Global Network on Electoral Justice

Observatory on Social Media

Disinformation campaigns cast doubt on the integrity of the process, electoral institutions, or electoral outcome

Version: October 2023

General Coordination: Board of the Observatory on Judicial Independence

Design Coordination: Technical Secretariat of the GNEJ

Research and Writing: Kristina Wilfore

Research Assistance: Sarah Hesterman

TABLE OF CONTENTS

GLOSSARY OF TERMS

Artificial Intelligence (AI): Artificial intelligence leverages computers and machines to mimic the problem-solving and decision-making capabilities of the human mind. Specific applications of AI include expert systems, natural language processing, speech recognition, and machine vision. Generative AI is a type of artificial intelligence technology that can produce various types of content, including text, imagery, audio and synthetic data. The recent attention around generative AI has been driven by the simplicity of new user interfaces for creating high-quality text, graphics, and videos in a matter of seconds.

Astroturfing: The practice of artificially creating the appearance of grassroots support or opposition for a particular cause or organization. It involves using deceptive tactics, such as creating fake social media accounts or websites, to manipulate public opinion.

Bot: An automated software program that performs tasks on the internet. In the context of misinformation, bots can be used to amplify or spread false information, often through social media platforms.

Clickbait: Sensational or misleading headlines and content designed to attract attention and generate website traffic. It often exaggerates or misrepresents information to entice users to click on a link.

Deepfake: Synthetic media that combines artificial intelligence (AI) with manipulated audio or visual elements to create realistic but fabricated content. Deepfakes can be used to create convincing fake videos or audios of people saying or doing things they never did.

Disinformation: Purposefully false or misleading information created and spread with the intent of doing harm. Harm could be directed at a person, social group, organization, or country. The goal of disinformation is to be believed. As such, disinformation can include some true facts, but are stripped of context or blended with falsehoods to support the intended message. Misinformation and disinformation can appear in political ads or social media posts. They can include fake news stories or doctored videos.

Definitions of misinformation or disinformation become complex when someone unknowingly shares a campaign message intended to suppress voter turnout, for example, such as a post with the wrong voting date. While the person sharing this

information may not realize it is false, they are unwittingly part of a disinformation campaign and often a key ingredient to the virality of false information.

Echo chamber: An online environment or social network where people are exposed to information and opinions that reinforce their existing beliefs or biases. In an echo chamber, dissenting or contradictory views are often ignored or suppressed.

Election related mis- and disinformation: False and misleading information pertaining to electoral processes, including information about voting times and locations, candidates and election officials, and the overall integrity of the election. This type of manipulation can have influence over voter behavior, disrupt electoral infrastructure, impact voting outcomes, mobilize voters based on lies, call into question the results, and undermine trust in democratic processes.

A common tactic to undermine elections is to confuse voters about the voting process (the time, place, and manner of the election) or to inflame security threats around voting so that people chose to stay home ("self-suppress") due to worries about intimidation, violence, or other consequences. Election disinformation can alter public perceptions about elections and their security, thereby impacting legislation and democratic norms in the long run.

Encrypted messaging applications (EMAs): EMAs disallow third-party access to information through end-to-end encryption, which means a sender's messages are coded to be protected from interception during transmission and then reverted back to the original text or file when received using cryptography algorithms. Popular EMAs are WhatsApp, Telegram and Signal.

Fact-checking: The process of verifying the accuracy and validity of information, claims, or statements. Fact-checking involves researching and analyzing evidence from reliable sources to assess the truthfulness of a claim.

Fake News: False or misleading information presented as legitimate news. It can be fabricated, distorted, or completely made up to resemble credible journalism, and is often shared through traditional or social media platforms. While some countries refer to mis- and disinformation as fake news, the concept has evolved to be understood as the specific phenomenon of creating and sharing fabricated stories intentionally created to look like legitimate sources of news, while mis- and disinformation are broader concepts encompassing many types of false information.

Filter bubble: The isolation of individuals within personalized online environments that present information aligned with their pre-existing preferences and interests. This can limit exposure to diverse perspectives and contribute to the reinforcement of existing beliefs.

Gendered disinformation: The spread of deceptive or inaccurate information and images against women political leaders, journalists, and public figures, following storylines that often draw on misogyny, as well as gender stereotypes around the role of women. Both state and non-state actors strategically use gendered disinformation to silence women, discourage online political discourse, and shape perceptions toward gender and the role of women in societies.[1]

Hoax: A deliberate deception or fabrication, typically spread through various channels, such as social media, emails, or websites. Hoaxes are created to trick and mislead people, often playing on their fears or desires.

Information pollution: Encompasses verifiably false, misleading and manipulated content on- and offline, which is created, produced, and disseminated intentionally or unintentionally, and has the potential to cause harm.[2]

Malinformation: When genuine information is shared with the intent to cause harm.[3] Examples include the intentional leakage of a politician's private emails, which is what happened during the presidential elections in both the U.S. and France in 2017.[4] So-called "revenge porn," is another example of malinformation. "Image-based abuse" is a more accurate catch-all term that entails the use of intimate or faked sexually explicit images without consent.

---

[1] U.S. Department of State, "Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors," March 27, 2023, https://www.state.gov/gendered-disinformation-tactics-themes-and-trends-by-foreign-malign-actors/; #ShePersisted, "The Problem," accessed September 14, 2023, https://she-persisted.org/the-problem.

[2] *Strategic Guidance: Information Integrity: Forging a pathway to Truth, Resilience and Trust* (UNDP, February 2022), https://www.undp.org/sites/g/files/zskgke326/files/2022-02/UNDP-Information-Integrity-Forging-a-Pathway-to-Truth-Resilience-and-Trust.pdf.

[3] California State University San Marcos, "Misinformation and Disinformation," accessed September 14, 2023, https://www.csusm.edu/elections/get-informed/misinformation.html .

[4] During the 2017 presidential elections in France, then-candidate Emmanuel Macron's campaign team reported hundreds of internal documents had been leaked in a hacking operation, noting the real documents were being shared alongside fake documents on social media. See: "Macron Campaign Says Massive Email Leaks Meant to Undermine It," Reuters, May 6, 2017, https://www.reuters.com/article/us-france-election-macron-response/macron-campaign-says-massive-email-leaks-meant-to-undermine-it-idUSKBN1812DA.

Sophisticated bad actors can use new technology, taking interference to the next level. Against the backdrop of Germany's 2021 race for a chancellor, Deutsche Welle explained how an election can be hacked — and what can be done to protect them.[5]

Misinformation: False or misleading information that was not purposefully created or spread with intent to harm, but can nevertheless lead to harm as "honest mistakes." Someone posting an article containing out-of-date information about voting times or voter registration without realizing it is wrong, is misinformation. The intent of the person sharing such information is what distinguishes misinformation from disinformation.

Mitigation: The reduction of something harmful or the reduction of its harmful effects. It may refer to measures taken to reduce the harmful effects of opposition, or to manage harmful incidents that have already occurred.

Online gendered abuse or technology-facilitated gender based violence: Online gendered abuse refers to a spectrum of activities and behaviors that involve technology as a central aspect of perpetuating violence, abuse, or harassment against (both cis and trans) women.[6]

Propaganda: Information, often biased or misleading, used to promote a particular political, religious, or ideological agenda. It aims to shape public opinion and influence behavior by appealing to emotions rather than rationality.[7]

Rumor: An unverified or unconfirmed piece of information or story that is circulated informally, often through word-of-mouth or social media. Rumors can spread quickly and may lack credibility or evidence.

Troll: An individual who deliberately posts inflammatory, offensive, or provocative content online to provoke and disrupt discussions. Trolls aim to generate emotional responses and cause discord rather than engage in genuine conversation. Trolls are oftentimes paid provocateurs orchestrating harassment campaigns and spreading disinformation.

Troll farms: Troll farms are professionalized groups that work in a coordinated fashion to spread mis- and disinformation through fake profiles and accounts that appear to belong to real people. Some campaigns are designed to amplify bogus support for

---

[5] DW News, "How to Hack an Election: Cyber Threats to Democracy | DW News," YouTube, September 5, 2021, https://www.youtube.com/watch?v=vrk2kO2eg4M&t=24s.

[6] UNFPA, "16 Days of Activism against Gender-Based Violence - The Background," accessed September 14, 2023, https://www.unfpa.org/thevirtualisreal-background#glossary.

[7] James Wallner, "Recognizing Propaganda in Politics," January 15, 2020, https://www.legislativeprocedure.com/blog/2020/1/15/recognizing-propaganda.

political ideas and meddle in elections; others create overall confusion and distrust of democratic institutions. Private firms straddling traditional marketing and the world of geopolitical influence operations are now selling services once conducted principally by intelligence agencies.[8]

Today's information environment is more chaotic and easier to manipulate than ever before. For more in the way of definitions, refer to the ***Verification Handbook for Disinformation and Media Manipulation*** *produced for journalists by the European Journalism Centre guiding information manipulation investigations.*[9]

[8] Max Fisher, "Disinformation for Hire, a Shadow Industry, Is Quietly Booming," July 25, 2021, https://www.nytimes.com/2021/07/25/world/europe/disinformation-social-media.html.

[9] *Verification Handbook For Disinformation And Media Manipulation* (Maastricht: European Journalism Centre, 2020), https://datajournalism.com/read/handbook/verification-3 .

INTRODUCTION & SUMMARY

The distortion of facts, data, and analysis in political and civil discourse heightened during election periods is one of the most prevalent challenges to democracy today. While citizens have direct access to more information than ever before, the rise of the internet and the advent of social media fundamentally alters the information ecosystem before, during, and after elections, blurring the lines between accurate information and false or low-quality content. Elections are best thought of as a continuous, integrated process made up of building blocks that interact with and influence each other, rather than a series of isolated events. Election-related mis-and disinformation should be evaluated within the same framework.[10]

Disinformation –false or intentionally misleading information that aims to achieve an economic or political goal – poses a significant risk to the integrity of elections, the primary system through which citizens exercise their right to vote and elect their representation in government.[11] While anti-democracy disinformation campaigns are not new, technology and the ubiquitous existence of social media have changed the scope and scale of efforts meant to undermine elections.

The volume of election-related mis-and disinformation in recent elections is unprecedented compared to previous electoral cycles, and inherently linked with social media platforms that facilitate the amplification of false information.[12] Citizens increasingly use social media platforms to access information about voting, candidates' platforms, and key political debates which "the ubiquity and speed of the internet enables information, including misinformation and disinformation, to spread rapidly and virally."[13]

When unmitigated, the information pollution resulting from disinformation exerts undue influence on political debate and election outcomes, hindering citizens' ability to make informed choices and participate in genuinely democratic electoral processes. A polluted information ecosystem around elections can influence voter behaviors, undermine the credibility of candidates and election officials, and erode trust in democratic process and

---

[10] USAID, "Supporting Free and Fair Elections," accessed September 14, 2023, https://www.usaid.gov/democracy/supporting-free-and-fair-elections.

[11] *Social Media, Disinformation and Electoral Integrity: IFES Working Paper* (IFES, August 2019), https://www.ifes.org/sites/default/files/migrate/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf.

[12] NDI, "Disinformation, Social Media, and Electoral Integrity," accessed September 14, 2023, https://www.ndi.org/disinformation-social-media-and-electoral-integrity.

[13] https://unesdoc.unesco.org/ark:/48223/pf0000370634

electoral institutions while exacerbating political polarization. Disinformation is aiding the global decline in trust in institutions (such as media and government) which used to be looked at as authoritative sources of factual information, making the challenges of combating mis-and disinformation even more complex.[14]

In recent elections, disinformation has posed significant harm to individuals involved in and facilitating elections, including election workers, election officials, election observers in the form of damaging reputations and credibility. Online attacks on individuals can ultimately erode trust in democratic processes and undermine democratic participation altogether when these narratives reach wide audiences. In 2019, Freedom House noted the utilization of "informational measures" and content manipulation by state and non-state actors to "distort the media landscape during elections" as the most popular tactic for digital election interference, and in 2022, International IDEA identified that since 2016, there has been an "ascending trend" of "cases of disinformation against elections," including that targeting "processes, organizations," and "individuals supporting the management of the processes."[15]

Digital platforms can be weaponized to simulate artificial momentum and run inorganic operations, with social media companies that have failed to dedicate resources to content moderation and tackling mis- and disinformation content, especially in non-English speaking countries where such faked efforts go unaverted.

As more and more sophisticated Artificial Intelligence (AI) technologies continue to emerge, these will only serve to strengthen the effectiveness and spread of election mis- and disinformation. New user-friendly applications allow people to generate inauthentic text, audio, visual, and audio-visual content that is inexpensive to create and increasingly realistic. Such tools provide a new avenue for malign actors seeking to exert influence over voters, target election officials and candidates, and spread false narratives about elections to erode trust in democratic processes and institutions.[16] Generative tools can be utilized to create fake but realistic looking conversations between poll workers discussing throwing out ballots, for example, or generate a deepfake video portraying a candidate engaging in corrupt activities, with misleading social media posts that can be quickly shared by users. Experts and industry leaders are fearful of generative AI's ability

[14] *Edelman Trust Barometer 2022* (Edelman, 2022), https://www.edelman.com/trust/2022-trust-barometer.

[15] *Freedom on the Net 2019* (Freedom House, 2019), https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf;, International IDEA, "The Information Environment Around Elections," accessed September 14, 2023, https://www.idea.int/our-work/what-we-do/elections/information-environment-around-elections.

[16] Mekela Panditharatne & Noah Giansiracusa, "How AI Puts Elections at Risk — And the Needed Safeguards," last updated July 21, 2023, https://www.brennancenter.org/our-work/analysis-opinion/how-ai-puts-elections-risk-and-needed-safeguards; Darrell M. West, "How AI Will Transform the 2024 Elections," May 3, 2023, https://www.brookings.edu/articles/how-ai-will-transform-the-2024-elections/.

to churn out convincing disinformation at unprecedented speeds, noting this will have a profound effect on future elections if unmitigated.[17] AI-generated disinformation is already starting to be deployed in elections this year, and is perceived by experts as posing immense risks that could generate widespread, global harm to electoral integrity in 2024.[18]

With this context, it is clear that the role of social media in elections must be urgently addressed through a multi-stakeholder approach, with specific attention paid to supporting the roles of election management bodies and electoral justices in tackling this pressing challenge to democracy.

This white paper uncovers the role of disinformation in impacting election integrity globally. It identifies mis- and disinformation in elections as negatively impacting electoral processes and democracy, and provides crucial insight into the role of foreign influence. The paper assesses actions that can be taken to curtail online mis- and disinformation harms, including policy responses, and the role of gendered disinformation in eroding election integrity. It features case studies of court proceedings addressing election disinformation and actions justices have taken on election-related cases. The paper provides key advice for how to identify information pollution and assess its impact on elections, and addresses the challenges of ruling on election disinformation. Finally, this paper provides advice to electoral justices as key stakeholders in ensuring the integrity of elections.

METHODOLOGY

This white paper is the result of desk research that included a critical examination of existing academic literature, reports, and publications related to disinformation in election contexts and its impact on democracy. This report is grounded in an understanding of international principles as applied to digital activity and election integrity and draws on the work of key stakeholders in the international development arena as practitioners. Analysis of news articles, opinion pieces, and media coverage related to election-related mis-and disinformation was conducted. Case studies throughout provide insights into the strategies employed to respond to election integrity attacks in the information environment. This report was commissioned by the Global Network on Electoral Justice

[17] Tiffany Hsu and Stuart A. Thompson, “Disinformation Researchers Raise Alarms About A.I. Chatbots,” February 9, 2023, https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html; Diane Bartz, Zeba Siddiqui, & Jeffrey Dastin, “OpenAI Chief Concerned About AI Being Used to Compromise Elections,” May 17, 2023, https://www.reuters.com/technology/openai-chief-goes-before-us-congress-propose-licenses-building-ai-2023-05-16/.

[18] Daniel Zuidijk, “Deepfakes in Slovakia Preview How AI Will Change the Face of Elections,” October 4, 2023, https://www.bloomberg.com/news/newsletters/2023-10-04/deepfakes-in-slovakia-preview-how-ai-will-change-the-face-of-elections; Jennifer Huddleston, “AI and the Future of Our Elections,” September 27, 2023

(GNEJ) and written by Kristina Wilfore, a campaign and elections expert and specialist in countering disinformation, with inputs and direction provided by the GNEJ.

PART 1: INTERNATIONAL PRINCIPLES OF ELECTION INTEGRITY

As the 2024 elections near, in which two billion people will be eligible to vote for the first time in history, election integrity must remain a foremost priority to preserve the rights of citizens to participate in elections free of undue influence around the world.[19] Democracy is at its strongest when citizens can have full faith in their ability to participate in credible and transparent electoral processes, and in electing representation uninfluenced by manipulation, deception, and foreign interference.

Social media platforms can aid in accessing real-time election information by mobilizing voters and campaigning for office through more intimate access to constituents. Yet their proliferation in the last twenty years has created new dynamics surrounding the spread of mis- and disinformation, undermining perceptions about the integrity of the election. Understanding the role of international law in maintaining election integrity and social media's role in affirming or undermining international principles for fair elections is thus an important step towards successfully safeguarding electoral processes free from information pollution that can influence voter behavior, lessen the credibility of candidates and election officials, and erode trust in democratic processes and institutions.

