists, EMBs, youth, etc. It is increasingly clear from academic research that critical thinking skills need to be developed from early childhood, but also that it's never too late to develop such skills. This kind of education is especially needed in countries that hail from former authoritarian regimes, where freedom of thinking was limited, and populations may need additional assistance to develop these skills.

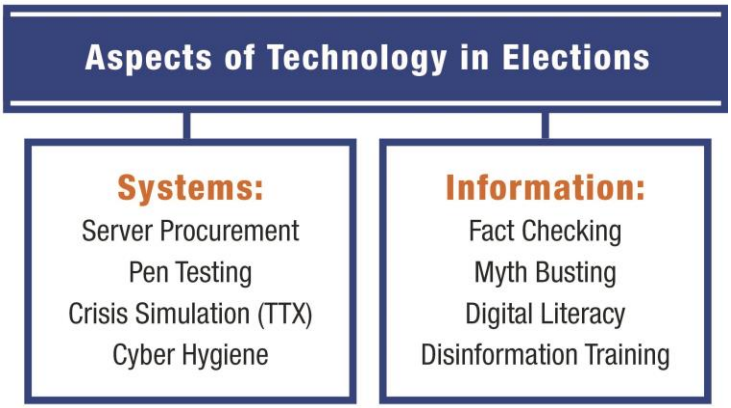
In addition, the state of research on actual and potential effects is still in its early stages, so solutions may evolve on the basis of empirical evidence on the effects. At the same time, it is important to separate the two conceptual questions methodically in terms of the technology (systems) and

[3] See forthcoming Consortium for Elections and Political Process Strengthening paper, *Violence Against Women in Elections Online: A Social Media Analysis Tool*

application (information). The two expert audiences are bifurcated, and although there has been cross-over, they are quite distinct.

Experts clearly opine that a clear differentiation should be made between cyber threats and cyber-enabled (technology) information operations. The main relevance lies in the proper allocation of resources for tackling each unique set of problems. This is in terms of human expertise, material resources, the strategies to be implemented, and the specific technologies that need to be developed and deployed. The mistake of putting both under the umbrella of cyber threats has been repeatedly made with obvious consequences. To ameliorate this situation, this paper draws the distinction clearly and this is presented graphically in Diagram 1, below.

Diagram 1 – Bifurcation Between Systems vs. Information Aspects to Technology in Elections[4]



IFES has been engaged in extensive development and programing on the systems side of this continuum in developing a holistic methodology for cybersecurity in elections assessments,[5] in developing playbooks for authorities, in engaging in procurement, where applicable, and in conducting extensive cyber hygiene training. This working paper, however, aims to present the latest innovative thinking, solutions and tools being applied both by IFES and other key organizations in this field on the informational side of the spectrum on disinformation and related issues and the recent developments in this field.

---

4 Pen Testing is penetration testing, a wide-spread technique in cybersecurity to determine vulnerabilities, TTX are table-top crisis simulation exercises that enable EMBs to practice their approaches in real-team scenarios, while disinformation hygiene training is an emerging area in which stakeholders are given basic tools to understand this space and attempt to mitigate threats and vulnerabilities.

5 See IFES' Holistic Exposure and Adaptation Testing (HEAT) Process for Election Management Bodies, https://www.ifes.org/publications/cybersecurity-elections

Literature Review[6]

A great deal of work has been done to understand disinformation, how it works, and what can be done about it. This review represents a snapshot of that work as it relates to elections and is divided into the following sections: sources and vectors of disinformation; how and why disinformation spreads; international law and standards; and programmatic responses to disinformation. Links and full references for all sources cited are available in the annexes at the end of the document. To save space, disinformation will be used throughout to mean dis-, mis- and malinformation unless otherwise noted (see Definitions and Concepts section below).

Sources and Vectors of Disinformation

One particularly useful and widely cited framework for understanding disinformation, and from which IFES has drawn to conceptualize the problem, is First Draft's "Information Disorder." In First Draft's framework, information disorder – and thus disinformation – "contaminates public discourse," working as a pollutant in the information ecosystem.[7] Information disorder is broken into three elements: agent, message and interpreter. Diagnosing a particular case of information disorder requires answering questions about each element: the agent responsible for the disinformation, including their motivation and intended audience; the message, including its duration and level of accuracy; and the interpreter, including what action, if any, was taken upon encountering the disinformation.[8] This section will primarily consider the agent; subsequent sections also cover the message and interpreter.

Agents include independent trolls ("human-controlled accounts performing bot-like activities" or harassing others online),[9] paid trolls, conspiracy theorists, disinformation websites, partisan media, politicians, foreign governments, influential bloggers, activists or government officials, and ordinary users gathered *en masse*.[10] Their intents and targets vary; for example, domestic partisan agents may use disinformation to win campaigns through smear tactics, hostile foreign state or nonstate authoritarian agents may intend to structurally undermine democracy by increasing intolerance and polarization, and disaffected anarchic agents may intend to dismantle state institutions and social order. While many are primarily concerned at the moment with automated/inauthentic means of amplifications, there is a growing need to also start addressing the role played by parties, politicians and hyperpartisan media in creating, disseminating and "endorsing" disinformation and divisive contents.

---

6 The authors would also like to thank IFES' Center for Applied Learning and Research for their assistance in conducting this literature review and in the development of definitions; in particular, Lisa Reppell, Russell Bloom and Erica Shein.

7 Wardle and Derakhshan, *Information Disorder: Toward an interdisciplinary framework for research and policymaking,* 10.

8 Ibid, 25-28.

9 Ibid.

10 Tucker et al., "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature," 22-28.

Among the new tools available to disinformation campaigns are deepfakes (digitally altered or fabricated videos or audio that are increasingly lifelike),[11] microtargeting (using consumer data, especially on social media, to send different information to different groups, often particular minority groups), manufactured amplification (artificially boosting the reach of information by manipulating search engine results, promoting hashtags or links on social media or other means), and bots ("social media accounts that are operated entirely by computer programs and are designed to generate posts and/or engage with content on a particular platform").[12]

Among the tactics agents use to spread disinformation online are selective censorship, hacking and sharing, manipulating search rankings, and using bots and trolls to share information directly. [13] Selective censorship is removing certain content from a platform to exclude it from the conversation. Hacking and sharing information is a common tactic and can have serious implications for EMBs, as has been seen in recent cases in Germany and the Philippines.[14]

The manipulation of search rankings can take place on search engines or on social media platforms. Groups of bots, called botnets, and coordinated groups of trolls promoting specific narratives, called troll farms, are deployed to generate online conversation and get stories trending,[15] using hashtags and sharing each other's content through "mutual admiration societies."[16] In the 2016 U.S. elections, 400,000 bots produced approximately 3.8 million tweets in the last month of the elections alone, for which Russian operations were largely responsible.[17] Many of the accounts and tweets in the 2016 elections have been traced back to the Internet Research Agency (IRA), a Russian troll farm.[18]

How and Why Disinformation Spreads

The role of the interpreter is increasingly central to attempts to understand the spread of disinformation. There are still important debates on some of the details of the relevant psychology and

---

11 See, e.g., computer-generated videos from the University of Washington: Langston, "Lip-syncing Obama: New tools turn audio clips into a realistic video." Although deepfakes are currently getting the lion's share of attention, other emergent technologies are also expected to be increasing the spread of computational propaganda in the future (artificial intelligence (AI), automated voice systems, interactive memes, machine learning, interactive memes, virtual reality (VR), augmented reality). This means that digital disinformation is likely to be more effective and harder to combat as future disinformation campaigns will most likely harness our senses through realistic sounding AI voices and VR, which will allow for multisensory propaganda experiences. In addition, we should expect to see increasing laundering narratives through peer-to-peer messaging services, which are harder to monitor due to encryption.

12 Wardle, "Information Disorder, Part 1: The Essential Glossary."

13 Tucker et al., 30.

14 See, e.g., the recent hack in Germany or the impeachment of Philippines Commission on Elections Chair Andres Bautista. Wemer, "Angela Merkel's Data Leaked"; and Cupin, "From resignation to impeachment: Chairman Bautista's longest day."

15 See, e.g., Metzel, "Alleged Troll Factory Mastermind Prigozhin, 12 Other Russians Charged with U.S. Election Meddling."

16 Tucker et al., 30.

17 Tucker et al., 32.

18 Popken and Cobiella, "Russian troll describes work in the infamous misinformation factory."

how disinformation spreads – such as the disputed "backfire effect"[19] – but there is agreement about the general picture. Humans did not evolve to process information and respond rationally; instead, they use mental shortcuts to simplify decision-making. These heuristics combine with another evolved feature, the need to belong to a group, to create vulnerabilities to the kind of systematic manipulation disinformation campaigns use. Our heuristics and biases dispose us to believe information when it is presented in certain ways and wanting to send the proper in-group signals lead people to spread information even if they don't necessarily trust it. Media consumption itself is individual, but "invariably, when we use social media to share news, we become performers," activating our biases and group commitments.[20] We also tend to remain committed to our prior beliefs long after seeing evidence to the contrary.[21] Confirmation bias, conformity and motivated cognition dispose us to accept some messages more credulously than others.[22] With algorithms distorting the content people see, the perception of our group can become skewed toward extremes.

