

# Combating Information Manipulation: A Playbook for Elections and Beyond

---



**Stanford** | Internet Observatory  
Cyber Policy Center

# Combating Information Manipulation: **A Playbook for Elections and Beyond**

---

September 2021



**Stanford** | Internet Observatory  
Cyber Policy Center

# Contents

<b>About Us</b> . . . . .	<b>1</b>	Fact-Checking . . . . .	32
<b>Acknowledgements</b> . . . . .	<b>2</b>	Social Media Platform Initiatives to Increase Access to Credible Information . . . . .	36
<b>Introduction</b> . . . . .	<b>3</b>	Facebook . . . . .	36
The Playbook Approach . . . . .	3	Twitter . . . . .	36
<b>Background: Understanding Information Manipulation</b> . . . . .	<b>4</b>	WhatsApp . . . . .	36
How To Use This Section . . . . .	4	Google . . . . .	36
What is Information Manipulation? . . . . .	5	YouTube . . . . .	37
Threat Actors . . . . .	5	Strategic Silence . . . . .	38
Content . . . . .	6	<b>Step 3: Build Resilience</b> . . . . .	<b>40</b>
Tactics . . . . .	6	A Whole-of-Society Approach for Resilience . . . . .	40
Vectors . . . . .	8	Public Awareness Campaigns . . . . .	42
Emerging Challenges for Information Manipulation . . . . .	9	Digital Literacy . . . . .	44
<b>Identifying, Responding to and Building Resilience against Information Manipulation</b> . . . . .	<b>10</b>	“Games to Discern” . . . . .	45
How to Use This Section . . . . .	10	Social Media Platform Digital Literacy Initiatives . . . . .	45
<b>Step 1: Identify</b> . . . . .	<b>11</b>	<b>Appendices</b> . . . . .	<b>47</b>
Mapping the Information Environment . . . . .	11	<b>Appendix A: Case Studies</b> . . . . .	<b>48</b>
Identifying Common Information Manipulation Narratives . . . . .	12	Mexico Case Study . . . . .	48
Identifying Ongoing Information Manipulation Efforts . . . . .	12	Background and Political Context . . . . .	48
Five Key Principles . . . . .	13	Information Manipulation in Mexico . . . . .	48
Developing a Workflow . . . . .	15	Interventions . . . . .	49
<b>Step 2: Respond</b> . . . . .	<b>16</b>	Lessons from Mexico for Civil Society’s Response to Information Manipulation . . . . .	50
Reporting . . . . .	16	Taiwan Case Study . . . . .	51
Reporting to Elections Management Bodies, Government Agencies and Law Enforcement . . . . .	17	Background and Political Context . . . . .	51
Reporting to Social Media Platforms . . . . .	20	Taiwan’s Whole-of-Society Response to Disinformation Campaigns . . . . .	51
User Reporting . . . . .	21	Lessons from Taiwan for a Whole-of-Society Response to Information Manipulation . . . . .	53
Other Ways to Engage with Platforms . . . . .	24	<b>Appendix B: Additional Information on Social Media Platforms</b> . . . . .	<b>54</b>
Engage with Platform Teams . . . . .	24	Overview of Social Media Platforms’ Policies . . . . .	54
Participate in Collaborative Cross-Industry Efforts . . . . .	26	Overview of Social Media Platforms’ Product Features and Interventions . . . . .	56
Strategic Communications . . . . .	27	<b>Appendix C: Additional Resources</b> . . . . .	<b>59</b>
Inclusive Communications . . . . .	30		

# About Us

---

## The International Republican Institute

The International Republican Institute (IRI) is a nonprofit, nonpartisan, nongovernmental organization committed to advancing democracy and freedom worldwide. IRI has supported civil society organizations, journalists, democratic governments and other democratic actors in more than 100 countries since 1983—in Africa, Asia, Eurasia, Europe, Latin America and the Caribbean, the Middle East, and North Africa. The IRI Technology and Democracy team works in every region of the world to help grassroots actors turn digitization and the technological revolution into a force for democratic progress, including a diversity of programs aimed at countering and building resilience to information manipulation worldwide.

## The National Democratic Institute

The National Democratic Institute for International Affairs (NDI) is a nonprofit, nonpartisan, nongovernmental organization that responds to the aspirations of people around the world to live in democratic societies that recognize and promote basic human rights. Since its founding in 1983 as one of the four core institutes of the National Endowment for Democracy, NDI and its local partners have worked to support and strengthen democratic institutions and practices by strengthening political parties, civic organizations and parliaments, safeguarding elections, and promoting citizen participation, openness and accountability in government. NDI is the leading organization working on implementing a diverse range of programs with critical information and communications technology (ICT) components, targeting democratic institutions and supporting democrats in general, particularly its INFO/tegrity initiatives supporting responses countering disinformation, hate speech and other harmful forms of content while promoting information integrity worldwide.

## The Stanford Internet Observatory

The Stanford Internet Observatory is a cross-disciplinary program of research, teaching and policy engagement for the study of abuse in current information technologies, with a focus on social media. The Stanford Internet Observatory was founded in 2019 to research the misuse of the internet to cause harm, formulate technical and policy responses, and teach the next generation how to avoid the mistakes of the past.

# Acknowledgements

## Authors

This Playbook was authored by Daniel Arnaudo, Samantha Bradshaw, Hui Hui Ooi, Kaleigh Schwalbe, Amy Studdart, Vera Zakem and Amanda Zink.

## Acknowledgements

We thank the many individuals from around the world who supported us in the development of this Playbook, including for their contributions as interviewees, roundtable participants and as peer-reviewers to this Playbook. We would also like to thank Renée DiResta, Elena Cryst, Josh Goldstein, Shelby Grossman, Sarah Moulton and Moira Whelan for their feedback and guidance on this report.

We also thank the National Endowment for Democracy for its support in the Playbook's development.



## Introduction

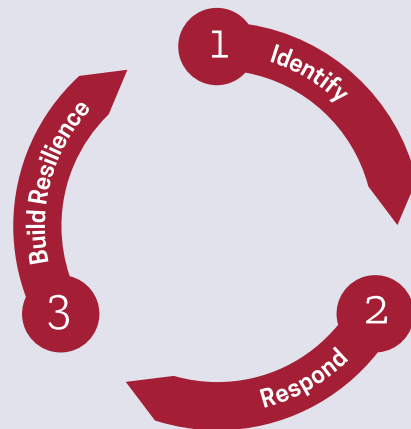
Over the past few years, the International Republican Institute (IRI), the National Democratic Institute (NDI) and the Stanford Internet Observatory (SIO) have observed efforts to undermine election-related information integrity in every corner of the world. Without concerted efforts to identify, respond to, and build long-term resilience to election-related information manipulation, attacks on information integrity threaten to delegitimize elections globally, reduce faith in elected governments, polarize societies and weaken democracies writ large.

Dealing with information manipulation around an election is a new and unfamiliar phenomenon for many countries. Civil society actors, journalists, governments, election management bodies and other democratic actors often end up scrambling to respond in the lead-up to an election. To address this challenge, IRI, NDI and SIO have joined forces to create this playbook, intended to help leapfrog the first six months of the electoral preparation process. The playbook lays out the basics of the problem and the core elements of a response, and points to trusted resources for those looking to do a deeper dive into a particular type of intervention or threat.

We hope this playbook will enable you and everyone dedicated to defending democracy to push back against efforts that undermine free and fair political competition. Since information manipulation is an ongoing challenge, this playbook will also be useful outside of an election cycle.

## The Playbook Approach

The playbook approach consists of how to (1) **identify** ongoing information manipulation campaigns; (2) develop real-time and short-term **responses**; and (3) **build long-term resilience** to information manipulation. While we outline three distinct steps in this playbook, the process for combating information manipulation is circular, with each step overlapping and reinforcing the others. Planning timelines will vary based on context, but—if at all possible—we encourage proactive rather than reactive planning to effectively counter electoral information manipulation. The playbook's three-part strategy can help you develop rapid and real-time responses, as well as establish long-term and sustainable approaches to building resilience in order to maintain the integrity of elections and strengthen democratic processes.



# Background: Understanding Information Manipulation

---

## How To Use This Section

This section lays out the components of information manipulation and defines commonly used terms.

## What is Information Manipulation?

Information manipulation is **a set of tactics involving the collection and dissemination of information in order to influence or disrupt democratic decision-making**. While information manipulation can co-opt traditional information channels—such as television broadcasting, print or radio—we focus on the digital aspects of information manipulation. Here, we explore how information manipulation campaigns co-opt different digital **vectors**, are led by different **actors**, and use a variety of **tactics** to distribute different kinds of **content**.

### Threat Actors

In most election environments, a number of different actors will likely engage in information manipulation. To make things more complicated, while some of those actors may operate independently, others may operate in coordination, be at cross-purposes, or benefit from the general chaos and lack of trust in the information environment. Different actors have different goals for engaging in information manipulation. A political campaign will focus on winning an election; the influence industry and commercial public relations firms want to make money; a foreign adversary might try to influence the election outcome, advance national interests, or sow chaos; or an extremist group might focus on advancing their political cause. Here, we have outlined common threat actors involved in information manipulation campaigns. Though this list is not exhaustive, it provides a starting point for thinking about the relevant actors in your own country's context.

- **Political parties and campaigns** use information manipulation to discredit the opposition, use false amplification to reach a wider audience or suggest that they have more public support than they do, or manipulate political discourse in a way that serves their campaign agenda. It is important to note that political campaigns can make use of information manipulation both outside of and during election cycles.
- **Hate and other extremist groups** use information manipulation to advance their social or political agenda, often by fomenting hate and political polarization; silencing, intimidating or otherwise disenfranchising target groups; and inciting violence. Their goals can include turning the majority electorate against a particular group, increasing support for extremist policies, and/or suppressing political participation.
- **Foreign governments** use information manipulation as a tool of statecraft and geopolitics. Information manipulation might be used to influence the outcome of an election in a strategically important country, advance the interests of the government or shape public perception of the state abroad. Information manipulation can be both covert (e.g., through the use of fake accounts) or overt (e.g., through state-backed media).
- **Domestic governments** use information manipulation to influence public attitudes and suppress the political participation or expression of certain users, such as activists, journalists or political opponents. Like foreign states, governments use both overt and covert information manipulation to achieve political goals, including the repression of human rights. Domestic governments can also more readily enact censorship as a form of information manipulation.
- **Commercial actors**, composed of social media platforms, public relations companies or strategic communication firms, use information manipulation as part of a business model, working with other actors to spread disinformation for profit. The influence industry often works with political campaigns, governments or foreign states to support their particular goals.
- **Non-independent media** with a specific political agenda or economic interest, or who are backed by a government or other political actor, may use information manipulation to influence public attitudes in line with the goals of their backers.

Determining who is behind information manipulation can be difficult, especially when the goals of different actors might overlap. For example, a foreign state actor might amplify content produced by domestic hate groups or conspiracy theorists. At the same time, there are market incentives for producing mis/disinformation where generating virality can also generate income for users who create appealing content and place advertisements on their pages.



There are many ways to categorize the kinds of threat actors involved in information manipulation campaigns. The DFRLab’s [Dichotomies of Disinformation](#)<sup>1</sup> can help you think about the kinds of actors and motivations behind information manipulation in more detail.

## Content

Information manipulation makes use of a variety of content to influence, disrupt or distort the information ecosystem. This content can be used to influence public attitudes or beliefs, persuade individuals to act or behave in a certain way—such as suppressing the vote of a particular group of people—or incite hate and violence. Many types of content can be involved in information manipulation. Here we outline a few key terms used throughout the report and by other researchers, activists and practitioners who study and combat information manipulation.

- **Misinformation** is false, inaccurate or misleading information, regardless of the intent to deceive.
- **Disinformation** is the deliberate creation, distribution and/or amplification of false, inaccurate or misleading information intended to deceive.
- **Malinformation** takes truthful or factual information and weaponizes it for persuasion. For example, this might include content that was released as part of a hack-and-leak operation, where private messages are shared publicly with the goal of undermining an adversary.
- **Propaganda** is information designed to promote a political goal, action or outcome. Propaganda often involves disinformation, but can also make use of facts, stolen information or half-truths to sway individuals. It often makes emotional appeals, rather than focusing on rational thoughts or arguments. Propaganda can be pushed by other actors, but in this report, we focus specifically on state-sponsored propaganda.
- **Hate speech** is the use of discriminatory language with

reference to a person or group on the basis of identity, including an individual’s religion, ethnicity, nationality, ability, gender or sexual orientation. Hate speech is often a part of broader information manipulation efforts. It is particularly present in election contexts where the goal of the information manipulation is to polarize political discourse and/or suppress the political participation of a particular group.

There are many additional ways to categorize the types of content involved in information manipulation. For additional resources, see [Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making](#), commissioned by the Council of Europe and produced in cooperation with First Draft and Harvard University’s Shorenstein Center on Media, Politics and Public Policy.<sup>2</sup>

## Tactics

Information manipulation makes use of a variety of tactics to spread, amplify or target messages to different audiences on social media. Many of these tactics exploit the features of digital and social networking technologies to spread different kinds of content. While media manipulation is not new, digital tactics can change the scope, scale and precision of information manipulation in various ways. Here, we provide definitions for some of the key tactics that researchers, journalists, activists and platform companies have identified.

- **AI-generated technology** is used in information manipulation to create fake profiles or content. Artificial Intelligence (AI) technologies, like Generative Adversarial Networks (GAN), use machine learning “neural networks” to create images or videos that look like real people but are completely fake. This includes “deepfake” videos, which use AI technologies to create realistic-looking videos that are entirely false.
- **Manipulated visual content** is used in information manipulation to photoshop images or edit videos. This can involve so-called “cheap fakes,” which do not use AI-generated technologies, but rather alter videos with a lower level of technical sophistication.

<sup>1</sup> Emerson T. Brooking, *Dichotomies of Disinformation*, (Atlantic Council’s Digital Forensic Research Lab, February 2020), <https://github.com/DFRLab/Dichotomies-of-Disinformation>.

<sup>2</sup> Claire Wardle, and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking* (Council of Europe, October 31, 2017), <https://shorensteincenter.org/information-disorder-framework-for-research-and-policymaking/>.

- **Search engine manipulation** uses tools from digital advertising—such as keyword placement—to exploit gaps in search results. These strategies attempt to place disinformation at the top of search engine queries, so that individuals looking for accurate information are more likely to come across disinformation.
- **Fake websites** are used to create the substance behind an influence manipulation campaign by creating “fake news” websites or content farms that publish large amounts of false, misleading or inaccurate stories, sometimes counterfeiting real news organizations.
- **Trolling** is the bullying or harassing of individuals to provoke an emotional reaction in the target. While anyone can be trolled online, certain communities experience trolling differently—and often more severely. This includes women, individuals with diverse gender identities, racial or ethnic minorities, or people of color.
- **Computational propaganda** involves the use of “bots” and other forms of automated technologies to amplify propaganda and other harmful content online. Bots are pieces of code designed to mimic human behavior by liking, sharing, retweeting or even commenting on posts. They can be used to falsely amplify certain kinds of content or accounts online.
- **Fake or “sock puppet” accounts** involve accounts, run by real people, who generate inorganic engagement. Like bots, fake or sock puppet accounts can like, share, retweet or comment on posts to falsely amplify certain kinds of content or accounts online. But rather than being automated, fake or sock puppet accounts are run by real people.
- **Hack-and-leak operations** involve hacking into private or sensitive information sources and strategically leaking information to the public in order to undermine the trust or integrity of a person or idea.

- **Account takeovers** involve hacking into the accounts of real people in order to impersonate them or spread mis/disinformation to large audiences.
- **Advertising and microtargeting** involve using online advertising platforms to collect data about users and targeting them with persuasive messaging.
- **Censorship** involves blocking, redirecting or throttling access to certain kinds of information online.

Many types of tactics can be used to manipulate the digital information ecosystem. These tactics will depend on the platform being used, the skillset of those involved and the unique country context where information manipulation is taking place. For further information about the kinds of tactics used in information manipulation campaigns, see the [USAID Disinformation Primer](#).<sup>3</sup>



### Coordinated Inauthentic Behavior and Information Operations

Social media platforms are trying to take more steps to combat information manipulation on their platforms. When they look for information manipulation, they use terms like Facebook’s “Coordinated Inauthentic Behaviour” or Twitter’s “information operations.” Although the terms for information manipulation and their tactics differ across platforms, social media platforms are increasingly taking action against networks of fake accounts that spread disinformation, incite violence or undermine the integrity of elections. One core component of platform definitions around information manipulation is the use of inauthentic or sock puppet accounts that pretend to be someone they are not—like a foreign state actor pretending to be a citizen of another country.

<sup>3</sup> USAID Center of Excellence on Democracy, Human Rights and Governance. “Center of Excellence on Democracy, Human Rights and Governance Disinformation Primer.” February 2021. [https://pdf.usaid.gov/pdf\\_docs/PA00XFKF.pdf](https://pdf.usaid.gov/pdf_docs/PA00XFKF.pdf).

Since 2018, Twitter and Facebook have released more data about information manipulation on their platforms in their company blogs or on [Twitter's Information Operation Transparency Center](#).<sup>4</sup> You can also find links to resources that can help you identify coordinated information manipulation campaigns in the Step 1: Identify section of this report. However, it is important to note that the definitions used by platforms to take down information manipulation have limitations. For example, when networks of real users share misleading information about an election, it can be much harder for platforms to take action against authentic users rather than inauthentic accounts. This is why it is important to additionally respond and build resilience to information manipulation through fact-checking, media literacy and the establishment of collaborative networks so that real users are less susceptible to sharing harmful, inaccurate or misleading information. You can read more about these strategies in the Step 2: Respond and Step 3: Build Resiliency sections of this report.

## Vectors

The information ecosystem has dramatically changed over the past three decades. The internet and social media in particular have created an environment in which information manipulation is massively scalable, very cheap and easy to experiment with.

While **popular social media platforms**— such as Facebook, Twitter and YouTube—are often highlighted as vectors for information manipulation, these activities also occur on **other social media platforms** like Reddit, Pinterest, Instagram, TikTok, Tumblr and WeChat. They also occur across **encrypted and non-encrypted messaging platforms** like LINE, Telegram, WhatsApp, Facebook Messenger, Signal or Viber. (For more information on conducting ethical investigations in closed environments, see the Step 2: Respond section.) Some information manipulation might target **internet search providers**, like Google, Yahoo or Bing. Others might **target niche communities** of users, like gamers, through platforms such as Twitch, Xbox Live, or PlayStation Online. As the major social media platforms have increased their efforts to set limits around the spread of harmful content, new social media platforms have been created. Some of those platforms are focused on creating unmoderated environments, and others have put in place moderation policies that explicitly facilitate the speech of one ideology over another, often with a focus on niche or extremist political views.

Information manipulation almost always occurs both online and offline: television, radio, print, academia and other aspects of the information ecosystem can be involved. For instance, journalists or the media may amplify content created as part of an information manipulation campaign if that content has been shared by an important political figure, or if it is particularly sensational and likely to attract audiences. A sophisticated actor that engages in information manipulation may capture prominent news outlets or give grants to research entities to produce analysis that supports their objectives.

<sup>4</sup> Twitter Transparency Center, "Information Operations," (Twitter, n.d.), <https://transparency.twitter.com/en/reports/information-operations.html>.

## Emerging Challenges for Information Manipulation

Information manipulation is constantly adapting to changes in the media ecosystem. As social media companies become better at detecting and removing information manipulation from their platforms, threat actors have also learned to modify their strategies, tools and tactics. While early researchers were concerned about the use of political bots to amplify mis/disinformation on social media platforms, today the distinction between automated bot accounts and human curated content is becoming less clear. The rise of various commercial actors offering disinformation as a service also makes it harder for social media companies to detect information manipulation and take action against them, as trolls-for-hire are paid to pollute the information sphere. At the same time, the distinction between foreign information operations and domestic extremism or terrorism is becoming less clear, as foreign meddling has increasingly co-opted domestic narratives to amplify pre-existing racial, gender or political divisions.

While platforms have policies for removing coordinated *inauthentic* behaviour (CIB), as of August 2021, there are no clear guidelines for managing coordinated *authentic* behaviour. When mainstream and globally ubiquitous platforms like Facebook, Twitter, TikTok or YouTube take action against content and accounts, sometimes these voices reappear on other platforms or private channels that lack the same standards for removing content or accounts that spread mis/disinformation or hate speech, or that incite violence. And some platforms, like WhatsApp or Signal, encrypt personal and group messages, making it much harder to detect information operations and counter the spread of mis/disinformation and other forms of harmful information. While encryption can protect the privacy and security of online activists and human rights defenders, malign actors have leveraged the security of these platforms to enhance the spread of harmful or misleading information online.

At the same time, not everyone experiences information manipulation the same way. Journalists, political activists and

members of the political opposition are frequently the targets of smear campaigns and harassment designed to undermine their credibility and legitimacy as professionals. These campaigns are often more severe for women, minorities, or people of color, who face greater levels of harassment, online threats and sexualization. Minority or marginalized populations are also often targets of online violence that can have real-world implications for their security and safety, as online speech can affect offline political violence.