International Frameworks for Credible Elections

Fifty years ago, less than half of the world's nations chose their leaders by elections; now, almost all countries do. The United Nations General Assembly has stated on many occasions that there is no single model of democracy, nor one size that fits all. While it is each country's sovereign right to choose how to conduct its elections, UN Member States have agreed to abide by a set of obligations and commitments to protect and promote the electoral rights of their citizens. "While different, the range of democratic systems does share one important similarity—an intricate link with the civil and political rights and obligations enshrined in the UN Charter and various UN and regional instruments" stated Jeffrey Feltman, who served as the UN Under-Secretary-General for Political Affairs from 2012 until 2018 and as the Focal Point for Electoral Assistance.[20]

---

[19] Odanga Madung, "Brazil, Kenya, the US – Tech Giants Are Putting Democracy in Peril the World Over," January 25, 2023, https://www.theguardian.com/global-development/2023/jan/25/brazil-kenya-the-us-tech-giants-are-putting-democracy-in-peril-the-world-over.

[20] *International Obligations for Elections: Guidelines for Legal Frameworks* (International IDEA, 2014), https://www.idea.int/sites/default/files/publications/international-obligations-for-elections.pdf.

A credible election is defined as one in which the election outcome reflects the free expression of the will of the people. This is best achieved through elections that are transparent, inclusive, and accountable with equitable opportunities for competition. Such principles are buttressed by several electoral process-related obligations, as well as a number of key rights and freedoms, each of which derives from public international law.

For example, Article 21 of the Universal Declaration of Human Rights, adopted in 1948, provides the foundation for international law stating, "the will of the people shall be the basis of the authority of government; this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures."[21] Of particular relevance to elections are General Comments 25 and 34 on the International Covenant on Civil and Political Rights, adopted by the UN General Assembly in 1966 and made effective in 1976, which pertain to "the right to participate in public affairs, voting rights and the right of equal access to public service" and safeguarding the freedom of expression, including a "free, uncensored and unhindered press," respectively.[22] These comments proclaim the rights of voters to develop opinions "free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind," the right to access accurate information to aid in making informed decisions, and the general obligation of electoral institutions to be accountable to making electoral information transparent and available to voters for this reason. Similar principles are also enshrined in the European Court of Human Rights and the Inter-American Commission on Human Rights.

The Lack of International Frameworks Governing Social Media

Technology is not inherently democratic or undemocratic. It does not automatically "level the playing field," nor give everyone a voice, or create the conditions for objective reality or credible elections. In fact, the technology currently embraced in modern society with design and growth focused on scalability, efficiency, and market potential can work in opposition to values such as equity, agency, and protection of vulnerable populations.[23]

While nation-states have gone to great lengths to define international law in support of election integrity, and commit to the principles of free elections, digital platforms as private

---

[21] United Nations, "Universal Declaration of Human Rights," accessed September 14, 2023, https://www.un.org/en/about-us/universal-declaration-of-human-rights.

[22] *General Comment No. 34: Article 19: Freedoms of Opinion and Expression* (International Covenant on Civil and Political Rights, Human Rights Committee, 2011), https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf; *General Comment 25: The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service* (Office of the High Commissioner for Human Rights, 1996), https://www.equalrightstrust.org/ertdocumentbank/general%20comment%2025.pdf.

[23] Janet Haven and danah boyd, "Philanthropy Must Rethink Its Support of Technology Solutions That Harm Democracy," November 30, 2020, https://www.philanthropy.com/article/philanthropy-must-rethink-its-support-of-technology-solutions-that-harm-democracy.

companies are not regulated by international law and in most countries, not regulated by domestic legislation. Outside of the European Union and a handful of countries in the West, most social media is not regulated in any fashion other than self-regulation. To fill this void, international bodies have developed basic principles and standards for the use of technologies, as outlined in UNESCO's draft Guidelines for Regulating Digital Platforms.[24] The recently approved European Union's Digital Service Act has set a new standard for holding social media companies accountable for the harms facilitated by their platforms, with regulatory measures encompassing illegal content, disinformation, data privacy violations, and more (further addressed in part 5 of this paper covering policy and regulatory responses to election-related information pollution).

Legal frameworks at the national and international level governing the role of traditional media (newspapers, radio, and broadcast networks) in elections can be learned from as the regulation of the online world takes shape. However, it is important to understand that media laws are overwhelmingly directed at regulating the behavior of governments in relation to the media, rather than in regulating the media themselves.[25] Furthermore, social media and election oversight standards are not well developed and are often hotly contested in environments with high political polarization.

It is imperative that standards evolve so governments can strike a thoughtful balance between protecting freedom of expression and utilizing regulatory approaches to preserve international election principles, as to mitigate the risks to democratic processes and electoral integrity posed by disinformation, or on the other hand, implicated by digital platform regulation that carries unintended consequences.

Social media is fundamentally different from traditional media in important ways, which makes the oversight of online election discourse even more complex. First, the medium itself provides massive reach through internet-based platforms, giving social media companies infinite bandwidth, with millions of accounts that can each target much wider as well as narrower audiences. Second, not only is traditional media limited by the number of news media networks, but broadcast licensing and oversight is typically regulated by the state with jurisprudence established decades ago. Third, traditional news content throughout election cycles is typically produced with editorial oversight by producers with executives, which makes it easier for companies to supervise the content that is shared on their platforms - as well as easier for third parties to hold companies accountable for lies and distortion of facts. This is in contrast to social media, in which platforms are

[24] UNESCO, "Guidelines for Regulating Digital Platforms," accessed September 14, 2023, https://www.unesco.org/en/internet-conference/guidelines.

[25] Administration and Cost of Elections Project, "Media and Elections," accessed September 14, 2023, https://aceproject.org/ace-en/topics/me/mea/default,

merely conduits for user-generated content that is subject to moderation provided by the companies.

Experts such as Nobel Peace Prize-winning journalist Maria Ressa argue that too much attention is focused on content moderation choices of social media companies, which is a downstream problem of social media and its impact on democracy. “Move further upstream to algorithmic amplification. That’s the operating system; that’s where the micro-targeting is. What is an algorithm? Opinion in code. That’s where one editor’s decision is multiplied millions and millions of times. And that’s not even where the problem is. Go further upstream to where our personal data has been pulled together by machine learning to make a model of you that knows you better than you know yourself, and then all of that is pulled together by artificial intelligence.”[26]

Awareness of the role of digital technologies in exacerbating threats to election integrity is increasing. For example, Freedom House’s research initiative, “Election Watch for the Digital Age,” applies an Election Vulnerability Index to countries assessing concerns regarding election-related online harms, government control over the internet and user content, the state of the electoral system and fairness in political participation, human rights and other relevant standards.[27] In the context of elections, the introduction of new voting methods or election technologies has inspired mis- and disinformation about voting machine malfunctions, results transmissions, or the vulnerability to hacking, for example. Such accusations can spread quickly without evidence or mitigation on social media platforms whose algorithms are designed to prioritize content that generates high engagement, particularly if it is sensational and controversial.[28] Even if users engage with misleading content on electronic voting methods positively, to debunk false information, for example, this may signal to algorithms that this is a topic worthy of prioritizing, further aiding in the amplification of harmful narratives concerning the reliability of electronic voting methods.

It is reasonable to assume that with more public knowledge of how social media algorithms work to increase engagement online, as well as a deepening understanding of the business model based on data collection and surveillance advertising, the same principles which have governed traditional media during elections can be judiciously

[26] Maria Ressa, “We’re All Being Manipulated the Same Way,” April 6, 2022, https://www.theatlantic.com/ideas/archive/2022/04/maria-ressa-disinformation-manipulation/629483/

[27] Freedom House, “About the Project: Election Watch,” accessed September 27, 2023, https://freedomhouse.org/report/election-watch-digital-age/about.

[28] *Hoax in the Machine: Disinformation Against Voting Systems Manufacturers and Technologies in the 2022 US Midterm Elections (*Recorded Future, 2022), https://go.recordedfuture.com/hubfs/reports/ta-2022-1107.pdf; Juliana Gragnani & Jake Horton, “Brazil Election: Do Voting Machines Lead to Fraud?,” October 3, 2022, https://www.bbc.com/news/63061930.

applied to social media.

In addition to addressing algorithmic transparency, it is necessary that robust social media reforms are embedded in key areas of wider technology regulation, such as those concerning privacy and data protection, competition, transparency and accountability, and cybersecurity. This may manifest as measures seeking to ensure the data of users is not misused, antitrust efforts to dilute the over-concentrated market power of major social media companies, mandating risk assessments, testing whether platforms are abiding by their terms of services and policies, and strengthening defenses against the domestic and foreign manipulation of platforms.

Regulations Governing Traditional Media

Most countries have adopted legal frameworks around media with the assumption that the state does not intervene in the news and programming operations of the broadcasters, which is the main source of critique for those who believe the self-regulation model of social media should prevail given concerns about an authoritarian minded government curtailing free expression of citizens online. Broadcasters' role during elections, for example, does not differ from their normal journalistic role during non-election periods. Where laws have developed, ethical considerations continue to apply.

A distinguishing feature of the election period is the obligation to achieve equitable coverage of political parties without abdicating news value judgments. This right is also undermined by the structure of social media, as platforms allow for the uneven coverage of candidates, political parties, and harmful, politically advantageous narratives due to algorithms that amplify sensational and divisive content and foster the creation of echo chambers with a design that has decentralized sources of election-related information. Thus, platforms can undermine the foundational principle of achieving equitable, unbiased coverage during election periods, which traditional media are obligated to uphold.

Italy has, for example, a media regulator who has the principal responsibility for supervising media coverage of elections. In that case, there are two separate bodies: a Parliamentary Oversight Committee that has responsibility for public broadcasting, and AGCOM, an independent non-government regulator for radio, television, and newspapers, which is responsible for the privately-owned media. Both institutions make regulations governing coverage by the respective media sectors in elections. As of yet, no country has established a social media regulator responsible for overseeing election-related content; instead, oversight is mostly provided by election administrators who lack the means to enforce standards for election integrity.

The UN, the Organization for Security and Co-operation in Europe (OSCE) and other regional organizations, such as the Council of Europe, and the European Union, have introduced initiatives to re-affirm that the protection of human rights and fundamental freedoms should apply as much to the online world as to the real world. Thus, access to the Internet and the use of social networks during elections has become a key topic of keen interest.

CASE STUDY: Liability Shields and Defamation

In the United States, broadcast news networks and their local affiliates are not allowed to censor or edit campaign ads directly from political candidates. Different rules apply for cable networks, however, which can choose what ads to air and can request edits. But in both cases, politicians are not required to provide factual statements in their ad campaigns. Political ads are considered political speech, which is protected under the First Amendment. The Federal Communications Commission (FCC) oversees campaign ads and enforces rules for political programming, like disclosing sponsorships and making sure legally qualified candidates get "equal time." This does not entail fact checking content. The caveat is that if a candidate or a political organization makes a false factual statement, that is defamatory, they could be held liable through existing law. In the case of candidate content, TV stations are not liable for such defamation suits, but rather whomever produced the advertisement is. In 1996, internet based social-media platforms were granted broad "safe harbor" protections against legal liability for any content users post on their platforms, shielding them from accountability. Those protections, spelled out in Section 230 of the Communications Decency Act were written a quarter century ago during a bygone age of what many consider "naïve technological optimism and primitive technological capabilities." When digital platforms are granted complete legal immunity for the content that their users post, this reduces their incentives to proactively remove content causing social harm.[29] While there is growing consensus that Section 230 should be updated, there is no agreement on how. Legal scholars have put forward a variety of proposals, almost all of which adopt a carrot-and-stick approach, by tying a platform's safe-harbor protections to its use of reasonable content-moderation policies. Social media platforms have dedicated large sums of money to lobbyists, academic institutions, and civil society to fight regulatory legislation, suggested by some critics as a form of soft capture; in 2022, AdImpact, a political ad tracking company, identified 72 million USD was spent to campaign against two bills aimed at diluting the power of social media companies in the market.[30]

[29] Michael D. Smith & Marshall Van Alstyne, "It's Time to Update Section 230," August 12, 2021, https://hbr.org/2021/08/its-time-to-update-section-230.

[30] Cat Zakrzewski, Will Oremus, Gerrit de Vynck, & Cristiano Lima, "With Clock Ticking, Battle Over Tech Regulation Intensifies," June 27, 2022, https://www.washingtonpost.com/technology/2022/06/27/antitrust-tech-battle-congress/;

PART 2: THE ROLE OF MIS- AND DISINFORMATION IN ELECTIONS

The free flow of information via the internet and social media platforms contributes to open debate and an exchange of ideas, two crucial tenets of democracy. With access to tools for two-way communication through social networking sites, election campaigns entered into an era where political discourse moved away from one-sided transmissions to wider channels in which voters can express opinions, engage with candidates, and bring transparency to election administration. However, the misuse of technology greatly exceeds the ability to govern it during a time of transformational technological change.[31] Decades into technological evolutions in society, the advent of the internet and proliferation of social media poses both opportunities and significant risk to elections. Election integrity must be prioritized in the direction of safeguarding candidates and election officials from delegitimizing attacks perpetrated through mis- and disinformation with the weaponization of digital tools, which can impose both individual harms and consequences for democracies at large.

How Social Media Challenges can Manifest in Election Disputes

Many countries struggle with a lack of clarity about how and where to apply pre-existing frameworks for election integrity to social media when disputes are surfaced around campaign finance regimes, hate speech regulations, party/campaign codes of conduct, and other campaign-related challenges. Courts may be asked to rule regarding whether an action that occurs on social media violates particular election codes of conduct in these three directions:

1 - Campaign finance and political advertising

Social media advertisements are a key tool for candidates to share information about their policy platform and appeal to voters ahead of an election. There are many aspects of political advertising that may warrant the attention of a court. For example, the dissemination of disinformation through paid political ads remains a pressing issue in light of platforms failing to enforce their content standards on advertising, meaning misleading information can be targeted through ads at a high number of users and influence voter behavior without mitigation. Furthermore, some countries may regulate online political advertising within campaign finance regulatory frameworks, such as mandating disclosure requirements around who is paying for the ad or instituting spending limits to ensure fairness, and imposing fines or even disqualification from the race if violated. In

Tech Transparency Project, "Tech Funding Database," September 14, 2023, https://django.techtransparencyproject.org/techfundingdb/.

[31] The Campaign for RAND, "Campaign Priority: We Must Govern Emerging Technologies and Guard Against Existential Risks," accessed September 14, 2023, https://campaign.rand.org/campaign-priority/governing-technology/.

jurisdictions where political advertising is not covered under such a framework, courts may lack the ability to hold political actors accountable for misleading advertisements online and those which facilitate voter suppression, creating a lack of clarity and consequently contributing to an environment more tolerant of such malign behavior. This is highly dependent on the jurisdiction and indicates the difficulty of developing high-level, widely applicable advice to electoral justices.

Access to ad libraries and data concerning who bought the advertisements, how much was spent, and the reach they achieved can be crucial in determining whether foul play took place. However, while platforms like Facebook have an ad library meant to enhance transparency, they often fail to disclose this information and limit search indicators, depending on the country. This can significantly hamper the ability of researchers or election officials to track political advertisements that do not follow legal requirements set out by the jurisdiction. To further complicate matters, addressing the paid influencers, bloggers, and other thought leaders who may not technically be producing advertisements but whose vocal support is classified as a campaign expenditure can cause uncertainty as to how to regulate this kind of content, and the boundaries of enforcement under existing legislation focused on campaign finance and/or political advertisements during election periods.

2 - Codes of conduct and accountability measures

In today's political and digital landscape, disinformation can be exploited by candidates and political campaigns to achieve support for their positions and ultimately obtain power through manipulating and deceiving voters. Sometimes, “supporters” of a candidate engage in spreading disinformation acting as an agent of a political campaign or public figure to mount vile or harmful information that formal contestants may not wish to associate themselves with publicly, while still benefiting from the outcomes of eroding voters’ trust in the opposition, causing confusion, and sowing division amongst the electorate. Thus, there is little incentive by political actors to combat disinformation, as doing so could potentially eliminate a strategic advantage. Civil society and individual electoral stakeholders may be brought to court if perpetrators are found to have violated election regulations or laws concerning disinformation, defamation, fraud, and violations of civil rights. Again, different jurisdictions will have varying approaches to ruling on political actors engaging in disinformation, but electoral justices should know how to use relevant existing legislation to its fullest extent to hold perpetrators accountable. It remains crucial that candidates and political parties take responsibility for using deceptive tactics to influence voters online and are disincentivized by using disinformation for their own gain. This can be achieved by developing codes of conduct existing outside of formal law

to introduce measures of accountability and encourage better behavior during election periods and beyond.

3 - Hate speech

Within the context of elections, hate speech ("discriminatory" speech, meaning biased, bigoted or intolerant, or "pejorative" speech, prejudiced, contemptuous or demeaning of an individual or group) may be deployed to attack political opposition, manipulate voters' perceptions of opposition, and attack marginalized groups to further garner support for an ideological agenda.[32] Policies used to curb hate speech risk limiting free speech and are inconsistently enforced across the globe. Countries such as the United States grant social media companies broad powers in managing their content and enforcing hate speech rules. Others, including Germany, can force companies to remove posts within certain time periods.[33] Electoral court rulings on hate speech can be made challenging by the need to balance free expression with mitigating harms to an election's integrity, targeted individuals, and groups. In some jurisdictions, electoral courts have the power to rule on hate speech perpetrated by political figures, particularly if it may incite violence or is discriminatory. While rulings on hate speech may be made to preserve voters' rights to making informed decisions free of manipulation, to ensure a race operates fairly, and to protect targets, they may be subject to criticism of stifling debate and inhibiting political discourse by free expression advocates or supporters of the perpetrator/s.