People are generally more attracted to news with false information than with true information. In a 2018 study on the spread of news stories on Twitter, the MIT Media Lab found that "falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information."[23] The truth took "about six times as long as falsehood to reach 1,500 people," and, controlling for relevant variables, falsehoods were "70% more likely to be retweeted than the truth."[24]

Even when controlling for bots, the difference between the spread of true and false tweets persisted. One explanation offered by the authors of the study was novelty (a difference in the message) – the content of false tweets was significantly more novel than that of true tweets.[25] In fact, the authors found that disinformation "dominates according to both metrics. It consistently reaches a larger audience, and it tunnels much deeper into social networks than real news does".[26] IFES' research has also revealed gendered dimensions to this finding: false or salacious information about women spreads further, faster and more intensely than disinformation about men. Advertising incentives and the online market structure offer another explanation for why false information spreads quickly online. Many social media and other online platforms have no membership fee and get their revenue from selling ad space and

---

19 See, e.g., Nyhan, Brendan and Jason Reifler, "When Corrections Fail: The persistence of political misperceptions"; and Wood, Thomas and Ethan Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence."

20 Wardle and Derakhshan*, Information Disorder: Toward an interdisciplinary framework for research and policy making,* 43.

21 See, e.g., the following blog entries by Dan Kahan: "Weekend update: You'd have to be science illiterate to think 'belief in evolution' measures scientific literacy"; and "What sorts of inferences can/can't be drawn from the 'Republican shift' (now that we have enough information to answer the question)?"

22 See, e.g., *Wardle and Derakhshan, Information Disorder: Toward an interdisciplinary framework for research and policy making,* 44; and Palmertz, "Theoretical foundations of influence operations: a review of relevant psychological research," 12-16.

23 Vosoughi et al., "The spread of true and false news online."

24 Ibid, 1148-49.

25 Ibid, 1150.

26 "The Grim Conclusions of the Largest-Ever Study of Fake News" (2018), *The Atlantic*, https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/

user data. The market structure incentivizes clickbait, and the selling of consumer data and the way ads are shown on these platforms allow for microtargeting, protected by the anonymity provided by ad policies and a lack of regulation.[27]

As an illustrative example of how platforms are adjusting, in 2018 Facebook introduced new rules for political and issue ads. According to the rules, any advertiser who wants to run political or issue ads must be verified on the platform and include "paid for" information with the advertising. In addition, Facebook has also created a searchable archive (Facebook Ad Library) that logs political and issue ad content from advertisers going back seven years. These new rules should make it more difficult for those attempting to spread disinformation on Facebook.

The precision with which individuals and groups can be targeted allows bad actors to exploit ideological and cultural divisions and raises additional concerns over hate speech and discrimination, such as the recent news that Russian attempts to influence the 2016 U.S. election particularly targeted African-American voters.[28] Russian actors have also used social media assaults to discredit, degrade and threaten female politicians and public figures in pro-Western regimes or contested spheres of influence. For example, IFES social media research in Ukraine identified up to 35 percent of overall online violence against in elections content around the general elections was posted by IP addresses in Russia. Ideological and cultural divisions are also commonly exploited, such as the IRA's targeting of gun lobby supporters.[29]

27 Bradshaw and Howard, "Why Does Junk Spread so Quickly Across Social Media?" 11-13.

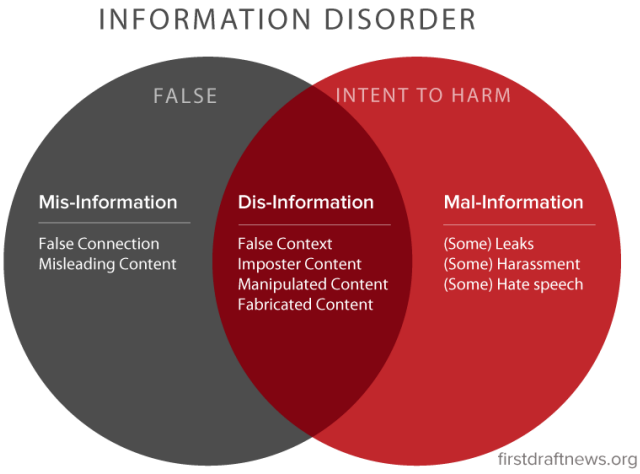
28 Shane and Frenkel, "Russian 2016 Influence Operation Targeted African-Americans on Social Media."

29 Kevin Poulsen, "Russian Troll Farm Internet Research Agency has New Meta-Trolling Propaganda Campaign."

Definitions and Concepts[30]

Foundational Definitions

- **Disinformation** is false or misleading information that is created or disseminated with the intent to cause harm or to benefit the perpetrator. The intent to cause harm may be directed toward individuals, groups, institutions, or processes.
- **Malinformation** is accurate information that is shared with the intent to cause harm or to benefit the perpetrator, often by moving private information into the public sphere.
- **Misinformation** is false or misleading information that is shared without the intent to cause harm or realization that it is incorrect. In some cases, actors may unknowingly perpetuate the spread of disinformation by sharing content they believe to be accurate among their networks.



- **Inauthentic Actors** are individuals or organizations working to mislead others about who they are or what they are doing.
- **Hate Speech,** despite or perhaps because of its ubiquity, has no consensus definition. For the purposes of this paper, *hate speech* encompasses polarizing expression or speech that promotes intolerance, hatred and incitement to violence by explicit or indirect reference to race, national or ethnic origin, religion, gender, sexual orientation, age or disability or other immutable groupings.[31]

Coordinated Actors

- **Influence campaigns** are "actions taken by governments or organized non-state actors to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome."[32] Influence campaigns increasingly deploy an array of disinformation tactics with the goal of manipulating public opinion and undermining the integrity of the information environment.
- **Internet trolls** are human users on internet platforms who intentionally harass, provoke or intimidate others, often to distract and sow discord. Trolls can act as individuals, and in this capacity

30 These definitions are replicated from: Reppell and Shein, *Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions*, International Foundation for Electoral Systems, 2019.

31 There is no consensus definition for this concept used by scholars, practitioners, and in legal instruments. As noted in Sellars, Andrew, "Defining Hate Speech," various definitions are crafted to support their application, whether in academic research or as the basis for a domestic law or international treaty. Some definitions and sources are available in Mohan, Vasu and Barnes, *Countering Hate Speech in Elections: Strategies for Electoral Management Bodies*, an IFES white paper.

32 Wardle (2017), p. 16.

share many characteristics with individual perpetrators of hate speech. However, trolls can also engage in coordinated inauthentic behavior.

Inauthentic Content

- **Junk news** includes the publication of propaganda and ideologically extreme, hyperpartisan or conspiratorial political news and information under the guise of providing credible information. The term includes news publications that present verifiably false content or commentary as factual news.[33]
- **Deepfakes** are digitally altered images and videos that use AI to combine real source material with manufactured content to create hyper-realistic portrayals of individuals saying or doing things that did not occur.

Manufactured Amplification

- **Computational propaganda** is "the use of algorithms, automation, and human curation to purposefully distribute misleading information over social media networks. Computational propaganda involves learning from and mimicking real people so as to manipulate public opinion across a diverse range of platforms and device networks."[34]
- **Bots** are simple computer codes that can simulate human beings and make posts online. Botnets are the coordinated efforts of multiple bots.
- **Content or click farms** are commercial enterprises that employ individuals to generate fraudulent profiles, posts and likes to promote specific narratives online. Coordinated efforts to direct the attention of internet trolls toward targets or in promotion of certain messages can use the same model as content farms and are referred to as troll farms.

International Standards and Good Practice

While the extent of international standards and good practice for traditional media is prolific, specific application to social media is still in its infancy. As well, there is an important distinction to be made between traditional mass media (TV, radio, etc.) and some online media pages that can rightly be classified as mass media (and thereby linked to journalistic practices), versus most social media that are not true media *per se* (individuals, parties, discussion groups, etc.). To this latter category, applicable regulation and standards may vary and are still at incipient stages. Nevertheless, general protections of fundamental freedoms should pertain, especially those contained in the Universal Declaration of Human Rights, Article 19 of the International Covenant on Civil and Political Rights (ICCPR), Article 10 of the European Convention on Human Rights (ECHR) and other key regional and good practice documents (see Annex 1).

---

33 This definition is adapted from the Oxford Internet Institute, https://newsaggregator.oii.ox.ac.uk/methodology.php

34 Woolley, Samuel C. & Philip N. Howard, "Computational Propaganda Worldwide: Executive Summary," p. 6, http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Casestudies-ExecutiveSummary.pdf.

Most relevant international conventions and standards were written for traditional media. The right to freedom of expression (ICCPR,[35] ECHR[36]) limits government action against influence campaigns. But disinformation impinges on the right to political participation "without undue influence or coercion of any kind which may distort or inhibit the free expression of the elector's will," (Gen. Com. 25, ICCPR, Article 25), which also states that "voters should be able to form opinions independently, free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind."[37]

The fact that online disinformation might violate Article 25 "has not been explored, especially not in its practical implications... an argument can be made that governments have an obligation to ensure that social media companies organize discourse on their platforms in a manner that does not unduly distort or allow manipulative interference in order to guarantee proper public participation in electoral processes."[38] Social media companies are in some cases establishing their own rules before being required to. In 2016, Facebook, Microsoft, Twitter and YouTube agreed with the European Commission to a code of conduct regarding online hate speech. In 2018, Instagram, Google+, Snapchat and Dailymotion announced their intent to join the code.[39] Yet, this question of social platform responsibility is a key element of the ongoing debate in a number of areas – type of regulation, transparency and enforcement, content moderation and removal, etc.