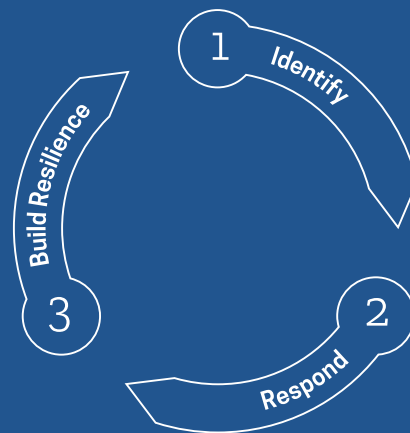
Finally, technology itself is constantly evolving, and new innovations are creating new opportunities for information manipulation. Artificial Intelligence is creating numerous opportunities for online deception: automated bot accounts can use machine learning algorithms (like GPT-3) to sound more human; generative adversarial networks (GANs) can be used to create fake profile pictures that look like real people, or other forms of synthetic media like “deepfake” videos. For example, deepfake videos are being used to falsely depict women in pornography, which can have damaging and lasting impacts on their mental health and career prospects. Innovations in data surveillance also introduce new challenges for information manipulation as it becomes much easier to target specific communities or individuals with persuasive messaging. Data about user likes and interests can be used to predict the values and behaviors of individuals or groups, and commercial actors are already building models to target communities of people with (de) mobilization messages. The data that can be used in information manipulation will only grow as the Internet of Things introduces more data points about users, from wearable devices to smart cars, appliances and sensors. We must evolve our responses to keep pace with these innovations and to build resilience to future information manipulation.

# Identifying, Responding to and Building Resilience against Information Manipulation

---

## How to Use This Section

Once you understand the basic aspects of information manipulation, the next critical step is to develop an understanding and skillset to identify future risks and ongoing operations in your own country's context. Identifying ongoing campaigns, as well as future risks, is one of the most difficult steps, because malign actors will often obscure their identity and create barriers for technical attribution. To help you with this process, this section provides a few key strategies. We have also compiled a list of useful and accessible resources to assist you throughout the process of identification.



## Step 1 Identify

### Mapping the Information Environment

The first step to identifying information manipulation is to map the information environment in order to identify the unique vulnerabilities for your election. You should conduct a risk assessment that identifies the various threat actors who might launch an information manipulation campaign and the channels—including digital, broadcast, radio or print—that might be used as part of their efforts. You will also want to identify various partners you will work with to combat threats that arise, such as social media company policy representatives, government officials, law enforcement agencies or other civil society organizations (CSOs). This section provides an overview of key questions to answer in order to map the information environment.

#### ● What is the media and information landscape?

The first step of *mapping the information environment* is to understand your current media landscape. Where do people get their political information? Where is information manipulation likely to take place? Here, you should consider traditional media entities, like television broadcasters, newspapers and radio stations, and assess the transparency of media ownership, correction policies, and the professional standards that media adhere to. You should also consider digital media, such as social media platforms, encrypted chat applications or web forums. Familiarize yourself with the policies of platforms where you suspect information operations could take place by reviewing their terms of service agreements and community guidelines as well as other country-specific measures that might have been announced in company blogs. You should also make yourself familiar with existing fact-checking initiatives and the role of other online influencers in shaping political discourse for certain communities of users.

#### ● Where are online audiences, and which communities of users are more vulnerable to information manipulation or negative implications from these campaigns?

Information manipulation affects users differently, and women, people of color, and people with diverse gender

identities and sexual orientation experience information manipulation more severely than others.<sup>5</sup> The second step of *mapping the information environment* is to understand your audiences and the groups of people who might be marginalized, suppressed or deeply affected by ongoing information manipulation efforts. This involves looking closely at small and local communities in your country's context.

#### ● Who are the likely threat actors?

The third step of *mapping the information environment* is to identify the various threat actors and understand their motivations for conducting information manipulation. Understanding who is, or could be, behind information manipulation will help you respond and build resilience to future operations. Ask yourself: Who are the main threat actors—are they domestic actors, foreign states or both? What could be the motivations behind these campaigns—are they for political disruption or economic gain? See the Background section for more information on threat actors and their motivations for conducting information manipulation.

#### ● Who are partners you can work with to combat information manipulation?

The fourth step of *mapping the information environment* is to identify partners who can assist you with combating information manipulation. Here, you should consider identifying relevant government and non-government partners, such as election management bodies, who may be able to assist in responding to ongoing information manipulation campaigns. Journalists and other CSOs in your country can work with you to fact-check or provide counter-messaging to information manipulation. You should also identify contacts at social media platforms who you can work with to take content or accounts off social media. It is important to note that not every partner will be relevant for every context. What is important is to identify who will bring skills, resources or capacity to assist you in responding and building resilience to information manipulation. To help you identify relevant partners, you can explore the Countering Disinformation Guide Intervention Database.<sup>6</sup>

<sup>5</sup> National Democratic Institute, Tweets that Chill: Analyzing Online Violence Against Women in Politics (NDI, June 14, 2019), <https://www.ndi.org/tweets-that-chill>.

<sup>6</sup> International Foundation for Electoral Systems, International Republican Institute, National Democratic Institute, "Database of Informational Interventions" (Consortium for Elections and Political Process Strengthening, 2021), <https://counteringdisinformation.org/index.php/interventions>.



- **What are the relevant regulations you should be aware of in your research and for reporting?**

Each country will have different laws or regulations relating to elections, campaigns and online speech. The fifth step of *mapping the information environment* involves developing an understanding of your country's relevant legal and regulatory landscape around issues related to the electoral information environment. Knowing these rules will prepare you to better work with government agencies in responding or building resilience to information manipulation, or to protect yourself and your colleagues and organization when deciding how best to respond. For example, some governments might have regulations that prevent campaigning three days before an election. There might also be rules or regulations in place about foreign ad purchases. Knowing these relevant regulations can help you report unlawful content to regulators and election management bodies, as well as to social media platforms for removal.



**Tip: Don't Forget Regional and Local Platforms**

In this playbook we list the major global social media platforms, but be aware that there are many other regional and local platforms as well. Cross-platform sharing of manipulated content is common, and you will observe information shared on one platform being shared across others. We advise that you do a deep dive on your local information ecosystem and observe which social media platforms are widely used and how they relate to each other. In addition to your own observation, a useful resource to assist in mapping the information landscape is the [We Are Social](https://wearesocial.com/digital-2020)<sup>7</sup> reports on social media use by country. The [Global Cyber Troops inventory](https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/)<sup>8</sup> also provides an overview of information manipulation by country and the various vectors used in information manipulation campaigns.

## Identifying Common Information Manipulation Narratives

After understanding the information environment you will be working in, you can start to think about the kinds of narratives or themes different actors might use in their information manipulation campaigns. We have outlined common narratives used in information campaigns during elections to assist you with identification.

- **Polarizing and divisive content** is sometimes used in information manipulation campaigns to inflame political, racial, religious, cultural or gender divides. These narratives often focus on pre-existing divisions within society, and use identity-based narratives to sow discord and discontent among the electorate.
- **Delegitimization narratives** spread content that undermines the integrity of the electoral process. This could be false claims about the security of voting machines, errors in ballot casting or tabulation, or other alleged irregularities. These narratives are designed to sow distrust in the processes that support elections. Delegitimization narratives can also focus on discrediting certain politicians or candidates, election officials, or civic entities.
- **Political suppression** narratives are used to discourage certain groups of people from participating in politics. These suppression strategies target democratic processes; this could include spreading disinformation about how and where to vote, or suggesting that certain communities of individuals are not allowed to vote or that there is violence at polling stations. They could also include narratives that pressure voters to attend, or not attend, political rallies or events, or that encourage voter fraud.
- **Hate, harassment and violence** is another form of suppression that uses harassment, slander or threats of violence to discourage certain users or communities from expressing their thoughts or opinions online or engaging in

<sup>7</sup> We Are Social, "Digital in 2020" (2020), <https://wearesocial.com/digital-2020>.

<sup>8</sup> Samantha Bradshaw, Hannah Bailey, and Philip Howard, "Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation," Computational Propaganda Research Project (Oxford Internet Institute, January 13, 2021), <https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/>.

debates necessary for a well-functioning democracy. Hate, harassment and violence create a culture of fear and can stifle political expression online.

- **Premature election results or claims of victory** are sometimes made on social media to erode trust in the outcome of the election. They are often made before ballot counting has been completed, and are especially likely if a political race is close and contentious.

Many kinds of narratives can emerge during an election, and many of them will be specific to your country's context. It is important to think about the kinds of narratives that might be used as part of information manipulation campaigns so that you can be more prepared to respond with counter-messaging or build resilience to narratives before they spread. The Computational Propaganda Project's [Global Cyber Troops Inventory](#)<sup>9</sup> describes other kinds of narratives or "communication strategies" that have been observed in information manipulation campaigns around the world.

## Identifying Ongoing Information Manipulation Efforts

Once you have mapped the information environment and understand the different kinds of narratives actors might use to undermine election integrity, you should begin monitoring the ecosystem for ongoing campaigns. Threat actors will often try to conceal their identity or their campaigns in order to avoid detection. However, there are a number of resources and best practices available for identifying ongoing campaigns, which we have compiled for you. You should also consider these five key principles when conducting investigations into information manipulation.

### Five Key Principles

- 1. Context Matters.** Every country and election will occur in a different media, cultural, social and economic environment. It is important to map your information ecosystem and the likely threats in order to focus on the relevant technologies and platforms that are prominently used in your country.
- 2. Know Your Limits.** Any research into information manipulation comes with limitations, and the data we collect about these kinds of campaigns are always imperfect. It is important to understand what you do and don't know about information manipulation based on the data you are working with, and not jump to conclusions about the authenticity of online information. It can be just as damaging to the legitimacy of an election if information manipulation is misattributed.
- 3. Behavior Over Content.** When identifying information manipulation, it is important to look at patterns in behaviour of accounts, rather than looking at a single piece of content. Platforms are better able to respond to coordinated inauthentic behaviour, and identifying large networks of accounts in coordination to manipulate the online information environment will provide a stronger basis for taking down content and accounts.
- 4. Do No Harm.** The collection, storage and use of online data can have implications for personal privacy and security, and it is important that any information collected to monitor information manipulation is conducted in an ethical way. Online data can come with expectations of privacy, and thinking about consent, security and privacy is an important part of your job as an investigator. Data that is improperly stored or anonymized can have negative consequences for personal privacy or the security of users who are participating in politics online. Thus, it is important to take necessary steps to do no harm, and to protect and secure any data you're working with.
- 5. Zero Tolerance for Hate, Suppression and Violence.** Hate or incitement to violence online can have real-world consequences not only for the integrity of elections, but for the security of citizens. These narratives do not always come from coordinated inauthentic accounts or as part of formalized information operations but might be shared by authentic or real users. However, any information that spreads hate, attempts to suppress political participation or speech, or incites violence should be immediately reported to the platforms and other relevant parties regardless of the source. More information on how to report content can be found in the Reporting section of Step 2: Respond.

<sup>9</sup> Bradshaw, Bailey, and Howard, "Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation."



### Tip: Identifying Information Manipulation Offline

Keep in mind that information manipulation campaigns can occur offline as well as through online platforms; mainstream media (such as television, radio and newspaper) are common vectors for the spread of false information, and you should keep the above principles in mind when consuming information from offline sources as well. Be sure to

verify the information that you hear or see before you share it with your trusted networks and members of your organization. Since mainstream media often has intentional biases, it can sometimes feature false information or “half-truths” (see Digital Literacy in the Step 3: Build Resilience section, page 44).

## Open Source Intelligence (OSINT) Tools for Identifying Information Manipulation

OSINT is the collection and analysis of information from public (open) sources. These resources can be used for tracking and identifying disinformation.



### **Bellingcat’s Online Investigation Toolkit:**

**(Resource List)** This easy-to-navigate Google Doc spreadsheet has different tabs for different tools for verifying information, such as image and video verification; social media content and accounts; phone numbers and closed messaging services; maps and location-based services; transport trackers; IP and website analysis; international companies; environment; tools for improving online security, privacy and data visualization; academic resources; and additional guidebooks.<sup>10</sup>



### **Data Journalism’s Verification Handbook for Disinformation and Media Manipulation:**

**(Guide)** This handbook helps you conduct OSINT research into social media accounts, bot-detection and image manipulation. It also provides resources for conducting investigations on the web and across platforms, as well as some tips and tools for attribution.<sup>11</sup>



### **The Beacon Project’s Media Monitoring Handbook:**

**(Guide)** This handbook helps you conduct data-driven analyses of disinformation narratives and their sources. The handbook is a good starting place for researchers interested in conducting media monitoring, but are not sure where to start, as well as those looking to ensure methodological best practices are being applied.<sup>12</sup>



### **CrowdTangle: (Tool)**

Facebook created CrowdTangle as a tool for identifying and monitoring trends on social media. The tool can track verified accounts, Pages and public Groups. The tool can also be used to monitor public accounts on Instagram and subreddit threads on the Reddit platform.<sup>13</sup>

You should review these resources, as well as additional tools in Appendix C on page 59, to determine what tools will be most useful to you and your organization for identifying information manipulation. Every campaign, organization and country context will be different and require a mix of tools, skills and partners, so developing an understanding of the tools that can help you identify and monitor ongoing campaigns will empower you to respond and build resilience.

<sup>10</sup> “Bellingcat’s Online Investigation Toolkit” (Bellingcat, 2021), <https://docs.google.com/spreadsheets/d/18rtqh8EG2q1xBo2cLNyhlDuK9jrPGwYr9DI2UncoqJQ/edit#gid=930747607>

<sup>11</sup> Craig Silverman, ed., Verification Handbook for Disinformation and Media Manipulation (European Journalism Centre n.d.), <https://datajournalism.com/read/handbook/verification-3/>.

<sup>12</sup> The Beacon Project, “Media Monitoring Handbook” (International Republican Institute, August 2021), <https://www.data-iribeaconproject.org/handbook/>.

<sup>13</sup> CrowdTangle (Facebook, n.d.), <https://www.crowdtangle.com>.

## Developing a Workflow

When tracking information manipulation, you will need to develop both short- and long-term monitoring strategies. When developing a workflow, you should consider:

- **What are your goals or main objectives?** Are you trying to reduce the impact of disinformation by fact-checking narratives? Or are you trying to build accountability around malign actors engaging in disinformation? Your goals will directly shape the scope of your monitoring, as well as the kinds of tools and partners you work with.
- **What is the scope** of your monitoring? Determining scope involves asking questions like:
  - What is the internet penetration in your country, and will social media be a source of information during elections?
  - What platforms are in scope for monitoring?
  - What are the biggest threats to electoral integrity as it relates to disinformation?
  - Who are the potential actors involved in information manipulation?
  - What election-related themes are considered in-scope for your monitoring, what languages will you work in, and what issues will be outside the scope of your investigations?
- **What tools will you use** to assist you with identification and monitoring?
- When gathering data about influence manipulation across digital, print or broadcast media, **how will data be collected, labeled and stored** to make analysis and triage more accessible to you and your organization? The process of tracking and monitoring influence campaigns can take weeks or even months, and establishing a system that allows for collection over time will be critical to your success.
- **Who will be responsible for monitoring** the information ecosystem? How will they work and **how will they be trained** in order to have a consistent approach in identifying influence manipulation online?

- Are there certain time periods that will require you and your organization to **ramp up monitoring activities**, such as before an election or important political referendum?

For more resources on developing your workflow and thinking about the scope of your identification processes, see [The European Union Guide for Civil Society on Monitoring Social Media During Elections](https://www.ndi.org/sites/default/files/social-media-DEF.pdf).<sup>14</sup>

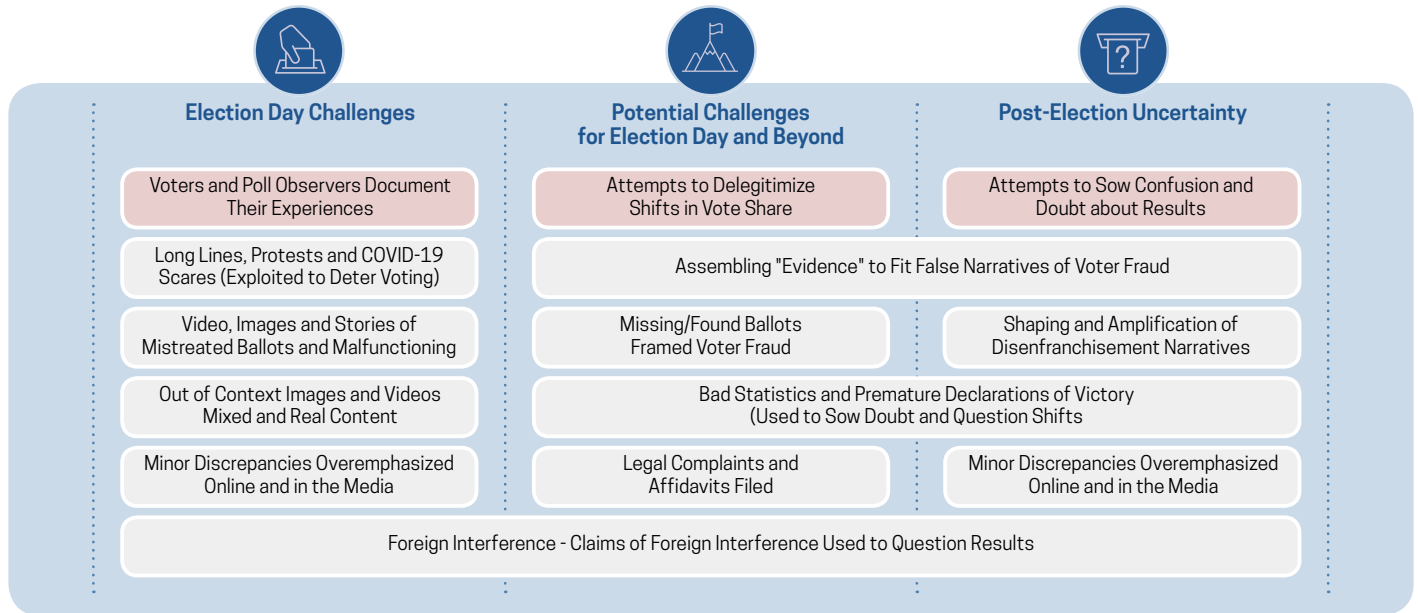
<sup>14</sup> *Guide for Civil Society on Monitoring Social Media During Elections* (European Union, August 2019), <https://www.ndi.org/sites/default/files/social-media-DEF.pdf>.

## Step 2 Respond

In this playbook, respond means reacting quickly to harmful election-related online activity. No matter how strong your defenses and emphasis on prevention might be, the reality is that those defending information integrity will always be playing catch-up. As such, it is critical to additionally focus on identification and

response in order to act swiftly and effectively against elections-related information manipulation once it occurs. This chapter will cover responses including reporting to elections management bodies (EMBs), government agencies, law enforcement and social media platforms; strategic communications; fact-checking; and strategic silence.

**Figure 1. What to Expect: Challenges for Election Day and Beyond**



**Source:** Model adapted from the Election Integrity Partnership's "What to expect on election night and days after." The initial graphic addressed a U.S. case study, which we have adapted to apply in a global context.<sup>15</sup>

## Reporting

Information manipulation can be reported to elections management bodies (EMBs), government agencies, law enforcement, social media platforms, international non-governmental organizations (INGOs), fact-checking organizations, or organizations that represent the issue area or targeted community.

Each entity has different, and sometimes overlapping, roles in responding to information manipulation. Social media platforms can investigate and take steps to reduce the spread of misinformation and hate speech; governments and election commissions can create legal frameworks that limit the capacity of malign actors to engage in information manipulation, as well as

launch information campaigns to share accurate information or debunk inaccurate information; fact-checking organizations can investigate the veracity of a claim and publicly debunk information manipulation; INGOs can work with local partners to ensure that concerns are taken seriously and that the capacity exists to address the situation; and organizations working on issue areas or with targeted communities can take steps to protect their communities and/or contribute to debunking efforts. The most effective efforts at reporting will likely involve engaging with a number of different partners on the basis of the local context and the specifics of the observed information manipulation. Note that responding to information manipulation is challenging for CSOs, EMBs or activists to approach alone; governments and technology companies must also step up to meet the challenges.

<sup>15</sup> Kate Starbird et al., "Uncertainty and Misinformation: What to Expect on Election Night and Days After," (Election Integrity Partnership, October 26, 2020), <https://www.eipartnership.net/news/what-to-expect>.

This section will help you develop an understanding of who plays what role, how to most effectively report information manipulation, and what you can expect once you report. It is important to understand that the guidance we provide may not work for every type of actor or information environment. When choosing the best tactics to follow, you need to consider your country's context, the types of relationships or partners you already have, and your group's mission, technical skills and expertise. For instance, not all groups will have the skills to conduct fact-checking or be able to report information manipulation to a government that is itself the source of the manipulation.

After going through the suggestions under **Step 1: Identify**, you should now consider the goals you would like to achieve by reporting information manipulation.

- Have the content taken down?
- Have users or pages banned?
- Prompt an investigation into coordinated inauthentic behavior or other violations of platform terms of service?
- Raise more attention and awareness on a specific event, trend or threat actor?
- Advocate for the government and social media platforms to take preemptive steps?

Once you consider the above questions, you will be better able to select which entities are most appropriate to report observed information manipulation to. You can report a similar issue or violation to several entities at once. We have grouped potential entities into three general categories:



### Government

Check if your government, elections commission, or other cyber or information agencies have a mis/disinformation reporting system. If yes, you should consider reporting the violating content to them based on the consideration of a number of factors (see page 17).