Causes of Mis/Disinformation During Elections

Mis- and disinformation during elections encompasses the deliberate dissemination of false, misleading and manipulated information.[34] Disinformation can be used to strategically manipulate public opinion, deceive voters, suppress turnout (particularly among historically marginalized communities), increase polarization, and sway electoral outcomes.[35] False information about candidates, their policies, or the electoral process itself can mislead voters, and compromise their ability to make informed decisions.

---

[32] To provide a unified framework for the United Nations to address the issue globally, the UN Strategy and Plan of Action on Hate Speech defines hate speech as…*"**any kind of communication** in speech, writing or behaviour, that **attacks** or uses **pejorative** or **discriminatory** language with reference to a person or a group on the basis of **who they are**, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor."* See: United Nations, "Understanding Hate Speech," accessed September 14, 2023, https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech.

[33] Zachary Laub, "Hate Speech on Social Media: Global Comparisons," last updated June 7, 2019, https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons.

[34] *Strategic Guidance: Information Integrity: Forging a pathway to Truth, Resilience and Trust* (UNDP, February 2022), https://www.undp.org/sites/g/files/zskgke326/files/2022-02/UNDP-Information-Integrity-Forging-a-Pathway-to-Truth-Resilience-and-Trust.pdf.

[35] *The Impact of Disinformation on Democratic Processes and Human Rights in the World* (Directorate-General for External Policies Policy Department, European Parliament, 2021), https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO_STU(2021)653635_EN.pdf.

Political parties, candidates, and others trying to influence an election through false means are simply doing what bad actors in elections have done for years - manipulating political behavior. What is new is that technologies have modernized digital methods for doing so. The ability to digitally microtarget voters has also aided the spread of disinformation, allowing both political entities and individuals to disseminate ads to targeted groups with great precision, using data collected by social media platforms. "In commercial settings, microtargeting has come under fire for enabling discriminatory advertising, depriving historically marginalized communities of opportunities for jobs, housing, banking, and more. Political microtargeting, meanwhile, has experienced similar scrutiny, especially due to the limited monitoring of political ad purchases," according to Samantha Lai, a Research Analyst from the Center for Technology Innovation at The Brookings Institution.[36]

In an era of social media, there are several factors that make emerging campaigning technologies additionally effective and harmful. The design of digital platforms like X (formerly known as Twitter), Facebook, YouTube, TikTok, and Instagram as well as messaging apps like WhatsApp and Telegram eases the spread and salience of election integrity threats.[37] Deborah Brown, a Senior Researcher and Advocate for Digital Rights, Technology and Human Rights at Human Rights Watch, notes, "combatting electoral misinformation and disinformation is particularly difficult for these platforms because they were designed to maximize clicks, likes, and shares of the most engaging content—not to deliver reliable and accurate election information."[38] Furthermore, actions like that of Elon Musk, the head of X, to roll back mechanisms for tackling election-related disinformation and dismantle the team focused on this issue, further compounds platform-facilitated threats to election integrity.[39] The design of social media platforms has made it easier than ever before to spread disinformation, and with over half of the world's population registered as users, technology-facilitated information pollution poses a growing and immediate threat to the integrity of elections globally.[40]

---

[36] Samantha Lai, "Data Misuse and Disinformation: Technology and the 2022 Elections," June 21, 2022, https://www.brookings.edu/articles/data-misuse-and-disinformation-technology-and-the-2022-elections/.

[37] *Defending Elections Against Malicious Spread of Misinformation* (Proceedings of the AAAI Conference on Artificial Intelligence, 2019), https://ojs.aaai.org/index.php/AAAI/article/view/4056.

[38] Deborah Brown, "Can Social Media Platforms Stop Electoral Disinformation and Respect Free Speech?," October 30, 2020, https://www.hrw.org/news/2020/10/30/can-social-media-platforms-stop-electoral-disinformation-and-respect-free-speech.

[39] Clothilde Goujard, "Musk Ousts X Team Curbing Election Disinformation," September 28, 2023, https://www.politico.eu/article/musk-ousts-x-team-curbing-election-disinformation; Josh Taylor, "X/Twitter Scraps Feature Letting Users Report Misleading Information," September 26, 2023, https://www.theguardian.com/technology/2023/sep/27/xtwitter-scraps-function-letting-users-report-misleading-information .

[40] Belle Wong & Cassie Bottorff, "Top Social Media Statistics And Trends Of 2023," https://www.forbes.com/advisor/business/social-media-statistics/.

Encrypted messaging applications (EMAs) that rely on end-to-end encryption such as Telegram, WhatsApp, and Signal, offer a level of intimacy and security that have made them remarkably popular among activists and others who want to communicate without fear of government surveillance. These qualities also make them a useful vector for disinformation, easing the spread of untraceable claims to users via trusted contacts in a secure environment.[41]

The nature of social media and the ability to spread content at one click of a button is not the only reason election disinformation is able to propagate online so easily. Tech employees of major social media platforms have come forward in recent years as whistleblowers to expose the failure of certain social media companies to adequately oversee fake engagement and foreign influence operations on their platforms, particularly around elections.

Sophie Zhang, a former data scientist in the "fake engagement" division of Facebook, exposed the platform's failure to combat political manipulation campaigns in an internal memo that went public in 2021.[42] Zhang's memo revealed that fake accounts enabled politicians to mislead the public and gain power, particularly during times of elections and political transitions, and how little Meta (Facebook's parent company) did to mitigate this problem despite Zhang's repeated efforts to bring it to the attention of executive leadership.

Zhang identified that in addition to commercial motivations, fake engagement was being used on what Facebook called "civic," or "political" targets. In testimony before the British Parliament, Zhang revealed that while removing fake accounts is part of Facebook's policy, "there was a perverse effect in that, if I found fake accounts that were not directly tied to any political figure, they were often easier to take down than if I found fake accounts that were." This effect, she said, "creates an incentive for major political figures to essentially commit a crime openly."[43] Meta claims it ultimately took action on the abuse uncovered and argued that the lag in enforcement was not an attempt to protect powerful people who use its service, but provided no evidence to this effect.

---

[41] Jacob Gursky & Samuel Woolley, "Countering Disinformation and Protecting Democratic Communication on Encrypted Messaging Applications," June 2021, https://www.brookings.edu/articles/countering-disinformation-and-protecting-democratic-communication-on-encrypted-messaging-applications/.

[42] Julia Carrie Wong, "How Facebook Let Fake Engagement Distort Global Politics: A Whistleblower's Account," April 12, 2021, https://www.theguardian.com/technology/2021/apr/12/facebook-fake-engagement-whistleblower-sophie-zhang.

[43] Eloise Berry, "Another Facebook Whistleblower Just Testified in British Parliament. Here's What to Know About Her Allegations," October 18, 2021, https://time.com/6107835/sophie-zhang-facebook-testimony/.

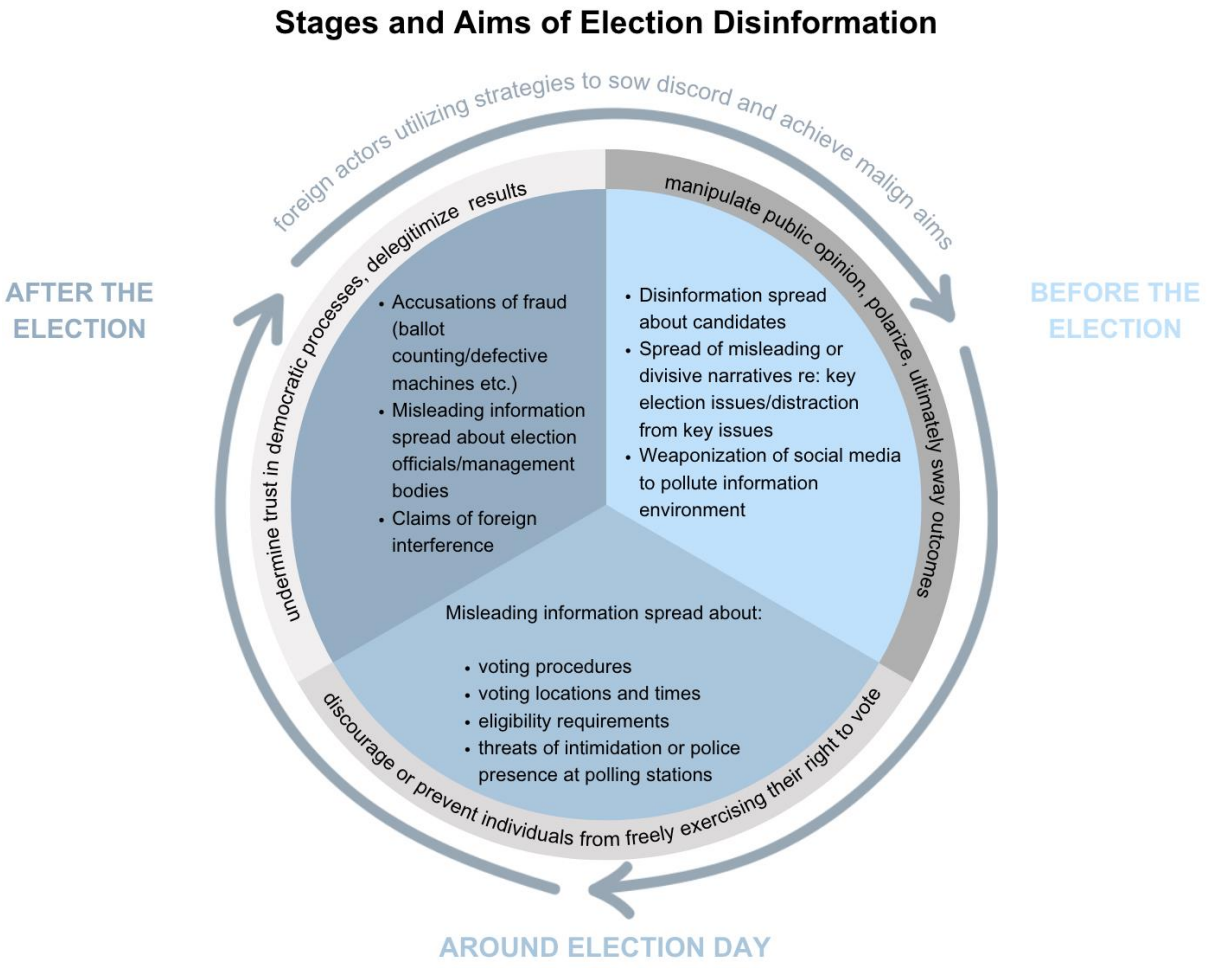
In 2021, data scientist, former Facebook employee, and whistleblower Frances Haugen asserted in her testimony to the U.S. House of Representatives that Meta's "leadership knows how to make Facebook and Instagram safer," but "they repeatedly chose to ignore these options, and continue to put their profits before people," posing a threat to "the integrity of our democracies."[44] The documents revealed devastating failures to act on critical election-related issues facilitated by the platform, with no clear policies to address post-election violence.

Stages of Election Mis- and Disinformation

The role of mis- and disinformation can become noteworthy at various stages of planning for and administering elections in a democracy. This multifaceted work of election management bodies (EMBs) encompasses areas like budget formulation, delineation of districts and voting boundaries, voter registration, nomination of candidates, procurement of vendors, selection of technology, logistics, management of Election Day operations, results tabulation and transmission, facilitation of voting from abroad, resolving election disputes, and offering specialized voting services. When executed with precision, electoral operations contribute to nurturing a sense of legitimacy and credibility in the eyes of both voters and candidates regarding the eventual election outcomes. However, at any point in an electoral cycle, disinformation attacks can be deployed to undermine the legitimacy of elections and the work of EMBs and election officials. One can consider this in stages to understand how the manifestations and aims of election-related disinformation vary depending on the period at which it is deployed.

The figure below depicts the stages and aims of election disinformation, but it is important to note that many points contained within the blue circle can also be spread unintentionally as misinformation. For example, false information about candidates and key election issues before the election, voting procedures and requirements around election day, and fraud, foreign interference, and election officials/EMBs after the election can be spread unintentionally, but still result in the same harms posed by the intentional spread of disinformation.

[44] *Written Testimony of Frances Haugen Before the United States House of Representatives Committee on Energy and Commerce Subcommittee on Communications and Technology* (U.S. House of Representatives,, 2021), https://docs.house.gov/meetings/IF/IF16/20211201/114268/HHRG-117-IF16-Wstate-HaugenF-20211201-U1.pdf.



45

The Impact of Election Mis/Disinformation

Manipulation of public discourse

Mis- and disinformation leads to increased polarization through the spread of misleading or divisive narratives. Through targeting the public discourse, voters are deceived and societal divisions are deepened. Disinformation reinforces pre-held biases and can further strengthen voters' belief in misinformed or harmful narratives, who may be more inclined to share content on social media that confirms their beliefs regardless of its veracity.[46] Through carefully crafted appeals to emotion and sensationalized narratives designed to elicit reactions, gain support for a particular ideology or position, or directly influence

[45] Adapted from *Combating Information Manipulation: A Playbook for Elections and Beyond* (NDI, 2021), p. 16, https://www.ndi.org/publications/combating-information-manipulation-playbook-elections-and-beyond.

[46] Mathias Osmundsen, Michael Bang Petersen, & Alexander Bor, "How Partisan Polarization Drives the Spread of Fake News," May 13, 2021, https://www.brookings.edu/articles/how-partisan-polarization-drives-the-spread-of-fake-news/.

election outcomes, disinformation can fuel the spread of conspiracy theories and amplify extremist viewpoints, fostering emotional connections to a shared identity and distance from those who do not subscribe to the same beliefs. The ability of information pollution to sway the beliefs of the public on a wide scale cannot be understated - "when large segments of the public are misinformed in the same direction, shared misperceptions can systematically bias collective opinion."[47]

Disinformation in particular can be weaponized to spread sensationalized and misleading information to ensure certain agendas gain more visibility, allowing perpetrators to exacerbate political polarization and further entrench socio-economic divides. When public discourse is manipulated in these ways, it can distort the issues at stake in an election and undermine a fair and balanced debate.

Increased polarization

The role of "information overload" in the propagation of disinformation has seen frequent discussion in the past few years.[48] As social media platforms provide users daily access to an onslaught of information, designed to allow content-based communication with unprecedented speed and reach, it becomes challenging to determine what is fact and fiction, and easy to seek comfort in familiar viewpoints. The unique qualities of social networking platforms as spaces where users are able to freely form communities unencumbered by physical boundaries enable the formation of echo chambers, in which members are repeatedly exposed to viewpoints aligning with their own. Social networking platforms act as "a polarized digital space where users tend to promote their favorite narratives, form polarized groups and resist information that does not conform to their beliefs," increasing how rapidly mis- and disinformation can spread between these groups.[49] Not always user-initiated, social media design also lends itself to the formation of echo chambers: the algorithmic design of social networking platforms results in a "curation of content based on users' past activity (cf. filter bubbles), which limits the novelty and diversity of the content that users are exposed to, and which—instead of contributing to viewpoint diversity—leads to online clustering and polarization."[50]

Such a design means that disinformation can reach wide viewership and even reach virality through algorithmic amplification and confirmation bias, through encouraging

---

[47] Jennifer Jerit & Yangzi Zhao, "Political Misinformation," 2020, *Annual Review of Political Science* Vol. 23, pp. 77-94, https://www.annualreviews.org/doi/10.1146/annurev-polisci-050718-032814.

[48] Filippo Menczer & Thomas Hills, "Information Overload Helps Fake News Spread, and Social Media Knows It," December 1, 2020, https://www.scientificamerican.com/article/information-overload-helps-fake-news-spread-and-social-media-knows-it/.

[49] Petter Törnberg, "Echo Chambers and Viral Misinformation: Modeling Fake News as Complex Contagion," 2018, *PLoS One* 13(9), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6147442/.

[50] Ludovic Terren & Rosa Borge-Bravo, "Echo Chambers on Social Media: A Systematic Review of the Literature," 2021, Review of Communication Research 9, pp. 99-118, https://rcommunicationr.org/index.php/rcr/article/view/94.

interaction with content that affirms pre-existing beliefs regardless of its truthfulness. The role of platforms in deepening political divisions and sowing polarization both domestically and globally thus must be considered. Polarization driven by mis- and disinformation exacerbates hostility between the public, impinges on citizens' right to making decisions based on accurate information, and diminishes the possibility of constructive debate and dialogue, a cornerstone of a thriving democracy.