The fact that international conventions and standards were created for traditional media affects a country's options for countering a foreign government's influence campaign in its elections. Important points to remember are that international conventions and standards were created for traditional campaigning (not only for media), international conventions and standards are addressed to nation states and not private entities, and international conventions and standards have traditionally tried to address a number of possible "undue influences" in campaigns.

Foreign influence campaigns and information theft fall under the category of cyber interference, a broader category than cyberattacks, which must meet the standards for force in Article 2(4) of the UN Charter. Cyber interference can mean physical destruction of equipment, meddling with vote counts, information theft, and information campaigns.[40] Article 2(4) of the UN Charter prohibits force, with an exception for self-defense against "armed attacks" established in Article 51. If cyber interference does not constitute an attack under international law, there is no legal route to respond with force under the UN Charter.

---

35 UN Human Rights Committee (HRC), CCPR General Comment No. 25: Article 25 (Participation in Public Affairs and the Right to Vote), The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service, 12 July 1996, CCPR/C/21/Rev.1/Add.7.

36 *European Convention on Human Rights*, art. 10.1.

37 Ibid.

38 Meyer-Resende, "A New Frontier: Social Media / Networks, Disinformation and Public International Law in the Context of Election Observation," 12-13.

39 "Countering illegal hate speech online," European Commission. For a good overview of other actions taken by internet companies and the human rights implications, see The NYU Stern Center for Business and Human Rights, "Harmful Content."

40 Van de Velde, "The Law of Cyber Interference in Elections," 17.

One possibility for defining an influence campaign as an attack is if it were seen to violate the ICCPR rights. This could mean the right to privacy in Article 17, the prohibition on "propaganda for war" in Article 20, and the right to hold "genuine periodic elections" without "unreasonable restrictions" in Article 25.[41] Disinformation campaigns could be countered with libel or slander laws, but states lack the resources and standing to sue. In principle, this could be a dangerous practice, where for instance an authoritarian regime uses libel laws to silence opposition views on the basis that they are false. Such suits are also limited to individual targets, not other states.[42] Disinformation campaigns are straightforward interference, but the intent is unclear with malinformation, since the intention of the actor concerned may not always be clear and may not be malicious in intent. Apart from the Council of Europe's 2006 Cybercrime Convention (Budapest Convention), there are no other binding international instruments at present that directly tackle the prevention of and punishment for cyberattacks.[43]

At the same time, the issues pertaining to social media are somewhat distinct and specific. There is an ongoing debate as to whether social networking services are media or not; this is central and no consensus on this exists to date. However, some of the rules applied to legacy media would justifiably apply to social platforms (e.g., the social page of a public broadcaster or newspaper). As well and more generally, while based on the same platforms, user-generated content and their producers are different (average citizens, trolls, media, candidates, etc.) and, as such, they are subject to different types of obligations.

Unlike traditional media, social media are usually interactive, use digital, online and mobile technology and are often audience-created and user-driven. They have an increasing impact on the public, including during elections, and enable politicians to pass on their messages directly to the electorate. Given the nature of online communication, it would be difficult to apply the same principles to social media platforms that the traditional media must respect during elections. As well, most social media are transnational in scope and do not predispose themselves in the same way to regulation that traditional media do. All of these issues are questions that are being teased out by international experts and organizations.

Nevertheless, some progress has been made. The Council of Europe's Commission for Democracy through Law (Venice Commission) is in the process of delineating a report of good practice in this sphere. At the same time, the European Union is exploring the possibility of developing a methodology for the observation of social media in the context of elections and Organization for Security and Co-operation in Europe's Office for Democratic Institutions and Human Rights (OSCE/ODIHR) is planning the same in the context of electoral campaigns. Key election observation organizations such as the European Union (EU) and National Democratic Institute are starting to deploy social media and disinformation analysts on their election observation missions.

The UN Special Rapporteur on the Promotion and Protection of the Right of Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, the Organization of American States (OAS)

41 Ibid, 22-24.

42 Ibid, 28. See, e.g., the recent prosecution of Russians by the US.

43 "Budapest Convention and related standards," Council of Europe.

Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access to Information have also recently made a Joint Declaration on Freedom of Expression and "Fake News," Disinformation And Propaganda, showing that interest in and engagement on such topics is pertinent in all geographic regions.[44]

Lastly, some states have already legislated in this space, particularly Germany and France.[45] It is important to ensure, however, that any regulation of social media platforms is not heavy-handed and will not result in undue restrictions on freedom of expression. Although case law is still forthcoming, it will be interesting to see how and on what basis national courts adjudicate such questions.

At the same time, the European Court of Human Rights' (ECtHR) jurisprudence regarding online media has developed in recent years. For example, in 2017, the ECtHR for the first time ruled on the online news media's liability for reporting serious allegations against a political candidate.[46] In its judgment in

---

44 See https://www.osce.org/fom/302796

45 The German parliament passed a law according to which social media platforms can face fines of up to €50 million for failing to remove hate speech. The law stipulates that the social media platforms must remove "obviously illegal" hate speech and other postings within 24 hours of receiving a notification. The intention of the law is to crack down on hate speech, criminal material and fake news on social networks, but it could also have consequences for free speech online. The companies affected by the law argued that the law could invite censorship and could have a chilling effect on free speech. The government made it clear that a fine would not necessarily be imposed after just one infraction, but only following a systematic refusal to act.

46 See *Olafsson v. Iceland:* https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-171974%22]} In this case, the applicant was an editor of an online news website *Pressan*, which published an article on a candidate in Iceland's 2010 Constitutional Assembly elections. In the article, two adult sisters alleged that the candidate, a relative of theirs, had "sexually abused them when they were children". The article was based on an interview with one of the women as well as a letter she had published on her own website. *Pressan* had also contacted the candidate before publication, and the article also included his denial of the allegations. Iceland's Supreme Court ruled that the website's editor was liable for "insinuations that the candidate was guilty of having abused children." Moreover, it held that the applicant "had a supervisory obligation which entailed that he should conduct his editorial duties in such a way that the published material would not harm anyone through its being defamatory." The Supreme Court agreed that candidates "have to endure a certain amount of public discussion" but ruled that this did not include being "accused of this criminal act in the media".

The ECtHR held that it had "no cause to question the Supreme Court's assessment that the allegations were defamatory." At the same time, the court noted that the article concerned matters of public interest, namely a candidate running in a general election, and the "issue of sexual violence against children," Secondly, because the candidate was running in a general election, he "must be considered to have inevitably and knowingly entered the public domain and laid himself open to closer scrutiny of his acts. The limits of acceptable criticism must accordingly be wider than in the case of a private individual." Importantly, the court reiterated the principle that a "general requirement for journalists systematically and formally to distance themselves from the content of a quotation that might insult or provoke others or damage their reputation is not reconcilable with the press's role of providing information on current events, opinions and ideas." The court then applied these principles to the website's editor and held that he had "acted in good faith" and ensured that the article had been written in compliance with "ordinary journalistic obligations to verify a factual allegation." Finally, the court stated that the interest in

this case, the court laid down important principles under Article 10 applicable to online media during elections, in particular where online media outlets follow good-faith journalism practices. In addition, the court also delivered important judgments concerning the online news media's liability for reader comments, which are also discussed below.

In 2015, the ECtHR ruled in the case concerning an online news website's liability for reader comments. Importantly, the court specified what factors should be taken into consideration when deciding whether an online news site should be liable for reader comments. These factors include the context of the comments, the measures applied by the applicant company to prevent or remove defamatory comments, the possible liability of the actual authors of the comments as an alternative to the applicant company's liability, and the consequences of the domestic proceedings for the applicant company. The court held that a commercial news site being held liable for reader comments that were "clearly unlawful speech" and "hate speech" did not violate Article 10 of the ECHR. The court concluded that "where third-party user comments are in the form of hate speech and direct threats to the physical integrity of individuals," liability may be imposed on online news sites "if they fail to take measures to remove clearly unlawful comments without delay."[47] While this decision concerned a webpage of a mass medium, it is arguable that a similar argument would hold for the social media pages of parties, candidates, influencers and other groups.

The same question pertains in a more recent case in 2017, the ECtHR highlighted an important principle, saying that "the liability for third-party comments may have negative consequences on the comment-related environment of an internet portal and thus a chilling effect on freedom of expression via [the internet]".[48] The case law shows the difficulty that the court has in reconciling two competing principles during elections: firstly, that it is particularly important in the period preceding an election that opinions and information of all kinds be permitted to circulate freely"; but secondly, and equally importantly, that during elections it may also be considered "necessary," to place "certain restrictions, of a type which would not usually be acceptable, on freedom of expression, in order to secure the "free expression of the opinion of the people in the choice of the legislature."[49]

Thus, although few international standards pertain to the area of disinformation, social media and elections,[50] such legal precedent begins to establish a norm that may become more widely applicable in the future.

---

protecting the candidate from the defamatory allegations was "largely preserved by the possibility open to him under Icelandic law to bring defamation proceedings against the sisters."