### Social Media Platforms

If the content or behavior is in violation of its hosted social media platform's policies and terms, you can report the content to the relevant platform (see page 20).



### Fact-Checkers

Consider sharing observed mis/disinformation to fact-checking groups in your country or region (see page 32).

## Reporting to Elections Management Bodies, Government Agencies and Law Enforcement

Most democratic countries have an Elections Management Body (EMB), Elections Commission, Elections Council or Elections Board<sup>16</sup> that oversees implementation of the elections process, as well as government agencies and law enforcement that help uphold the elections-related regulations of any country.

Many EMBs do not have resources, structures or mechanisms in place to address election-related information manipulation or to protect themselves and the country's elections from electoral information manipulation narratives. For those who do, very few have reporting mechanisms created for citizens to report elections-related information manipulation observed online.<sup>17</sup> Furthermore, EMBs typically do not have the mandate to develop rigorous regulations around online campaigning nor the ability to enforce existing regulations. However, some EMBs have created disincentives to deter malign actors from taking part in electoral information manipulation by establishing campaigning codes of conduct and collaborating with social media platforms to regulate the behaviors of political parties and electoral candidates.

<sup>16</sup> The formal names vary by country and their electoral models can be independent, mixed, judicial or executive, but we will use the term Elections Management Body (EMB) in this playbook.

<sup>17</sup> For example, Israel's Central Election Commission's website provides contact numbers for police hotlines and the National Center for Cyber Incidents and Information Security for citizens to report attempts to manipulate voters through fake profiles and the like, instead of handling and investigating the violations itself.



If you would like your EMB to explore solutions to disincentivize electoral information manipulation or create codes of conduct, you should directly advocate to your EMB. However, keep in mind that some EMBs are not independent bodies, so they may not be impartial in the structures and policies they enact or the actions they take against violators.



### Available Resources for EMBs

The International Foundation for Electoral Systems (IFES) has a number of resources and programs designed to help election commissions or management bodies effectively preempt and respond to information manipulation. If you work for an EMB or would like to learn more about the role that EMBs can play in responding to election-related information manipulation, see the Consortium for Elections and Political Process Strengthening (CEPPS)<sup>18</sup> Countering Disinformation Guide's [section](#)<sup>19</sup> on EMB approaches to countering disinformation.

In addition, some governments have created agencies to combat cyberattacks and other digital threats, some of which also function to safeguard electoral infrastructure, e.g., the United States' Cybersecurity and Infrastructure Security Agency (CISA) and Indonesia's National Cyber Encryption Agency (BSSN). Check to see if these types of agencies exist in your country and if they have set up citizen reporting mechanisms for elections-related information manipulation. If not, consider advocating to your elected government officials to put these in place. Review

the CEPPS Countering Disinformation Guide's [Advocacy Toward Governments section](#)<sup>20</sup> for guidance and examples of how a CSO might advocate for its government to take action. It is important to be aware that authoritarian regimes have frequently cracked down on freedom of expression through newly developed cybersecurity agencies and legislation, using laws that identify opposition content as harmful "fake news or misinformation."

If you or your organization work on elections campaigns, you need to carefully consider the legal frameworks, law enforcement, independent oversight bodies and other regulatory agencies that are or could be involved in the social media and larger information space. In certain cases, local or federal police agencies have teams specifically dedicated to enforcement of laws online. When they are honest brokers and can be relied upon, these law enforcement bodies form viable avenues for reporting and oversight of harmful campaigns. The judicial system can also play a role in governing the online space and can order measures to stop the dissemination of information manipulation online.

Independent oversight bodies might be an anti-corruption agency, a political finance oversight body, or a media oversight body. In the aforementioned [legal and regulatory section](#)<sup>21</sup> of the CEPPS Countering Disinformation Guide, four kinds of regulatory approach are outlined in more detail that address both platforms and domestic actors, including measures to restrict online content and behaviors and promote transparency, equity and democratic information during campaigning and elections. These are all potential avenues that may present useful foci for policy advocacy, which are explored in more detail in the Guide, but should be carefully considered for each given national context. In other circumstances, particularly when the government does not respect democratic norms or is otherwise compromised, law enforcement or regulatory bodies can often play an

<sup>18</sup> Established in 1995, CEPPS pools the expertise of three international organizations dedicated to democratic development: the International Foundation for Electoral Systems (IFES), the International Republican Institute (IRI), and the National Democratic Institute (NDI). CEPPS has a 25-year track record of collaboration and leadership in democracy, human rights, and governance support, learning from experience, and adopting new approaches and tools based on the ever-evolving technological landscape. The groups operate as a consortium to provide USAID and other donors with the capacity to deliver complex democracy, rights and governance (DRG) programming at scale across the entire spectrum of political contexts and geographic regions.

<sup>19</sup> USAID and National Democratic Institute, "Election Management Body Approaches to Countering Disinformation" in *Countering Disinformation: A Guide to Promoting Information Integrity*, (Consortium for Elections and Political Process Strengthening, 2021), <https://counteringdisinformation.org/topics/embs/0-overview-emb-approaches>.

<sup>20</sup> USAID and National Democratic Institute, "Building Civil Society Capacity To Mitigate And Counter Disinformation" in *Countering Disinformation: A Guide to Promoting Information Integrity*, (Consortium for Elections and Political Process Strengthening, 2021), <https://counteringdisinformation.org/topics/csos/5-advocacy-toward-governments>.

<sup>21</sup> USAID and National Democratic Institute, "Legal and Regulatory Responses to Disinformation" in *Countering Disinformation: A Guide to Promoting Information Integrity*, (Consortium for Elections and Political Process Strengthening, 2021), <https://counteringdisinformation.org/node/2704/>.

actively harmful role, developing and applying laws that limit legitimate political speech in the name of combating information manipulation. It is more problematic and often counterproductive to report to those actors.

Countries that are generally free (see the Tools to Evaluate a Government's Openness box) can create legal frameworks for reporting and responding to information manipulation in ways that protect and open the space for democratic discourse, while guarding against information manipulation. If a country is not free or partly free, you should proceed with caution in engaging with regulatory, judicial or other governmental organizations. At all levels of freedom, a country's oversight bodies may be weak or ineffective in this domain, even in strong democracies. You should evaluate these bodies and survey regulations carefully, and potentially engage experts and review resources on their efficacy, trustworthiness and ambition.

If the rule of law or transparency of these bodies is low, other options (detailed below) should be considered. For additional guidance on reporting to law enforcement bodies, reference the [CEPPS Countering Disinformation Guide's section on Legal Frameworks and Enforcement](#).<sup>25</sup>

As you seek to report election-related information manipulation to government agencies, elections commissions and other law enforcement bodies, some critical factors to consider are:

- Do you trust these agencies to act **impartially**?
- Does your government have the **capacity and capability** to take action on the reported content?
- Have there been **actions taken** by your government agency upon receiving reports of harmful online content leading to the takedown of those posts?
- Does your government **regulate speech online** or have laws against pre-elections smear campaigns, character attacks and information manipulation campaigns? Would those laws



### Tools to Evaluate A Government's Openness

One measure of a government's political posture is Freedom House's [Freedom in the World index](#),<sup>22</sup> which measures whether a country is free, partly free or not free based on factors such as political rights and civil liberties. Freedom House's [Freedom on the Net Index](#)<sup>23</sup> surveys a government's regulations related to the internet, alongside various other factors, as it assesses how open or closed a country's national internet is, as well as the structure of the government agencies that regulate it and any relevant laws. It assigns each country a score that helps to classify the countries based on a number of factors, references and further details that should be reviewed as a component of an assessment of the information space and legal framework. These rankings are updated annually, but they need to be weighed in with the current situation in your country, for it may change quickly. Another measure is the [Varieties of Democracy \(V-Dem\) Institute](#),<sup>24</sup> which has established a robust, multidimensional dataset that accounts for democracy's complex systems and allows users to assess how a particular democracy has fared over time. These represent just three methods of gauging a country's openness and respect for the rule of law, but understanding this component is a critical early step.

be used against opposition voices and political parties running for elections?

- What is the general **political posture** of the government? Is it generally open and democratic, or trending authoritarian?

<sup>22</sup> "Freedom in the World Index" (Freedom House, 2021), <https://freedomhouse.org/report/freedom-world/2021/democracy-under-siege/countries-and-regions>.

<sup>23</sup> "Freedom on the Net Index" (Freedom House, 2021), <https://freedomhouse.org/countries/freedom-net/scores>.

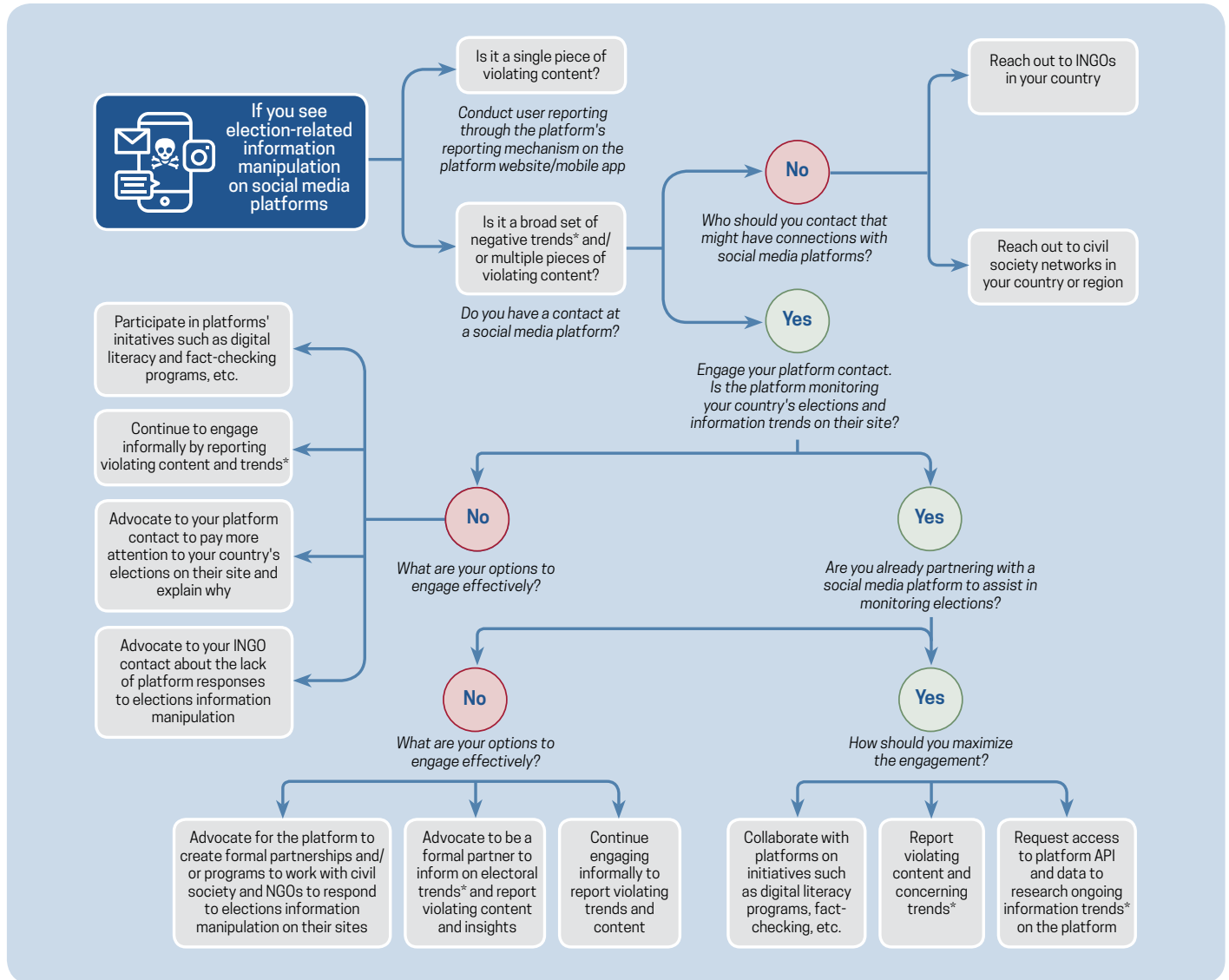
<sup>24</sup> "V-Dem: Global Standards, Local Knowledge" (Varieties of Democracy, n.d.), <https://www.v-dem.net/en/>.

<sup>25</sup> USAID and National Democratic Institute, "Legal and Regulatory Responses to Disinformation" in *Countering Disinformation: A Guide to Promoting Information Integrity*. <https://counteringdisinformation.org/topics/legal/6-enforcement#EnforcementMandate>.

- Does your government have a **history of suppressing opposition** voices and criticisms, especially ahead of elections?
- Are there government agencies that are themselves **agents of information manipulation** (both foreign and domestic)?

For a comprehensive list of examples of actions taken by governments around the world to defend against misinformation, including efforts ranging from legitimate attempts by democratic governments to ensure information integrity, to efforts by authoritarian regimes to censor speech they do not like, visit Poynter’s [Guide to Anti-Misinformation Actions Around the World](#).<sup>26</sup>

## Reporting to Social Media Platforms



\* If you are an individual or group conducting social media monitoring to identify and report information manipulation trends, see page 15 for detailed instructions on short- and long-term monitoring strategies. For more specific guidance on using Facebook's CrowdTangle data to identify trends, refer to First Draft's resource [here](#).<sup>27</sup>

<sup>26</sup> Daniel Funke and Daniela Flamini, "A Guide to Anti-Misinformation Actions Around the World," (Poynter, April 13, 2021), <https://www.poynter.org/ifcn/anti-misinformation-actions/>.

<sup>27</sup> Carlotta Dotto, "How to Analyze Facebook Data for Misinformation Trends and Narratives." (First Draft, May 7, 2020), <https://firstdraftnews.org/articles/how-to-analyze-facebook-data-for-misinformation-trends-and-narratives/>.

Social media platforms and online web hosts track and respond to information manipulation through a variety of mechanisms: user reporting, partnerships with civil society to identify local trends and risks, engagement with experts, back end in-house and external threat intelligence and digital forensics, cross-industry coordination, and engagement with governmental organizations. Depending on what type of organization you are and what you wish to communicate to the platforms, some of these options will be more relevant to you, and others less so.



**Tip: Don't Make Reporting to Social Media Platforms Your Only Approach**








Note that reporting violating content to social media platforms is a necessary but insufficient step. Platforms are unlikely to be responsive to user reporting in a timely manner, and it can take days or weeks to receive a response. Content that you find threatens electoral integrity might not be against a platform's policy or community standards and could be left up. Often, platforms are also unprepared to handle a country's electoral information environment on their sites or do not understand the local information space and threats. Hence, reporting to platforms should not be your only step and needs to be taken together with our other recommended actions. Social media platforms are rapidly evolving and require all those working on improving electoral information integrity to constantly monitor and, in some cases, advocate for product and policy changes and adjust their strategies in interacting with platforms.

The methods that platforms undertake to handle reporting about information manipulation, counter mis/disinformation, moderate content, and collaborate internally and externally vary widely and depend on where the company was founded, how long it has been operating, its finances, and its relationships with external stakeholders and the governments, among other considerations.

## User Reporting

User reporting is the most accessible way of raising concerns about specific pieces of content that violate the policies of the social media platform on which the content is being shared. User reporting is usually as simple as flagging a specific piece of content within the platform and giving an explanation as to why it is harmful. Note that user reports are usually reviewed by automated systems, human content moderators, and—in rare instances—other units within a company, on the basis of whether they violate existing company standards or policies. Those processes frequently suffer from a lack of societal or political context and knowledge of local languages. User reporting is not an effective way to draw attention to concerning trends or to a large-scale information manipulation campaign. However, it is effective at removing single pieces of content or social media accounts that clearly violate platform policies. For more details about each platform's community policies and guidelines for how reportable content is defined and other elections-specific platform interventions, refer to Appendix B on page 55.

The table below provides guidance on key platforms' reporting processes. We listed the major social media platforms here due to their large number of users and global reach.

Platform	How to report
Facebook 	If you identify content and/or accounts on Facebook that you suspect are spreading harmful content ahead of elections, follow the links below. <ul style="list-style-type: none"> <li>● <a href="#">Mark a Facebook Post as False News</a><sup>28</sup></li> <li>● <a href="#">How to Report Things</a><sup>29</sup></li> </ul> Appeals can be referred to the Oversight Board. See the Facebook’s Oversight Board box on page 23.
Instagram 	To submit reports of harmful mis/disinformation around elections, go to the <a href="#">Reducing the Spread of False Information on Instagram</a> page. <sup>30</sup>
Google 	Google’s different products have individual terms of service that contain restrictions on hateful and misleading behavior and content, and Google’s processes for reporting mis/disinformation and other harmful content on its platform are also product specific. However, Google Search is most relevant in the case of this playbook. The tool for requesting removal of information from Google Search can be found on <a href="#">this page</a> . <sup>31</sup>
Snapchat 	To file a report of suspected elections-related mis/disinformation and other harmful content, use Snapchat’s <a href="#">in-app reporting feature</a> <sup>32</sup> or complete <a href="#">this form</a> <sup>33</sup> on its website.
TikTok 	To report a video, comment, user, hashtag, etc., suspected of mis/disinformation and other harmful content, see detailed instructions on TikTok’s <a href="#">Report a Problem</a> site. <sup>34</sup>
Twitter 	To report Tweets, Lists and Direct Messages that you suspect are spreading harmful content about your country’s elections, follow the instructions <a href="#">here</a> . <sup>35</sup> Twitter defines harmful content under its <a href="#">Twitter Rules</a> <sup>36</sup> that can help you understand what is off limits and reportable under its definitions.
YouTube 	To report mis/disinformation and other harmful content that appears on YouTube through its videos, playlist, thumbnail, comment, channel, etc., use its in-platform mechanism that can be found on the <a href="#">report inappropriate content page</a> . <sup>37</sup>
WhatsApp 	To report harmful content to WhatsApp, follow the instructions <a href="#">here</a> . <sup>38</sup> Note that WhatsApp is a closed, encrypted messaging app, so monitoring content on this app is different from the other social media platforms listed above.

<sup>28</sup> Facebook Help Center, “How do I mark a Facebook post as false news?” (Facebook, n.d.), <https://www.facebook.com/help/572838089565953>.

<sup>29</sup> Facebook Help Center, “How to Report Things” (Facebook, n.d.), <https://www.facebook.com/help/1380418588640631>.

<sup>30</sup> Instagram Help Center, “Reducing the Spread of False Information on Instagram” (Instagram, n.d.), <https://help.instagram.com/1735798276553028>.

<sup>31</sup> Google Support, “Removing Content from Google” (Google, n.d.), <https://support.google.com/legal/troubleshooter/1114905>.

<sup>32</sup> Snapchat Support, “Report Abuse on Snapchat” (Snapchat, n.d.), <https://support.snapchat.com/en-US/a/report-abuse-in-app>.

<sup>33</sup> Snapchat Support, “Contact Us” (Snapchat, n.d.), <https://support.snapchat.com/en-US/i-need-help>.

<sup>34</sup> TikTok Support, “Report a Problem” (TikTok, n.d.), <https://support.tiktok.com/en/safety-hc/report-a-problem>.

<sup>35</sup> Twitter Help Center, “Report a Tweet, List, or Direct Message” (Twitter, n.d.), <https://help.twitter.com/en/safety-and-security/report-a-tweet>.

<sup>36</sup> Twitter Help Center, “The Twitter Rules” (Twitter, n.d.), <https://help.twitter.com/en/rules-and-policies/twitter-rules>.

<sup>37</sup> YouTube Help Center, “Report Inappropriate Content” (YouTube, n.d.), <https://support.google.com/youtube/answer/2802027>.

<sup>38</sup> WhatsApp Help Center, “How to Stay Safe on WhatsApp” (WhatsApp, n.d.), <https://faq.whatsapp.com/general/security-and-privacy/staying-safe-on-whatsapp>.

Besides information presented in the table above, you can also find the Mozilla Foundation’s [detailed instructions](#)<sup>39</sup> on in-app reporting. Be aware that the response time and investigative

processes differ by platforms. Typically, after a platform user reports violating content through the platform’s online system, **the review process flows as below:**



The platform’s automated system checks for obvious violations within the reported content—child pornography, known slurs, etc.—and removes violating posts.



If the automated system is unable to provide definitive answers, then the reported violation is reviewed by content moderators, who may or may not be conversant in the local context and language.



If content moderators find the posts to be in violation of the platform’s policies and community standards, the post is removed. In cases where it is not clear, the post is escalated to other teams within the company, e.g., the policy team, trust and safety team, etc.



Depending on the severity and political impact of the issue, the specific post may be shared with the leadership of the company for more detailed considerations and review before a decision is made.



### Facebook’s Oversight Board

Facebook created the [Oversight Board](#) to help it answer some of the most difficult questions around freedom of expression online, what to take down, what to leave up and why.<sup>40</sup> The Oversight Board also provides an appeals process for people to challenge content decisions on Facebook or Instagram. If you already requested that Facebook or Instagram review one of its content decisions and you disagree with the final decision, you can appeal to the board. The process is detailed [here](#).<sup>41</sup> Not all submitted cases will be selected to undergo the appeals process, and the timeline of the process is fairly lengthy.