The period of euphoria about the possibility that social media might usher in a golden age of global democratization has now been met with widespread concern among media, scholars, the philanthropic community, civil society, and politicians themselves about the impact on elections. Refer to "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature" for a comprehensive overview of the scholarly literature on the relationship between three factors that may be undermining the quality of democracy: social media usage, political polarization, and the prevalence of disinformation online.[51]

Voter suppression

Democracies bestow voting as a fundamental right upon citizens, ensuring unencumbered political engagement and promoting the role of citizens in fairly determining their representation. Mis- and disinformation, however, can be utilized as a tool to suppress voters and diminish equal participation. Voters can be intentionally misled, sowing confusion and doubt, and malicious actors can discourage or prevent certain groups of people from freely exercising their right to vote, undermining the principle of universal suffrage.[52]

The Center for Democracy & Technology notes that "much voter suppression activity is motivated by partisan interests, and targets demographic groups that are presumed to be planning to vote for the opposition," opting to "stop them from voting at all."[53] Election-related disinformation deployed with the aim of achieving voter suppression manifests in a variety of ways, spreading false information about:

- Voting procedures, such as claiming electronic voting machines are malfunctioning or rigged, undermining the legitimacy of mail-in ballots, sharing false information about registration deadlines, and claims of voter fraud targeting ballot counting

[51] Joshua A. Tucker, et al., *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature* (Hewlett Foundation, March 2018), https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf.

[52] According to the Carter Center, "the obligation to ensure universal suffrage appears in the International Covenant on Civil and Political Rights and other international treaties and requires that the state guarantee that the broadest pool of voters be allowed to cast ballots." See here: Election Standards at the Carter Center, "Universal Suffrage," accessed September 14, 2023, https://eos.cartercenter.org/obligations/10.

[53] Emma Llansó, "Online Voter Suppression: A Guide for Election Officials on How to Spot & Counter," October 15, 2020, https://cdt.org/insights/online-voter-suppression-a-guide-for-election-officials-on-how-to-spot-counter/.

- Locations and times of polling stations or registration centers, such as sharing on social media, that voting is taking place at an incorrect location and/or time
- Eligibility requirements, such as spreading misleading advice on forms of identification needed and conditions that must be met to register for voting or cast a ballot
- Threats of intimidation or police presence, such as claiming voters who arrive to cast a ballot in person may face risks to their physical safety, or overstating the presence of law enforcement to sway voters from marginalized communities from showing up[54]

It must be noted that the above-identified manifestations of election-related disinformation often disproportionately affect individuals with marginalized and intersecting identities, as "voter suppression is often targeted at specific, vulnerable communities."[55] This can have a proportional impact on the amount of people who decide not to vote by convincing them their vote doesn't matter or that they cannot trust the procedures, or misleading them around polling station closing times, or voter registration, or ID rules, which inhibits citizens' ability to access their right to participate in a free and fair election and vote for their direct representation in government. Furthermore, this stifles the voices of marginalized communities and impedes them from electing representation that may address their unique needs. Therefore, election disinformation represents a key obstacle to voter turnout, particularly within marginalized communities who have faced systematic and historic discrimination, and further widens socio-economic inequalities.

Foreign influence campaigns

Mis- and disinformation can be employed as part of broader influence campaigns aimed at shaping public opinion and interfering in the electoral process. These efforts may be orchestrated by foreign actors, with documented cases demonstrating campaigns originating in Russia or China, seeking to undermine democratic systems, sometimes in collaboration with domestic entities with vested interests.[56]

Researchers from the Department of Political Science at the University of British Columbia identified foreign actors who utilize "digital techniques" to target "fair opportunities for citizen participation (such as voting, running for office, or contributing to

[54] Ibid.

[55] Daniel Arnaudo et al.,*Combating Information Manipulation: A Playbook for Elections and Beyond* (International Republican Institute, September 2021), https://www.iri.org/resources/combating-information-manipulation-a-playbook-for-elections-and-beyond/.

[56] Adrian Shahbaz & Allie Funk, "Digital Election Interference," 2019, https://freedomhouse.org/report/freedom-on-the-net/2019/the-crisis-of-social-media/digital-election-interference; Naja Bentzen, Foreign Interference in Democracies: Understanding the Threat, and Evolving Responses (European Parliament, 2020), https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/652082/EPRS_BRI(2020)652082_EN.pdf.

public debates); public deliberation that enables citizens to share and understand each other's insights and perspectives; and key institutional actions by electoral commissions, political parties, and other organizations, including the enforcement of electoral regulations."[57] By flooding the information space with deceptive content, disinformation campaigns can distort public perceptions, erode trust in institutions, and manipulate electoral outcomes. Foreign actors may utilize an array of tactics to interfere in the election processes of other countries so as to sow division, foster distrust in democratic processes, and perpetuate beneficial narratives. They may opt to conduct wide scale influence operations to weaponize social networking platforms and coordinate sophisticated disinformation campaigns, in which false and sensationalized narratives targeting candidates seek to influence public opinion and voter decisions, or those targeting election officials seek to encourage doubt in the electoral process or election outcomes.

Foreign actors may deploy astroturfing, or creating an illusion of a grassroots movement in support of an issue or candidate while in fact being orchestrated by malign actors with the aim of further influencing public opinion or strengthening disinformation efforts, or employ bots to perpetuate their narratives on a massive scale.[58] Furthermore, they may create misrepresentative accounts, or "sockpuppets," impersonating electoral agencies or officials to misinform voters on key issues or pertinent election information, facilitating voter suppression and preventing citizens from engaging fairly in the electoral process.[59]

Erosion of trust in democratic processes

When information pollution becomes pervasive, it can erode public trust in the electoral process. If people no longer believe that elections are fair, transparent, or free from manipulation, they may disengage from the political process or reject the legitimacy of electoral outcomes.[60] The prevalence of disinformation and its impact on elections is within itself a factor that may erode this trust in the process: with the knowledge that disinformation is rampant and ever-increasing, facilitated by amplification on social media, citizens may feel incredulous about participating in seemingly unfair elections tainted by deception and view their ability to act on informed decisions eroded. This can result in

---

[57] Chris Tenove et al., *Digital Threats to Democratic Elections: How Foreign Actors Use Digital Techniques to Undermine Democracy* (Centre for the Study of Democratic Institutions, University of British Columbia, 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3235819.

[58] "Election Interference: A Unique Harm Requiring Unique Solutions." in *Defending Democracies: Combating Foreign Election Interference in a Digital Age* (Oxford: Oxford University Press, 2021), eds. Jens David Ohlin & Duncan B. Hollis, https://academic.oup.com/book/39306/chapter-abstract/338903882?redirectedFrom=fulltext.

[59] Ibid.

[60] "Electoral Integrity Matters: How Electoral Process Conditions the Relationship Between Political Losing and Political Trust," *Qual Quant* 56, pp. 1709-1728, https://link.springer.com/article/10.1007/s11135-020-01050-1; *The Impact of Disinformation on Democratic Processes and Human Rights in the World* (Directorate-General for External Policies Policy Department, European Parliament, 2021), https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO_STU(2021)653635_EN.pdf.

feeling their participation is inconsequential if electoral outcomes are so heavily influenced, and thus retreating from democratic participation altogether. Information overload, and consequently, disinformation overload, can be overwhelming and further entrench ideological positionings, making voters increasingly distrustful of any election-related information, whether it comes from fellow social media users or government institutions. The presence of influence operations may also hinder citizens' confidence in their electoral processes, prompting skepticism of systems perceived as being gamed by foreign actors, and thus corrupted or particularly vulnerable to external influence. This erosion of trust weakens democratic institutions and can have long-term consequences for the stability and functioning of a democracy.[61]

[61] Michael Tomz & Jessica L. P. Weeks, "What Americans Really Think About Foreign Interference in the U.S. Elections," June 19, 2019, https://www.washingtonpost.com/politics/2019/06/19/what-americans-really-think-about-foreign-meddling-us-elections/.

CASE STUDY: Election Denialism in the US 2020 Election

In the United States 2020 presidential election, foreign actors made attempts to spread false or misleading information and influence the election, yet these efforts were less impactful than domestic mis- and disinformation. For example, a St. Petersburg-based troll farm, the Internet Research Agency, created a bogus news website that solicited pieces on divisive political topics from bona fide reporters to push to American audiences.[62] Meanwhile, Iranian actors tried to threaten Americans to vote for Donald Trump while posing as members of the far-right white nationalist U.S. group Proud Boys.[63] Yet, despite these examples, there were no foreign disinformation campaigns or hack-and-leak operations on a scale comparable to the 2016 presidential election, in part attributed to greater public awareness of foreign influence operations and better preparation by U.S. government agencies.[64] However, The "Stop the Steal" movement — whose adherents spread the lie that election officials acted to "steal" the election from former President Donald Trump — was one of the fastest-growing disinformation campaigns ever executed on Facebook. While extremists mobilized first on alternative and fringe online platforms, false claims of election fraud and violent, angry rhetoric spread aggressively across larger mainstream platforms, benefiting from social media platforms' lax standards and algorithmic boosts of such content.[65] Users coordinated to spread false claims of election fraud over YouTube, X (formerly known as Twitter), Facebook, Instagram, and TikTok with content that was rife with incitement to violence, threats, hate speech, and misinformation about the election.[66] Facebook, for example, had no explicit policy against election denial. Its systems for detecting violent rhetoric were also noted as unreliable, therefore it took down relatively few of these groups before the January 6th insurrection.

As reported in "Social Media & the January 6th Attack on the U.S. Capitol" prepared for the Select Committee to Investigate the January 6th Attack on the United States Capitol, "Despite the knowledge that these groups had ties to violent actors, employee recommendations that the company take the problem more seriously were ignored or outright rejected." The report highlights how the attack on the U.S. Capitol was driven by the radicalization of a smaller subset of users on social media, not the result of political polarization generally, In fact, political polarization may have contributed to the weak response of social media companies, with the authors concluding that "Major platforms' lax enforcement against violent rhetoric, hate speech, and the big lie stemmed from longstanding fear of scrutiny from elected officials and government regulators. Many of these voices called for stronger platform action and greater corporate responsibility; but on the right side of the spectrum, critics made largely baseless accusations that platform integrity efforts were designed to somehow

suppress or censor conservative political speech."

According to nonpartisan election observers, partisan influencers, politicians, and activists alike continue to use the popularity of election denialism in the US to build audiences and profit off lies about voting and elections[67].

While it was possible for election-related conspiracies to spread from one country to another 20 years ago, it was more difficult and thus rare. With the advent of social media, election denialism has gone global since 2020 and was a feature in elections in France, Germany, Australia, Brazil, and other countries, with analysts concluding that election deniers are hedging their bets online when they do not have the electoral majorities needed to gain power.[68]

[62] *(U) Report of the Select Committee on Intelligence United States Senate on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election: Volume 2: Russia's Use of Social Media With Additional Views* (U.S. Senate, Senate Intelligence Committee, 2019), https://intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

[63] Ellen Nakashima et al., "U.S. Government Concludes Iran Was Behind Threatening Emails Sent to Democrats," October 22, 2020, https://www.washingtonpost.com/technology/2020/10/20/proud-boys-emails-florida/.

[64] Alistair Somerville & Jonas Heering, "The Disinformation Shift: From Foreign to Domestic," December 1, 2020, https://isd.georgetown.edu/2020/12/01/the-disinformation-shift-from-foreign-to-domestic/.

[65] Cooper Gatewood, "Why Voter Fraud Narratives Remain Persistent," December 8, 2020, https://www.isdglobal.org/digital_dispatches/why-voter-fraud-narratives-remain-persistent/.

[66] Michael Baldassaro, Katie Harbath, & Michael Scholtens, *The Big Lie and Big Tech: Misinformation Repeat Offenders and Social Media in the 2020 U.S. Election* (The Carter Center, 2021), https://www.cartercenter.org/resources/pdfs/news/peace_publications/democracy/the-big-lie-and-big-tech.pdf.

[67] Emma Steiner, *Under the Microscope: Election Disinformation in 2022 and What We Learned for 2024* (Common Cause Education Fund, 2023), https://www.commoncause.org/resource/under-the-microscope/.

[68] Jiore Craig, Cécile Simmons & Rhea Bhatnagar, "How January 6 Inspired Election Disinformation Around the World," January 13, 2023, https://www.isdglobal.org/digital_dispatches/how-january-6-inspired-election-disinformation-around-the-world/.

The Institute for Strategic Dialogue analyzed online activity around elections in Germany, France, Australia, and the US midterms in 2022 and found three recurring themes related to election denial. First, social media product features are amplifying election disinformation and assisting the organizing efforts of those pushing the false claims. Next, election denial conspiracies consistently have a mutually beneficial relationship with conspiracies based in hate, white supremacy, racism, homophobia, xenophobia, and sexism. Finally, the success of election denial efforts online is mixed. Evidence suggests that election denialism alone is not an effective tool to speak to voters' top concerns and is only sometimes effective in mobilizing and radicalizing. This efficacy is boosted when election denial narratives are mixed with hate-based conspiracies and are diminished in countries with central election authorities and strong accountability measures.

Interventions: The Role of Election Management Bodies

EMBs play a crucial role in providing accurate information before, during, and after an election. In the new information environment, it is even more important for EMBs to gain the public's trust as the utmost authority on election information, and act to affirm their legitimacy in order to foster citizens' trust in electoral processes and their management.

EMBs that are more involved in taking action on disinformation come from higher performing democracies, according to International IDEA.[69] Initially devised within the context of EMBs in Asia, researcher and professor Netina Tan proposed an index to evaluate the digital readiness of EMBs that may have wider applications: Tan's index is based on criteria that takes into consideration the independence and autonomy of the EMB, the existing legal frameworks around "online political communication, campaign finance, and disinformation," the level of "respect for the rule of law as an indicator of the confidence in the capacity of the EMB and government to enforce the regulatory framework," and the EMB's access to technical tools and capabilities to deal with technologically facilitated threats.[70] Assessing the baseline readiness of EMBs to respond to digital threats to election integrity like mis/disinformation may serve as an important first step before evaluating what actions can be taken to mitigate these harms.

[69] International IDEA, "The Information Environment Around Elections," accessed September 14, 2023, https://www.idea.int/our-work/what-we-do/elections/information-environment-around-elections.

[70] Netina Tan, "Electoral Management of Digital Campaigns and Disinformation in East and Southeast Asia," *Election Law Journal: Rules, Politics, and Policy* 19(2), pp. 111-261, https://www.liebertpub.com/doi/epdf/10.1089/elj.2019.0599.

According to a report prepared by the National Democratic Institute (NDI), the International Republican Institute, and Stanford University, "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," and "very few have reporting mechanisms created for citizens to report elections-related information manipulation observed online," and "typically do not have the mandate to develop rigorous regulations around online campaigning nor the ability to enforce existing regulations," although some 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."[71] Countering Disinformation and the Center for Technology & Democracy outline key approaches EMBs can take to address information pollution:[72]

Before election day:

- NDI recommends EMBs conduct a Preliminary Assessment of the Information Environment by examining the nature, vulnerabilities, mitigating factors, and opportunities around the electoral information environment, online and otherwise. This can vary significantly from country to country. A checklist of actions is available for such assessments.[73]
- It is recommended that EMBs establish a presence on social media and in traditional media to share accurate and updated information on election procedures, eligibility requirements, and polling times and locations ahead of election day.

According to Countering Disinformation, "the INEC of Nigeria deploys its longstanding institutional investment in public communication as a bulwark against disinformation. During electoral periods, the INEC provides daily televised briefings, participates in live TV interviews, issues regular press statements to explain the policies and decisions of the commission, and runs the INEC Citizens Contact Centre (ICCC) to provide the public with access to the commission and communicate with critical stakeholders."[74]

---

[71] Daniel Arnaudo et al., *Combating Information Manipulation: A Playbook for Elections and Beyond* (International Republican Institute, September 2021), https://www.iri.org/resources/combating-information-manipulation-a-playbook-for-elections-and-beyond/.

[72] Countering Disinformation, "Election Management Body Approaches to Countering Disinformation: Complete Document - EMB Approaches," accessed September 14, 2023, https://counteringdisinformation.org/topics/embs/complete-document-emb-approaches.

[73] *Disinformation and Electoral Integrity: A Guidance Document for NDI Elections Programs* (NDI, 2019), https://www.ndi.org/sites/default/files/Disinformation%20and%20Electoral%20Integrity_NDI_External_Updated%20May%202019%20%281%29.pdf

[74] Countering Disinformation, "Election Management Body Approaches to Countering Disinformation: Complete Document - EMB Approaches," accessed September 14, 2023, https://counteringdisinformation.org/topics/embs/complete-document-emb-approaches.

- Create a code of conduct or declaration of principles pertaining to election periods to "define how political parties, candidates, media or the electorate at large should behave during the electoral period." This can help guide efforts to tackle mis- and disinformation by instituting common practices, and allows EMBs to communicate expected appropriate behavior to election stakeholders. Common elements include affirming a "commitment to freedom of expression," banning election disinformation, "restricting deceptive online behaviors used to promote campaign content," "prohibitions against incitement to violence and hate speech," and a "proactive obligation to share correct information."
- Initiate conversations with social media companies to amplify distribution of trustworthy information and limit the proliferation of mis/disinformation during elections, and urge them to take more decisive action. It should be noted this may be difficult for EBMs in jurisdictions where social media companies have historically dedicated little resources or attention, and that these companies have a history of failing to deliver on promises of regulating their platforms to mitigate the harms of election-related misinformation, especially in markets they deem as less valuable.