47 See *Delfi AS v. Estonia*: https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-155105%22]}

48 See *Pihl v. Sweden*: https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-172145%22]}

49 European Audiovisual Observatory, *Media coverage of elections: the legal framework in Europe*, Strasbourg, 2017,http://www.obs.coe.int/documents/205595/8714633/IRIS+Special+2017-1+Media+coverage+of+elections+-+the+legal+framework+in+Europe.pdf/b9df6209-651b-456a-bdf5-1c911d6768cc.

50 Although important steps are being made. See, for instance, "International Standards and Comparative National Approaches to Countering Disinformation in the Context of Freedom of the Media (on the

Regulation versus Self-Regulation

Regarding the issue of regulation versus self-regulation, it must be explicitly considered whether the activities of private companies toward other entities should be regulated by governments. For example, should decisions by Facebook vis-à-vis an organization that it claims are involved in hate speech be judged by state institutions or as an internal business matter? Does Facebook's status as a *de facto* monopoly mean that its activities could or should be regulated? There have also been discussions at the political level suggesting that the issue might be best dealt with from an anti-trust or monopoly framework and that breaking up big technology companies might be worth exploring. The dynamic has also recently shifted with Mark Zuckerberg, the head of Facebook, calling for regulation of the internet in specific spheres.[51]

It is important to consider to what extent should citizens look to governments to regulate these areas, not least as some governments are themselves already taking steps to restrict online communication and social media use for free communication. Another emerging trend seems to be co-regulation, in which industry is drawn into discussions with regulators about appropriate actions.

In the domain of self-regulation, algorithms can be and are already being used to identify and take down unsuitable content, but so far, they are not without fault. Thus, there is a real danger of limiting the freedom of online expression that needs to be avoided. As well, there is the concomitant issue of transparency with regard to the self-regulation of tech companies. This is especially the case with regard to content curation, access to data, implementation of mandatory content removal, algorithms, and other areas. These need to be sufficiently analyzed and understood if an appropriate model of self-regulation is to take hold.

Some level of regulations or self-regulation based on good practice should be considered. This could include a decision that if a company has a certain share of communication in a certain area in a country, they will be required to establish a legal presence in that country, as they have done in certain countries.

Thus, it is important to continue to put public pressure on big technology companies. At the moment, they rarely divulge their algorithms, although they do at times admit data leaks , seen most evidently with the Cambridge Analytica scandal during the 2016 U.S. elections.

In late October 2017, U.S. senators announced a new bill that would regulate online political ads. The new bill, called the Honest Ads Act, requires companies like Facebook and Google to keep copies of political ads and make them publicly available. Under the act, the companies would be required to release information on whom were those ads targeted to, as well as information on the buyer and the

---

request of the Russian Federation)", Office of the OSCE Representative on Freedom of the Media. Vienna, March 2019

51 See https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html?noredirect=on&utm_campaign=mbrf_02042019-1&utm_content=27918&utm_medium=email&utm_source=mediabrifing&utm_term=.504f82ddec60 where Zuckerberg himself calls for state regulation of the internet in four areas – harmful content, election integrity, privacy and data portability.

rates charged for the ads. These provisions would bring disclosure rules more in line with how political ads are regulated in traditional media and apply to any platform with more than 50 million monthly viewers. The companies would be required to keep and release data on anyone spending more than $500 on political ads in a year.

There seems to be a built-in mechanism that only scandal or serious problems will surface the risks and dangers to the wider public, creating the need for accountability, especially algorithmic. As a complement to this, IFES could be open to working together with governments around the world for which it is providing democratic assistance on regulation questions and with these companies, while being clear about the intentions and focus of everyone involved.

Constitutional Provisions and Legal Framework (Constitutions, Electoral Laws, Media Laws, Political Party Laws)

In general, since this field is a relatively new subject of inquiry, much of the framework issues pertain to traditional media. Nevertheless, this working paper attempts to tease out both what is relevant to social media from these traditional sources, but also to catalog any new relevant instruments to the discussion. In the interest of timeliness, this paper summarizes preliminary research being undertaken by IFES to conduct a global survey of legal and regulatory responses to disinformation. This future research will include a deeper analysis of cases under each of the below categories. At present, this section identifies new and emerging areas where laws and regulation to address disinformation are evolving.

Most of the public and private actions taken against disinformation thus far have been domestic, not international. Governments' primary means of intervention are regulatory and legislative. In many countries, there is an ongoing discussion on whether online platforms should be held to the same standards as traditional media.[52] Regulatory and legal responses available to governments to date have included:

- Regulation of content,[53] including:
    - Removal of content
    - Criminalizing the dissemination of disinformation
    - Criminalizing dissemination on grounds of defamation or hate speech
- Regulatory or legal mandates to monitor social media
- Classifying social media as traditional media

---

52 In the U.S., for example, a series of court cases in the 1990s led to Section 230 of the Communications Decency Act, which made internet platforms "not liable for the content they host… even when they try to intervene." Fitzgerald, "The Court Case that Enabled Today's Toxic Internet."

53 The issue of content regulation and its tension with freedom of expression is a central issue. A good starting point to delineate the parameters of such a discussion is the recent UN Special Rapporteur on the Promotion and Protection of the Right of Freedom of Opinion and Expression Report on Content Moderation and Free Expression (April 2018) and the Joint Declaration on Media Independence and Diversity in The Digital Age.

- Data privacy laws
- Political finance regulation that pertains to social media
- Regulation of dissemination methods
- Regulation of specific technology companies or social media platforms
- Voter education mandates

The push toward regulation in some contexts has been synonymous with a push to regulate content on social media platforms. Anti-disinformation legislation that takes this approach can run the very real risk of infringing on human rights, especially freedom of speech. Some approaches that have passed or are currently under debate, such as those taken in Germany[54] and India,[55] can provide governments with the right to require social media companies or internet service providers to remove content that is termed to be in violation of set standards. Another regulation of content approach is the criminalization of dissemination, such as the now-repealed law in Malaysia criminalizing fake news,[56] or those in place or under debate in Belarus,[57] Egypt,[58] Cambodia[59] and others. Extending defamation or hate speech laws to disinformation concerns can also be a form of content regulation. Such measures raise considerable censorship concerns in places where they are used to crack down on any dissenting speech.

Other countries are extending existing law or regulation that pertains to traditional media as a response to disinformation. This includes establishing regulatory or legal mandates to monitor social media, which may dovetail with existing mandates to monitor traditional media in electoral contexts. In Denmark, for example, a task force within the Ministry of Foreign Affairs was set up to track new pieces of misinformation online. Another approach to extending the existing regulatory framework is to classify social media in the same way as traditional media – specifically in areas like advertising, opinion polling and political finance norms – which allows for the existing legal framework governing traditional media to be extended to social media. In Albania, for example, changes to the electoral code in 2012 made political advertisements on digital platforms subject to the same regulation as traditional media.[60] Nevertheless, it should be kept in mind that the association between political ads and disinformation may be contentious as it seems to suggest that states would then have the right to regulate the content of campaigns.

While data privacy laws are adopted for many reasons, the limiting of personal data to interrupt the targeted dissemination of disinformation is one goal. For example, the EU General Data Protection

[54] https://www.washingtonpost.com/news/theworldpost/wp/2018/02/20/netzdg/?noredirect=on&utm_term=.0bbbab9d23dc

[55] https://www.nytimes.com/2019/02/14/technology/india-internet-censorship.html

[56] Yeung, "Malaysia appeals controversial fake news law." For a thorough repository of national legislation and case law, see Vargas Valdez, "Study on the Rile of Social Media and the Internet in Democratic Development."

[57] https://cpj.org/2018/06/belarus-moves-to-prosecute-fake-news-control-the-i.php

[58] https://www.reuters.com/article/us-egypt-politics/egypt-targets-social-media-with-new-law-idUSKBN1K722C

[59] https://www.theguardian.com/world/2018/jul/06/cambodia-fake-news-crackdown-prompts-fears-over-press-freedom?CMP=share_btn_tw

[60] https://www.osce.org/albania/159501?download=true

Regulation (GDPR) is a measure by which regulators could disrupt how bad actors spread and target disinformation.[61]

The realm of political finance regulation is ripe for consideration of how these sets of laws can be expanded or altered to mitigate or disrupt the flow of disinformation. Examples include the proposed Honest Ads Act in the U.S.[62] and 2018 French legislation that, among other things, requires social media platforms to make public who paid for and promoted ads.[63] Extension of campaign restrictions, such as campaign silence periods on social media, is another approach. In Mexico, for example, the Electoral Tribunal of the Federal Judicial Branch ruled that a political party's use of Twitter during the election silence period constituted a part of the party's campaign strategy and, therefore, fell within the legal provision governing silence periods.[64]

Another emerging regulatory area is the regulation of dissemination methods. Germany has explored anti-botnet legislation[65] (which raises an important questions as to the need to define in broad terms what are "acceptable" forms of regulation in this area, as bots are not necessarily wrong *per se*) and Ireland has drafted a law criminalizing the use of a bot to post political content through multiple fake accounts.[66] South Korea has proposed multiple amendments to existing laws to curb disinformation, including one that criminalizes the use of bots to manipulate online comment sections.[67]

Regulation or the threat of regulation against specific technology companies or social media platforms has also effected changes to platform's policies. India is one prominent example: the government made demands of WhatsApp following abuse on the platform. WhatsApp complied with all them except the demand of "adding a feature that could help trace origin of a message."[68] There has been another round of pushback, with proposed regulation requiring WhatsApp to get rid of its current encryption.