### A Growing Trend: Information Manipulation in Closed Groups and Encrypted Messaging Applications

As platforms update their policies on information manipulation and increase and improve their efforts on content moderation and removal, malign actors are increasingly moving their information manipulation efforts to sites that are more difficult to monitor, specifically

closed groups and encrypted messaging applications such as Facebook Groups, WhatsApp, Telegram, Signal, LINE and WeChat. Many of these apps are encrypted, and there is no effective way to monitor or proactively remove the spread of malign information. To counter harmful forms of content

<sup>39</sup> Audrey Hingle, “Misinfo Monday: How to Report Election Misinformation” (Mozilla, October 12, 2020), <https://foundation.mozilla.org/en/blog/misinfo-monday-how-report-election-misinformation/>.

<sup>40</sup> Oversight Board, “Ensuring Respect for Free Expression, through Independent Judgment” (Facebook Oversight Board, n.d.), <https://oversightboard.com>.

<sup>41</sup> Oversight Board, “Appealing Content Decisions on Facebook or Instagram” (Facebook Oversight Board, n.d.), <https://oversightboard.com/appeals-process/>.



like disinformation, some apps have updated products and policies. For example, WhatsApp set up message forwarding limits to help “slow down the spread of rumors, viral messages and fake news.”<sup>42</sup> Journalists and researchers have attempted to report from encrypted messaging apps by joining closed groups and setting up tip lines to encourage the public to send in content. However, these methods also pose many challenges for those who attempt to report violating content from encrypted messaging apps, particularly ethical challenges.<sup>43</sup> Others from civil society have launched public awareness campaigns by sharing accurate information on WhatsApp. The effect of these types of campaigns remains to be seen, and results are difficult to measure. However, there have been some successful efforts to combat information manipulation on closed messaging platforms. In Taiwan, a collaboration

between LINE and Cofacts allows for viral messages to be fact-checked by volunteers and debunked in chat without intrusions on privacy. In Spain, the fact-checking organization Maldita.es added an automated chatbot to its existing WhatsApp tipline in July 2020 to improve response time and build a database to track misinformation trends.<sup>44</sup>

If you’d like more information on how to monitor and report inside closed groups and messaging apps, refer to the European Journalism Centre’s Verification Handbook<sup>45</sup> (specifically Chapter 7), First Draft’s Essential Guide on closed messaging apps and ads,<sup>46</sup> and Brookings Institution’s policy brief on Countering Disinformation and Protecting Democratic Communication on Encrypted Messaging Applications.<sup>47</sup>

## Other Ways to Engage with Platforms

### Engage with Platform Teams

Most platforms have a variety of teams that can serve as touch points in the lead-up to, during and after elections. Those teams have a variety of incentive structures, roles, and interests, and are sometimes unaware of each other. Some may be locally based, while others are based in regional hubs or at company

headquarters. The table below gives a broad overview of the roles related to elections and information manipulation that may exist in a given company. Note that identifying the right staff can be difficult. Some of these roles may be occupied by the same team or person, and newer platforms—even those with a large user base or impact on the information space—may have limited field presence, country representatives, or staff in these functions.

<sup>42</sup> WhatsApp Help Center, “About Forwarding Limits” (WhatsApp, n.d.), <https://faq.whatsapp.com/general/chats/about-forwarding-limits/>.

<sup>43</sup> Connie Moon Sehat, Tarunima Prabhakar, and Aleksei Kaminski, *Ethical Approaches to Closed Messaging Research: Considerations in Democratic Contexts* (MisinfoCon and The Carter Center, March 15, 2021), <https://www.dropbox.com/s/rkchyrtdkn5buw9/FINAL-Ethical-Approaches-to%20Closed-Messaging-Research.pdf?dl=0>.

<sup>44</sup> Harrison Mantas, “WhatsApp Can Be a Black Box of Misinformation, but Maldita May Have Opened a Window” (Poynter, June 9, 2021), <https://www.poynter.org/fact-checking/2021/whatsapp-can-be-a-black-box-of-misinformation-but-maldita-may-have-opened-a-window/>.

<sup>45</sup> Silverman, *Verification Handbook for Disinformation and Media Manipulation*.

<sup>46</sup> Carlotta Dotto, Rory Smith, and Claire Wardle, “Closed Groups, Messaging Apps & Online Ads” (First Draft, November 2019), [https://firstdraftnews.org/wp-content/uploads/2019/11/Messaging\\_Apps\\_Digital\\_AW-1.pdf?x11129](https://firstdraftnews.org/wp-content/uploads/2019/11/Messaging_Apps_Digital_AW-1.pdf?x11129).

<sup>47</sup> Jacob Gursky and Samuel Woolley, *Countering Disinformation and Protecting Democratic Communication on Encrypted Messaging Applications* (Brookings Institution, June 2021), [https://www.brookings.edu/wp-content/uploads/2021/06/FP\\_20210611\\_encryption\\_gursky\\_woolley.pdf](https://www.brookings.edu/wp-content/uploads/2021/06/FP_20210611_encryption_gursky_woolley.pdf).

Platform Teams	Roles
<b>Public Policy and Government Relations</b>	<p>The public policy and government relations teams are usually responsible for engaging with regulatory agencies and other government bodies. Their predominant role is to ensure a favorable regulatory environment for the platform. Companies tend to have public policy representatives located in the capitals of countries that are significant markets. Public policy teams can be good entry points for those attempting to address information manipulation, but you should be aware that they have multiple competing priorities and incentives—particularly in instances where a national government is a bad actor in the information space—and thus may not always see dealing with information manipulation as consistent with their role. Some companies have community outreach/partnerships teams that specifically engage with civil society, advocacy groups and academia.</p>
<b>Content Policy/ Human Rights</b>	<p>Established social media platforms who have faced significant issues relating to online harm will generally have a variety of teams working on mitigating that harm. Those teams will oversee the creation of policies that determine what is and is not allowed on the platform; develop and track specific policies related to human rights; and, often, develop partnerships with civil society to help inform the company’s approach to content and user behaviour.</p> <p>Some platforms have dedicated teams explicitly charged with ensuring election integrity. In some instances, those teams are permanently in place, but in other instances they may be established on a temporary basis to respond to a specific election of significance to the platform.</p>
<b>Content Moderators</b>	<p>Often contractors, not employees within the company, content moderators review user-reported content and decide if it aligns with the platforms’ policies and community standards. These contractors do not have the scope to change platforms’ standards.</p>
<b>Product</b>	<p>Product teams are responsible for launching platforms’ products and improving or changing products to prevent platform abuse and limit the spread of mis/disinformation (i.e., limiting forwarding of messages on WhatsApp, etc.).</p>
<b>Threat Intelligence</b>	<p>Researchers who conduct deep dive investigations into threats manifesting on a platform typically look at coordinated inauthentic or other types of behavior.</p>

If you are not already in touch with the platforms of relevance to your information space, most global social media companies have established partnerships with major INGOs or coalitions, such as the [Design 4 Democracy Coalition](https://d4dcoalition.org),<sup>48</sup> and other local or regional civil society networks, who can work with you to ensure you are communicating with the best point of contact within each company.

<sup>48</sup> Design for Democracy Coalition, <https://d4dcoalition.org>.



### Tip: An Introduction to the Design 4 Democracy Coalition

The [Design 4 Democracy Coalition \(D4D\)](#), led by NDI, IRI, IFES and International IDEA, is an international group of democracy and human rights organizations, from a diverse collection of regions, political ideologies and backgrounds, that is committed to ensuring that the technology industry embraces democracy as a core design principle. By developing a forum for coordination and support within the democracy community on technology issues, and by creating an institutional channel for communication

between the democracy community, civil society organizations and the tech industry, D4D is working to strengthen democracy in the digital age. The Coalition could serve as a useful resource for you; [contact information](#) is available on the D4D website. The D4D coalition also developed the TRACE Tool, a form that allows you to request access to training or tools provided by D4D's technology partners, or to flag content or profile issues that need to be addressed through expedited means.<sup>49</sup>

## Participate in Collaborative Cross-Industry Efforts

Most of the major online platforms have established processes and programs for partnering with civil society, independent media and academia on issues related to information manipulation. Those include mechanisms for gaining insight into local context and linguistic issues; formal partnerships with fact-checkers, journalists and civil society; and rapid escalation channels for select groups in crisis situations. Partners can preemptively provide contextual knowledge and flag issue areas and potential events that might be subject to information manipulation and cause real-life violence. This contextualization enables platforms to take immediate actions either by removing content or by

taking proactive actions in modifying their products, policies and resources to prevent the platform's facilitation of violence or anti-democratic behavior. This tactic is particularly effective for CSOs, journalists or activists who are victims of state-sponsored disinformation and harassment.

These mechanisms are continuously adapted and may or may not be active in your country. The INGOs detailed under the D4D coalition above can help you navigate which programs are active in your country and how you can participate in those programs. You can find examples of successful initiatives in **Step 3: Build Resiliency** on page 40.

## How social media companies approach information manipulation

Social media platforms typically prefer not to rely only on users reporting election-related mis/disinformation spreading on their platforms. According to the CEPPS [Countering Disinformation](#) guide,<sup>50</sup> platforms have enacted policies, product interventions and enforcement measures

to limit the spread of mis/disinformation. Most platforms also have some form of content moderation tools in place; you can find an inventory of these tools in the [Toolkit for Civil Society and Moderation Inventory](#)<sup>51</sup> developed by [Meedan](#).<sup>52</sup>

<sup>49</sup> Design for Democracy Coalition, <https://d4dcoalition.org>; Design for Democracy Coalition, "Contact Us" (D4D, n.d.), <https://d4dcoalition.org/index.php/contact-us>; Design for Democracy Coalition, "D4D Trace Tool" (D4D, n.d.).

<sup>50</sup> USAID and National Democratic Institute, *Countering Disinformation: A Guide to Promoting Information Integrity*.

<sup>51</sup> Kat Lo, *Toolkit for Civil Society and Moderation Inventory* (Meedan, November 18, 2020), <https://meedan.com/reports/toolkit-for-civil-society-and-moderation-inventory/>.

<sup>52</sup> Meedan, <https://meedan.com>.

Platforms also limit the spread of election-related mis/disinformation through the design and implementation of **product features and technical or human interventions**. This is highly dependent on the nature and functionality of specific platforms—traditional social media services, image and video sharing platforms, messaging applications, and search engines. Both Twitter and Facebook use automation<sup>53</sup> for detecting certain types of mis/disinformation and enforcing content policies. The

companies similarly employ technical tools to assist in the detection of inauthentic activity on their platforms and then publicly disclose their findings in periodic transparency reports that include data on account removals. You can find more details about different types of platforms' efforts to limit the spread of mis/disinformation through product features and technical/human intervention in Appendix B on page 57 or in the CEPPS Countering Disinformation Guide's topical section on platforms.<sup>54</sup>

## Strategic Communications

Communications in response to or in preparation for election-related information manipulation are typically broken down into two approaches: proactive and responsive.

- **Proactive Communication:** This approach aims to provide accurate, reliable, consistent and concise information regarding an election *before* any false narratives take hold, in an effort to create a trustworthy information space for citizens.
- **Responsive Communication:** This approach aims to counter false narratives once they have already gained traction, frequently involving directly identifying a false narrative and its objectives and responding to those inaccuracies with the truth (See the *Fact-Checking* section on page 32).

In advance of developing a strategic communications campaign, and before the election itself, take the time to consider common mis/disinformation and false narratives that inevitably emerge before, during and after elections so you will be prepared with positive, strategic communications well in advance. Common information manipulation campaigns include content that spreads confusion about voting procedures or technical processes,

such as false voting times; content that may result in direct voter suppression, such as falsely reporting unsafe voting environments or inefficient or closed voting locations; and content that may delegitimize the election, such as narratives of widespread voter fraud, broken voting infrastructure or large-scale conspiracy theories.<sup>55</sup>

As you consider potential false narratives, begin planning for and agreeing to ready-made, well-crafted narratives and messaging that can be sustainably and quickly deployed throughout the election. Keep the best practices listed below in mind as you plan your strategic communication campaign to be sure messaging is clear from the outset and remains consistent and easy to understand throughout the election.

- Carefully **research your audience** and plan your campaign.
  - Who is your audience(s)?
  - What is the purpose of your message?
  - What will resonate with your audience? How can you build inclusive messaging?
- Clearly define—and remain consistent with—your **messaging goals**.

<sup>53</sup> @Vijaya and Matt Derella, "An Update on Our Continuity Strategy during COVID-19," Twitter Company (blog) (Twitter, April 1, 2020), [https://blog.twitter.com/en\\_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19](https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19).

<sup>54</sup> USAID and National Democratic Institute, "Platform Specific Engagement for Information Integrity" in *Countering Disinformation: A Guide to Promoting Information Integrity*, (Consortium for Elections and Political Process Strengthening, 2021), <https://counteringdisinformation.org/node/2722/>.

<sup>55</sup> Election Integrity Partnership, "Election Official Handbook: Preparing for Election Day Misinformation" (October 20, 2020), <https://www.eipartnership.net/news/how-to-prepare-for-election-day-misinformation#Common%20Narratives>.

- **Create content.** Concentrate on messaging that provides action steps for your audience. What are you asking them to do?
- **Select the platforms and tactics** to share your counter-narrative. Consider accessibility and inclusivity in your platform choice; diversify communication channels to include those that are accessible in low-bandwidth areas; and use graphics, songs, theater plays or other innovative communication methods for those who may not be literate.
- **Evaluate the impact.** Continue to conduct research and monitor dialogue while keeping your original communication goals in mind. If conditions change, consider adapting your message to ensure your original communication goals are met.

While not all information manipulation necessitates a response (see the *Strategic Silence* section on page 38), there will be myriad false narratives that require attention as they gain public notice and influence. Any public communications regarding information manipulation—whether proactive or responsive in nature—should uphold **truthfulness, openness, fairness and accuracy**.<sup>56</sup> To communicate about disinformation effectively, you should address:

- **Timeliness.** Speed is critical in countering information manipulation effectively; this means developing protocols for strategic communications that balance speed and accuracy, with clear guidelines on necessary approvals and communication steps. The longer disinformation goes unanswered, the more likely it will be effective.
- **Messaging.** All communications should be accurate, values-driven and compelling enough to compete (see *Why Does Disinformation Go Viral?*). Your messaging should be empathetic to concerns and follow “Easy Read” protocols described here.<sup>57</sup> Additional guidance on developing a compelling messaging campaign can be found in the *Co/Act Toolkit*.<sup>58</sup>

- **Avoid Accidental Amplification.** If communications are directly countering a falsehood, the message must be framed in a way that ensures amplification of the truth rather than accidentally attracting more attention to the falsehood. Framing an unproven assertion between two truths better emphasizes accurate information rather than simply stating it.
- **Partnering/Networks.** Often other groups or networks have the same interests, and working together increases efficiency and strengthens the credibility of the information when shared by multiple sources. Consider partnering with existing networks or influencers—those with large numbers of followers who are able to reach broad groups of the population—to amplify your messages and build bridges with skeptical audiences (see here<sup>59</sup> for an example of Finland’s use of influencers to spread true information about elections).

### Why Does Disinformation Go Viral?

To effectively out-communicate fake news, it’s important to understand how and why it often spreads so quickly. In the era of social media, the extent to which mis/disinformation gains traction is often linked to the emotional response the underlying narrative is able to evoke. Evoking emotions such as disgust, surprise, anger, fear and contempt can play a vital role in how quickly news is shared. Disinformation is often crafted in a way that plays into those emotions, capitalizing on vulnerabilities in how we form our opinions and exacerbating existing divisions and biases to encourage emotion to overwhelm reason or logic. Using humor, empathy, creativity, and interesting images or graphics in your communications may help truthful messages compete with disinformation.

<sup>56</sup> James Pamment et al., *RESIST: Counter-Disinformation Toolkit* (Government Communication Service, 2019), <https://3x7ip91ron4ju9ehf2unqrm1-wpengine.netdna-ssl.com/wp-content/uploads/2020/03/RESIST-Counter-Disinformation-Toolkit.pdf>.

<sup>57</sup> People First, “A Guide to Making Easy Read Information” (New Zealand: Office for Disability Resources, Ministry of Social Development, n.d.), <https://www.od.govt.nz/guidance-and-resources/a-guide-to-making-easy-read-information/>.

<sup>58</sup> Co/Act, *Human Centered Design for Activists* (Co/Act, National Democratic Institute, n.d.), [https://www.ndi.org/sites/default/files/Co\\_Act%20Toolkit.pdf](https://www.ndi.org/sites/default/files/Co_Act%20Toolkit.pdf).

<sup>59</sup> Jon Henley, “Finland Enlists Social Influencers in Fight Against Covid-19,” *The Guardian*, April 1, 2020, <https://www.theguardian.com/world/2020/apr/01/finland-enlists-social-influencers-in-fight-against-covid-19>.

Many studies have indicated that the best predictor of whether people will believe a rumor is the number of times they are exposed to it.<sup>60</sup> Using this same principle in promoting accurate news, communication campaigns should emphasize the repetition of a clear, focused message to most effectively spread the truth.<sup>61</sup> Focus communications on sharing what the government is doing to organize and prepare for the elections, refuting mis/disinformation, advancing the truth, and seeking to develop relationships with key audiences and constituencies.

Given the typically high volume of disinformation narratives surrounding elections, and the frequently limited capacity of democracy actors—including EMBs, CSOs, and even mass media sites—to dedicate resources to address this challenge, focus on countering the objectives of information manipulation campaigns, which are often aimed at exploiting existing divisions or changing public opinion about a political candidate or party, rather than countering individual narratives. EMBs, CSOs and other democracy actors should focus on proactive communication strategies, dedicating resources to promoting the truth rather than countering falsehoods.

In general, consider these steps as you plan your communications:

1. Identify the **key facts** related to the election that are most critical to continually reaffirm as true—consider the who, what, where, when and how of elections—and use your messaging to establish ground truth facts as much as possible.
2. Decide on the **most trusted information channels** and partners to help convey the message; provide clear messaging to them along with guidance to communicate the message.
3. As you regularly share your message, continue to **monitor media coverage**, including social media, and establish a feedback loop for how your messages are picked up and responded to.
4. **Modify your message** if conditions change (such as a shift on election day or outbreaks of violence) to demonstrate responsiveness, but be sure to maintain clear communications objectives and message consistency.

### Key factors fueling the infodemic

Citizens seek **clear, definitive information** in evolving, uncertain circumstances

**False and misleading information** is spread within **closed networks** (though this information is not necessarily trusted)

**Disinformation** is **increasingly well disguised** and **decreasingly challenged** by audiences with **low media literacy skills**

Citizens must **navigate** and **evaluate** an **overload** of often **conflicting information**



### Implications for tackling the infodemic via public communication initiatives

Providing clearer, more definitive information through official channels and established media outlets

"Pre-bunk" or warn about potential disinformation before it occurs, as part of communication and public information campaigns

Maintaining transparent communication about the situation, government action and risks to restore trust in public institutions and in the information and guidance they relay

Ensuring consistency, even if information is tentative, and messaging discipline across public authorities to speak with a single voice and reduce information overload

**Source:** Image adapted from the Organization for Economic Co-operation and Development's (OECD) report *Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new Coronavirus*.<sup>62</sup>

<sup>60</sup> Lisa Fazio, David Rand, and Gordon Pennycook, "Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements," *Psychonomic Bulletin & Review* 26, no. 5 (October 2019): 1705-1710, <https://doi.org/10.3758/s13423-019-01651-4>.

<sup>61</sup> Norbert Schwarz and Madeline Jalbert, "When (Fake) News Feels True: Intuitions of Truth and the Acceptance and Correction of Misinformation," *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation*, ed. Rainer Greifeneder, Mariela E. Jaffé, Eryn Newman, and Norbert Schwarz (Routledge: August 14, 2020), <https://library.oapen.org/viewer/web/viewer.html?file=/bitstream/handle/20.500.12657/46921/9781000179033.pdf?sequence=1&isAllowed=y>.

<sup>62</sup> "Transparency, Communication and trust: the role of public communication in responding to the wave of disinformation about the new Coronavirus," (OECD, July 3, 2020), <https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-the-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus-bef7ad6e/>.



For more thorough guidance on communications aimed at countering information manipulation, refer to the resources below.



The RESIST Counter Disinformation Toolkit Annex E: Strategic Communication (Tool) is a step-by-step guideline for deploying the FACT and OASIS models for effective strategic communication.<sup>63</sup>



Countering Information Influence Activities: A Handbook for Communicators, (Guide) published by the Swedish Civil Contingencies Agency, includes extensive guidance on how to choose the best communications response according to the information manipulation occurring.<sup>64</sup>



Information Manipulation: A Challenge for Our Democracies (Guide) offers useful case studies and suggestions based on previous strategic communication campaigns.<sup>65</sup>

## Inclusive Communications

It is important to adapt your messaging for different contexts and to intentionally work to reach diverse audiences. Marginalized social and issue-focused groups—including women, immigrants and minorities, among others—are often the primary targets of mis/disinformation attacks. Information manipulation campaigns often work to exploit existing socioeconomic divisions; voter suppression is often targeted at specific, vulnerable communities; and biases and prejudices are often amplified to sow discord, confusion and disenfranchisement. You can find more information on how to understand the gendered dimensions of disinformation in particular in the Gender and Disinformation chapter of the CEPPS Countering Disinformation guide.<sup>66</sup>

Given the often explicitly divisive goal of information manipulation campaigns, it is absolutely critical for communication campaigns to prioritize reaching marginalized and targeted groups and provide accurate information that empowers those groups to mitigate potential impacts. As such, we recommend that democracy actors implement a wide range of communication strategies before, during and after the election. Consider diversifying your approach by partnering with trusted networks and community leaders, and tailor your message to appeal to unique audiences; in doing so, you will reach a broader proportion of the population and more vulnerable groups, which will increase the chances of outcompeting inaccurate information in the spaces where it is most widely promoted.