In 2022, India's Chief Election Commissioner Rajiv Kumar stated "election management bodies (EMBs) expected social media sites to use their 'algorithm power' to proactively flag" information pollution in India (this is frequently described incorrectly as fake news, a more specific type of mis- and disinformation) to facilitate "credible electoral outcomes." According to Kumar, "more early or deeper red-flagging of fake news based on known modus operandi and genres is not an unfair expectation from the EMBs."In 2023, he noted that EMBs "can work together on pressing challenges, including countering fake narratives which are trying to derail election integrity worldwide.[75]

Before Mexico's 2018 elections, the country's National Institute of Elections (INE) entered into agreements with Facebook and Google to train staff on how to monitor platforms and to bolster them as the authoritative source of election information, respectively.[76]

[75] The Indian Express, "Election Bodies Expect Social Media Sites to Proactively Flag Fake News: CEC Rajiv Kumar," November 1, 2022, https://indianexpress.com/article/india/election-bodies-social-media-sites-fake-news-rajiv-kumar-8240534; The Economic Times, "Poll Management Bodies Can Work Together to Counter Fake Narratives: CEC Rajiv Kumar," last updated July 12, 2023, https://economictimes.indiatimes.com/news/politics-and-nation/poll-management-bodies-can-work-together-to-counter-fake-narratives-cec-rajiv-kumar/articleshow/101695980.cms?from=mdr/.

[76] Leonie Rauls, "How Latin American Governments Are Fighting Fake News," October 19, 2021, https://americasquarterly.org/article/how-latin-american-governments-are-fighting-fake-news/.

- Coordinate with civil society "to enhance the reach of their messaging or extend their capacity to engage in time and labor-intensive activities such as fact-checking or social listening."
- Work with state entities to coordinate efforts and communications and increase access to resources for counteracting mis- and disinformation.

The U.S. Cybersecurity and Infrastructure Security Agency's "Rumor vs. Reality" program was implemented to combat mis- and disinformation and complement "election officials' voter education and civic literacy efforts," and "help them build resilience against foreign influence operations and disinformation narratives about election infrastructure."[77]

In 2018, Swedish officials at the Civil Contingencies Agencies coordinated with other agencies, trained civil servants, collaborated with both traditional and social media, and conducted media monitoring to proactively tackle information pollution and prevent the undermining of the upcoming election.[78]

- Engage in exchange and dialogue with other EMBs to share best practices, inform, and strengthen strategies for countering information pollution.
- Establish a clear mechanism to address complaints related to dis- or misinformation.

In Pakistan, the Election Commission made accessible on their website official complaint forms, which "featured timely summaries of the numbers and types of complaints submitted to the ECP," and allowed complainants "to look up the status of their case," which "increased the transparency of the process, while providing a degree of credibility and professionalism that had not existed previously" to the commission.[79]

On election day:

- Prepare to counteract mis- and disinformation surrounding voting locations, times, and accessibility to polls in real-time, establishing a system to identify and refute this information.

[77] U.S. Cybersecurity and Infrastructure Security Agency, "Election Security Rumor vs. Reality," accessed September 14, 2023, https://www.cisa.gov/rumor-vs-reality.

[78] Gordon LaForge, *Sweden Defends Its Elections Against Disinformation, 2016–2018* (Innovations for Successful Societies, Princeton University's School of Public and International Affairs, 2020), https://successfulsocieties.princeton.edu/publications/sweden-defends-its-elections-against-disinformation-2016-%E2%80%93-2018.

[79] Chad Vickery (ed.), *Guidelines for Understanding, Adjudicating, and Resolving Disputes in Elections (GUARDE)* (IFES, 2011), https://www.ifes.org/publications/guidelines-understanding-adjudicating-and-resolving-disputes-elections-guarde.

Throughout an election period and beyond:

- Employ strategic communications and educate voters on how to identify mis- and disinformation - don't amplify propaganda, but utilize the unique characteristics of social media platforms that allow a direct dialogue with voters and quickly impart necessary information.
- Create a strategy for crisis communications when a significant threat posed by mis- and disinformation has been identified, to ensure efficiency when tackling information pollution.
- Conduct social listening "to understand and respond to disinformation threats - establish a system for monitoring mis- and disinformation that affects your jurisdiction and learn how to report it, set up social listening to inform a rapid incident response system or to inform strategic and communication planning."

PART 3: THE ROLE OF GENDERED DISINFORMATION IN ELECTIONS

Gendered disinformation denotes the spread of deceptive or inaccurate information and images against women political leaders, as well as journalists, and other female public figures. It is a global phenomenon, weaponized to undermine the credibility of women in public-facing roles and ultimately undermine democracy. It is especially pronounced when targeting women from racial, ethnic, religious, or other minority groups.[80] As of 2020, 85% of women globally have faced online violence, with misinformation and defamation maintaining a prevalence rate of 67%, according to the Economist Intelligence Unit.[81]

Gendered disinformation narratives are largely underpinned by gendered stereotypes, such as expectations of how a woman should look or behave, whether it be politically, sexually, or morally.[82] A 2023 report from the UN's Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression identifies political aims of gendered disinformation as including "creating a more polarized electorate," impeding informed decision-making, and exploiting topical, highly-visible events like elections to

[80] Lucina Di Meco and Kristina Wilfore, "Gendered Disinformation is a National Security Problem, March 8, 2021, https://www.brookings.edu/techstream/gendered-disinformation-is-a-national-security-problem/.
[81] The Economist Intelligence Unit, "Measuring the Prevalence of Online Violence Against Women," accessed October 13, 2023, https://onlineviolencewomen.eiu.com/.
[82] Shmyla Khan and Amna Khan, "Locating Gender in the Disinformation Landscape," April 28, 2022, https://www.boell.de/en/2022/04/28/locating-gender-disinformation-landscape.

"achieve greater influence."[83] Thus, the integrity of elections worldwide is put at risk by gendered disinformation, specifically that which targets women candidates, who are in the public eye by virtue of running for office, or election workers, and election officials, who may be thrust into the public eye when targeted by gendered disinformation campaigns aimed at undermining election integrity.

Gendered disinformation perpetuates misogynistic tropes and stereotypes by exploiting discriminatory attitudes towards women, reinforcing ideas of women's role in society and shaping public perception of candidates and women in positions of leadership based on their gender. Sexism can thus become a vehicle with which to spread ideologically motivated and illiberal disinformation. Women who maintain intersecting identities are the target of some of the most violent, vicious gendered disinformation, and online hate campaigns. Gendered disinformation makes online spaces less safe for women, as they may feel intimidated to stop exercising their freedom of expression in the hopes of avoiding future attacks, ceasing to engage in political discourse by self-censoring, or feeling the need to leave social media platforms altogether to avoid further psychological and reputational harm.

While major social media companies did not invent misogyny, they are responsible for facilitating the malign use of their platforms. In a 2022 report, NDI proposed several recommendations for platforms to mitigate online violence targeting women in politics, including working in partnership with fact-checkers to identify misleading gendered content and civil society to address misinformative gendered content that "may not contain clearly fact-checkable claims but nonetheless amplify gender norms that increase discrimination and hate towards women," and developing adaptable mechanisms for moderating harmful gendered content to "respond to evolving threats," particularly relevant during election periods when abuse and violence can be heightened and more frequent.[84] The former is a reactive effort, while the latter is proactive. Actions to counter gendered disinformation should prioritize proactivity to thwart attacks before they can achieve impact and prevent the ability of malign actors to weaponize digital platforms to target women.[85]

---

[83] United Nations, General Assembly, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan,* A/78/288 (7 August 2023), https://srfreedex.org/wp-content/uploads/2023/10/A.78.288.pdf, pp. 4, 13.

[84] *Interventions for Ending Online Violence Against Women in Politics* (National Democratic Institute, 2022), https://www.ndi.org/sites/default/files/NDI%20Interventions%20to%20End%20OVAW-P.pdf.

[85] Countering Disinformation, "Understanding the Gender Dimensions of Disinformation," accessed September 14, 2023, https://counteringdisinformation.org/topics/gender/3-current-approaches-countering-gendered-disinformation-and-addressing-gender.

Impact: Damaging Reputation and Credibility

The immediate objective of these campaigns is to discredit female-identifying political contestants and portray them as untrustworthy, unqualified, unintelligent, and unlikeable, and further undermine them using sexualized and character-based attacks. The ultimate goals are to reverse progress made towards gender equality, undermine democratic principles and institutions, and shape perceptions of women's ability to lead or participate in the public sphere. Through targeting women involved in election processes, malign actors can exploit societal divisions and gender biases to manipulate public opinion, influence election outcomes, disrupt electoral infrastructure, and foster cynicism towards women's participation in democratic institutions.

When disinformation campaigns target individuals involved in the electoral process, they are employed through weaponizing gender to damage their reputation and credibility. False allegations, manipulated images or videos, or misleading information ultimately undermines their chances of success or public trust. These narratives can be utilized to question women's competence, integrity, and suitability for leadership roles. They may accuse women of corruption or being influenced by foreign actors to portray them as untrustworthy, or too unqualified, thus incapable of performing the duties required of their role. Character attacks based on intelligence are common, along with sexual objectification online. A common and constant position of gendered disinformation is to accuse women of being unlikeable for their stances on certain issues or for personal characteristics, such as not being a mother, appearing "cold" or "bossy." These attacks can have direct ramifications for how the public perceives the target, and greatly influence their opinion regarding the target's reputation and credibility.

Amplifying Gender-Related Attacks

Disinformation can exploit gender-related issues to create divisions and polarize public opinion. It can exaggerate or distort debates on topics like reproductive rights, gender equality or identity, or sexual harassment, using inflammatory content to manipulate voters' emotions and deepen societal rifts. These issues can be highly contentious in public debate, and for many individuals, hold significant emotional weight as they have direct effects on offline safety. Because they are hot-button topics with strong polarization, individuals may be more likely to engage in these debates or to share their beliefs, thus allowing mis- and disinformation around these issues to be amplified more so than less-contested issues.

Disrupting Electoral Infrastructure

Gendered disinformation campaigns may also seek to disrupt the operational aspects of the electoral process. By spreading false information about voting procedures,

registration requirements, or polling locations, they can confuse or deter potential voters, particularly those from marginalized or vulnerable communities. This can lead to disproportionate impacts on women and their further suppression in contexts where they have historically had impeded access to exercising their right to vote, reducing the overall integrity and inclusivity of the election.

Women election workers have faced ideologically driven and sexist, in addition to racist, online abuse and disinformation campaigns aimed at discrediting them and undermining legitimate election results. Such was the case of poll-workers Ruby Freeman and Shaye Moss, who during the 2020 Presidential elections in the United States were the subjects of an edited video which falsely depicted them as tampering with ballots and committing voter fraud, inciting a deluge of racist harassment and horrific threats. In this case, Freeman and Moss' reputation was tarnished and their credibility as poll workers diminished to achieve the result of undermining the outcomes of the elections. The Brennan Center for Justice notes that in their study on attacks facing election officials in the United States, "for women and election workers of color, the threats were particularly graphic and often laced with racist and gendered insults," noting an uptick in "gendered and racist attacks."

Foreign influence operations and their use of gender

Foreign actors with malign intentions deliberately exploit anti-gender narratives through disinformation, reinforcing conservative gender norms, and perpetuating discrimination against those who hold non-heteronormative identities. When attempting to shape public opinion and interfere in electoral processes, foreign actors may weaponize gender to perpetuate notions about women in positions of leadership, or to further undermine the credibility of election officials based on misogynistic stereotypes, and thus ultimately encourage doubt in the electoral process or election outcomes. For example, when Foreign Minister Annalena Baerbock ran for Chancellor of Germany in 2021, she faced an onslaught of gendered disinformation attacks, including those seeking to undermine her qualifications and manipulated images depicting her nude to damage her credibility and reputation.[86] Pro-Russian actors were one of the most vocal groups targeting Baerbock, amplifying negative sexist narratives and disparaging allegations to exert foreign influence over German political processes by weaponizing her gender to cause disruption.[87] Disinformation targeting men involved in electoral processes can also reaffirm harmful gender norms, enforcing restrictive beliefs around masculinity and discriminatory attitudes towards sexual minorities.

[86] Kristina Wilfore, "The Gendered Disinformation Playbook in Germany Is a Warning for Europe," October 29, 2021, https://www.brookings.edu/articles/the-gendered-disinformation-playbook-in-germany-is-a-warning-for-europe/.

[87] Julia Smirnova et al., *Digitale Gewalt und Desinformation gegen Spitzenkandidat:innen vor der Bundestagswahl 2021* (Institute for Strategic Dialogue, 2021), https://www.isdglobal.org/wp-content/uploads/2021/09/Digitale-Gewalt-und-Desinformation_v5.pdf

PART 4: APPROACHES TO STEMMING MIS- AND DISINFORMATION

The impact of mis- and disinformation on elections has increasingly gained attention from media, international organizations, and national governments seeking to address the destabilizing effect information pollution has on democratic processes. With over half of the world's population registered on social media platforms, addressing election mis- and disinformation is necessary to maintain the integrity of democratic processes and to ensure they function smoothly and without undue influence. As outlined in this section, responses must be led by a multifaceted and multi-stakeholder approach.

Approaches to Stemming Election Integrity Threats

Approaches to strengthening election integrity on social media tend to focus in four directions:

- **Promoting authoritative information about the election process.** In some localities, social media companies support distribution and algorithmic boost of authoritative information from the EMB or other sources of election administration. This may entail collaboration between EMBs and social media companies to strengthen the former's authoritative voice on election-related information, although the amount of effort put in by different platforms may vary, and jurisdictions in smaller markets face challenges in securing support from companies compared to those with larger platform user bases.[88] Furthermore, such impact is modest given that social media platforms view boosting authoritative content in competition with the content generated by users and advertisers.

- **Regulating political advertising.** Political advertising is a form of campaigning that allows candidates to directly convey their message to voters and influence the political debate. By running ads on various types of media, candidates can reach audiences that otherwise may not have been paying attention to the election and build name recognition, highlight important issues, and draw contrast with their opponents. There is value in political advertising in offering voters information and choices. However, without fact checks on candidate ads, this creates an environment where false information can spread unchecked.

- **Curtailing misinformation/disinformation.** Approaches to strengthening election integrity on social media also focus heavily on curtailing mis- and

[88] Countering Disinformation, "Election Management Body Approaches to Countering Disinformation," accessed September 14, 2023, https://counteringdisinformation.org/topics/embs/7-emb-coordination-technology-and-social-media-companies.

disinformation. While platform rules about actions taken to address foreign interference tend to be strong, identification and enforcement have proven to be weak. How domestic disinformation policies are organized varies across digital platforms. Malinformation approaches also vary across platforms. Platforms have distinct approaches on how to handle misinformation from prominent politicians, which is where platform bias toward incumbent ruling parties take effect, as evidenced in India.[89] Refer to the Methods section for more details on approaches to curtailing election-related mis/disinformation.

- **Balance legal approaches with election integrity threats.** Effective responses to disinformation require government regulators and social media companies to strike a balance between nuanced legal principles that govern digital expression while meaningfully addressing online harms to elections. The International Covenant on Civil and Political Rights (ICCPR) lays out three critical principles relevant to the information environment:[90]
    - Article 19 affirms the right to free expression, which entails freedom to seek and receive information, to impart information, and to hold opinions. It also helps governments assess when restrictions on free expression may be legitimate.
    - Article 20 defines when expression should be prohibited, such as advocacy for war or expression of national, religious, or racial hatred that incites violence or discrimination.
    - Article 17 affirms the right to privacy and freedom from its arbitrary or unlawful interference.
    - Article 25 affirms the right to democratic participation, including the right to vote and to freely choose a representation in government.

  Decisions around the oversight of social media during elections that balance out these potentially conflicting provisions are country specific and vary in approach. Social media companies often give preference to the freedom to impart information over protection of privacy or freedom to form opinions, and they have failed to consider the impact of disinformation on rights to democratic participation. Refer to the “policy responses” chapter for more specifics.

[89] Jeff Horwitz & Newley Purnell, “Facebook Executive Supported India’s Modi, Disparaged Opposition in Internal Messages,” August 30, 2020, https://www.wsj.com/articles/facebook-executive-supported-indias-modi-disparaged-opposition-in-internal-messages-11598809348.

[90] International Covenant on Civil and Political Rights (United Nations General Assembly, 1966), https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights.

Methods for Stemming Mis/Disinformation

Fact-checking

Fact-checking plays a crucial role in verifying the accuracy of information. Fact-checkers carefully scrutinize information presented in news articles, political advertisements, social media posts, and other sources to identify mis- and disinformation and verify claims. The process entails identifying potentially misleading or false claims, research, such as cross-referencing with other sources on the topic at hand, and ultimately evaluating the accuracy of the information. Social media companies often collaborate with fact-checkers to flag or label content that has been identified as false or misleading. Meta works alongside the International Fact-Checking Network to "identify, review and take action" on misinformation, and Twitter allows users to provide "additional context," and state they "only intervene if content breaks our rules."[91] Both companies provide the opportunity to label content publicly so users are informed of a post's veracity. However, both companies have also faced significant scrutiny for failing to address election mis- and disinformation on their platforms, rolling back key initiatives to address specific online harms around elections in the U.S., and dedicating even fewer resources to fact-checking in non-English speaking contexts.[92]

> ***Limitations of fact checking*** Fact-checking may be vital for media literacy, discouraging politicians from lying and correcting the journalistic record, but research shows fact checks can oversimplify and distort political conflicts. Brookings Institution researchers found fact-checking mostly influences the politically uncommitted — those who do not have much information about an issue, rather than those who have inaccurate information.[93] Debunking mis- and disinformation can also backfire if the facts threaten the viewers' worldview. A 2020 meta-analysis, a study that systematically combines dozens of research findings, concluded that fact-checking's impact on people's beliefs is "quite weak."[94] While an important tool for attempting to refute mis- and disinformation, it should not be

[91] Meta, "About Fact-Checking on Facebook and Instagram," accessed September 14, 2023, https://www.facebook.com/business/help/2593586717571940?id=673052479947730; Twitter (X), "How We Address Misinformation on Twitter," accessed September 14, 2023, https://help.twitter.com/en/resources/addressing-misleading-info.