Facebook has also been expanding measures it launched in 2018 in India, the U.S. and the United Kingdom, introducing measures requiring political advertisers to undergo checks to prove they live in the country they are targeting. Moreover, ads will be labeled with disclosures identifying who paid for them and their contact details, in case they are a business and not an individual. Unregistered ads will be blocked. Facebook also introduced an Ads Library that will enable users to search political ads in a public database to find information on the number of times an ad was viewed and demographic data about who saw it.

---

61 https://www.cfr.org/blog/could-europes-new-data-protection-regulation-curb-online-disinformation

63 http://www.assemblee-nationale.fr/15/ta/tap0190.pdf

64 Vargas Valdez, "Study on the Rile of Social Media and the Internet in Democratic Development."

65 https://www.rsis.edu.sg/wp-content/uploads/2018/03/PR180416_Countering-Fake-News.pdf

66 https://web.archive.org/web/20180427203938/http:/www.oireachtas.ie/documents/bills28/bills/2017/15017/b15017d.pdf

67 https://www.poynter.org/fact-checking/2018/in-east-and-southeast-asia-misinformation-is-a-visible-and-growing-concern/

68 https://timesofindia.indiatimes.com/business/india-business/whatsapp-agrees-to-meet-all-govt-demands-except-message-traceability/articleshow/65508688.cms

The ad policies have had mixed results – reporters at Vice News were recently able to purchase ads on Facebook under the names of public figures and groups, and while they were required to provide a verified address and ID to Facebook, the public saw only the fake names.[69]
Flagging content on platforms faces difficulties as well: the verification quality can be poor, and even if it is well-executed, the evidence is mixed and often negative on the potential to change readers' opinions.[70] The tools available to social media platforms also vary by the nature of the service. WhatsApp's end-to-end encryption, for example, means that it cannot access or alter user messages, even when ordered to. Indonesian officials threatened to ban WhatsApp for lewd GIFs available on the app, and the situation was remedied only by the company that produced the GIFs changing its policy, not by WhatsApp changing content shared by users.[71]

Another legal and regulatory area of future interest will be if and how countries include or alter voter or citizen education mandates that address the issue of disinformation. Governments may choose to adopt explicit mandates or curricula changes that require an array of state institutions to educate the electorate or students about risks to information integrity and build media literacy.

Understanding of how to counter disinformation is inchoate and there are important gaps in the research. The effects of exposure to disinformation on individual beliefs and behaviors are still unknown, as are the likely effects of regulatory responses.[72] Given the impact on elections that disinformation has had and will continue to have, EMBs and governments cannot wait for a fully developed set of research results to guide them. They will need to act using the best knowledge available and with full respect of human rights.

Current Responses of EMBs to Disinformation

From among the countries where IFES is currently working, responses to disinformation threats on the part of EMBs have been varied. Specific responses noted by IFES include the following:

- Direct EMB coordination with technology companies;
- Fact-checking;
- Voter education;
- Identification of and strategic and crisis planning for disinformation threats; and
- EMB coordination with other state agencies.

EMBs and CSOs can also work directly with technology companies to provide local capacity these companies typically lack, including help with local languages. In Mexico, the National Electoral Institute signed a Memorandum of Understanding and worked during the 2018 elections with Facebook,[73] and in Estonia, the State Electoral Office "is monitoring online media in cooperation with other authorities and

69 Turton, "Facebook's political ad tool let us buy ads 'paid for' by Mike Pence and ISIS."
70 Tucker et al., 38.
71 Tan, "Indonesia wanted to block WhatsApp because people are sending 'obscene GIFs.'"
72 Tucker et al., 54 and 62.
73 Bassante, "Protecting the Mexican Election from Abuse."

Panama

The Electoral Tribunal launched a Digital Ethical Pact in June 2018 that invited social media users to voluntarily commit to sharing information responsibly and be wary of information that could be fake news, fake accounts, or malicious bots. The pact encourages respectful debate about ideas, and was launched in public events and a social media campaign.

Other branches of government have also focused on working together with citizens to increase media literacy and disarm fake news. For example, in August 2018, the Government Ministry hosted the first Digital Forum and Training Workshop on Fake News and Digital Journalism.

private companies such as Facebook and Google."[74] Actions available to technology companies include flagging or removing false content, rating content based on the findings of fact-checkers, changing their algorithms to promote truthful information and hide false information, and ad disclosure policies. Facebook, for example, has begun changing their ad disclosure policies and has experimented with flagging false content. These changes are not instituted globally but rather adapted to each country's context.[75] Prior to Nigeria's elections in 2019, Facebook worked with local CSO fake-checking organizations to halt the spread of fake news on the platforms. The local CSOs were able to assess the accuracy of the news shared on the platform using locally contextualized knowledge.

One response available to many electoral stakeholders is fact-checking, and EMBs can also provide or partner with organizations that attempt to check disinformation by providing accurate information to citizens. Fact-checking initiatives can focus on debunking false stories or rating content by truthfulness and use a wide range of methods. Many fact-checking groups were created by news organizations or independent journalists, such as *Mafindo* in Indonesia and *Chequeado* in Mexico,[76] and often receive governmental or international support. There are ongoing efforts to develop fact-checking algorithms, but such attempts are nascent.[77] The efficacy of fact-checking is unclear, and the external

74 Estonia: Parliamentary Elections, March 3, 2019, Office for Democratic Institutions and Human Rights *Needs Assessment Mission Report*, Organization for Security and Co-operation in Europe, 5.

75 Dave, "Exclusive: Facebook brings stricter ads rules to India ahead of general election."

76 Mexico, in particular, seems to present a case of good practice. The Mexican EMB has engaged with organizations like the Venice Commission and specific countries like Tunisia to offer a good practice model, covering both disinformation and campaign finance regulation.

77 One example is The Duke Tech and Check Cooperative, a project to automate fact-checking. For a discussion of the issues these efforts face, see Funke, "Automated fact-checking has come a long way. But it still faces significant challenges."

validity of psychological studies on how we process debunking and corrective information, as well as the studies' results, are mixed.[78] EMBs can also call upon their voter education and voter information mandates to address the challenges of disinformation. Another possible area of intervention may involve cooperation with credible legacy media to create a counter-narrative against disinformation on elections. In many countries traditional media contribute to amplifying disinformation and polarizing narratives, rather than debunking them.

Strategic and crisis planning to help EMBs anticipate disinformation threats and map out responses is another prospective avenue for EMBs to develop their disinformation responses. The use of media monitoring, social listening or other tools to monitor disinformation and identify demographics and social cleavages vulnerable to exploitation by disinformation actors, for example, can be used for planning how to address disinformation threats as they emerge.

Mexico

Mexico's EMBs have kept up a focus on innovating in the realm of disinformation.

The National Electoral Institute (INE) is the first EMB in the world to have memoranda of understanding with Facebook, Twitter and Google.

The 2014 reforms to Mexico's electoral process gave the Electoral Tribunal of the Federal Judicial Branch (TEPJF) jurisdiction over complaints related to electoral campaigning and the media.

The TEPJF must adopt binding decisions on cases related to the use of social media, disinformation strategies and other new features of democracy in the digital era, even where rules do not exist. The TEPJF made several notable decisions in this area during the 2018 electoral process, including protecting freedom of expression while encouraging access to fact checkers as a response to fake news.

EMBs might also consider measures to protect their technology infrastructure against hacks on EMB databases that compromise sensitive voter data or take control of EMB websites or social media accounts to use those platforms to publish false information about elections or to diminish the EMB's credibility.

EMB partnerships with other state agencies to monitor or counter disinformation are also an avenue being used. Australia has established an Electoral Integrity Task Force to assist the EMB with

[78] See, e.g., the following studies: Pennybook and Rand, "The Implied Truth Effect"; Pennycook et al., "Prior Exposure Increases Perceived Accuracy of Fake News"; and Nyhan et al., "Real Solutions for Fake News?"

Ukraine

The absence of regulations and enforcement mechanisms for social media for political advertisement decreases the overall transparency of political and campaign financing, weakens accountability of electoral (and other) stakeholders, and increases opportunities for disinformation.

Ahead of Ukraine's 2019 presidential and parliamentary elections, IFES established a dialogue across civil society, campaign finance oversight bodies and social media representatives, leading to an increase in overall transparency of campaign finance on and beyond social media.

The CEC also established direct contact and dialogue with social media representatives/Facebook and increased its online presence/communication ahead of and through election day.

disinformation,[79] and Denmark has launched an initiative that includes cooperation between the EMB and other government agencies (e.g., data protection, media councils, etc.).[80]

79 https://www.reuters.com/article/us-australia-security-elections/australia-forms-task-force-to-guard-elections-from-cyber-attacks-idUSKCN1J506D

80 http://um.dk/en/news/newsdisplaypage/?newsid=1df5adbb-d1df-402b-b9ac-57fd4485ffa4

Future Innovations and Areas of Research

Having analyzed the current approaches in this field, including with a panel of U.S. and European experts in context of the Regional Elections Administration and Political Process Strengthening (REAPPS) II project,[81] the main facets of development seem to lie in three areas, presented below in terms of time and impact horizons for better efficacy in application:[82]

1. Short-Term: Exposing Disinformation

- Supporting civil society in fact-checking and myth-busting
- Media monitoring and analysis
- Ground-zero source attribution of fake news

One of the most popular responses to the phenomenon of social media and disinformation has been a wide variety of fact-checking and myth-busting responses being used by both the journalistic community and by CSOs (Stop-Fake, etc.). This is a much-touted avenue that has received much attention, and naming and shaming is still considered to be a major avenue for countering this threat.