No matter the diverse channels or messages you use, inclusion, accessibility and transparency should be at the forefront of your strategy. Text messages, radio and traditional media outreach may reach a wider part of the population in spaces where internet accessibility is expensive or low. Following Easy Read principles to reach lower literacy populations as well as creating content in multiple languages, including indigenous languages, is key to accessible and far-reaching communications.<sup>67</sup> Diversifying communications across platforms such as Facebook, Twitter, Instagram and WhatsApp, among others, will increase the segments of the population you're able to reach. Finally, aim to produce easily shareable content for maximally efficient communication, such as using easy to understand graphics, wherever possible.

<sup>63</sup> James Pamment et al., "Annex E: Strategic Communication" in RESIST: *Counter-Disinformation Toolkit* (Government Communication Service, 2019), <https://gcs.civilservice.gov.uk/publications/resist-counter-disinformation-toolkit/#Annex-E-Strategic-communication>.

<sup>64</sup> *Countering Information Influence Activities: A Handbook for Communicators* (Swedish Civil Contingencies Agency, March 2019), <https://www.msb.se/RibData/Filer/pdf/28698.pdf>.

<sup>65</sup> Jean-Baptiste Jeangène Vilmer et al., *Information Manipulation: A Challenge for Our Democracies*, report by the Policy Planning Staff (CAPS) of the Ministry for Europe and Foreign Affairs and the Institute for Strategic Research (IRSEM) of the Ministry for the Armed Forces (August 2018), [https://www.diplomatie.gouv.fr/IMG/pdf/information\\_manipulation\\_rvb\\_cle838736.pdf](https://www.diplomatie.gouv.fr/IMG/pdf/information_manipulation_rvb_cle838736.pdf).

<sup>66</sup> USAID and National Democratic Institute, "Understanding the Gender Dimensions of Disinformation" in *Countering Disinformation: A Guide to Promoting Information Integrity* (Consortium for Elections and Political Process Strengthening, 2021), <https://counteringdisinformation.org/topics/gender/O-overview-gender-disinformation>.

<sup>67</sup> People First, "A Guide to Making Easy Read Information." (Office for Disability Issues). <https://www.odi.govt.nz/guidance-and-resources/a-guide-to-making-easy-read-information/>.



### Case Study: Supporting Access to Independent Media for Low-Income Populations in Latin America

Over the last decade, IRI has conducted programs throughout Latin America to support independent media as they produce and disseminate independent, reliable news content in fast-paced and ever-changing sociopolitical contexts, often targeting lower-income sectors that are less likely to receive this information elsewhere. Through its work, IRI has found that grassroots approaches and the establishment of alliances between civil society and the media, as well as careful consideration of security concerns, have been critical to successfully engage

with a broad range of stakeholders to improve the dissemination of truthful news. If your organization is seeking to share truthful information in an accessible way to lower-income sectors, consider supporting a wide range of independent media outlets through a network of online, printed, radio, television and social media content, and establish a dissemination network using tools such as radio broadcasts, mass messaging, and targeted digital advertisements, or alternative methods such as public screenings, street theaters or humor shows.



### Tip: Avoid Amplifying False Information

Repeating false information in order to correct it can sometimes result in an increased belief in the false information. As such, your communications campaigns should focus on repeating accurate information without reference to the

information that you are trying to debunk. For instance, an effective campaign would say "Election Day is XXX. Only trust information from official sources," and not "Election Day is not XXX. Do not trust information from XXX."



### Case Studies of Communication Tactics

#### Case Study: Communications by the Independent National Electoral Commission of Nigeria

During electoral periods, the INEC of Nigeria provides daily televised briefings, participates in live TV interviews, and regularly issues press statements to explain the policies and actions of the commission.<sup>68</sup> This regular, proactive communication not only informs the general public of the activities of the Commission, but also creates transparency and trust around the election process. INEC's activities continue beyond electoral periods as well, with the creation of accessible

voter education resources that explain details of how and where to vote, how to register, and voters' rights and responsibilities. INEC also regularly generates a newsletter and press releases that transparently contain updates to election processes. These efforts have actively amplified the facts about electoral processes in Nigeria, helping to proactively counteract any information manipulation around Nigerian elections.

<sup>68</sup> Independent National Electoral Commission Nigeria, <https://www.inecnigeria.org>.



### Case Study: Taiwan's Fast Fair Fun Campaign

Key to Taiwan's success in managing the COVID-19 pandemic was a unique, and ultimately extremely effective, communications strategy that may be emulated in elections-related contexts. Rather than responsively countering or fact-checking false narratives about the virus, the Taiwanese government launched a communications campaign centered on three elements: fair, fast and fun.

- **Fast:** As soon as citizens began reporting outbreaks and concerns, the Taiwanese government immediately enacted policies to suspend travel to and from China, indicating trust and transparency between the state and its citizens. Additionally, facts, often shared by way of memes, were rapidly deployed to promote the truth before mis/disinformation could spread.

- **Fair:** In order to maximize transparency and full information, the state took action to make health-related data public, including data regarding mask supply. This equipped all citizens with access to critical information and ensured fair access for everyone.
- **Fun:** With a mentality of "humor over rumor," the state created communication campaigns using humor to dispel rumors about mask supply, how COVID-19 spreads, etc., including a "spokesdog" to deliver safety guidelines to the public in an approachable and entertaining way. This campaign demonstrated how factual humor spreads faster than rumor.

## Fact-Checking

Fact-checking is a process that seeks to verify information and provide accurate, unbiased analysis of a claim. Although fact-checking alone can be ineffective in protecting the integrity of the information environment surrounding an election—the direct impact of corrections is often very limited—it can prove useful in correcting key pieces of election-related information manipulation.

If your organization seeks to develop skills beyond reporting concerns to fact-checkers and would like to develop the capacity to regularly and sustainably conduct fact-checks yourself, consider the guidance below. As outlined in the [comprehensive guide developed by Poynter](#), fact-checking information related to elections is guided by one main question: "**How do we know that?**"<sup>69</sup>

<sup>69</sup> Alexios Mantzarlis, "Module 5: Fact-checking 101," in Journalism, "Fake News," and Disinformation (UNESCO, 2018), [https://en.unesco.org/sites/default/files/module\\_5.pdf](https://en.unesco.org/sites/default/files/module_5.pdf).

Generally speaking, fact-checking is composed of three steps:<sup>70</sup>



### 1 Find fact-checkable claims

by monitoring social media, mainstream media and political statements discussing election-related information to identify a dubious or incorrect claim that can be objectively verified. In selecting a claim, consider:

- a) How viral is the claim (What is its extent, reach and spread)?
- b) What is the source of the claim? (Who shared it?)
- c) What is the nature of the claim? (Can it lead to violence? Is it provocative?)



### 2 Find the facts

once you select a claim, by gathering the best available evidence regarding the claim, being sure to evaluate the reliability of your sources. Available tools for this step include:

- a) **Google Image Search** to determine the origin of photos or videos.<sup>71</sup>
- b) **TinEye Reverse Image Search** to determine how long and how often an image has been available and how it has been edited.<sup>72</sup>
- c) **Google Fact Check Explorer** to find existing fact-check results regarding a person, topic or issue.<sup>73</sup>
- d) **Amnesty International YouTube DataViewer** to determine if a video or parts of a video have been previously uploaded online.<sup>74</sup>
- e) **The Global Disinformation Index** to find the probability of disinformation on a specific media outlet.<sup>75</sup>

You can find a robust list of additional resources [here](#).<sup>76</sup>



### 3 Correct the record

by evaluating the claim in light of the best available evidence, usually on a scale of truthfulness: true, mostly true, half true, mostly false, false, and pants on fire.

<sup>70</sup> “How to Fact-Check Like a Pro” (The Public Library of Albuquerque and Bernalillo County, n.d.), <https://abqlibrary.org/FakeNews/FactCheck>.

<sup>71</sup> Google Images (Google, n.d.), <https://images.google.com>.

<sup>72</sup> TinEye Reverse Image Search (TinEye, n.d.), <https://tineye.com>.

<sup>73</sup> Google Fact Check Explorer (Google, n.d.), <https://toolbox.google.com/factcheck/explorer>.

<sup>74</sup> YouTube DataViewer (Amnesty International, n.d.), <https://citizenevidence.amnestyusa.org>.

<sup>75</sup> Global Disinformation Index (GDI, n.d.), <https://disinformationindex.org>.

<sup>76</sup> “Tools That Fight Disinformation Online” (RAND Corporation, n.d.), <https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html>.

For more robust guidance and training content to develop skills to effectively identify, evaluate and fact-check claims, and to train others in how to fact-check as well, consider referencing the following resources:

- [First Draft’s free library of training content](#), including online courses, toolkits and resources.<sup>77</sup>
- [Learn to Discern \(L2D\) Curriculum and Online Game](#) to strengthen media and information literacy skills.<sup>78</sup>
- [Trust and Verification](#), a free online course, teaches how to build trust as a journalist or content creator in an age of information manipulation.<sup>79</sup>

Fact-checking itself is an imperfect strategy. As your organization considers evaluating the veracity of facts in the hopes of countering or debunking false information, keep in mind the biases, not only of your audience but also of yourself, that might influence perceptions of the truth. Furthermore, if you suspect false fact-checkers might be clouding the information space, reference the [International Fact-Checking Network’s fact-checker’s code of principles](#) to determine if the behavior of the possibly false fact-checker is untrustworthy.<sup>80</sup>

### Partnering with the Media

Important to any election protection efforts is working with local media outlets to provide authoritative content in advance of and during elections and ensuring that they are part of your partnership team to counter, prebunk and debunk false narratives and content. As such, ensuring that independent media are equipped with best practices to identify, respond to and expose false narratives and are able to quickly provide authoritative content will support the overall mission of countering information manipulation during elections. There are a number of resources for the media, including free training courses available on [First Draft’s website](#).<sup>81</sup>

<sup>77</sup> First Draft Training (First Draft, n.d.), <https://firstdraftnews.org/training/>.

<sup>78</sup> Learn to Discern, “Media Literacy Training” (IREX, n.d.), <https://www.irex.org/project/learn-discern-l2d-media-literacy-training>.

<sup>79</sup> Craig Silverman, Instructor, “Trust and Verification in an Age of Misinformation,” online course (Knight Center for Journalism in the Americas) <https://journalismcourses.org/es/course/trustandverification/>.

<sup>80</sup> International Fact-Checking Network, “Fact-Checkers’ Code of Principles” (Poynter, September 15, 2016), <https://www.poynter.org/ifcn-fact-checkers-code-of-principles/>.

<sup>81</sup> First Draft, “Training Resources,” <https://firstdraftnews.org/training/>.

### Collaborative Fact-Checking Efforts

While journalists are often the leading actors in fact-checking efforts, many of the most successful fact-checking initiatives have been the result of collaboration across stakeholder groups; CSOs, NGOs and even EMBs can complement journalists' efforts by acting as reliable sources of information and offering supplementary expertise. Below are examples of fact-checking collaborations that involve multiple democratic actors.

- **StopFake** is a fact-checking organization founded by Ukrainian professors and students to identify and investigate fake information about events in Ukraine.<sup>82</sup>
- **Africa Check** is Africa's first independent, non-profit organization covering Kenya, Nigeria, Senegal and South Africa, analyzing important public statements and publishing fact-checking reports to guide public debate.<sup>83</sup>
- **Chequeado** is a nonpartisan, nonprofit organization dedicated to the verification of public discourse and to countering mis/disinformation. Chequeado convenes all stakeholder groups in their efforts to combat mis/disinformation.<sup>84</sup>
- **The International Fact-Checking Network (IFCN)** is a unit of the Poynter Institute that convenes fact-

checkers worldwide and that actively promotes best practices and exchanges in this field, in addition to providing training and fellowships.<sup>85</sup>

- **Verificado** is a collaborative fact-checking platform that aims to combat disinformation and fake news surrounding Mexican elections, as well as verify reports on the electoral process (*see the Mexico Case Study in Appendix A on page 49 for additional details*).<sup>86</sup>
- **Third-party fact-checkers** have partnered with Facebook to review and rate the accuracy of Facebook articles and posts. In countries such as Colombia, Indonesia, and Ukraine, as well as various members of the EU, Facebook has commissioned groups—through what is described as “a thorough and rigorous application process” established by the IFCN—to become trusted fact-checkers who vet content, provide input into the algorithms that define the News Feed, and downgrade and flag content that is identified as false.<sup>87</sup>

Global fact-checking resources may also be useful, such as Claim Buster and AP Fact Check,<sup>88</sup> among others.

<sup>82</sup> StopFake (Media Reforms Center, n.d.), <https://www.stopfake.org/en/main/>.

<sup>83</sup> Africa Check, (Africa Check, n.d.), <https://africacheck.org>.

<sup>84</sup> Chequeado, (La Voz Pública Foundation, n.d.), <https://chequeado.com>.

<sup>85</sup> The International Fact-Checking Network (Poynter, n.d.), <https://www.poynter.org/ifcn/>.

<sup>86</sup> Verificado, (Verificado, n.d.), <https://verificado.com.mx/tag/fact-checking/>.

<sup>87</sup> Facebook Business Help Center, “Fact Checking on Facebook” (Facebook, n.d.), <https://www.facebook.com/business/help/2593586717571940>; Tessa Lyons, “Hard Questions: How is Facebook’s Fact-Checking Program Working?” *Hard Questions* (blog), Facebook, June 14, 2018, <https://about.fb.com/news/2018/06/hard-questions-fact-checking/>; The International Fact-Checking Network (Poynter, n.d.), <https://ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles>.

<sup>88</sup> ClaimBuster (IDIR Lab, University of Texas at Austin, n.d.), <https://idir.uta.edu/claimbuster/>; AP Fact Check (AP, n.d.), <https://apnews.com/hub/ap-fact-check>.



As your organization monitors the information space for concerning claims surrounding the election, keep an eye out specifically for the key facts, described below, that may be manipulated to confuse or dissuade voters.

Key Facts During an Election Process:	
<b>Who?</b>	The entities and people who make elections happen.
<b>What?</b>	The machines, systems and ways that we vote.
<b>When?</b>	The day(s), times and deadlines that guides registration and voting timelines.
<b>Where?</b>	The locations we gather to vote.
<b>How?</b>	How voting happens.

If your organization detects misleading information related to the who, what, when, where and how of an election, take the time to report the claim.

## Social Media Platform Initiatives to Increase Access to Credible Information

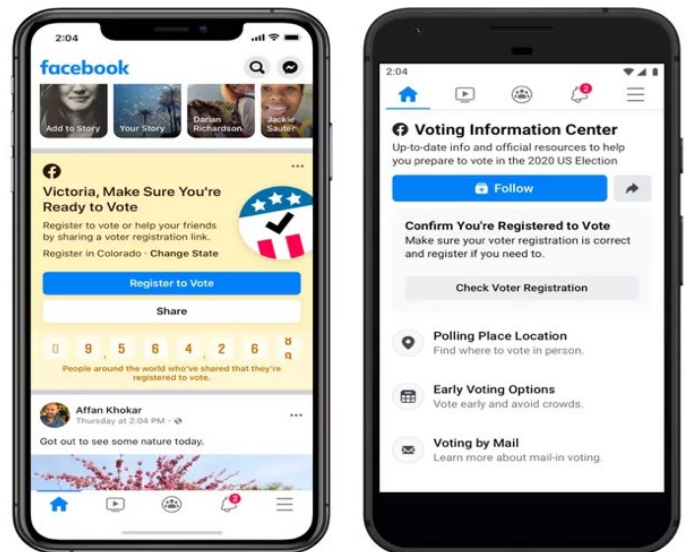
A key piece of fact-checking is having access to credible, reliable and unrestricted information on and offline. In response to advocacy efforts and complaints over complicity by social media platforms, some platforms have launched initiatives to increase access to credible information, ranging from redirecting users to reliable sources of election-related news, to expanding Application Programming Interface (API) access to enable research, to improving product features to discourage the sharing of false information. As your organization seeks to fact-check information around an election, keep these efforts in mind to either capitalize on existing initiatives or to advocate for similar initiatives if they do not yet exist in your country.

### Facebook

Facebook has implemented a number of initiatives to improve access to data and authoritative information, both for fact-checkers and researchers. One feature affixes information labels or buttons to posts that reference certain topics vulnerable to information manipulation, such as COVID-19, vaccines and elections. For example, the company labels content referencing

“ballots” or “voting” (irrespective of the content’s veracity) during an election, directing Facebook users to official voting information. These labels were used extensively during the 2020 U.S. presidential election and have been used during elections in other countries, as well, including Colombia, the UK and Germany, among others.<sup>89</sup>

For example, in preparation for the 2019 local elections in Colombia, Facebook partnered with Colombia’s National Electoral Council (CNE) to provide citizens with credible information about voting by creating Election Day reminders and an Informed Voter button, which redirected the user to the local election authority for voter information about where and when they could vote. These features have been used in other elections around the globe. Examples of Facebook’s voter information features are depicted below:



Facebook has also begun labeling certain state-controlled media to provide greater transparency on the sources of information on the platform. These labels currently appear on Pages and on the platform’s ad libraries; they will eventually be expanded to be more widely visible. The labels build on transparency features already in operation on Facebook Pages, which include panels that provide context on how the Page is administered (including information about the users who manage the Page and the countries from which they are operating), as well as information about whether the Page is state-controlled.<sup>90</sup>

<sup>89</sup> Hannes Grasegger, “Facebook Says its ‘Voter Button’ is Good for Turnout. But Should the Tech Giant be Nudging Us At All?” *The Guardian*, April 15, 2018, <https://www.theguardian.com/technology/2018/apr/15/facebook-says-it-voter-button-is-good-for-turn-but-should-the-tech-giant-be-nudging-us-at-all>.

<sup>90</sup> Colin Crowell and @misskaul, “Protecting the Integrity of the Election Conversation in India” (Twitter, February 21, 2019), [https://blog.twitter.com/en\\_in/topics/events/2019/election-integrity](https://blog.twitter.com/en_in/topics/events/2019/election-integrity).

## Twitter

Twitter has developed a number of policies, campaigns and product features to provide users with access to credible and authoritative information. In 2019, ahead of India's election, Twitter undertook substantial efforts to provide users with access to credible information about the election, while also evolving its product, updating rules and addressing information manipulation on its service affecting India overall.<sup>91</sup> These efforts also included additional product features and enhancements to prevent users from sharing misleading information about voting. More recently, Twitter announced a partnership with the Associated Press (AP) and Reuters to expand its efforts to better highlight reliable news as well as to add more context to the news and trends circulating on its platform.<sup>92</sup>

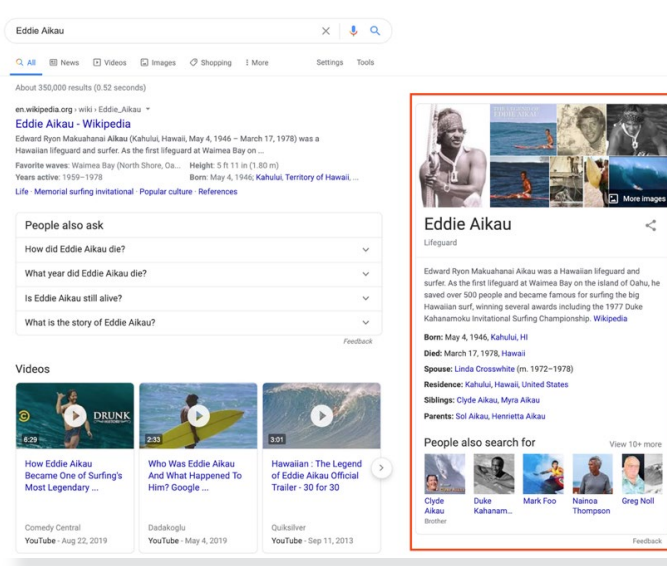
## WhatsApp

As an encrypted messaging platform, WhatsApp has limited information available to users and researchers about activities on its services. However, WhatsApp has provided access to its API in order to support certain research initiatives. The company has expanded API access through the Zendesk system—particularly for groups connected to the First Draft Coalition, such as Comprova in Brazil and CrossCheck in Nigeria.<sup>93</sup> This approach has been used to collect data on political events, the spread of false information and hate speech, and other research goals. The International Fact-Checking Network has also developed a collaboration with WhatsApp that includes access to the API for certain kinds of research.

## Google

Google's Knowledge Panels are boxes of information that appear

when users search for people, places, things and organizations that are in the Knowledge Graph, Google's database of facts.<sup>94</sup> These automatically generated boxes of information, depicted below, provide a snapshot of information on a particular topic. While knowledge panels were created to provide information and address information manipulation, they have been the cause of magnifying some disinformation.<sup>95</sup>



## YouTube

In order to provide users with accurate information, YouTube provides Breaking News and Top News features, which elevate information from verified news sources.<sup>96</sup> As part of the company's ongoing efforts, YouTube has indicated that it is expanding the use of information panels to provide users with additional context from fact-checkers.<sup>97</sup>

<sup>91</sup> Colin Crowell y @misskaul, "Protección de la integridad de la conversación electoral en India" (Twitter, 21 de febrero de 2019), [https://blog.twitter.com/en\\_in/topics/events/2019/election-integrity](https://blog.twitter.com/en_in/topics/events/2019/election-integrity).