[92] Amanda Seitz, "Meta Quieter on Election Misinformation as Midterms Loom," August 5, 2022, https://apnews.com/article/2022-midterm-elections-technology-social-media-voting-0ab5375951df71093d240a6631edb9da; Ali Swenson, "False Claims of a Stolen Election Thrive Unchecked on Twitter Even as Musk Promises Otherwise," May 18, 2023, https://apnews.com/article/elon-musk-twitter-trump-misinformation-election-lies-5137a88a58eaaca0e45ba043db911d15.

[93] Jianing Li & Michael W. Wagner, "When Are Readers Likely to Believe a Fact-Check?," May 27, 2020, https://www.brookings.edu/articles/when-are-readers-likely-to-believe-a-fact-check/.

[94] Nathan Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," *Political Communication* Volume 37, 2020, pp. 350-375, https://www.tandfonline.com/doi/abs/10.1080/10584609.2019.1668894?journalCode=upcp20.

over relied on, given "the direct impact of corrections is often very limited" according to NDI.[95] Furthermore, some analysis suggests that the impact may be further lessened by the ideologies held by an individual even after exposure to results of fact-checking that challenge their beliefs.[96] Misinformation often continues to influence people's thinking even after they receive and accept a correction—this is known as the "continued influence effect." Attacks targeting candidates and election officials also often feature discriminatory elements concerning the individual's identity, such as their race, gender, sexual orientation, and religious beliefs. For example, when disinformation is gendered, fact-checking may not account for embedded stereotypes and underlying biases, further complicated by differing manifestations in gendered disinformation due to varying cultural contexts.[97] This represents a key challenge for fact-checking approaches, as such mis- and disinformation cannot be scrutinized using traditional methods of debunking, and require analysis that is contextualized and conducted with an understanding that these attacks may contain nuance and subtlety that is impossible to detect with this method.

Content Moderation

Content moderation is an oft-discussed tool for addressing election-related mis- and disinformation on social media. Depending on the platform and in the context of disinformation, this can entail "labels, warnings or removal of content," "promoting access to the most authoritative sources, restricting the financial incentives of disinformation by demonetizing content, making disinformation less visible in newsfeeds, timelines or search results, and reducing its reach by penalizing clickbait."[98] The process can be carried out both through automatic moderation, using machine learning algorithms to identify mis- and disinformation, and through human moderation, which relies on manual identification and flagging. The former risks the overlooking of crucial cultural and linguistic context and underlying discriminatory elements, such as racism and sexism, often necessitating the latter, although this can be a laborious and bias-laden process.[99]

---

[95] *Combating Information Manipulation: A Playbook for Elections and Beyond* (International Republican Institute, September 2021), https://www.iri.org/resources/combating-information-manipulation-a-playbook-for-elections-and-beyond/.

[96] Nathan Walter et al., "Fact-Checking: A Meta-Analysis of What Works and for Whom," *Political Communication* Volume 37, 2020, pp. 350-375, https://www.tandfonline.com/doi/abs/10.1080/10584609.2019.1668894?journalCode=upcp20.

[97] Kristina Wilfore, "Security, Misogyny, and Disinformation Undermining Women's Leadership," in Gender and Security in Digital Space (London: Routledge, 2022), https://www.taylorfrancis.com/chapters/edit/10.4324/9781003261605-11/security-misogyny-disinformation-undermining-women-leadership-kristina-wilfore.

[98] United Nations, "Countering Disinformation," accessed September 14, 2023, https://www.un.org/en/countering-disinformation.

[99] Ibid.; *Disinformation and freedom of expression Submission in response to the call by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression…* (Association for Progressive

Social media companies have promoted the use of "misinformation classifiers"—automated systems trained on machine learning to detect and take down posts with harmful falsehoods. However, they are not developed enough to recognize and take action on millions of multilanguage disinformation, thus their applicability to the majority of the world is limited.

Limitations of Content Moderation Social media companies have different policies and standards on content moderation surrounding elections, but critical consensus finds most major platforms have previously demonstrated failure in stemming the spread of election mis- and disinformation, particularly in non-English speaking contexts, citing a lack of transparencies on content moderation policies and decisions and little accountability or incentive for privately owned and operated companies to take definitive action. As a result of the COVID-19 pandemic and the abundance of mis- and disinformation that followed, "there has been increased pressure on internet companies, particularly social media platforms, to monitor, identify, and filter 'untruthful' content circulating on their networks," marking a key moment in which companies were made to face how information pollution is handled on their platforms.[100] However, the risks posed by mis- and disinformation spread on social media platforms long predates the COVID-19 pandemic, requiring swifter and more dedicated action to mitigating the harms it poses to democratic principles. Furthermore, automated content moderation systems are vulnerable to failing to detect all mis- and disinformation and other harmful content, underpinned by biases based on the data they are trained on by design, and failing to apply the contextual analysis and understanding of cultural and linguistic nuances that is often required to make determinations on content that is racist, sexist, and otherwise harmful. While content moderation overseen by humans following automated moderation is frequently proposed as a solution to catching anything that was missed due to the aforementioned vulnerabilities, it is often already too late, as the harmful content has already been pushed to audiences and amplified by platforms' algorithmic designs. While moderation may provide a "quick fix" to addressing the harmful disinformative content, to address the long-term consequences of information pollution, it is a "far more effective approach" to "address the malign behavior used to disseminate false content rather than the content itself."[101]

Communication, 2021), https://www.ohchr.org/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/APC-Disinformation-Submission.pdf.

[100] Agustina Del Campo, "Disinformation Is Not Simply a Content Moderation Issue," October 19, 2021, https://carnegieendowment.org/2021/10/19/disinformation-is-not-simply-content-moderation-issue-pub-85514.

[101] Julian Jaursch et al., "Tackling Disinformation: Going Beyond Content Moderation," November 15, 2019, https://www.institutmontaigne.org/en/expressions/tackling-disinformation-going-beyond-content-moderation.

CASE STUDY: East Africa Content Moderation Lawsuits

In February 2022, the Nairobi, Kenya, office of Sama, an organization that won a contract to provide content moderation for Facebook, was exposed as allegedly utilizing abusive labor practices for content moderation by the actions of whistleblower Daniel Motaung. Sama was the "epicenter of Facebook's content moderation operation for the whole of Sub-Saharan Africa," but workers were being paid as little as $1.50 per hour, and faced "a workplace culture characterized by mental trauma, intimidation, and alleged suppression of the right to unionize."[102] Content moderators in Kenya have described this work as "torture" and report being pressured to watch over a 1,000 videos a day which included child molestation, torture, and live murder. Over 200 former employees are suing the local contractor Sama and Facebook in a suit that may ripple worldwide.

Facebook claims to have spent over $5 billion on safety measures in 2021 and contracts over 15,000 content moderators globally, mostly through third-parties like Sama. Employees were told they would only be exposed to false information, but were instead exposed to traumatic and disturbing content. Ultimately, the case of Sama is an example of how major digital platforms, like Facebook, outsource trauma to developing countries, where labor is much cheaper.

A related class-action lawsuit was filed in Nairobi (where the content moderation hub for Eastern and Southern Africa was based), accusing the company of monetizing the viral potential of hate and violence in conflict-torn Ethiopia, in violation of more than 10 articles of Kenya's Constitution. It also alleges the company does not devote enough resources to content moderation on the continent compared to the United States.

Media Literacy Initiatives

The health of democracies depends on citizens with skills to navigate a polarized media and information landscape, recognize emotional manipulation, and identify falsehoods, manipulation, and conspiracies. Also described as "digital literacy," media literacy is geared toward providing skills for citizens to succeed in an increasingly digital world by helping people analyze and evaluate information online from its sourcing and watch for

[102] Billy Perrig, "Inside Facebook's African Sweatshop," last updated February 17, 2022, https://time.com/6147458/facebook-africa-content-moderation-employee-treatment/.

the hallmarks of manipulation. In a period of great partisan division, and multiple interpretations of freedom of speech in different
democracies, proponents of digital literacy argue that it enables and empowers the individual and avoids debates over censorship of the community. Inoculation methods are an oft-discussed avenue of addressing individuals' ability to discern the accuracy of content they encounter in digital spaces. According to RAND, media literacy "teaches participants to consider the implications of message construction from numerous angles, such as how the motivations of those disseminating information could influence content selection and framing and how different kinds of media and other technologies affect the nature of communication."[103] The use of such approaches suggests that "by providing knowledge and skills to refute media messages, literacy interventions may help audiences to resist the influence of harmful media content."[104] Proponents of media literacy purport it fosters critical thinking skills and cynicism and have packaged approaches from various countries in applying teaching methods for such initiatives.[105]

Limitations of Education and Digital Literacy Initiatives. In many cases, media literacy is regarded as a quick-fix solution or misplaced shuffling off responsibility from the state — away from corporate accountability — to the individual. Media literacy has no impact on digital platforms that amplify sensationalized and misleading content for profit. The efficacy of media literacy based on its aims has prompted contentious scholarly debate, with some coming to media literacy's defense and others arguing for its demise.[106] Media theorist danah boyd asserts that the project of media literacy has backfired. The same consumption and production practices that media literacy advocates celebrate, boyd argues, are now being weaponized in reactionary agendas for climate denialism and white supremacy.[107] In other words, promoting a posture of generalized skepticism can be used to dismiss credible information as easily as it can support propaganda or racist conspiracies. Furthermore, inoculation methods against mis- and disinformation have achieved varying levels of influence on individuals' ability to discern fake news or mis/disinformation. These efforts prioritize treating reactions to mis- and disinformation over targeting the actors behind it.

---

[103] Alice Huguet, Exploring Media Literacy Education as a Tool for Mitigating Truth Decay (RAND Corporation, 2019), https://www.rand.org/pubs/research_reports/RR3050.html

[104] Se-Hoon Jeong, & Hyunyi Cho, & Yoori Hwang, "Media Literacy Interventions: A Meta-Analytic Review," *Journal of Communication*, 62(3), 2012, pp. 454–472, https://doi.org/10.1111/j.1460-2466.2012.01643.x

[105] *Learn to Discern: Media Literacy Trainer's Manual* (IREX, 2020), https://www.irex.org/resource/learn-discern-media-literacy-trainers-manual.

[106] T. Philip Nichols & Robert Jean LeBlanc, "Media Education and the Limits of "Literacy": Ecological Orientations to Performative Platforms," *Curriculum Inquiry* Volume 51, 2021, https://www.tandfonline.com/doi/full/10.1080/03626784.2020.1865104.

[107] danah boyd, "You Think You Want Media Literacy… Do You?," March 9, 2018, https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2.

Counter Speech - Debunking and Inoculation

Counter speech is a communication tactic underpinned by the notion that disinformation and hate speech should be counteracted by a surplus of credible information and speech.[108] It may be deployed in response to mis- and disinformation through fostering alternative dialogues, challenging harmful narratives, and educating people and organizations to debunk and counter these narratives. Counter speech can be utilized by a number of actors, such as the intended targets of disinformation, as well as ordinary citizens and people in authoritative roles.[109] In the context of elections, nonpartisan organizations and media outlets often attempt to stop disinformation from taking root in the first place—to "inoculate" audiences against false or misleading information. Numerous studies have shown that when individuals are provided accurate information about a topic from a trusted messenger, it reduces the impact of disinformation. Research indicates that voters most at risk from election disinformation are new voters, and infrequent voters who do not have as much experience navigating complex election systems. Increasingly, EMBs and nonpartisan voting rights organizations produce digital content that combines voter information and messaging with "prebunking" to stop disinformation before it breaks out. Election officials and judges are important sources of trusted information; however, elections officials are often underfunded, understaffed, and have limitations on the reach of the content they produce.

The **Debunking Handbook 2020** summarizes the current state of the science of misinformation and its debunking. It was written by a team of 22 prominent scholars of misinformation and its debunking, and it represents the current consensus on the science of debunking for engaged citizens, policymakers, journalists, and other practitioners.[110]

> *Limitations of Counter Speech and Debunking.* Placing the burden on groups to counteract information pollution and harmful speech allows platforms to circumvent taking responsibility for such content, especially when it is violative of their own community standards and terms of service.[111] Disinformation experts have also noted the use of "censorship by noise" to manipulate how certain narratives are shaped by inundating social media platforms with conflicting and

[108] Joshua Garland, "Impact and Dynamics of Hate and Counter Speech Online," *EPJ Data Science* Volume 11, 2020, https://doi.org/10.1140/epjds/s13688-021-00314-6.

[109] *Bianca Cepollaro, Maxime Lepoutre, & Robert Mark Simpson, "Counterspeech," Philosophy Compass, 18(1), 2022, https://doi.org/10.1111/phc3.12890.*

[110] Stephan Lewandowsky et al., *The Debunking Handbook 2020* (Skeptical Science, 2020), https://www.climatechangecommunication.org/wp-content/uploads/2020/10/DebunkingHandbook2020.pdf

[111] Bianca Cepollaro, Maxime Lepoutre, & Robert Mark Simpson, "Counterspeech," Philosophy Compass, 18(1), 2022, https://doi.org/10.1111/phc3.12890; Governing Hate Speech by Means of Counterspeech on Facebook (Japan: 66th ICA Annual Conference, 2016), https://www.researchgate.net/publication/303497937_Governing_hate_speech_by_means_of_counterspeech_on_Facebook.

disinformative information surrounding the targeted topic, so as to undermine unabated discussion and "drown out all other viewpoints."[112] While the purpose of counter speech is often framed as to combat mis- and disinformation through an abundance of pushback, it can unintentionally result in the same consequences as those achieved by actors sharing increased amounts of content to mislead and cause harm.

Design Feature Changes

Design feature changes include changes to the visual representation of information on platforms, such as providing context to information presented in social media posts or warning labels that indicate the veracity of a post's content. This feature is particularly relevant during election periods, when there is an uptick in mis- and disinformation surrounding the electoral process, political issues, and candidates.[113] Adding "friction" to tech platform design involves methods and tools that can tilt users toward more reflective forms of interaction by aiming to put small hurdles into online interactions in order to prevent damaging reactive communication and behavior. Contextual information and warning labels can provide immediate notification to social media users of the truthfulness of a post, which may limit its proliferation and influence users from further amplifying false or misleading information. In the case of messaging apps, limits placed on the number of users one can forward to, on WhatsApp, for instance, appear to have slowed the proliferation of disinformation in India and elsewhere.[114] Initially, WhatsApp users could forward a message to up to 256 groups at once; that number was cut to 20 in 2018 and 5 in 2019, a move first tested in the wake of Indian mob violence. In 2020, the limit dropped to one, but only for messages that had already been forwarded five or more times.

Limitations of Design Features There are arguments about which are the right frictions to add and which ones might be too onerous or too restrictive of particular opinions or actions. Not all frictions work equally well, requiring lots of testing and data. "Amid all the debates over deplatforming — which tend to involve individual users, often with their own fan bases — it is good to also make room for more

[112] CBC Radio, "World Leaders Enact 'Censorship Through Noise' in the Digital Era, Says Author," last updated January 31, 2020, https://www.cbc.ca/radio/asithappens/as-it-happens-tuesday-edition-1.5277837/world-leaders-enact-censorship-through-noise-in-the-digital-era-says-author-1.5277847; Tactics of Disinformation (Cybersecurity and Infrastructure Security Agency, 2022), https://www.cisa.gov/sites/default/files/publications/tactics-of-disinformation_508.pdf.

[113] International IDEA, "The Information Environment Around Elections," accessed September 14, 2023, https://www.idea.int/our-work/what-we-do/elections/information-environment-around-elections.

[114] Joshua Benton, "Whatsapp Seems Ready to Restrict How Easily Messages Spread in a Bid to Reduce Misinformation," April 4, 2022, https://www.niemanlab.org/2022/04/whatsapp-seems-ready-to-restrict-how-easily-messages-spread-in-a-bid-to-reduce-misinformation/.

structural changes within a platform that can approach the problem from a different angle" according to the Neiman Lab.[115] Other critics of design feature changes have identified the possibility of further entrenching ideologies through a "backfire effect," which "occurs when an evidence-based correction is presented to an individual and they report believing even more in the very misconception the correction is aiming to rectify," although it is debated whether this phenomenon has been empirically established.[116] Furthermore, it must be noted that this is a band-aid solution to a problem social media companies are perpetuating themselves: the algorithmic amplification of mis- and disinformation is profitable, and thus the platform-facilitated spread of information pollution must be addressed at the root.