This issue should also encompass commentary, and the old argument that websites providing space for user comments are not responsible for the comments left by users is no longer tenable. At the same time, while access to general information is not seen as a problem, the storage, the processing and the accessing of personal data by the state and citizens has raised questions regarding data ownership and privacy. Recent scandals about data leaks in administration and technology companies and the processing of personal data stored in social networks resulted in regulations such as the GDPR, which clearly state who has access to data and how it can be used. Moreover, transparency is not only about accessing more information but also about the quality of information. Nowadays, access to data means accountability can be more efficient but also more vulnerable. We have access to a massive amount of data with which to hold politicians accountable, but disinformation can spread quickly and lead to uncontrollable and undesirable effects. There is therefore not only the risk that citizens will get lost in a sea of data and information, but that they will get lost in a sea of information that does not even relate

---

81 The REAPPS II programming (2018-23) funded by the U.S. Agency for International Development will focus on developing and strengthening democratic leadership skills among politically active young people from political parties and civil society; increasing government oversight, political participation, and especially institutional and public resilience to anti-democratic foreign influences; and strengthening election administration leadership and effectiveness. The overall goal of this regional program is to promote cross-border learning and networking on a wide range of existing initiatives and to develop new capacity in political participation, government transparency and election administration. Further details can be found at https://www.ifes.org/REAPPS

82 While these are specific actions that EMBs or INGOs can take, they are by no means comprehensive. Other areas of development include accountability mechanisms being derived from case law, developments in evidence gathering standards, bilateral Memorandums of Understanding being developed by specific EMBs, work being done through international groups like Design 4 Democracy and convening powers to discuss and bridge the gap between social media providers, EMBs, courts, regulators, INGOs, etc.

to the truth. In the future, new forms of intelligence will have an even greater impact on transparency and accountability, making algorithmic accountability a necessity. This is equally true of the new challenges presented by "deepfakes" and similar strategies in this space must be developed.

Nevertheless, the efficacy of this short-term avenue must be taken with a degree of skepticism. The latest academic research indicates that "the phenomenon of belief perseverance, when false beliefs persist despite corrections, combines with other features of our psychology to limit the impact of fact-checking."[83] However, with this limitation in mind, programming with an element of exposing disinformation may have an impact as part of a larger and more comprehensive initiative.

In concert with this type of fact-checking and myth-busting methodology, social media monitoring has become a crucial element of this strategy. A number of countries, including Georgia, Nigeria, and Ukraine, have been pioneering new methodologies in this space and while still a work in progress, lessons learned from these pilots should be drawn. In particular, INGOs can continue to assist CSOs to develop more methodologically sound approaches to social media disinformation monitoring and give them a global voice (e.g., on disinformation in electoral processes, election violence, hate speech and discrimination, etc.).

Finally, another technique that could be considered is ground-zero source determination and attribution of disinformation. For technology platforms such as Facebook and Twitter, it has become increasingly possible to trace back threads to their originator.[84] With possible cooperation between such platforms, INGOs and CSOs, it may be possible to design strategies to give notoriety to who originated various fake news and to publicize this as well as the path of dissemination. While it may be compelling for a user to believe certain information if she thinks it comes from her colleague, she may look differently upon it if she is presented with information that the originator of the information thread, for instance, is a hacker in Russia and that it has been amplified in its spread by a series of bots rather than real humans. Similarly, users may look upon certain information differently if they know that 50 percent of certain disinformation has originated from a single source (such statistics are increasingly coming to light). Thus, this area of assistance to both EMBs and CSOs may lead to the development of important mitigating measures.

2. Medium-Term: Risk Identification and Mitigation

- Introduction of disinformation criteria and analysis into assessment methodologies
- Work with EMBs, particularly in developing a new generation of media assistance, including questions of legal/procedural reform as well as technical/information technology (IT) tools that may help EMBs to identify cases of disinformation in different forms

---

83 See, e.g., Thorson 2015 and Flynn et al. 2016.

84 Examples include the use of Social Network Analysis to examine information networks and identify prominent information actors and leading influencers in disinformation campaigns; the use of cyber forensics to identify the media approach that is used for the disinformation campaign, while understanding the patterns associated with dissemination of information and how they build engagement with their audiences and frame the narrative; the use of Focal Structure Analysis to identify coordinated disinformation campaigns and which individuals are connected and have the most influence and to build these tools into the design of programs from the onset.

- Mainstreaming disinformation issues in EMB crisis simulations such as tabletop exercises (TTX)
- EMBs engagement with social media companies to preserve electoral integrity

The particular strength of INGOs offering technical assistance to EMBs may lie in helping them to identify and then mitigate risks stemming from disinformation threats in a more systematic and strategic manner. While assessment criteria have tended to focus on traditional issues of electoral integrity, it is important that these new challenges are incorporated into INGOs' methodologies and approaches. In this space, which is quickly moving and changing, it is important to be aware of what has already been done to not be duplicative. At the same time, global experience and knowledge can add comparative value in precisely delimiting the factors in this new space that have a fundamental impact on electoral integrity questions.

Having adjusted the methodological frameworks, INGOs should then work with their EMB partners to face disinformation campaigns more strategically. This is necessary to build resilience and mitigate the risks identified from such social media and disinformation threats. With a proper analysis in hand, work can then be undertaken to mainstream this through strategic planning mechanisms, communication, outreach strategies and crisis simulation exercises.
This will allow EMBs to understand in advance of disinformation campaigns occurring what types of disinformation they may expect and to be transparent and forward-leaning in presenting an accurate narrative of challenges and events in advance. Just a risk-mitigating approach has been shown to work in the latest strategic approaches in the U.S. and Sweden.[85] As well, such solid initial analysis could form the basis for further legal or procedural reform necessary to deal with these issues.

In many ways, advance inoculation has been shown to have more impact than attempting to take corrective measures after disinformation has already spread. As a result, mainstreaming such strategies through social media programs, similar to how traditional media strategies were developed, may be the best way to mitigate these risks more coherently. There may also be a role for developing crisis simulation TTX to help EMBs game out threats and solution scenarios in real-time and develop reaction resilience to such disinformation campaigns.

Lastly, there may also be scope for INGOs to triangulate their work between EMBs and social media companies to better preserve electoral integrity as well as their institutional independence, which is key. INGOs can help EMBs to develop clear and more evidence-based strategies on how they might effectively consult or cooperate with organizations such as Facebook, Twitter and Instagram to develop threat matrices and reaction strategies to known disinformation threats. INGOs could also help both to identify emerging good practices from which lessons can be drawn; examples like those of Mexico or Panama may be good places to start.

[85] See particularly the work of the Election Assistance Commission in the U.S., in concert with the Belfer Center and the National Association of Secretaries of State and the Swedish Civil Contingencies Agency.

3. Long-Term: Education and Awareness-Raising

- Increase support for good-quality journalism
- Build disinformation resilience and greater digital literacy through civic education and journalist training
- Support civil society on voter education in the disinformation space

Most experts agree that an antidote to the virus of disinformation continues to be good-quality journalism and there is a fundamental question as to whether societies can get back to this level. Most people globally still primarily rely on television and radio for news. As well, of the people who go to social media for their news, increasing evidence shows that many do not trust it, with trust in traditional media still being higher.[86] Discussions need to take place about how good-quality journalism needs to be based on old-style approaches to basing reporting on facts, but it must also address the huge amounts of data that are now available and that need to be sifted through and analyzed. Furthermore, journalists need to find new and creative ways of how to better monetize the value of information; it undoubtedly remains a valuable asset, but the media sector is currently not capable of fully monetizing that value to their overall sustainability.

At the same time, the latest research from Scandinavian countries (Finland and Norway) seem to indicate that the most effective tool to combat disinformation is to provide citizens from an early age with the critical and analytical tools to be able to process all forms of information appropriately and in context in the form of media literacy education. This is often embedded in wider-ranging civic education initiatives, provided through the education system. This education can be conducted by CSOs or by state agencies but might ultimately be most effective when incorporated into schools' curricula.[87] In a media literacy training pilot conducted in Ukraine, participants were more likely to "demonstrate sophisticated knowledge of the news media industry" and 82 percent reported that they cross-checked their news 18 months after the training.[88] If shown to be impactful, these could be further developed and rolled out on a more global basis.

These same ideas and strategies can be incorporated by INGOs in the assistance they provide to CSOs. While many CSOs may be well-versed in traditional aspects of electoral cycles – candidate and voter registration, Election Day procedures, counting and tabulation – very few have had the resources (IT, data analysis, sociology, etc., given the multidisciplinary nature of this phenomena) to closely look at the issue of disinformation and electoral integrity. Thus, assistance from INGOs in this area may also have an impact.

---

86 See Hunt and Gentzkow, 13.

87 Jankowicz, "The Disinformation Vaccination."

88 Murrock, et al., "Winning the war on state-sponsored propaganda," 4, 14.

Conclusions

In conclusion, this working paper has sought to establish a baseline in its analysis and evaluation of the imminent threat of social media disinformation and electoral integrity. It has underscored the novelty of this space and its distinction from traditional media, both in terms of academic inquiry and international obligations and good practice.