<sup>92</sup> Sarah Perez, "Twitter Partners with AP and Reuters to Address Misinformation on Its Platform," TechCrunch, August 2, 2021, <https://techcrunch.com/2021/08/02/twitter-partners-with-ap-and-reuters-to-address-misinformation-on-its-platform/>.

<sup>93</sup> "Zendesk Introduces WhatsApp for Zendesk" (Zendesk, August 16, 2019), <https://www.zendesk.com/company/press/zendesk-introduces-whatsapp-zendesk/>; First Draft, "Introducing the First Draft Coalition" (First Draft, June 18, 2015), <https://medium.com/1st-draft/introducing-the-first-draft-coalition-e557fdacd1a6>; First Draft, "Comprova" (First Draft, n.d.), <https://firstdraftnews.org/tackling/comprova/>; First Draft, "Cross-Check Nigeria" (First Draft, n.d.), <https://firstdraftnews.org/tackling/crosscheck-nigeria/>.

<sup>94</sup> Knowledge Panel Help, "About Knowledge Panels" (Google, n.d.), <https://support.google.com/knowledgepanel/answer/9163198?hl=en>; Knowledge Panel Help, "How Google's Knowledge Graph Works" (Google, n.d.), <https://support.google.com/knowledgepanel/answer/9787176?hl=en>.

<sup>95</sup> Barry Schwartz, "Google adds new knowledge panel to provide information about news publishers," Search Engine Lab (November 7, 2017), <https://searchengineland.com/google-adds-new-knowledge-graph-learn-news-publishers-286394>; Lora Kelley, "The Google Feature Magnifying Disinformation," Atlantic (September 23, 2019), <https://www.theatlantic.com/technology/archive/2019/09/googles-knowledge-panels-are-magnifying-disinformation/598474/>.

<sup>96</sup> YouTube Help Center, "Breaking News and Top News on YouTube" (YouTube, n.d.), <https://support.google.com/youtube/answer/9057101?hl=en>.

<sup>97</sup> PTI, "Fighting fake news: YouTube to show 'information panels' on news-related videos," The Economic Times (March 7, 2019), <https://economictimes.indiatimes.com/magazines/panache/fighting-fake-news-youtube-to-show-information-panels-on-news-related-videos/articleshow/68302365.cms>.

## Strategic Silence

As mentioned earlier in this playbook, not all information manipulation necessitates a response. Even as skills related to spotting false information improve, it is critical, when deciding to debunk a falsehood, to avoid amplifying the very message you are trying to correct. Actively deciding against debunking a false claim is strategic silence. In considering whether or not a falsehood warrants a reply, evaluate the following:

- What is the risk level associated with the claim? Might it lead to violence or physical harm? Does it threaten to significantly undermine elections and/or voter confidence in the process or outcomes?
- What are the levels of engagement?
- How widespread is the attention?
- Who created the falsehood? Are they an established voice who may be considered credible? How much influence do they have?
- Has the falsehood already had a demonstrated effect?

If engagement levels with the claim are low, attention is not widespread, there is no demonstrated effect, or the claim has not or is unlikely to have an impact on voter behavior and beliefs, it is unlikely the falsehood requires an intervention. In a situation where the claim does not yet rise to the level of necessitating a response, we recommend that you log the incident and add it to any existing monitoring routines in case of increased relevancy or engagement. While it may seem counterintuitive to allow a falsehood to remain unchecked, countering a claim that has not garnered much attention nor achieved much influence can have the unintended impact of amplifying or reinforcing a falsehood simply as a result of repeating it.



### Thinking About Timelines

While there's no exact timeline for how long you or your organization should remain silent, continual monitoring of the identified false narrative(s) or manipulated piece(s) of information will help determine when communication is needed. If a false or misleading narrative or content begins to gain traction, whether that be quickly over just a few days or over the course of a few weeks, adjusting your strategy to address the false narrative may become prudent or even necessary.



## Resources for Responding to Disinformation

Once you have identified disinformation, these tools and resources listed below can help you counter or respond to disinformation.

- **Demtech/Comprop Navigator (*Resource List*):** As part of the project on Computational Propaganda, the Oxford Internet Institute developed the Demtech Navigator as an online guide for civil society organizations that provides tools, information and resources from a variety of sources with strategies for dealing with disinformation, fake news, cybersecurity and online harassment.<sup>98</sup>
- **RAND Corporation Database of Tools that Fight Disinformation (*Resource List*):** The RAND Corporation compiled a database of tools developed by nonprofit organizations in the U.S. for combating disinformation, especially on social media. These are product or resource-related tools, rather than resources that provide general information. The database includes tools for fact-checking, bot trackers and image verification.<sup>99</sup>
- **CEPPS Countering Disinformation Guide (*Guide*):** Commissioned by USAID, the Consortium for Elections and Political Process Strengthening (CEPPS)—composed of the National Democratic Institute, the International Republican Institute and the International Foundation for Electoral Systems—developed the CEPPS Disinformation Guide as a resource for civil society organizations, governments and election management bodies. The guide provides research on countering disinformation and a searchable database of initiatives by civil society organizations and other stakeholders from around the globe to target disinformation.<sup>100</sup>
- **Digital Sherlocks (*Network*):** The Atlantic Council launched Digital Sherlocks as a program to train a network of individuals on open-source tools for countering disinformation. To date, the Atlantic Council has trained over 1,500 people through 50 workshops in six continents to support worldwide digital resilience.<sup>101</sup>

<sup>98</sup> DemTech Navigator (Programme on Democracy and the Internet, Oxford University Institute, n.d.), <https://navigator.oii.ox.ac.uk>.

<sup>99</sup> “Tools That Fight Disinformation Online” (RAND Corporation, n.d.), <https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html>.

<sup>100</sup> International Foundation for Electoral Systems, International Republican Institute, National Democratic Institute. “Database of Informational Interventions” (Consortium for Elections and Political Process Strengthening, 2021). <https://counteringdisinformation.org/index.php/interventions>.



<sup>101</sup> 360/Digital Sherlocks (Atlantic Council, Digital Forensic Research Lab, n.d.), <https://www.digitalsherlocks.org>.

## Step 3 Build Resilience

In order to build a vibrant and strong information environment, existing and nascent democracies must prioritize building democratic processes that are resilient to information disorders, including information manipulation. In this playbook, what we mean by resilience is *the ability of citizens to participate in and contribute to democratic processes such as elections*. Citizens need to have the skills to find, identify, think critically about and evaluate the election-related information they are consuming online and offline, while public, private and civil society institutions need to ensure that citizens have access to credible resources and information. In this chapter, we highlight approaches to building resilience such as public awareness campaigns, civil society advocacy, digital literacy and cyber hygiene.




### A Whole-of-Society Approach for Resilience

Building a resilient society requires an understanding of global, regional and country-specific whole-of-society responses and interventions to counter information manipulation, such as those in the examples below. While governments, digital platforms, the private sector, academia and civil society each have their own approaches to mitigation, no one sector can address these challenges alone.

<p><b>Global</b></p> 	<p><b>Paris Call for Trust and Security in Cyberspace</b></p>	<p>The <u>Paris Call</u> consists of a group of 79 countries, 35 public authorities, 391 organizations and 705 companies that have come together to align around a set of nine principles to create an open, secure, safe and peaceful cyberspace. The Paris Call reaffirms these countries' commitment to international humanitarian law and customary international law to provide the same protections for citizens online as these laws provide offline. In creating this call, governments, civil society and the private sector, including social media companies, adhere to providing safety, stability and security in cyberspace, as well as increased trust and transparency to citizens. The call has created a multistakeholder forum process for organizations and countries to come together to increase information sharing and collaboration.<sup>102</sup></p>
<p><b>Regional (Europe)</b></p> 	<p><b>The European Union Code of Practice on Disinformation</b></p>	<p>The <u>EU Code of Practice on Disinformation</u><sup>103</sup> is one of the more multinational and well-resourced regional initiatives, as it has the support of the entire European bloc and includes signatories from Facebook, Google, Twitter and Mozilla, as well as advertisers and parts of the advertising industry. The Code is centered on five pillars: enhance transparency of online news; promote media and information literacy to counter disinformation; develop tools for empowering users and journalists to tackle disinformation; safeguard the diversity and sustainability of the European news media ecosystem; and promote continued research on the impact of disinformation in Europe to evaluate and adjust response measures.</p>

<sup>102</sup> Paris Call for Trust and Security in Cyberspace (November 12, 2018), <https://pariscall.international/en/>.

<sup>103</sup> Shaping Europe's digital future, "Code of Practice on Disinformation," (European Commission, n.d.) <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>.

<p><b>Regional (Latin America)</b></p> 	<p><b>Fundamedios</b></p>	<p>Founded in 2007, <u>Fundamedios</u> is an organization dedicated to promoting freedom of expression, upholding human rights, and monitoring aggressions and risks faced by journalists across Latin America. Fundamedios has actively worked to establish a network—spanning Ecuador, Bolivia, Argentina, Honduras and the U.S.—of civil society, media and international organizations to monitor and train journalists and other civil society stakeholders to better identify, understand and combat mis/disinformation. Additional regional collaborations include advocacy to governments to advance access to information, as well as generating media outlets to spread truthful content.<sup>104</sup></p>
<p><b>Country-Specific (Nigeria)</b></p> 	<p><b>Abuja Accord</b></p>	<p>Leading up to the 2015 general elections in Nigeria, multiple stakeholders affirmed their commitment to a peaceful electoral process through the signing of a five-point agreement, the <u>Abuja Accord</u>. Signatories—including presidential candidates, representatives of the EMB, and security agencies—committed themselves to increase the security of Nigeria’s elections, including agreeing to take proactive measures to prevent electoral violence, committing to fully abide by regulations as determined by the legal framework for elections in Nigeria, and placing national interest above partisan concern, among others.<sup>105</sup></p>
<p><b>Country-Specific (Argentina)</b></p> 	<p><b>Ethical Digital Commitment</b></p>	<p>In 2019, the National Electoral Council (CNE: Cámara Nacional Electoral) of Argentina launched an initiative to engage stakeholders of all types—including political parties as well as representatives of technology and social network companies—to sign an <u>Ethical Digital Commitment</u>. The aim of the commitment was to prevent the dissemination of fake news and any other mechanisms of information manipulation that may negatively affect elections. The commitment included collaboration across sectors of society, as signatories included people from various political parties; representatives of Google, Facebook, Twitter, and WhatsApp; and directors of the Association of Digital Journalism (ADEPA), among others.<sup>106</sup></p>

<sup>104</sup> Fundamedios (Fundamedios, n.d.), <https://www.fundamedios.org>.

<sup>105</sup> “Abuja Accord on the Prevention of Violence and Acceptance of Election Results by Presidential Candidates and Chairpersons of Political Parties Contesting the 2015 General Elections” (Nigeria, 2015), <https://www.idea.int/sites/default/files/codesofconduct/Abuja%20Accord%20January%202015.pdf>.

<sup>106</sup> “Ethical Digital Commitment” (Argentina: National Electoral Council, May 30, 2019), <https://www.electoral.gob.ar/nuevo/paginas/pdf/CompromisoEticoDigital.pdf>.



## Public Awareness Campaigns

While strategic communications are necessary in the lead-up to and during elections, for long-term resilience, continued efforts between election periods is also critical. Public awareness campaigns help citizens understand that the information environment is manipulated in ways that might undermine their capacity to exercise their democratic rights. It is important to enable and empower your targeted audience to think critically about the information they consume and to have the necessary toolkit to communicate and engage with their trusted networks of friends, family and colleagues, so that they in turn can share this understanding. In your efforts to raise awareness about the threat of information manipulation, consider the steps below.

Before your campaign:

- Identify the segment of the population and audience you want to appeal to at national, subnational and local levels.
- Identify other partner civil society and organizations to actively include in your awareness campaign or those who can help you amplify your messages.
- Determine *how* you will run your awareness campaign and what channels you will use to dispel falsehoods and manipulated information. Examples include public service announcements, press releases, social media, television radio, and word-of-mouth channels.

During your campaign:

- Use proactive communication to identify potential risks of information manipulation and its consequences both during and between election cycles.
- Raise your constituents’ and partners’ awareness about the types of information manipulation they may experience and see online and offline (falsehoods, “half-truths,” hate speech, state-sponsored propaganda, etc.).
- Share with your targeted population and constituents where to go to find skills, resources and programming on digital literacy, cyber hygiene and ways to respond to information manipulation.
- Share insights on when to stay silent in order to avoid spreading manipulated information.

After the campaign:

- Meet with your team to determine lessons learned and steps to improve follow-on iterations.
- Repeat successful public awareness campaigns for disparate targeted segments of the population (older populations, marginalized communities, etc.).

The table below includes examples of public awareness campaigns that successfully increased their targeted citizenry’s knowledge. While some of these examples focus on raising awareness around COVID-19, the tactics used are transferable to electoral contexts as well.

Democracy Actor	Examples
Government	“ <b>Stop the Spread</b> ” is a global campaign aimed at raising awareness about the risks of misinformation around COVID-19, encouraging the public to double-check information with trusted sources such as the WHO and national health authorities. <sup>107</sup>
Government	In Timor-Leste, government officials partnered with IRI to connect citizens and members of parliament through a talk show called “ <b>Koalia Ba Hau/Talk to Me!</b> ” Government officials are able to proactively share truthful information on topics such as COVID-19, and citizens are able to participate by asking questions and making comments. Koalia Ba Hau is broadcast on national television with a reach of 9,500 viewers, as well as on radio stations across the country. This form of proactive communication has proven to have better reach than traditional town halls and roundtables.

<sup>107</sup> World Health Organization, “Countering Misinformation about COVID-19: A Joint Campaign with the Government of the United Kingdom” (updated May 13, 2020), <https://www.who.int/news-room/feature-stories/detail/countering-misinformation-about-covid-19>.

Democracy Actor	Examples
EMB	In 2020, the Brazilian Superior Electoral Court (TSE) augmented its traditional public outreach strategies through the creation of “ <b>e-Título</b> ,” a mobile app that helps voters identify their polling stations and facilitates direct communication between voters and the TSE. <sup>108</sup>
EMB	The National Election Board of Ethiopia (NEBE) created a <b>#AskNebe campaign</b> <sup>109</sup> on Twitter for voters to directly communicate with the board and to ask questions about the election process and how to get credible information.
CSO	<b>Matsda2sh</b> (“do not believe”) is an Egyptian fact-checking CSO that has used Facebook to reach over 500 thousand followers with <b>awareness videos</b> on the dangers of disinformation. <sup>110</sup>

### The Importance of Advocacy

Civil society organizations can be a strong voice and advocate for greater transparency from local governments, political parties and EMBs in order to push for regulatory and legal changes to better protect future elections and ensure stronger commitments, auditability, and accountability measures from social media and other technology companies. Such appeals for transparency, accountability and reform will often require CSOs to design a proactive advocacy campaign.

When developing a government-focused advocacy campaign, consider the best practices listed below.

- Conduct a **situational and landscape analysis** on why an advocacy campaign may be warranted: What are the current laws and regulations on information manipulation, online hate, harassment, and freedom of expression within the society?
- Map the stakeholders and **build a coalition**: Who are the stakeholders you need to engage with from local governments, your partners, and legal and technical experts to build an advocacy coalition and a successful advocacy campaign? Since information

manipulation, online hate, and harassment often include marginalized communities, make sure to include the diversity of these voices and perspectives.

- Consider the **central issue** or set of issues to build your policy advocacy campaign around, and ensure alignment on the issue across your stakeholder ecosystem.
- Create **communication and messaging strategies**. This may include content development, a website, and social media and traditional media presence. Identify your validators who can amplify your campaign message.
- Identify the **government stakeholders** you want to advocate with and the best tactics for government advocacy. This may include engaging directly with governments, writing letters, submitting recommendations and partnering with internal allies who can carry out the message.

Additional information can be found in the [Open Internet for Democracy Advocacy Playbook](#).<sup>111</sup>

<sup>108</sup> Brazilian Superior Electoral Court, “e-Título” (app), <https://apps.apple.com/us/app/e-t%C3%ADtulo/id1320338088>.

<sup>109</sup> #AskNebe campaign (Twitter Campaign), <https://twitter.com/nebethiopia/status/1357311115257143298?lang=en>.

<sup>110</sup> Matsda2sh (Facebook Page), <https://www.facebook.com/matsda2sh/>.

<sup>111</sup> Open Internet for Democracy, Advocacy Playbook: Strategies to Build Coalitions & Tactics (OID, n.d.), <https://openinternet.global/sites/default/files/2020-10/Open%20Internet%20for%20Democracy%20Playbook%20%283April2019%20Release%29.pdf>.

## Digital Literacy

Digital literacy initiatives focus on building citizen capacity to operate in a highly digitized world. While digital literacy programs should be tailored to their audience, they typically include helping people learn how to quickly discern fact from fiction and develop an understanding of how information spreads online. In a digital literacy initiative, your partners, employees and citizens can learn these critical lessons:

- How to think critically about the information they consume both on social media and through traditional media outlets.
- The functions of social and mainstream media, including how information is curated and spread.
- How to identify credible content (i.e., can it be verified by multiple credible sources, has it been verified by credible fact-checking organizations, is the headline sensationalized, etc.).
- How to verify images and videos through programs such as [Google image search](#), [reverse image search](#), and [fake video news debunker](#).<sup>112</sup>
- How to avoid contributing to disinformation by not [sharing or commenting](#) on unverified content.<sup>113</sup>
- How to report false or harmful content in open and closed networks to social media platforms.
- Acknowledge how bias, “group think,” and cultural, religious, and social norms affect one’s ability to identify and evaluate credible content.

Any digital literacy initiative for citizens, partners and your own organization should also include lessons on cyber hygiene.

- Use strong passwords and two-factor authentication.

- Use encrypted messaging to communicate sensitive information.
- Use a [Virtual Private Network \(VPN\)](#) to establish private network connections to safely communicate and conduct the business of your organization.<sup>114</sup>
- Review privacy and security settings on your social media accounts.

IREX’s comprehensive [Learn to Discern \(L2D\)](#) program has extensive resources for building a digital literacy program on its [Resources for Learning & Impact](#) website.<sup>115</sup> [UNESCO’s Media and Information Literacy Resources](#) page also lists several useful resources.<sup>116</sup>



**Tip: Tailoring digital literacy programs for marginalized groups** is particularly important, as information manipulation often spreads in those communities, frequently targeting women and girls.



### Tip: Don’t Forget Traditional Media Sources

Remember that digital literacy should include understanding how information spreads offline through more traditional information sources (newspapers, radio, etc.), as well. Be sure to remind your participants that the same digital literacy skills apply when identifying credible information in newspapers and on the radio and that information can travel between offline traditional media and online trusted networks.

<sup>112</sup> Google Images (Google, n.d.), <https://www.google.com/imghp?hl=en>; Squobble.com, “RevEye Reverse Image Search” (app), <https://chrome.google.com/webstore/detail/reveye-reverse-image-sear/keaacclcjhehbbapnphnmpiklalfhelgf?hl=en>; InVID and WeVerifyFake, “News Debunker” (app), <https://chrome.google.com/webstore/detail/fake-news-debunker-by-inv/mhccpoafgdgbhjfkhcmgkndkeenfhe?hl=en>.

<sup>113</sup> ReFrame and PEN America, *Disinfo Defense Toolkit for Organizers and Advocates* (ReFrame and PEN America, n.d.), <https://pen.org/wp-content/uploads/2020/12/disinfo-defense-toolkit-v2-compressed.pdf>.

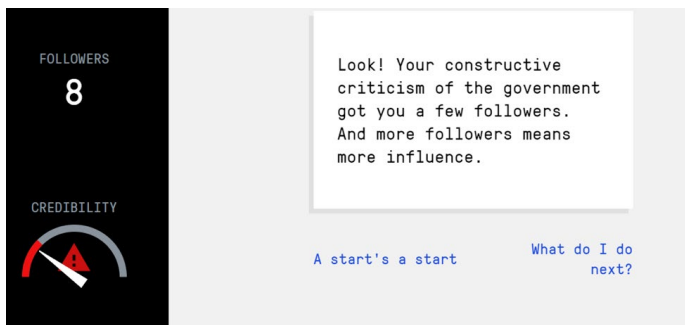
<sup>114</sup> Techopedia, “What is a Virtual Private Network (VPN)?” (Techopedia, updated November 14, 2016), <https://www.techopedia.com/definition/4806/virtual-private-network-vpn>.

<sup>115</sup> Learn to Discern, “Media Literacy Training” (IREX, n.d.), <https://www.irex.org/project/learn-discern-l2d-media-literacy-training>.

<sup>116</sup> “Media and Information Literacy - Resources” (UNESCO, n.d.), <https://en.unesco.org/themes/media-and-information-literacy/resources>.

## “Games to Discern”

Your digital literacy initiative can incorporate unique learning approaches. Games can teach how fake news spreads and how to identify and discern credible information and debunk false narratives or falsehoods. For example, psychologists from the University of Cambridge partnered with the Dutch media collective DROG to create the [Bad News Game](https://www.getbadnews.com/#intro), which seeks to build psychological resilience to disinformation.<sup>117</sup> The game incorporates active experiential learning by asking players to create a fake persona, attract followers and set up credibility as a fake news site. In other words, the game allows players to familiarize themselves with the mindset of threat actors who seek to spread disinformation. Players build resilience to disinformation by better understanding threat actors and their tactics, including impersonation, emotion, polarization, conspiracy theory peddling, discreditation of facts and trolling. The screenshot below shows the player’s experience.