Independent Research & Investigations on Election Mis/Disinformation

Independent investigators and researchers, largely from academic or nonpartisan civil society, can play a vital role in holding social media platforms accountable for curtailing online harms. They identify and document various forms of online harms, exposing policy and enforcement gaps that allow harmful content to proliferate. Through expert analysis and recommendations, they inform platforms, policymakers, and regulators on effective strategies to mitigate these harms. Independent investigators and researchers may collaborate with platforms and regulators, sharing their findings and expertise to shape policies and practices, or engage in advocacy efforts to raise public awareness about online harms, while promoting ethical practices and accountability within the industry. Their work drives improvements in platform policies, practices, and industry standards, contributing to the creation of safer digital environments, and is particularly crucial as social media companies largely lack economic incentives to strengthen their responses to harms facilitated by their platforms, both due to the profitability of sensationalized content and the state of the market: according to legal scholar Jack M. Balkin, "market competition won't produce the kind of culture and knowledge necessary for democratic self-government, democratic culture, or the growth and spread of knowledge," but rather "will overproduce conspiracy theories and speech that undermines democratic institutions," necessitating external pressure to meet the challenges posed by information pollution.[117]

Limitations of Independent Investigations Independent investigators and researchers face challenges when attempting to provide third-party guidance and

[115] Ibid.

[116] Briony Swire-Thompson, Joseph DeGutis, & David Lazer, "Searching for the Backfire Effect: Measurement and Design Considerations," *J Appl Res Mem Cogn.* 9(3), 2020, pp. 286–299, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7462781/.

[117] Jack M. Balkin, "How to Regulate (and Not Regulate) Social Media," March 25, 2020, https://knightcolumbia.org/content/how-to-regulate-and-not-regulate-social-media.

opinions, as they may be granted minimal access to necessary data needed to make informed determinations.[118] Furthermore, they typically have access too much fewer resources to support social listening around elections and digital forensic investigations compared to social media companies, who have the substantial access to data needed to conduct in-depth research on the harms facilitated by their platforms, and to rectify their shortcomings. In some contexts, election officials may also feel skepticism towards civil society research on social media, concerned that such efforts are in fact not independent but rather swayed by perceived partisan affiliations. Such apprehension can be made worse in jurisdictions experiencing high levels of political polarization, where there is a distrust of civil society pushed by government leaders, or where accusations of undue influence may be particularly damaging. Therefore, platforms must be incentivized to provide full transparency on their policies and actions to curtail online harms, and relationships underpinned by trust and collaboration between election officials and CSOs conducting research strengthened, so independent investigators and researchers have the ability to hold companies fully accountable.

PART 5: POLICY AND REGULATORY RESPONSES

There is a rising demand for more oversight of the social media space from government regulators, particularly in the face of the platforms' reluctance to self-regulate. Self-regulation refers to the process of platforms holding themselves accountable for abiding by their own terms of service and policies, and utilizes internal governance to ensure accordance with their own guidelines on various forms of harmful content and malign uses of their services, such as illegal content, mis- and disinformation, hate speech, data privacy, inauthentic activity, and more, in addition to ensuring reporting mechanisms and algorithms operate in a way that reduces harms.

Regarding the threat to information integrity and elections specifically, many major social media companies have increasingly outlined a commitment to tackling mis- and disinformation on their platforms through changes to policies, bolstering "authoritative information" such as through content labeling, sharing key data with researchers and CSOs, promoting digital literacy, and strengthening community responses to mis- and disinformation.[119] Despite these outward-facing efforts self-regulation has largely failed as an adequate response. Critics contend there is a natural disincentive for private

[118] Nathaniel Persily & Joshua A. Tucker, "How to Fix Social Media? Start With Independent Research.," December 1, 2021, https://www.brookings.edu/articles/how-to-fix-social-media-start-with-independent-research/.
[119] Countering Disinformation, "Platform Specific Engagement for Information Integrity," accessed September 14, 2023, https://counteringdisinformation.org/topics/platforms/0-overview-platforms.

companies who profit off of such content to abide by their own standards, thereby necessitating independent oversight.

Thematic Areas for Digital Platform Reform

While there is no one-size fits all approach to social media regulation, the three thematic areas of reform that analysts believe address the major problems driving mis/disinformation include:[120]

- Establishment of online safety standards
- Protection of personal information / data and surveillance capitalism[121]
- Placing limitations on market power of large companies

In order to counter election-related disinformation, governments should incentivize private firms to actively promote a healthier information ecosystem, and regulate them to ensure they do when incentives fall short. Generating policy on curtailing online harms can be the most effective path for accountability. Platforms should commit to work with regulatory and oversight bodies to enforce accountability of their actions, as appropriate. Social media platforms should aim to inform the public about how their policies explicitly prevent the spread of disinformation, as well as provide metrics on how it has done so at regular intervals. The evidence shows that without controls, the cooperation and transparency of the companies is mixed at best, particularly in the Global South.

Large digital platforms control where the majority of citizens get information about elections and occupy a dominant market position globally, which accentuates the already profound inequalities in basic media access within nations and among continents. The lack of transparency and safety standards – even in cases when independent researchers and media have surfaced evidence of major foreign influence operations during elections – is a troubling indicator of the problems inherent in the status quo approach to social media and elections. However, there are efforts being designed in many countries to strengthen accountability and oversight through new online safety legislation. The United Kingdom's Online Safety Bill (recently passed through Parliament) sets out a broad framework for regulation including putting a powerful, independent regulator in charge of overseeing risk management regimes of all social media and introducing a sanctions regime to hold private companies to account if they do not prevent harmful content from reaching people with respect to topics such as child abuse and terrorsim, fraudulent or

[120] Reset.tech, accessed September 14, 2023, https://www.reset.tech/.

[121] John Laidler, "High Tech Is Watching You," March 4, 2019, https://news.harvard.edu/gazette/story/2019/03/harvard-professor-says-surveillance-capitalism-is-undermining-democracy/.

harmful advertisements, and illegal content. Provisions to bolster protecting freedom of speech and privacy were built into the design of the legislation.[122]

Platform investments in policy, safety, and integrity must be determined by the level of risk they pose to human rights, not just by the commercial value of a particular country or whether they are located in jurisdictions with enforceable regulatory powers. Requiring private sector companies to commit to human rights assessments is consistent with the United Nations Guiding Principles on Business and Human Rights[123] and shifts the burden of responsibility on companies to demonstrate that their products and policies advance human rights and freedom of expression. Requiring risk assessments for Very Large Online Platforms (VLOPs) as defined by the European Union (such as X/Twitter, Instagram, Facebook, and YouTube) covering illegal content, personal harms, and manipulative and inauthentic use of platforms is a key facet of new regulations in the digital marketplace enacted by the European Parliament in August 2023, further described in part 5.[124]

The tragic impacts of viral hate speech in Ethiopia, Myanmar, and countless other places, as well as Kremlin-backed efforts to distort public opinion related to the war in Ukraine, demonstrate the need for properly and equitably resourced approaches in different linguistic environments.[125] Tech platforms must provide resourced moderation teams in all languages, including both cultural and linguistic competency. While global elites may be better connected everywhere, the same is not true of those who work for them. Media systems offer tremendous communication resources to people who can function in Western languages, are able-bodied and have the necessary buying power.[126] As pushback against labor violations stemming from the global domination of largely Western-based tech companies, more than 150 workers whose labor underpins the AI systems of Facebook, TikTok, and ChatGPT established the first African Content Moderators Union in 2023.[127] The formation of tech worker rights and protections could have significant consequences for the businesses of some of the world's biggest tech

[122] UK Parliament, "Online Safety Act 2023," accessed November 1, 2023, https://bills.parliament.uk/bills/3137.

[123] *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework* (UN Office of the High Commissioner for Human Rights, 2011), https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf.

[124] Guide to the Digital Serves Act, "Article 26 - Risk Assessment," accessed September 14, 2023, https://digitalservicesact.cc/dsa/art26.html.

[125] *Digital Services Act: Application of the Risk Management Framework to Russian Disinformation Campaigns* (European Commission, 2023), https://op.europa.eu/en/publication-detail/-/publication/c1d645d0-42f5-11ee-a8b8-01aa75ed71a1/language-de.

[126] The Conversation, "Why the Media Is a Key Dimension of Global Inequality," November 28, 2016, https://theconversation.com/why-the-media-is-a-key-dimension-of-global-inequality-69084.

[127] Billy Perrigo, "150 African Workers for ChatGPT, TikTok and Facebook Vote to Unionize at Landmark Nairobi Meeting," May 1, 2023, https://time.com/6275995/chatgpt-facebook-african-workers-union/.

companies.[128] Outsourced human-based content moderation has left many content moderators suffering from PTSD with jobs that are some of the lowest-paid in the global tech industry.

CASE STUDY: Brazil & Digital Platform Regulation

In the absence of comprehensive regulation of digital platforms, the Brazilian Supreme Court acted proactively to curtail threats against the integrity of the election in 2022. Tragically, even after traumatic events to democracies over the world, social media companies did not carry through on their commitments to effectively address election disinformation, polarizing hate speech, and political attack ad campaigns in the lead up to the Brazilian election.

A Memorandum of Understanding (MOU) between major digital platforms and the Superior Electoral Court (TSE) was signed months in advance of the election to establish a program to combat mis- and disinformation involving the judiciary and to disseminate credible information about the 2022 general elections. Civil society and academic organizations also demanded the adoption of more effective measures and adequacy of community guidelines against harmful content that could affect the Brazilian electoral process.

While the platforms had policies to combat disinformation against the integrity of the electoral process, they were administered inconsistently and with little transparency. Although the MOU included specific language that the platforms would shut down lies against the voting machines, disinformation about voting machines in previous elections was allowed. Furthermore, disinformation against candidates remained with few restrictions, either because of the absence of specific policies (as in the case of Twitter and YouTube) or because of exceptions given to politicians and candidates (in the case of Facebook and Instagram). Meta platforms did not have policies that determined action in the face of demonstrably false content that alleged electoral fraud, despite the TSEs efforts to hold them to account.

Rio de Janeiro Federal University's Netlab showed that 53% of the problematic content the Court sent to Meta was still circulating online just weeks before the election. Global Witness studies also showed how Google and Meta allowed ads to be published from

[128] Foxglove, "Tech Workers Rising: German Workers Elect First TikTok Works Council," October 21, 2022, https://www.foxglove.org.uk/2022/10/21/german-tiktok-works-council/.

abroad carrying disinformation about Brazilian elections and with local accounts calling for incitement of violence against the election results.

In the waning days of the election the Electoral Court was given the power to force companies to remove disinformation, which was hotly debated. The Court merely took action against what is legally defined as irregular advertising in the election, which is a longstanding and historical position within Brazilian law. The Judiciary affirmed their right to reinterpret competencies established by law and accelerate procedures and forms of enforcement of digital platforms to overcome their inefficiency and inaction.

Tech company executives were called upon in Brazil, well in advance of the election, to fix the problems linked to their products. They failed to respond to the problem or refused to remove harmful content, placing all responsibility precisely on the judiciary, contributing to the attempts to overthrow the government on January 8th.

Since the election, a package of bills were introduced to place severe new restrictions on what social networks can promote online, including liability for platforms that spread "untrue facts." The bill also has a new "must-carry" clause that obliges platforms to host public interest announcements.[129] Brazilian civil society and the digital rights movement noted that while they have long pressured tech companies for transparency and accountability, "these actors
have filled the civic space with empty promises and inefficient mechanisms" and instead are advocating for a framework for responsibility and transparency for digital platforms in Brazil that can be a baseline for a more democratic, safe, and healthier internet.[130]

Free Expression Considerations

It is essential to recognize that social media sites are important venues for users to exercise free speech rights protected by international law and enshrined in many nations' constitutions. Despite valid concerns around the power and use of social media platforms, they can have an extraordinarily positive effect on freedom of expression, facilitating public debate and strengthening social movements and trust in election processes.

[129] Steven Grattan, "Social Media Failing to Keep Up With Brazil Electoral Disinformation, Rights Groups Say," October 28, 2022, https://www.reuters.com/world/americas/social-media-failing-keep-up-with-brazil-electoral-disinformation-rights-groups-2022-10-28/.

[130] *Protecting Against Harm: International Organizations Stand in Solidarity With Brazil To Hold Big Tech Accountable* (AI Forensics et al., 2023), https://desinformante.com.br/wp-content/uploads/2023/05/Brazilian-2630-Bill_International-solidarity-Open-Letter_Version-to-signers.pdf.

The speech standards decided upon by private companies have various impacts in very different political contexts such that it is hard to ascertain what is 'too much' or 'too little' speech. Unregulated speech can be easily weaponized by authoritarian actors hoping to rollback democratic rights in the form of disinformation campaigns, as was seen in Myanmar.[131] Despite the importance of regulation, NDI notes, "care is needed to not subvert freedom of expression while trying to protect the integrity of the information space in elections and beyond them."[132]

In countries where traditional media is more restricted, under the control of governments or corporate agendas, social media often provides a unique space for expression, if not always as freely as it should. Countries like Iran and China have blocked access to certain social media networks in their entirety, while others have engaged in targeted blocking.[133] More commonly, countries use repressive laws to criminalize the posting and sharing of dissenting or controversial content.[134] The weaponizing of speech regulations is particularly prevalent in countries that have weak or failing democratic institutions. This includes governments with a history of institutional corruption, or societies with cultures of racism, sectarian tensions and/or religious violence. This places platforms in the tedious position of determining whether to avoid repercussions for failing to comply or continuing operations over safety and preserving the integrity of democracy.

In order to mitigate concerns about curtailing free expression, democratic governments can provide a basic framework for regulation that requires companies to be transparent, establish due process, and not discriminate in their enforcement. Inauthentic behavior online — in the form of bots, fake profiles, or hired provocateurs to amplify misinformation, disinformation, lies, hate speech, or conspiracy theories — should be the foundation in which policies are formed. Manipulated content through malign behavior should not be considered a threat to free expression, in fact the presence of fraudulent accounts and false narratives online undermine free expression. Content moderation focused on malign behavior rather than malign content/speech— in the form of companies banning, stopping amplification, or labeling content posted on their platforms—can be a way to tackle inauthentic behavior and illegal content and avoid limiting the free expression of individuals online.

---

[131] Reuters, "Protecting Against Harm: International Organizations Stand in Solidarity With Brazil To Hold Big Tech Accountable," September 12, 2022, https://www.reuters.com/world/asia-pacific/un-investigator-says-facebook-provided-vast-amount-myanmar-war-crimes-2022-09-12/.

[132] *Disinformation and Electoral Integrity: A Guidance Document for NDI Elections Programs* (NDI, 2019), https://www.ndi.org/publications/disinformation-and-electoral-integrity-guidance-document-ndi-elections-programs.

[133] Committee to Protect Journalists, "10 Most Censored Countries," September 10, 2019, https://cpj.org/2015/04/10-most-censored-countries/.

[134] Article 19, "Regulating Social Media: We Need a New Model That Protects Free Expression," April 25, 2018, https://www.article19.org/resources/regulating-social-media-need-new-model-protects-free-expression/.

In the debate over what counts as free expression online, critics of digital platform regulations often lose sight of how the platforms already prioritize the speech of powerful groups. This, combined with the algorithmic amplification of hate speech,[135] leads to a chilling freedom of expression for groups already heavily discriminated against and silenced in society, such as ethnic, religious, or sexual minorities.[136] This is true even in notionally 'free' discursive contexts. Pre-existing biases and institutional hegemony mean that some barriers to expression are invisible. These unofficial barriers to expression must be dismantled in favor of a more equitable regulatory environment. However, it is also true that ill-intended regulation can serve authoritarian actors just as well. Different forms of state-sponsored censorship can have a negative impact on these same targeted communities.[137]

[135] Audun Fladmoe & M. Nadim, "Silenced by Hate? Hate Speech as a Social Boundary to Free Speech," in *Boundary Struggles : Contestations of Free Speech in the Norwegian Public Sphere* (Cappelen Damm Akademisk, 2017), https://www.semanticscholar.org/paper/Silenced-by-hate-hate-speech-as-a-social-boundary-Fladmoe-Nadim/74011dee3446a8324bd7d6c6d0c28a6e2e8f4c89.

[136] Nesrine Malik, "The Myth of the Free Speech Crisis," September 3, 2019, https://www.theguardian.com/world/2019/sep/03/the-myth-of-the-free-speech-crisis.

[137] Access Now, "Internet Shutdowns in 2021: The Return of Digital Authoritarianism," last updated March 17, 2023, https://www.accessnow.org/internet-shutdowns-2021/.

CASE STUDY: Turkey and X (Twitter) in the 2023 Presidential Election

Turkey's president, Recep Tayyip Erdoğan, and his Justice and Development Party (AKP) have infamously utilized social media platforms to target opposition and further cement his power, turning to Twitter in particular which has over 16 million users in the country.[138] With a history of weaponizing the platform to run influence operations documented by Stanford University, troll campaigns were used "to shape public opinion and counter government critics on social media" and organize "'online lynching' campaigns targeting journalists, politicians and government critics." The AKP's youth faction was also singled-out by Twitter for utilizing inauthentic methods to push pro AKP narratives and spread misleading information targeting the political opposition.[139]

Since 2020, the government has enacted legislation concerning their power over content shared on social media, providing them with "wide-ranging powers" to do so and threats of penalties and lessening access to platforms if they failed to comply, and another bill passed in 2022 empowers the government to target individuals spreading what they deem as disinformation on social media platforms.