To get a better understanding of the issue, the working paper has reviewed the extant literature related to this area of enquiry and has provided an overview of the current state of play regarding constitutional and legal solutions being implemented by particular states, as well as the strategies and mitigation measures currently being explored by EMBs in trying to navigate this transmuting space.

The paper has also offered directions in which INGOs and EMBs can attempt to get a handle on this threat, looking at short-, medium- and long-term assistance areas and strategies that can be explored and implemented. With such a more cohesive roadmap in place, it is hoped that technical assistance in this space may be more firmly grounded and, thus, more impactful in its results.

Given the analysis and scenarios presented above, the question arises of what can INGOs, such as IFES, do to assist EMBs and CSOs – and possibly privacy authorities, media regulators, and authorities responsible for regulating campaign finance – in developing better strategies to cope with social media disinformation and its impact on electoral integrity. The most promising results appear to be in the longer term, through digital literacy and civic education initiatives, and in promoting better-quality journalism. These strategies, however, may take decades to yield effect – although this should not be a rationale for not engaging in this space. Thus, more medium-term strategies, such as developing assessment methodologies in this sphere and mainstreaming them through programming, may yield more immediate results. This should include assisting EMBs to a better understanding of how such social media disinformation operates and helping to train EMBs to develop strategic plans and responses to better tackle false information and well as possibly engaging with technology companies on better informed basis.

Lastly, more immediate assistance, such as fact-checking, myth-busting, or originator identification may also provide some more immediate results, especially when an election is imminent and more long-term, strategic assistance may not have been possible. This could also involve identifying and disseminating success stories, as it seems that voters are becoming increasingly tired of only hearing negative examples. In total, and throughout all these activities, it is important to analyze and identify the key causes and contributing factors and identifying appropriate solutions.

While this space is rapidly changing and developing and while constant updates to our thinking in this space will be necessary, an initial baseline is essential. This IFES working paper is published with this goal in mind, to better educate current thinking and the way forward on this controversial issue.

Annex 1 - International Standards and Good Practice

International Standards and Good Practice – Social Media and Elections

Treaty Obligations

Universal

- United Nations (UN) International Covenant on Civil and Political Rights (Article 19) (link)
- UN Convention on the Rights of Persons with Disabilities (Articles 21, 29) (link)

Regional

- European Convention on Human Rights (Article 10) (link)
- Commonwealth of Independent States' Convention on Human Rights and Fundamental Freedoms (Article 11) (link)
- Convention on the Standards of Democratic Elections, Electoral Rights and Freedoms in the Member States of the Commonwealth of Independent States (CIS) (Articles 7,9,10,13,19) (link)
- Council of Europe's Framework Convention for the Protection of National Minorities (Article 9) (link)

Political Commitments

Universal

- UN Universal Declaration of Human Rights (Article 19) (link)

Regional

- American Convention on Human Rights (link)
- African Charter on Human and Peoples' Rights (link)
- Organization for Security and Co-operation in Europe's (OSCE) 1990 Copenhagen Document (Articles 7,9,10) (link)
- OSCE 1990 Charter of Paris for a New Europe (Human Rights, Democracy and Rule of Law) (link)
- OSCE Document of the Moscow Meeting of the Conference on the Human Dimension (Paragraph 26) (link)
- OSCE 1999 Istanbul Summit Declaration (Paragraphs 26,27) (link)
- OSCE 2018 Ministerial Council Decision No. 3/18 on the Safety of Journalists (link)
- Declaration on Fundamental Principles Concerning the Contribution of the Mass Media to Strengthening Peace and International Understanding, to the Promotion of Human Rights and to Countering Racism, Apartheid, and Incitement to War (link)

Office of the UN High Commissioner for Human Rights General Comments

- General Comment 25: The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service (Paragraph 12) (link)
- General Comment 10: Freedom of expression (Paragraphs 1-3) (link)

Other International Standards and Good Practices

Universal

- UN General Assembly Resolution A/RES/55/96 (2001) – Promoting and Consolidating Democracy (Paragraph iv) (link)
- UN Special Rapporteur on the Promotion and Protection of the Right of Freedom of Opinion and Expression Reports (1999-2009) (link)
- UN Special Rapporteur on Freedom of Expression Report on Content Moderation and Free Expression (April 2018)
- Joint Declaration (2002) - UN Special Rapporteur on the Promotion and Protection of the Right of Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, Organization of American States (OAS) Special Rapporteur on Freedom of Expression (link)
- Joint Statement on the Media and Elections (2009) - UN Special Rapporteur on the Promotion and Protection of the Right of Freedom of Opinion and Expression , OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information (link)
- Joint Declaration on freedom of expression and "fake news," disinformation and propaganda (2017) – UN Special Rapporteur on the Promotion and Protection of the Right of Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression and the ACHPR Special Rapporteur on Freedom of Expression and Access to Information (link)
- Joint Declaration on Media Independence and Diversity in the Digital Age.

Regional

- European Commission (2018) Code of Practice on Disinformation (link)
- European Commission (2018) Action Plan on Disinformation (link)
- Council of Europe Recommendation CM/Rec (2007)15 of the Committee of Ministers to member states on measures concerning media coverage of election campaigns (2007) (link)
- Parliamentary Assembly of the Council of Europe (PACE): Resolution 2254 (2019) Media freedom as a condition for democratic elections
- PACE: Resolution 1636 (2008) Indicators for media in a democracy
- Council of Europe Declaration on freedom of political debate in the media by the Committee of Ministers (2004)
- Venice Commission's Code of Good Practice in Electoral Matters (2002)

Regional Treaties and Charters

- European Convention for the Protection of Human Rights and Fundamental Freedoms (link)
- Inter-American Convention on Human Rights (link)
- Inter-Parliamentary Union, Declaration on Criteria for Free and Fair Elections (link)
- African Charter on Human and Peoples' Rights (link)
- Declaration on the Principles Governing Democratic Elections in Africa (link)
- Convention on the Standards of Democratic Elections, Electoral Rights and Freedoms in the Member States of the CIS (link)

Relevant Case Law

- European Court of Human Rights (2017), *Olafsson v. Iceland* (link)

Annex 2 – Bibliography

White Papers/Guidebooks

IFES

- Vasu Mohan (2018), *Countering Hate Speech in Elections: Strategies for Electoral Management Bodies*, http://www.ifes.org/sites/default/files/2017_ifes_countering_hate_speech_white_paper_final.pdf
- Lisa Reppell and Erica Shein (2019), *Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions*, https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf

Non-IFES

- United Kingdom House of Commons (2019), Digital, Culture, Media and Sport Committee, *Disinformation and 'fake news': Final Report*, Eighth Report of Session 2017-19
- Council of Europe, "Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes" (Adopted by the Committee of Ministers on February 13, 2019, at the 1337th meeting of the Ministers' Deputies), https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b
- Kofi Annan Foundation (2019), "Digital Dangers to Democracy."
- Michael Meyer-Resende (2018), *A New Frontier: Social Media / Networks, Disinformation and Public International Law in the Context of Election Observation.* Democracy Reporting International, https://democracy-reporting.org/wp-content/uploads/2018/10/A-new-frontier_social-media_election-observation_Briefing-Paper-by-Michael-Meyer-Resende.pdf
- Election Observer Guidelines
    - European Union, forthcoming.
- European Audiovisual Observatory (2017), Media Coverage of Elections: The Legal Framework in Europe, Strasbourg, https://www.obs.coe.int/en/web/observatoire/-/media-coverage-of-elections-the-legal-framework-in-europe
- International IDEA (2019), Cybersecurity in Elections Models of Interagency Collaboration by Sam van der Staak and Peter Wolf, https://www.idea.int/publications/catalogue/cybersecurity-in-elections
- Office of the OSCE Representative on Freedom of the Media Vienna (2019), International Standards and Comparative National Approaches to Countering Disinformation in the Context of Freedom of the Media (on the request of the Russian Federation), https://www.osce.org/representative-on-freedom-of-media/424451?download=true
- Social Media Frameworks
    - Stakeholder Democracy Network (2019), Social Media Monitoring Methodology (forthcoming), http://www.stakeholderdemocracy.org/elections2019/?fbclid=IwAR1WI2VhKcKBzKf1SFnih7Ns7scBVoabYdYMqjTOvY8TLNShgtsWMk7vfPQ