Similar games exist for global, regional and country-specific audiences. Games that might be used globally include [PolitiTruth](https://www.cinqmarsmedia.com/politifact/index.html), [Be Internet Awesome](https://beinternetawesome.withgoogle.com/en_us/interland/landing/tower-of-treasure), [Factitious](http://factitious-pandemic.augamestudio.com/#/) and [Fakey](https://fakey.osome.iu.edu).<sup>118</sup> Other games, such as [Harmony Square](https://harmonysquare.game/en/) and [Fake It to Make It](https://www.fakeittomakeitgame.com), have been

created specifically for country contexts such as the U.S. and the Netherlands, respectively.<sup>119</sup> Review these existing tools to see if any might be useful in your digital literacy efforts. Many of these games may serve as useful starting points to conceive of a game to fit your own country’s context. The use of existing games will likely require translation, contextualization and other adjustments to be relevant to your country’s context.

## Social Media Platform Digital Literacy Initiatives

Social media platforms and other private sector partners have invested in building resiliency efforts globally through private-public-civic partnerships. Many of the resulting programs—detailed below—are available in multiple languages and could be useful resources as you work to develop the digital literacy of your organization and communities.

- **Facebook** and Asia-Pacific experts collaborated on the [We Think Digital](https://wethinkdigital.fb.com) program, which fosters digital literacy in the region through the creation of public guides to user actions, digital learning modules, video and other pedagogical resources.<sup>120</sup>
- **Twitter** partnered with UNESCO to create the [Teaching and Learning with Twitter](https://about.twitter.com/content/dam/about-twitter/en/tfg/download/teaching-learning-with-twitter-unesco.pdf) handbook, which helps educators around the world enable young people to think critically about the information they consume.<sup>121</sup>
- **Google** and **YouTube** have significantly invested in digital responsibility and [media literacy](https://www.blog.google/out-reach-initiatives/google-org/digital-and-media-literacy-education-korea/) in order to develop citizen and youth resilience.<sup>122</sup> Most recently, Google invested €25 million to help launch the [European Media and Information Fund](https://calouste.gulbenkian.pt/en/european-media-and-information-fund/).<sup>123</sup> This effort is meant to educate, train and support citizens in

<sup>117</sup> DROG, “Bad News Game” (online game), <https://www.getbadnews.com/#intro>.

<sup>118</sup> PolitiFact, “PolitiTruth” (app) <https://www.cinqmarsmedia.com/politifact/index.html>; Be Internet Awesome, “Interland” (online game), [https://beinternetawesome.withgoogle.com/en\\_us/interland/landing/tower-of-treasure](https://beinternetawesome.withgoogle.com/en_us/interland/landing/tower-of-treasure); AU Game Lab and JoLT, “Factitious 2020” (online game), <http://factitious-pandemic.augamestudio.com/#/>; Observatory on Social Media, “Fakey” (online game), <https://fakey.osome.iu.edu>.

<sup>119</sup> Global Engagement Center (GEC), Cybersecurity and Infrastructure Security Agency (CISA), DROG and University of Cambridge, “Harmony Square” (online game), <https://harmonysquare.game/en/>; Amanda Warner, “Fake It to Make It” (online game), <https://www.fakeittomakeitgame.com>.

<sup>120</sup> Facebook, “We Think Digital” (Facebook, n.d.), <https://wethinkdigital.fb.com>.

<sup>121</sup> Twitter, “Teaching and Learning with Twitter” (Twitter, n.d.), <https://about.twitter.com/content/dam/about-twitter/en/tfg/download/teaching-learning-with-twitter-unesco.pdf>.

<sup>122</sup> Jacqueline Fuller, “Bringing Digital and Media Literacy Education to More Schools in Korea” (Google.org, March 28, 2019), <https://www.blog.google/out-reach-initiatives/google-org/digital-and-media-literacy-education-korea/>.

<sup>123</sup> Matt Briton, “Google’s €25 million Contribution to Media Literacy,” The Keyword (Google blog), March 21, 2021, <https://blog.google/around-the-globe/google-europe/googles-25-million-contribution-to-media-literacy/>; “European Media and Information Fun,” Calouste Gulbenkian Foundation, <https://calouste.gulbenkian.pt/en/european-media-and-information-fund/>.

strengthening media literacy skills; support and scale the work of fact-checkers; and strengthen the expertise and research surrounding information manipulation in its varied forms.

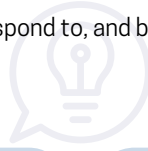
- **Microsoft** has also partnered with research institutions and CSOs globally, including the University of Washington, Sensity and USA Today, on its Defending Democracy Program, to build resilience and promote media and digital literacy, to help the public decipher falsehoods from half-truths and facts.<sup>124</sup> The ultimate goal of this initiative has led to a more engaged citizenry.<sup>125</sup>

<sup>124</sup> Sensity, <https://sensity.ai>; Tom Burt, "Announcing the Defending Democracy Program," Microsoft on the Issues (blog), April 13, 2018, <https://blogs.microsoft.com/on-the-issues/2018/04/13/announcing-the-defending-democracy-program/>.

<sup>125</sup> Tom Burt, "New Steps to Combat Disinformation" Microsoft on the Issues (blog), September 1, 2020, <https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/>.

## Key Tips for Combating Information Manipulation

The following are key tips for your organization to identify, respond to, and build resilience to information manipulation.



### Have a plan in place

Do not wait until you have experienced or witnessed electoral information manipulation narratives in your local online information ecosystem to begin identifying best approaches and strategies to counter them. Prepare proactively.

### Not all governments have good intentions

Proceed with caution if and when you decide to report election-related information manipulation to governments, as many do not respect democratic norms or are not impartial. Review resources and past actions the governments have taken to avoid doing more harm.

### Partner up to produce better results

Engage and collaborate with a diverse set of entities working on elections, such as EMBs and CSOs working on voter education; community and religious leaders who are trusted sources of information in their communities; social media platforms; journalists; and fact-checkers, etc.

### Adjust your expectations of social media platforms' behavior and actions

Electoral information manipulation is rampant on social media platforms. Familiarize yourself with the platforms' policies and community standards and understand that many platforms are not immediately responsive to user reports or prepared to tackle a country's electoral information environment on their sites.

### Mix, match and tailor your approaches

Responding to electoral information manipulation requires a combination of approaches to ensure successful outcomes, and the effectiveness of these approaches will vary across country contexts.

### Quick response and long-term resilience go hand-in-hand

Short-term responses to electoral information manipulation need to be complemented by long-term resilience building in areas such as digital literacy, public awareness campaigns and a whole-of-society approach to foster a well-informed public.

### See or hear something, say something

If you see or hear manipulated information directed at your organization, partner organizations, and/or segments of the population you work with, report it to government authorities, social media platforms and the media for investigation when appropriate.

### Digital literacy increases critical thinking

Educate your citizens, organization, partners and trusted networks about identifying false narratives and content. Encourage them to stay vigilant and think critically about the information consumed online and offline.



# Appendices

## Appendix A: Case Studies



### Mexico Case Study

#### Background and Political Context

In Mexico, information manipulation occurs in an environment with relatively high internet penetration rates, with approximately two-thirds of the country online, and high levels of social media use. Most people (86 percent) get their news from online sources, with Facebook (70 percent), YouTube (44 percent) and WhatsApp (39 percent) as the three largest platforms for online news. Despite the high rates of social and online media use, 60% of people in Mexico are concerned about the online information ecosystem and the spread of false information online.<sup>126</sup>

Further exacerbating this trend is the fact that Mexico's online environment is only "partially free"; this, combined with growing polarization and identity politics, has created ripe conditions for the spread of false information online. The government has also undertaken efforts to impede freedom of speech, freedom of the press, democratic practices and other fundamental human rights through surveillance, restrictive laws and information manipulation.<sup>127</sup>

#### Information Manipulation in Mexico

Within this political and online context, Mexico has had a long history of information manipulation by a variety of malign actors, including political candidates, the influence industry and other

local actors who have used social media to spread disinformation about politics. These campaigns are often characterized by highly automated accounts—sometimes called political bots—that have played a large role in amplifying disinformation online.

The use of bots first came to public attention during the 2012 presidential elections in Mexico, where researchers identified the use of "Peñabots" by the Institutional Revolutionary Party (PRI)<sup>128</sup> to support the campaign of then-candidate Enrique Peña Nieto. Since then, academic studies and journalistic investigations have identified the continued use of bots to disrupt online communication, political discourse and protest activities in Mexico.<sup>129</sup>

The 2018 general elections marked unprecedented challenges for the online information ecosystem. As the largest election in Mexico's history—with more than 3,400 seats open at the local, state and federal levels—social media became one of the main fronts for information manipulation. Although there were concerns that Russian information operations would pollute the information ecosystem, most of the disinformation originated within Mexico. In the lead-up to the vote, bot accounts generated viral hashtags to exacerbate political divides and amplify conspiracies around fraud and corruption.<sup>130</sup> Fake news distributors and fake pollsters blended fact with fiction to undermine the credibility of professional news organizations and hamper citizens' ability to access accurate information about the election, candidates and their campaigns.<sup>131</sup> Investigators found that many of these activities were fueled by commercial firms who were hired by

<sup>126</sup> Nic Newman, et al., *Reuters Institute Digital News Report 2020* (Reuters Institute for the Study of Journalism, 2020), [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR\\_2020\\_FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf).

<sup>127</sup> Freedom House, "Mexico: Freedom on the Net 2020 Country Report," Freedom House (2020), <https://freedomhouse.org/country/mexico/freedom-net/2020>.

<sup>128</sup> Luis Daniel, "Rise of the Peñabots," *Data and Society: Points* (blog), Data and Society Research Institute (February 24, 2016), <https://points.datasociety.net/rise-of-the-peñabots-d35f9fe12d67>.

<sup>129</sup> Luiza Bandeira et al., *Disinformation in Democracies: Strengthening Digital Resilience in Latin America* (Atlantic Council Digital Forensic Research Lab, March 2019), <https://www.atlanticcouncil.org/in-depth-research-reports/report/disinformation-democracies-strengthening-digital-resilience-latin-america/>; Samantha Bradshaw, Hannah Bailey, and Philip Howard, *Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation*, Computational Propaganda Research Project (Oxford Internet Institute, January 13, 2021), <https://demotech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/01/Cyber-Troop-Report-2020-v.2.pdf>; Pablo Suárez-Serrato et al., "On the Influence of Social Bots in Online Protests. Preliminary Findings of a Mexican Case Study," in E. Spiro and YY Ahn (eds) "Social Informatics. SocInfo 2016," *Lecture Notes in Computer Science*, vol. 10047 (Springer, Cham), [https://doi.org/10.1007/978-3-319-47874-6\\_19](https://doi.org/10.1007/978-3-319-47874-6_19).

<sup>130</sup> Luiza Bandeira et al., *Disinformation in Democracies: Strengthening Digital Resilience in Latin America*; Monika Glowacki et al., "News and Political Information Consumption in Mexico: Mapping the 2018 Mexican Presidential Election on Twitter and Facebook" (Computational Propaganda Project, 2018), <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/06/Mexico2018.pdf>.

<sup>131</sup> Jorge Buendia, *Fake Polls as Fake News: The Challenge for Mexico's Elections* (Wilson Center, 2018), <https://www.wilsoncenter.org/publication/fake-polls-fake-news-the-challenge-for-mexicos-elections>.

politicians and businesses to distort the information ecosystem in their favour.<sup>132</sup>

The 2018 general elections in Mexico were also characterized by high levels of political violence, with more than 100 politicians killed in the lead-up to the vote.<sup>133</sup> While many of these murders have been associated with organized crime and the long-fought war on drugs, political violence has also been exacerbated by information manipulation. During a contentious gubernatorial race in the province of Puebla, the Atlantic Council found automated accounts amplifying competing hashtags that made premature claims of victory for opposing candidates.<sup>134</sup> The province of Puebla reported high levels of violence, with citizens murdered and ballots stolen or set on fire.

Information manipulation does not only occur during elections in Mexico. In 2019, a Signa\_Lab investigation identified a network of Twitter accounts that were attacking journalists and news outlets that criticized the new president.<sup>135</sup> Information manipulation creates many challenges for freedom of speech and freedom of the press in Mexico: journalists, activists and political opponents disproportionately experience harassment, threats, rumor and slander on social media.<sup>136</sup> Like many other countries, women are particularly vulnerable to online smear campaigns, where fake accounts have shared manipulated videos and images to sexualize, question and degrade the credibility and legitimacy of professional women.<sup>137</sup> Although Mexico has gender parity rules for political parties, female politicians still face a disproportionate amount of harassment online.

## Interventions

In order to identify, respond to and build resilience to information manipulation during the 2018 presidential elections, Mexico's federal National Electoral Institute (INE) collaborated with social media platforms and CSOs to enhance information integrity throughout the country.

The three major social media companies (Facebook, Twitter and Google) worked directly with INE to make information about the election more easily accessible to citizens.<sup>138</sup> Since more than 60 million citizens in Mexico use the internet—many of whom use it for news discovery and curation—all three platforms livestreamed the Mexican presidential debates and official election announcements for the first time. Twitter established formal hashtag discussions around the presidential debates, which created a forum for real-time professional and journalistic commentary on the topics being debated. Facebook and Google also worked with INE to implement interactive buttons that would direct users to INE's election hub, help users find polling stations, and spread get-out-the-vote messages. Overall, these platform collaborations helped citizens find and access accurate information about the candidates as well as electoral procedures and information.<sup>139</sup>

In addition to working with platforms, civil society organizations also coordinated with INE to help identify, respond to and build resilience against information manipulation during the 2018 elections. One of the most notable interventions was Verificado 2018, which brought together more than 80 partners to identify and respond to information manipulation in real time.<sup>140</sup> Verificado

<sup>132</sup> Ben Nimmo et al., "#ElectionWatch: Trending Beyond Borders in Mexico," Medium, June 28, 2018, <https://medium.com/dfrlab/electionwatch-trending-beyond-borders-in-mexico-2a195ecc78f4>.

<sup>133</sup> Natasha Turak, "More than 100 Politicians Have Been Murdered in Mexico Ahead of Sunday's Election," CNBC, June 26, 2018, <https://www.cnbc.com/2018/06/26/more-than-100-politicians-murdered-in-mexico-ahead-of-election.html>.

<sup>134</sup> Bandeira et al., *Disinformation in Democracies: Strengthening Digital Resilience in Latin America*.

<sup>135</sup> Signa\_Lab, "Democracia, Libertad de Expresión y Esfera Digital. Análisis de Tendencias y Topologías En Twitter. El Caso de La #RedAMLOVE" (2019), [https://signalab.iteso.mx/informes/informe\\_redamlove.html](https://signalab.iteso.mx/informes/informe_redamlove.html).

<sup>136</sup> ARTICLE 19, "Mexico: Report shows silencing of journalists and media freedom" (ARTICLE 19: April 17, 2019), <https://www.article19.org/resources/mexico-report-shows-silencing-of-journalists-and-media-freedom/>.

<sup>137</sup> Freedom House, "Mexico: Freedom on the Net 2020 Country Report." <https://freedomhouse.org/country/mexico/freedom-net/2020>.

<sup>138</sup> Kofi Annan Commission on Elections and Democracy, "Protecting Electoral Integrity in the Digital Age," (January 2020), <https://www.kofiannanfoundation.org/our-work/kofi-annan-commission/the-kacedda-94nfyd3mj9q9phewncbtf5tgcgitlh/>.

<sup>139</sup> Kofi Annan Commission on Elections and Democracy, "Protecting Electoral Integrity in the Digital Age." [https://www.kofiannanfoundation.org/app/uploads/2020/05/85ef4e5d-kaf-kacedda-report\\_2020\\_english.pdf](https://www.kofiannanfoundation.org/app/uploads/2020/05/85ef4e5d-kaf-kacedda-report_2020_english.pdf).

<sup>140</sup> Bandeira et al., "Disinformation in Democracies: Strengthening Digital Resilience in Latin America." <https://www.atlanticcouncil.org/in-depth-research-reports/report/disinformation-democracies-strengthening-digital-resilience-latin-america/>.

created a central hub of election resources and produced informational videos—which registered 5.4 million visits—to help citizens understand the election process.<sup>141</sup> They also set up a series of processes for users to inquire about the veracity of content on social media platforms and receive trustworthy and timely replies from fact-checkers. On Twitter and Facebook, Verificado accounts had more than two hundred thousand followers. Verificado also operated a WhatsApp group where users could send in requests for fact-checks. Within the first week of operation, the group received more than 18,000 messages, 13,800 of which were answered by four Verificado staff. In total, Verificado’s group had more than 9,600 subscriptions and over 60,000 interactions.

## Lessons from Mexico for Civil Society’s Response to Information Manipulation

### Establish collaborative relationships with platform companies and civil society actors.

Social media is increasingly a source of news and information, and working with platform partners to help streamline access to official electoral information can help build confidence and trust in an election. By livestreaming debates that were traditionally only broadcast on television, INE and the social media companies helped users find and watch the debates through communication mediums used by millions of citizens. Formal hashtag discussions that promoted journalistic and professional commentary helped expose citizens to additional information and opinions about the debates, so they could formulate their own ideas and opinions. Voting centers and information about how and where to vote helped encourage citizens to get out and vote on the day of the election. By providing accurate information about voting processes and candidates, these collaborative relationships with platforms can help build resilience to information manipulation campaigns.

### Timing matters: Increase the speed and scale of fact-checking.

Being able to address information manipulation in real time and before narratives go viral is incredibly important to combating the spread of harmful mis/disinformation. Verificado was so successful largely due to the speed and scale at which it operated. By working with multiple trusted partners, staff could identify information manipulation as it emerged on social media,

and respond to user inquiries about the truthfulness of content quickly and easily. They also operated on online platforms where mis/disinformation was spreading in order to reach audiences with corrected information and counter-messaging. In the lead-up to the election, Verificado’s Twitter, Facebook and WhatsApp group reached hundreds of thousands of voters with responses to individual inquiries about the veracity of content. Developing these trusted channels that could respond with speed and accuracy and on platforms where users were finding misinformation helped Verificado debunk rumors and build trust among citizens and voters, which contributed to the overall success of the initiative.

### Create a clear and consistent brand for professional fact-checked information.

What helped make Verificado a successful intervention was not only its real-time approach to mis/disinformation narratives as they emerged, but the way it established a clear and central brand for accurate, trusted and professional journalistic content about the election. Despite working with more than 80 trusted partners, Verificado allowed different organizations to lend resources and expertise under a single trusted brand that gained recognition among users and citizens. Verificado’s fact-checks and information were also picked up and broadcast on local television stations and print media in addition to its online work. The trusted brand also helped Verificado establish itself as a neutral and professional source of factual information in an environment characterized by highly polarizing domestic disinformation.

<sup>141</sup> Bandeira et al., “Disinformation in Democracies: Strengthening Digital Resilience in Latin America.”



## Taiwan Case Study







### Background and Political Context

Taiwan was taken off guard by disinformation during its 2018 local elections and referendums. Over the next two years, the country developed a whole-of-society response to build social cohesion, counteract disinformation and ensure a successful 2020 presidential election. In 2018, Taiwan’s government was primarily responding to disinformation campaigns unilaterally, without any coordination with third-party fact-checkers or social media companies. At that time, relationships with social media companies and third-party fact-checkers who could collaborate with the government to further amplify credible stories did

not yet exist.<sup>142</sup> Yet, in under two years, Taiwan established relationships, communication and coordination channels between government, civil society, social media platforms and independent fact-checkers. This coordination enabled Taiwan to identify and respond to disinformation with speed and efficacy during the 2020 elections and to establish the foundation for long-term resilience to disinformation (see this report).<sup>143</sup>

### Taiwan’s Whole-of-Society Response to Disinformation Campaigns

Below are highlights of key interventions Taiwan implemented to establish a whole-of-society response to information manipulation around the 2020 presidential elections.

 Government	 Social Media Platforms	 Civil Society Organizations
 <b>Response:</b> The Political Warfare Bureau of the Ministry of National Defense established a “ <u>rapid handling team</u> ” to quickly identify and respond to disinformation from China and pro-China domestic outlets and individuals, use big data to analyze CCP disinformation campaigns, and amplify readable content on social media and via press briefings. <sup>144</sup>	 <b>Partnership:</b> LINE established a private-public-civic partnership with Taiwan’s Executive Yuan and the FactCheck Center, Cofacts, MyGoPen, Doublethink Lab, and others in the <u>Digital Accountability Project (DAP)</u> . <sup>145</sup> Through this partnership, LINE incorporated fact-checking into its service and developed a chatbot to make users aware of disinformation campaigns, enable users to submit content to be analyzed by well-regarded fact-checking organizations, and provide content from credible news sources to the users.	 <b>Grassroots Activism:</b> Civil society organizations developed trusted networks with independent fact-checkers such as Cofacts, Taiwan’s FactCheck Center and others, as well as with social media platforms in order to identify, respond to and build resilience to disinformation campaigns leading up to the 2020 elections by using closed group chats and in-person neighborhood and religious groups. This approach was particularly effective, as CSO representatives were able to collaborate and share information with trusted networks to facilitate trust. <sup>146</sup>







<sup>142</sup> Representative of the Central Election Commission (CEE) of Taiwan in discussion with author, April 2020.