Academic Research and Reports

- Alén-Savikko, Anette, Apa, Ernesto, Bassini, Marco, Javier Cabrera Blázquez, Francisco et.al (2018), *Media reporting: facts, nothing but facts? IRIS Special, European Audiovisual Observatory, Strasbourg, 2018. – P. 111-118. https://rm.coe.int/media-reporting-facts-nothing-but-facts/16808e3cda*
- Allcott, Hunt and Gentzkow, Matthew (2017) *Social Media and Fake News in the 2016 Election* (Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 211–236), https://web.stanford.edu/~gentzkow/research/fakenews.pdf
- Arnaudo, D. (2017). *Computational Propaganda in Brazil* (Computational Propaganda Working Paper Series No. 2017.8). Oxford, United Kingdom: Oxford Internet Institute, University of Oxford.
- Bradshaw, S., and Howard, P. N. (2017). *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation.* The Computational Propaganda Project. Retrieved from http://comprop.oii.ox.ac.uk/research/troops-trolls-and-trouble-makers-a-global-inventory-of-organized-social-media-manipulation/
- Bradshaw, S., and Howard, P. N. (2018). *Why does Junk News Spread So Quickly Across Social Media? Algorithms, Advertising and Exposure in Public Life.* Knight Foundation Working Paper. Retrieved from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/142/original/Topos_KF_White-Paper_Howard_V1_ado.pdf
- Bradshaw, S., Neudert, L.-M., and Howard, P. (Forthcoming). *Government Responses to Social Media Manipulation. Computational Propaganda Project Working Paper.*
- Earl, J., Martin, A., McCarthy, J. D., and Soule, S. A. (2004). *The use of newspaper data in the study of collective action*. Annual Review of Sociology, 65-80.
- Edwards, F., Howard, P. N., and Joyce, M. (2013). *Digital Activism and Non-Violent Conflict*. Digital Activism Research Project.
- Flynn, D.J., Brendan Nyhan, & Jason Reifler. 2016. The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs about Politics. Political Psychology 38(S1), 127-150. Retrieved from http://www.dartmouth.edu/~nyhan/nature-origins-misperceptions.pdf
- Greenwood, S., Perrin, A., and Duggan, M. (November 11, 2016). *Social Media Update 2016*. Retrieved January 13, 2018, from http://www.pewinternet.org/2016/11/11/social-media-update-2016/
- Herring, S. C. (2009). "Web Content Analysis: Expanding the Paradigm." In J. Hunsinger, L. Klastrup, and M. Allen (Eds.), *International Handbook of Internet Research* (pages 233-249). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_14
- Howard, P. N., and Hussain, M. M. (2013). *Democracy's Fourth Wave?: Digital Media and the Arab Spring*. New York, NY: Oxford University Press.
- Howard, P. N., Woolley, S., and Calo, R. (2018). *Algorithms, bots, and political communication in the U.S. 2016 election: The challenge of automated political communication for election law and administration*. Journal of Information Technology & Politics, 1-13. https://doi.org/10.1080/19331681.2018.1448735
- Howard, P., and Woolley, S. (2016). *Political Communication, Computational Propaganda, and Autonomous Agents*. International Journal of Communication, 10(Special Issue), 20. 24

- Joyce, M., Antonio, R., and Howard, P. N. (2013). Global Digital Activism Data Set. ICPSR. Retrieved from http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34625/version/2
- Pennycook, Gordon, Tyrone D. Cannon, & David G. Rand. (2018). Prior Exposure Increases Perceived Accuracy of Fake News. Journal of Experimental Psychology: General 47(12), 1865-1880. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958246&download=yes
- Sanovich, S. (2017). Computational Propaganda in Russia: The Origins of Digital Misinformation (Computational Propaganda Working Paper Series No. 2017.3). Oxford, United Kingdom: Oxford Internet Institute, University of Oxford.
- Strange, A., Parks, B. C., Tierney, M. J., Dreher, A., and Ramachandran, V. (2013). *China's Development Finance to Africa: A Media-Based Approach to Data Collection* (Working Paper No. 323). Retrieved from https://www.cgdev.org/publication/chinas-development-finance-africa-media-based-approach-data-collection
- Thorson, Emily. 2015. "Belief echoes: The persistent effects of corrected misinformation." Political Communication. https://www.tandfonline.com/doi/abs/10.1080/10584609.2015.1102187
- Vosoughi Soroush, Roy Deb, Aral Sinan (2018). "The spread of true and false news online," Science 359, 1146-1151.
- Woolley, S. C. (2016). *Automating Power: Social Bot Interference in Global Politics*. First Monday, 41(4). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/6161/5300

Mis/Disinformation

- *Accountability and Media Literacy Mechanisms as Counteraction to Disinformation in Europe* Journal of Digital Media and Policy, 2019. No. 3.
- The European External Action Service East Stratcom Task Force, which publishes a "Disinformation Review," predominantly aimed against disinformation from Russia. https://euvsdisinfo.eu/disinfo-review/
- *Disinformation in the Media under Russian Law*, IRIS Extra, European Audiovisual Observatory, Strasbourg, 2019, pgs. 1-27, https://rm.coe.int/disinformation-in-the-media-under-russian-law/1680967369
- *Foreign influence operations in the EU*, a July 2018 report by the European Parliamentary Research Service, focusing on misinformation and hybrid threats. http://www.europarl.europa.eu/RegData/etudes/BRIE/2018/625123/EPRS_BRI(2018)625123_EN.pdf
- "The Grim Conclusions of the Largest-Ever Study of Fake News" (2018), *The Atlantic*, https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/
- Atlantic Council's Digital Forensic Research Lab runs a program called "digital sherlocks," focused on disinformation, often focusing on Russian activities. https://www.digitalsherlocks.org/about
- "Bots and the Swedish election," a study by the governmental Swedish Defence Research Agency on automated accounts on Twitter used in relation to the September 2018 elections. https://www.foi.se/rapportsammanfattning?reportNo=FOI%20MEMO%206466
- *Countering information influence activities: A handbook for communicators*, a 2018 report by the Swedish governmental Swedish Civil Contingencies Agency. https://www.msb.se/sv/Produkter--

tjanster/Publikationer/Publikationer-fran-MSB/Countering-information-influence-activities--A-handbook-for-communicators/

- *Fake News: National Security in the Post-Truth Era*, report from January 2018 by academics at the S. Rajaratnam School of International Studies, Singapore. https://www.rsis.edu.sg/wp-content/uploads/2018/01/PR180313_Fake-News_WEB.pdf
- *Fake News and Freedom of the Media*, J. INT'L MEDIA & ENTERTAINMENT LAW. Vol. 8, No. 1, 2018/2019, https://www.swlaw.edu/sites/default/files/2019-03/JIMEL%208.1%20032919%20845AM.pdf
- *Closing digital loopholes on foreign ads*, by a University of Wisconsin scholar. https://www.issueone.org/wp-content/uploads/2018/04/04-16-18-CLC-IO-Issue-Brief-Young-Mie-Report-FINAL.pdf
- "The making of disinformation," by *CNN*. https://edition.cnn.com/2017/10/11/opinions/the-making-of-a-russian-disinformation-campaign-opinion-weiss/index.html
- "The troll-factory," by *Moscow Times*. https://themoscowtimes.com/news/mueller-charges-russians-with-pro-trump-anti-clinton-meddling-60544
- "Deepfakes and the New Disinformation War: The Coming of Age of Post-Truth Geopolitics," by Robert Chesney and Danielle Citron, in *Foreign Affairs* https://www.foreignaffairs.com/print/1123492
- "Swedish election (2018) notes high incidence of 'junk news'," by Oxford Internet Institute. http://www.ox.ac.uk/news/2018-09-06-swedish-election-second-only-us-proportion-%E2%80%98junk-news%E2%80%99-shared?fbclid=IwAR0HZgHr0t7nvy1_Di_uR8F8q7Vi-JPt-gSln47gPiQY9MC9KdE3tub1qQo
- Gu, Lion et al., "The Fake News Machine," Trend Micro, 2017, https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf.

General

- American Bar Association's "Legal Fact Check" deals not exclusively with foreign influence, but does address foreign influence over U.S. electoral processes and related issues. https://abalegalfactcheck.com/
- "How we are Exposing Foreign Interference in Ukraine's Elections," from Atlantic Council. https://www.atlanticcouncil.org/blogs/ukrainealert/how-we-are-exposing-foreign-interference-in-ukraine-s-elections?fbclid=IwAR3mQqnCoFuWwGvRsdcd5fRhZFx_p432HxmDWMgdCzDHduk73H2nWdhATUM
- "Technology favours tyranny," *The Atlantic*. https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/
- "DoS to revamp public diplomacy," *Foreign Policy*. https://foreignpolicy.com/2018/10/19/state-department-considering-public-diplomacy-overhaul/
- Lawson, Anton E. and William A. Worsnop, "Learning about evolution and rejecting a belief in special creation," *Journal of Research in Science Teaching:* 29(2), February 1992, https://onlinelibrary.wiley.com/doi/abs/10.1002/tea.3660290205

- Kahan, Dan, "Weekend update: You'd have to be science illiterate to think 'belief in evolution' measures scientific literacy," The Cultural Cognition Project at Yale Law School, May 24, 2014, http://www.culturalcognition.net/blog/2014/5/24/weekend-update-youd-have-to-be-science-illiterate-to-think-b.html
- Kahan, Dan, "What sorts of inferences can/can't be drawn from the 'Republican shift' (now that we have enough information to answer the question)?" The Cultural Cognition Project at Yale Law School, January 6, 2014, http://www.culturalcognition.net/blog/2014/1/6/what-sorts-of-inferences-cancant-be-drawn-from-the-republica.html
- Ohio State University, "Reliance on 'gut feelings' linked to belief in fake news," *EurekAlert!,* September 18, 2017, https://www.eurekalert.org/pub_releases/2017-09/osu-ro091817.php.
- Palmertz, Bjorn, *Theoretical foundations of influence operations: a review of relevant psychological research*, Swedish Defence University, 12-16, https://www.msb.se/Upload/Om%20MSB/Forskning/Kunskapsoversikt/Theoretical%20foundations%20of%20influence%20operations.pdf

Global Expertise. Local Solutions.
Sustainable Democracy.

IFES | 2011 Crystal Drive | 10th Floor | Arlington, VA 22202 | www.IFES.org