<sup>143</sup> Aaron Huang, *Combating and Defending Chinese Propaganda and Disinformation: A Case Study of Taiwan’s 2020 Elections* (Belfer Center for Science and International Affairs, 2020), <https://www.belfercenter.org/sites/default/files/files/publication/Combating%20Chinese%20Propaganda%20and%20Disinformation%20-%20Huang.pdf>.

<sup>144</sup> Jude Blanchett, et al., *Protecting Democracy in an Age of Disinformation: Lessons from Taiwan* (Center for Strategic & International Studies, January 2021), [https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/210127\\_Blanchette\\_Age\\_Disinformation.pdf](https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/210127_Blanchette_Age_Disinformation.pdf).

<sup>145</sup> Elizabeth Lange and Doowan Lee, “How One Social Media App is Beating Disinformation,” *Foreign Policy* (November 23, 2020), <https://foreignpolicy.com/2020/11/23/line-taiwan-disinformation-social-media-public-private-united-states/>.

<sup>146</sup> g0v community contributor in discussion with the author, December 2020.

Government	Social Media Platforms	Civil Society Organizations
<p> <b>Amplification:</b> The government created “<u>meme engineering</u>” teams within each government agency to use humor campaigns to respond to disinformation in an engaging manner.<sup>147</sup></p>	<p> <b>Response:</b> As a result of advocacy and persistent engagement from the government and CSOs, Facebook launched its real-time “<u>war room</u>” to prepare for Taiwan’s 2020 elections, enabling its family of apps to take down inauthentic behavior and false content in real time, and significantly strengthening the information environment as compared to the 2018 elections.<sup>148</sup></p>	<p> <b>Resilience:</b> CSOs launched digital literacy, public awareness and education campaigns to strengthen critical thinking skills and empower citizens with credible information while avoiding hyperpartisanship and politicization of the content.<sup>149</sup> This included:</p> <ul style="list-style-type: none"> <li>● Breaking down echo chambers by fostering intergenerational dialogues with <u>older adults</u>.</li> <li>● The “<u>Take back the TV remote</u>” initiative, where students refused to watch television that disproportionately covered pro-China stories.<sup>150</sup></li> </ul>
<p> <b>Resilience:</b> The government established a digital literacy curriculum for students and called for the <u>media literacy committee</u> to ensure the digital literacy curriculum is appropriately implemented.<sup>151</sup></p>	<p> <b>Resilience:</b> Facebook held digital literacy events in partnership with third-party fact-checking organizations such as Taiwan FactCheck Center, Cofacts, MyGoPen and Doublethink Lab to educate citizens about deciphering credible information from false content.</p>	
<p> <b>Legal:</b> The government enacted regulations, including penalties for spreading disinformation or rumors and interfering in local elections, such as amending the <u>Presidential and Vice Presidential Election and Recall Act</u><sup>152</sup> in May 2020 and passing the <u>Anti-Infiltration Law</u>.<sup>153</sup></p>		

<sup>147</sup> Anne Quito, “Taiwan is Using Humor as a Tool Against Coronavirus Hoaxes,” Quartz (June 5, 2020), <https://qz.com/1863931/taiwan-is-using-humor-to-quash-coronavirus-fake-news/>.

<sup>148</sup> Tzu-ti Huang, “Facebook Releases Report on Fight Against Disinformation in Run-Up to Taiwan Elections,” *Taiwan News* (October 6, 2020), <https://www.taiwannews.com.tw/en/news/4024275>.

<sup>149</sup> gOv community contributor in discussion with the author, December 2020.

<sup>150</sup> Olivia Yang, “Defending Democracy through Media Literacy,” *Taiwan Democracy Bulletin* 3, no. 6 (October 9, 2019), <https://bulletin.tfd.org.tw/tag/fake-news-cleaner/>.

<sup>151</sup> Nicola Smith, “Schoolkids in Taiwan Will Now Be Taught How to Identify Fake News,” TIME (April 7, 2017) <https://time.com/4730440/taiwan-fake-news-education/>; Sam Robbins, “Taiwan’s Push for Media Literacy - Is it All Fake News?” *Taiwan Insight* (March 27, 2020), <https://taiwaninsight.org/2020/03/27/taiwans-push-for-media-literacy-is-it-all-fake-news/>.

<sup>152</sup> Presidential and Vice Presidential Election and Recall Act, Amended May 6, 2020, Republic of China (Taiwan) Ministry of Interior, <https://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=D0020053>.

<sup>153</sup> Mainland Affairs Council, Republic of China (Taiwan), “Legislative Yuan Passes Anti-Infiltration Bill to Strengthen Defense for Democracy and Preserve Stable and Orderly Cross-Strait Exchanges” (press release), December 31, 2019, [https://www.mac.gov.tw/en/News\\_Content.aspx?n=2BA0753C-BE348412&s=88E5E1EF1343B1B8](https://www.mac.gov.tw/en/News_Content.aspx?n=2BA0753C-BE348412&s=88E5E1EF1343B1B8). It is important to note that while Taiwan passed a number of laws and regulations to counter disinformation and prevent foreign election interference, legal routes must be reviewed carefully to ensure they are in accordance with democratic principles and human rights.



## Lessons from Taiwan for a Whole-of-Society Response to Information Manipulation

As other civil society organizations, governments and citizens think about how to counter information manipulation during elections and civil unrest, Taiwan's whole-of-society approach provides best practices that can be replicated in other country and regional contexts.

### **Collaboration and private-public-civic partnerships are critical.**

Taiwan's approach centered on building strong private-public-civic partnerships in a short amount of time to enable the rapid identification, response and resilience to information manipulation. While launching initiatives such as digital literacy can take time, others, such as forming partnerships with fact-checking organizations and social media companies, can happen with expediency in order to disrupt external and internal information manipulation and debunk false content and narratives.

### **Know your audience and communicate with empathy.**

Taiwan has shown that a one-size-fits-all approach does not work in combating information manipulation and strengthening resilience. Taiwanese civil society and government created public awareness campaigns and resilience efforts to educate different segments of the population across generations. Key to these efforts was empathy, compassion and breaking down societal barriers that often decrease internal cohesion.

### **Creativity and innovation are critical to success.**

Taiwan has shown that creativity and innovation are critical to identifying, responding to and building resilience to information manipulation. The use of memes indicated that humor is a creative and powerful tool to debunk false narratives and amplify credible content. Memes are inexpensive and highly effective, and they can be replicated by other civil society actors and governments.

## Appendix B: Additional Information on Social Media Platforms

### Overview of Social Media Platforms’ Policies

The table below provides links to and highlights of key social media platforms’ policies relevant to their efforts to limit the spread of elections-related mis/disinformation.



Platform	Key Highlights
<p><b>Facebook Community Standards</b><sup>154</sup></p> 	<p>Facebook’s Community Standards do not presently ban mis/disinformation generally but do prohibit <u>content that misrepresents information about voting or elections, incites violence, and promotes hate speech</u>. Also, the Community Standards prohibit “<u>Coordinated Inauthentic Behavior</u>,” which is defined to generally prohibit activities that are characteristic of large-scale information operations on the platform.<sup>155</sup></p> <p>The company also has a responsibility to reduce the spread of “<u>false news</u>.” To operationalize this, Facebook commits to algorithmically reduce (or downrank) the distribution of such content, in addition to taking other steps to mitigate its impact and disincentivize its spread. The company has also developed a policy of removing particular categories of <u>manipulated media</u> that may mislead users; however, the policy is limited in scope. It extends only to media that is the product of artificial intelligence or machine learning and includes an allowance for any media deemed to be satire or content that edits, omits or changes the order of words that were actually said.<sup>156</sup></p> <p>In May 2021, Facebook launched a new <u>Transparency Center</u> that contains resources about its integrity and transparency efforts to users. This <u>new effort</u> shows how Facebook detects violations by using technology and review teams, and explains Facebook’s three-part approach to content enforcement: remove, reduce and inform.<sup>157</sup></p>

<sup>154</sup> Facebook, “Community Standards” (Facebook, n.d), <https://www.facebook.com/communitystandards/>.

<sup>155</sup> Facebook, “Community Standards: Coordinating Harm and Publicizing Crime” (Facebook, n.d), [https://www.facebook.com/communitystandards/coordinating\\_harm\\_publicizing\\_crime](https://www.facebook.com/communitystandards/coordinating_harm_publicizing_crime); Facebook, “Community Standards: Violence and Incitement” (Facebook, n.d), [https://www.facebook.com/communitystandards/credible\\_violence](https://www.facebook.com/communitystandards/credible_violence); Facebook, “Community Standards: Hate Speech” (Facebook, n.d), [https://www.facebook.com/communitystandards/hate\\_speech](https://www.facebook.com/communitystandards/hate_speech); Facebook, “Community Standards: Inauthentic Behavior,” [https://www.facebook.com/communitystandards/inauthentic\\_behavior/](https://www.facebook.com/communitystandards/inauthentic_behavior/).

<sup>156</sup> Facebook, “Community Standards: False News” (Facebook, n.d), [https://www.facebook.com/communitystandards/false\\_news](https://www.facebook.com/communitystandards/false_news); Facebook, “Community Standards: Manipulated Media” (Facebook, n.d), [https://www.facebook.com/communitystandards/manipulated\\_media](https://www.facebook.com/communitystandards/manipulated_media).

<sup>157</sup> Facebook Transparency Center (Facebook, n.d), <https://transparency.fb.com>; Facebook Transparency Center, “How We Enforce Our Policies” (Facebook, n.d), <https://transparency.fb.com/enforcement/>.

Platform	Key Highlights
<p><b>Twitter</b> <b>Rules</b> <sup>158</sup></p> 	<p>While there is no general policy on misinformation, Twitter’s Rules do include several provisions to address false or <u>misleading content</u> and behavior in specific contexts. Twitter’s policies prohibit disinformation and other content that may suppress participation or mislead people about when, where or how to participate in a <u>civic process</u> and content that includes hate speech or incites violence or harassment. Twitter also prohibits <u>inauthentic behavior and spam</u>. Related to disinformation, Twitter has updated its hateful conduct policy to prohibit language that dehumanizes people on the basis of race, ethnicity and national origin.<sup>159</sup></p> <p>Twitter’s <u>policies on elections</u> explicitly prohibit misleading information about the voting process. However, inaccurate statements about an elected or appointed official, candidate or political party are excluded from this policy.<sup>160</sup> Under these rules, Twitter has removed postings that feature disinformation about election processes, such as promoting the wrong voting day or false information about polling places—content that EMB election observers and others are increasingly working to monitor and report.</p>
<p><b>YouTube</b> <b>Community</b> <b>Guidelines</b> <sup>161</sup></p> 	<p>YouTube follows a three-strike policy that results in the suspension or termination of offending accounts related to disinformation. YouTube’s Community Guidelines include several provisions relevant to disinformation in particular contexts, including content that aims to mislead voters about the time, place, means, or eligibility requirements for voting or participating in a <u>census</u>; that advances false claims related to the <u>eligibility requirements</u> for political candidates to run for office and elected government officials to serve in office; or that promotes violence or hatred against or harassment of individuals or groups based on <u>intrinsic attributes</u>. In addition, YouTube has also expanded its anti-harassment policy that prohibits video creators from using hate speech and insults on the basis of gender, sexual orientation or race.<sup>162</sup></p> <p>YouTube has also developed a policy regarding manipulated media, which prohibits content that has been technically manipulated or doctored in a way that misleads users (beyond clips taken out of context) and may pose a serious risk of egregious harm. To further mitigate risks of manipulation or disinformation campaigns, YouTube also has policies that prohibit account impersonation, misrepresenting one’s country of origin or concealing association with a government actor. These policies also prohibit <u>artificially increasing engagement metrics</u>, either through the use of automatic systems or by serving up videos to unsuspecting viewers.<sup>163</sup></p>

<sup>158</sup> Twitter, “The Twitter Rules” (Twitter, n.d.), <https://help.twitter.com/en/rules-and-policies/twitter-rules>.



<sup>159</sup> Twitter Help Center, “Civic Integrity Policy” (Twitter, n.d.), <https://help.twitter.com/en/rules-and-policies/election-integrity-policy>; Twitter Help Center, “Platform Manipulation and Spam Policy,” (Twitter, n.d.), <https://help.twitter.com/en/rules-and-policies/platform-manipulation>.

<sup>160</sup> Twitter Help, “Civic Integrity Policy.” <https://help.twitter.com/en/rules-and-policies/election-integrity-policy>

<sup>161</sup> YouTube Help, “YouTube’s Community Guidelines” (YouTube, n.d.), <https://support.google.com/youtube/answer/9288567>.

<sup>162</sup> YouTube Help, “Spam, Deceptive Practices and Scams Policies” (YouTube, n.d.), <https://support.google.com/youtube/answer/2801973?hl=en>; YouTube Help, “Hate Speech Policy” (YouTube, n.d.), [https://support.google.com/youtube/answer/2801939?hl=en&ref\\_topic=9282436](https://support.google.com/youtube/answer/2801939?hl=en&ref_topic=9282436).

<sup>163</sup> YouTube Help, “Spam, Deceptive Practices and Scams Policies”; YouTube Help, “Impersonation Policy” (YouTube, n.d.), <https://support.google.com/youtube/answer/2801947?hl=en>; YouTube Help, “Fake Engagement Policy” (YouTube, n.d.), <https://support.google.com/youtube/answer/3399767?hl=en>.

Platform	Key Highlights
<p>TikTok <b>Community Guidelines</b><sup>164</sup></p> 	<p>In August 2020, TikTok updated its Community Guidelines prohibiting content that “misleads people about elections or other civic processes, content distributed by disinformation campaigns, and health misinformation.”<sup>165</sup> TikTok added a policy that “prohibits synthetic or manipulated content that misleads users by distorting the truth of events in a way that could cause harm.” This includes banning deepfakes in order to prevent the spread of disinformation. TikTok also increased the transparency of its policy around <u>coordinated inauthentic behavior</u>.<sup>166</sup></p>
<p>Snapchat <b>Community Guidelines</b><sup>167</sup></p> 	<p>In January 2017, Snapchat created policies to combat the spread of disinformation for the first time. Snapchat implemented policies for its news providers on the platform’s Discover page in order to combat disinformation and regulate information that is considered inappropriate for minors. These new guidelines require news outlets to <u>fact-check</u> their articles before they can be displayed on the platform’s Discover page.<sup>168</sup></p> <p>In an op-ed, Snapchat CEO Evan Spiegel described the platform as different from other types of social media and many other platforms, saying “content designed to be shared by friends is not necessarily content designed to deliver accurate information.” There isn’t a feed of information from users on Snapchat like there is with many other social media platforms—a distinction that makes Snapchat more comparable to a <u>messaging app</u>. With Snapchat’s updates, the platform makes use of <u>human editors</u> who monitor and regulate what is promoted on the Discover page, preventing the spread of false information.<sup>169</sup></p>

## Overview of Social Media Platforms’ Product Features and Interventions

Type of Platform	Key Example of Platform Efforts through Product Features and Technical/Human Interventions
<p><b>Traditional Social Media Companies</b></p>	<p><b>Facebook</b> uses algorithmic strategies to downrank false or disputed information, decreasing the content’s visibility in the News Feed; applies distribution limits against Pages and websites of repeat offenders; and employs notifications to users who have engaged with mis/disinformation.</p>

<sup>164</sup> “TikTok Community Guidelines” (TikTok, n.d.), <https://www.tiktok.com/community-guidelines?lang=en#37>.

<sup>165</sup> Vanessa Pappas, “Combating Misinformation and Election Interference on TikTok,” (TikTok, August 5, 2020), <https://newsroom.tiktok.com/en-us/combating-misinformation-and-election-interference-on-tiktok>.

<sup>166</sup> Nick Statt, “TikTok is Banning Deepfakes to Better Protect Against Misinformation,” The Verge, (August 5, 2020), <https://www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference>; Vanessa Pappas, “Combating Misinformation and Election Interference on TikTok.”

<sup>167</sup> Snap Inc., “Snapchat Community Guidelines” (Snap Inc., n.d.), <https://www.snap.com/en-US/community-guidelines>.

<sup>168</sup> Zameena Meja, “Snapchat Wants to Make on its Platform Disappear, Too,” Quartz (January 23, 2017), <https://qz.com/892774/snapchat-quietly-updates-its-guidelines-to-prevent-fake-news-on-its-discover-platform/>.

<sup>169</sup> Evan Spiegel, “How Snapchat is Separating Social from Media,” Axios, (November 29, 2017), <https://www.axios.com/how-snapchat-is-separating-social-from-media-2513315946.html>; Jamie Condliffe, “Snapchat Has a Plan to Fight Fake News: Ripping the ‘Social’ from ‘Media,’” *MIT Technology Review* (November 29, 2017), <https://www.technologyreview.com/2017/11/29/147413/snapchat-has-a-plan-to-fight-fake-news-ripping-the-social-from-the-media/>; Daniel Funke, “Here’s Why Snapchat’s Latest Update Further Insulates it from Fake News” (Poynter, December 1, 2017), <https://www.poynter.org/fact-checking/2017/heres-why-snapchats-latest-update-further-insulates-it-from-fake-news/>.

Type of Platform	Key Example of Platform Efforts through Product Features and Technical/Human Interventions
<b>Traditional Social Media Companies</b>	<p><b>Twitter</b> uses automated prompts cautioning users against sharing links they have not opened, intending to “promote informed discussion” and encourage users to evaluate information before sharing it. This follows the introduction of content labels and warnings, which the platform has affixed to tweets that are not subject to removal under the platform’s policies (or under the company’s “public interest” exception) but which nonetheless may include misinformation or <u>manipulated media</u>.<sup>170</sup></p> <p><b>Instagram</b> removes content identified as misinformation from hashtags and from its Explore page and makes accounts that repeatedly post misinformation harder to find by filtering content from that account from <u>searchable pages</u>.<sup>171</sup></p> <p><b>TikTok</b> uses technology to augment its content moderation practices, particularly to assist in identifying inauthentic behavior, patterns and accounts dedicated to spreading misleading or spam content. The company notes that its tools enforce its rules and make it more difficult to find harmful content, like misinformation and conspiracy theories, in the platform’s recommendations or search features.</p> <p><b>YouTube</b> similarly employs technology, particularly machine learning, to augment its efforts.<sup>172</sup> As the company notes in its policies, “machine learning is well-suited to detect patterns, which helps us to find content similar to other content we’ve already removed, even before it’s viewed.”</p>
<b>Messaging Applications</b>	<p><b>WhatsApp</b> introduced limits on message forwarding in 2018—which prevent users from forwarding a message to more than five people—as well as visual indicators to ensure that users can distinguish between forwarded messages and original content. In the context of the COVID-19 pandemic, WhatsApp further limited forwarding by announcing that messages that have been forwarded more than five times can subsequently only be shared with one user at a time. WhatsApp also developed systems for identifying and taking down automated accounts that send high volumes of messages. WhatsApp is currently experimenting with methods to detect patterns in <u>messages</u> through homomorphic encryption evaluation practices.<sup>173</sup> These strategies may help to inform analysis and technical interventions related to disinformation campaigns in the future.</p>

<sup>170</sup> Yoel Roth and Nick Pickles, “Updating our Approach to Misleading Information,” *Twitter Product* (blog), May 11, 2020, [https://blog.twitter.com/en\\_us/topics/product/2020/updates-our-approach-to-misleading-information](https://blog.twitter.com/en_us/topics/product/2020/updates-our-approach-to-misleading-information).

<sup>171</sup> Guy Rosen et al., “Helping to Protect the 2020 US Elections,” (Facebook, updated January 27, 2020), <https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/>.

<sup>172</sup> YouTube Help, “YouTube’s Community Guidelines.”

<sup>173</sup> Himanshu Gupta and Harsh Taneja, “WhatsApp has a fake news problem—that can be fixed without breaking encryption,” *Columbia Journalism Review*, August 23, 2018, [https://www.cjr.org/tow\\_center/whatsapp-doesnt-have-to-break-encryption-to-beat-fake-news.php](https://www.cjr.org/tow_center/whatsapp-doesnt-have-to-break-encryption-to-beat-fake-news.php).

Type of Platform	Key Example of Platform Efforts through Product Features and Technical/Human Interventions
Search Engines	<p><b>Google</b> changed its search algorithm to combat fake news dissemination and conspiracy theories. In a blog post, Google Vice President of Engineering Ben Gomes wrote that the company will “help surface more authoritative pages and demote low-quality content” in <a href="#">searches</a>.<sup>174</sup> In an effort to provide improved search guidelines, Google is adding real people to act as evaluators to “assess the quality of Google’s search results—give us feedback on our <a href="#">experiments</a>.”<sup>175</sup> Google will also provide “direct feedback tools” to allow users to flag unhelpful, sensitive or inappropriate content that appears in their searches.</p>

## Appendix C: Additional Resources

There are many resources available to aid in the identification, response, and building of resilience to disinformation. Please reference this spreadsheet, which will be continuously updated, for a growing list of tools and resources:

[The Information Manipulation Resource Annex](#)

<sup>174</sup> The Keyword, “Our latest quality improvements for search” (Google, April 25, 2017) <https://blog.google/products/search/our-latest-quality-improvements-search/>.

<sup>175</sup> Ben Gomes, “Our Latest Quality Improvements for Search,” The Keyword (blog), Google (April 25, 2017), <https://blog.google/products/search/our-latest-quality-improvements-search/>.





**Stanford** | Internet Observatory  
Cyber Policy Center