The Turkish government led the charge on having the most takedown requests issued to Twitter for content they perceived as noncompliant with their legislation. After acquisition of the platform by business magnate Elon Musk, Twitter has complied with nearly 90% of these requests, allegedly due to fear of the Turkish government revoking citizens' access to the platform, which would be a large blow as the country is its seventh largest market.[140] Ahead of the May 2023 presidential elections, Twitter faced widespread backlash for actioning content and accounts critical of Erdoğan, including those of activists and journalists, to avoid this exact scenario.[141] The move was largely planned as one which shows deference to a leader whose authoritarian tactics to stifle opposition to his rule, which some security analysts refer to as a form of conflict profiteering, a strategy that exploits instability and crises through complying with authoritarian regimes to bend to their requests thereby making the platform party in censorship and the stifling of legitimate political debate via content removal.[142]

---

[138] Rob Minto, "How Turkey's Government Gamed Twitter," May 15, 2023, https://www.newsweek.com/twitter-turkey-content-removal-pro-government-accounts-1800328.

[139] Shelby Grossman et al., *Political Retweet Rings and Compromised Accounts: A Twitter Influence Operation Linked to the Youth Wing of Turkey's Ruling Party* (Stanford Freeman Spogli Institute, 2020), https://fsi.stanford.edu/publication/june-2020-turkey-takedown.

[140] Rob Minto, "How Turkey's Government Gamed Twitter," May 15, 2023, https://www.newsweek.com/twitter-turkey-content-removal-pro-government-accounts-1800328.

[141] Perry Stein, "Twitter Says It Will Restrict Access to Some Tweets Before Turkey's Election," May 13, 2023, https://www.washingtonpost.com/technology/2023/05/13/turkey-twitter-musk-erdogan/.

Policy Approaches

As prefaced above, policy responses to curtailing online harms outside of during elections vary. Approaches to regulating social media platforms to date have typically targeted content moderation on mis- and disinformation, transparency, and accountability. In some countries, if companies fail to comply with regulatory policy, they can face fines and sanctions, or block access.

The concern about overstepping citizens' and users' rights to share their viewpoints free from censorship and punishment has thwarted the development of policies in the nascent field of social media policy. However, mis- and disinformation has severe consequences for exercising democratic freedoms, such as the withdrawal of voters from political participation and engaging in public debate, and dissuading them from running for political office or working in election facilitation roles. Therefore, it is crucial policy responses evolve to address election-related online harms and incentivize platforms against facilitating the spread and impact of mis/disinformation, harassment, and abuse.

When shaping a new regulatory environment, analysts recommend starting with behavior that is illegal offline as an extension of what should be considered illegal online. Death threats, sexual violence, and incitement of violence is already criminalized in many countries and therefore should not be allowed online in these countries, yet evidence shows they are flourishing in jurisdictions without oversight.

Misinformation that may fall short of a criminal standard can nonetheless be harmful to users because of its nature, intensity, or repetition or because some users may be more sensitive to that content, such as younger social media users or because it is targeted at people with particular characteristics or vulnerabilities. Much of this type of content has already been recognized as harmful content by the platforms themselves, which is embedded in their own terms of service, yet lacks a consistent response.

While progress has been made in recent years, social media companies still have much to do to reduce the spread of disinformation and combat malicious activity during elections. Some political parties have compiled a comparative policy analysis to present social media companies with additional potential solutions.[143] This is a useful practice which can be utilized in many contexts to tackle information pollution on social media platforms.

---

[142] Megan Cerullo, "Twitter Under Fire for Restricting Content Before Turkish Presidential Election," May 16, 2023, https://www.cbsnews.com/news/twitter-censoring-content-recep-tayyip-erdogan-turkish-presidential-election/.

[143] Democratic National Convention, "DNC Recommendations for Combating Online Misinformation," accessed September 14, 2023, https://democrats.org/who-we-are/what-we-do/disinfo/comparative-social-media-policy-analysis/.

CASE STUDY: European Union's Digital Services Act

The adoption of the European Union's Digital Services Act (DSA) in 2022 marked a landmark effort to curtail harms facilitated by online platforms through robust regulations and compliance requirements. The act outlines approaches to platform regulation such as mandating users receive more information about content removal and their ability to oppose these decisions, mandating transparency on terms of services, policies, and algorithmic designs, mandating the mitigation of risks including those posed by "disinformation or election manipulation, cyber violence against women, or harms to minors online," banning advertisements that are targeted through the use of personal data, and strengthening reporting mechanisms for users.[144]

A unique feature of the DSA is the mandating of annual risk audits independently led by "very large platforms," requiring reporting to encompass transparency on their content moderation activities and the status of their efforts to mitigate the aforementioned risks so as to determine compliance with the act's provisions.[145] Following their own audit, an independent outside party is hired to substantiate their findings and conduct their own compliance assessments, and throughout this process, platforms must provide data to regulators and external auditors to allow for in-depth evaluations of their compliance with the DSA's regulations and to conduct "further investigate their systemic risks," as noted by Claire Pershan, the Mozilla Foundation's EU Advocacy Lead.[146] Such audits serve as an important opportunity for regulators to determine if it is necessary to impose penalties, and can provide insight into how to shape future regulatory efforts to respond to the shortcomings of VLOPs.

In 2023, the European Commission released a comprehensive report investigating pro-Kremlin disinformation and the role of social media companies in amplifying these campaigns in the EU.[147] The report assesses how the regulatory measures of the DSA can be applied to the platforms' responses to this issue, and evaluates the performances of platforms utilizing the new framework, looking in particular at measures around risk audits and assessments.[148] The findings clearly indicate a previous failure to act in alignment with these new standards. Rigorous efforts to determine compliance shed light on the past shortcomings of social media companies to mitigate the spread of information pollution on their platforms, and thus place increased pressure on them to step up to obligations mandated by the DSA and shoulder the responsibility of ensuring their own compliance. This effort signals the EU's very serious intent to regulate, and goes beyond simply addressing product safety failures by thoroughly applying the compliance standards of the DSA to identify gaps between how companies have been responding to information pollution and what is now required.

The DSA is a fundamental step towards ensuring accountability and transparency from digital platforms, and can serve as a model to other jurisdictions where comprehensive social media regulation does not yet exist. Key principles of the DSA, such as obligations around transparency, mitigating the spread of information pollution, increasing obtainability of otherwise hard-to-access data, and the imposition of penalties to hold platforms accountable for failing to comply with its measures, can be adopted at a national level and tailored to fit many different contexts globally, so as to foster much-needed, comprehensive regulatory action targeting social media platforms.

Areas of Regulation

There are many areas that must be addressed when considering crafting policy around curtailing online harms, including:

- **Transparency measures,** focused on requiring accessibility to information on how algorithms are designed to promote certain content, on data practices, and on advertising policies. The choice of what to access and consume on social media platforms is less of a personal decision and more of algorithmic recommendations and curation. Social media platforms collect and collate large amounts of data from their users, oftentimes without consent or transparency in the handling, protection or use of the data. This data is manipulated algorithmically, often privileging and amplifying sensational news including mis/disinformation, hate speech, polarizing and extremist content. Platforms are often criticized for not doing enough to address the prevalence of these online harms and for lacking transparency in how their algorithms function.

- **Data protection and privacy laws**: Since platforms collect vast amounts of data, there should be greater transparency and accountability on how the data is

[144] European Commission, "Questions and Answers: Digital Services Act*," accessed September 14, 2023, https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348; European Commission, "The Digital Services Act: Ensuring a Safe and Accountable Online Environment," accessed September 14, 2023, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.

[145] "Digital Services Act: Commission Designates First Set of Very Large Online Platforms and Search Engines," April 25, 2023, https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413.

[146] Claire Pershan, "As the Digital Services Act Takes Shape, Are Platform Accountability Experts at a Crossroads?," May 26, 2023, https://foundation.mozilla.org/en/blog/digital-services-act-and-platform-accountability/.

[147] European Commission, *Digital Services Act.*

[148] Ibid.

gathered, stored and used, including safeguards against data mishandling and breaches. This requires legislative frameworks on usage to prevent exploitation of sensitive user information, and promote user rights to privacy and security.

- **Content moderation regulations,** such as those which pertain to enforcing terms of services, either requiring adaptation to meet emerging challenges or stricter adherence to following the rules established by companies themselves, or provisions that target the use of automated moderation to ensure they operate with minimal error.
    - Online harassment and cyberbullying measures: platforms should explicitly address harassment and bullying in their terms of services and community guidelines, and should provide clear and accessible channels for reporting. Platforms should take immediate action on reported harms and engage in full transparency on how they are tackling harassment and bullying. Discriminatory harassment and bullying, such as that which is racist and sexist, should be clearly denoted as a violation of terms of service and community guidelines.
    - Banning political deepfakes, which can have a particular impact for women candidates and election officials: "deepfakes are often used to discredit women candidates and public officials, so sanctioning the creation and/or distribution of deepfakes, or using existing legal provisions to prosecute the perpetrators of such acts, could have an impact on disinformation targeting women that serve in a public capacity."[149]

- **Antitrust and Competition Measures:** a handful of social media companies drive the digital technology market. Antitrust legislation targeting social media companies has been debated in countries like the U.S., with the intent to soften the power of a few companies' corner on the market.[150]

- **International Cooperation and Standardization:** Global frameworks can be established through collaborative efforts to reach consensus on regulatory approaches to countering disinformation. This is achieved through the cross-border sharing of knowledge, research, and best practices to stem the spread of online information pollution, and joint efforts to obtain full transparency and accountability from social media companies.

[149] Countering Disinformation, "Legal and Regulatory Responses to Disinformation," accessed September 14, 2023, https://counteringdisinformation.org/topics/legal/2-measures-restrict-online-content-and-behaviors.

[150] Marcy Gordon, "House Approves Antitrust Bill Targeting Big Tech Dominance," September 30, 2022, https://apnews.com/article/2022-midterm-elections-technology-business-lobbying-congress-6e49cfc65668b99c633647898d114a8b.

Key Recommendation: Governments and election administrators should work to ensure a regulatory framework is fit for purpose through transparency requirements, where responsibility is outlined (duty of care), behavior that is illegal offline is illegal online, controls are issued on legal but harmful content, effective complaints systems are established, accountability through an independent regulator and the courts is ensured, and consequences for offenses for bad actors, platforms, and senior officials within the company are defined.

Policy Responses on Election Integrity and Political Advertising Regulations

As previously established, policy plays a key role in responding to risks to election integrity posed by information pollution. Political advertising plays a large role in influencing voters, and should be addressed in tandem to other issues of election integrity to ensure trust in democratic processes and to stifle the spread of election-related mis- and disinformation.

- Policies should address the spread of disinformation that contributes to voter suppression, including specific provisions penalizing allowance of content that misleads users on voting procedures, voting locations and times, and eligibility requirements.
- Policymakers can require platforms to partner with fact-checkers in their country to ensure that the identifying and flagging of mis- and disinformation is informed by local knowledge and that adequate resources are appropriated to non-English speaking contexts.
- Policymakers can require social media companies to have transparent strategies for counteracting election-related information pollution in their jurisdictions, or face penalties.

> In 2019, the Canadian government "called on social media platforms to do more to combat disinformation ahead of the election. The move comes in tandem with Bill C-76, legislation that aims to compel tech companies to be more transparent about their anti-disinformation and advertising policies."[151]

- Microtargeting should be regulated, preventing the ability of advertisers to direct their ads at segmented parts of the population determined by user data.[152] Microtargeting can strengthen impact on voters by exploiting facets of their identity

[151] Daniel Funke & Daniela Flamini, "A Guide to Anti-misinformation Actions Around the World," accessed September 14, 2023, https://www.poynter.org/ifcn/anti-misinformation-actions/.
[152] U.K. Information Commissioner's Office, "Microtargeting," accessed September 14, 2023, https://ico.org.uk/for-the-public/be-data-aware/social-media-privacy-settings/microtargeting/.

and further entrenching ideological positions through the careful crafting of an appeal meant to evoke a reaction from a specific audience.[153] According to U.S.-based non-profit MapLight, platforms like Facebook allow advertisers to "target platform users on the basis of personal information such as age, gender, education, income, multicultural affinity, ZIP code, or interests," which "can also enable foreign interference and voter suppression."[154] A review of the Facebook Papers in 2021 found Facebook failed to act on these risks despite the fact staffers had acknowledged microtargeting could be exploited by politicians "to spread misinformation and target vulnerable users."[155]

- Transparency in political advertising policies can be mandated, requiring social media companies to report on who is paying for advertisements, how they are being monetized, and targeted audiences.

A 2019 investigation by the Australian Broadcasting Corporation identified that "the Australian Electoral Commission notified Twitter and Facebook they must comply with notifications of illegal ads on their platform," threatening them with "court injunctions if they do not comply."[156]

- Regulations can address the allowance of political advertising, ensuring companies abide by their terms of service, especially if they state they do not allow political advertising but fail to prevent it.

In November 2018, the French parliament passed legislation aiming "to empower judges to order the immediate removal of 'fake news' during election campaigns."[157]

Social Media Platform Bans

Calls to ban certain digital platforms such as TikTok can be understood against the need to rein in the unregulated and ungoverned online space. Yet, discussions about platforms'

[153] Ciara Torres-Spelliscy, "A Lie Just for You in 2020," September 21, 2020, https://www.brennancenter.org/our-work/analysis-opinion/lie-just-you-2020.
[154] Ibid.
[155] Cristiano Lima, "Facebook Knew Ads, Microtargeting Could Be Exploited by Politicians. It Accepted the Risk," October 26, 2021, https://www.washingtonpost.com/politics/2021/10/26/facebook-knew-ads-microtargeting-could-be-exploited-by-politicians-it-accepted-risk/.
[156] Daniel Funke & Daniela Flamini, "A Guide to Anti-misinformation Actions Around the World," accessed September 14, 2023, https://www.poynter.org/ifcn/anti-misinformation-actions/.
[157] Michael-Ross Fiorentino, "France Passes Controversial 'Fake News' Law," November 22, 2018, https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law.

accountability should not be based on a binary choice — to ban or not to ban. Rather, the solutions must be measured and based on open and inclusive multi-stakeholder dialogues with relevant sections of society including academia, civil society, tech companies, government, and the media. The resultant actions to enhance platform accountability and safety among users must be informed by a human rights- and user-centric approach in line with the constitution and international human rights standards to ensure the protection of citizens' fundamental rights while addressing legitimate concerns surrounding social media use.

PART 6: RECOMMENDATIONS

Electoral justices, as key stakeholders in ensuring the integrity of elections, can take several steps to be aware of the role of misinformation and disinformation in elections. International IDEA purports that an election justice system is "a key instrument of the rule of law and the ultimate guarantee of compliance with the democratic principle of holding free, fair and genuine elections."[158] Endowed with the ability to make key decisions to promote election integrity, justices should be fully informed of the goals, manifestations, and impacts of election-related information pollution. The ability to set precedents on how mis-and disinformation is legally addressed, and to take decisive action that can have important implications for democracy, should not be underestimated. The decisions made by electoral justices are crucial to ensuring platforms and perpetrators are held to account for the grave risks to fundamental democratic principles. Wielding the power to impose penalties for facilitating the spread of mis- and disinformation means providing clarity on what will and will not be tolerated.

- **Voter education and awareness**
    - Electoral justices or those whose work pertains to election matters can take a definitive stand and educate the public on the dangers of information pollution through initiating campaigns to educate voters on the risks of election disinformation

> The French 2022 presidential election was first to involve "the Court of Audit (Cour des comptes) in the prevention of election disinformation." The court conducts "legislative and financial audits of public and private institutions, including the government itself," and the court's president announced in October 2021 that by the end of the year, they would publish

[158] *Electoral Justice: An Overview of the International IDEA Handbook* (International IDEA, 2010), https://www.idea.int/sites/default/files/publications/chapters/electoral-justice-handbook/electoral-justice-handbook-overview.pdf.

> "12 memos on the most likely major themes of the campaign (such as pensions, energy policy, industry, police, etc.)," "to offer a counterpoint to 'caricatures' and 'disinformation.'"[159]

- Electoral justices can publicly support election authorities in gaining legitimacy by declaring them the authoritative source on accurate information on voting procedures, eligibility requirements, and voting locations and times
- Electoral justices can support the implementation of fact-checking initiatives specifically focused on elections in their jurisdiction to quickly identify mis- and disinformation circulating on platforms in real-time throughout the election period

- **Judicial education and awareness**
  - Electoral justices should educate themselves on the scope of laws concerning mis- and disinformation in their jurisdictions to understand how they may enforce them within the context of elections, including those that may pertain to:[160]
    - Regulating social media platforms in spreading mis- and disinformation on social media platforms, which may encompass transparency mandates, removal of harmful content, or content labeling.
    - Criminal/financial/political penalties for spreaders of harmful mis- and disinformation that may disenfranchise voters, whether it be social media users, candidates, or members of an opposition party.

> In 2018, Brazil state legislator Fernando Francischini "broadcast a live video on Facebook, viewed six million times, promoting false narratives about the voting machines being rigged against Bolsonaro." He was impeached in October 2021 for "misusing the media and abusing his position of power" and banned from running for election in the eight years following his last election.[161]

[159] William T. Adler & Dhanaraj Thakur, *A Lie Can Travel: Election Disinformation in the United States, Brazil, and France* (Center for Democracy & Technology, 2021), https://cdt.org/wp-content/uploads/2021/12/2021-12-13-CDT-KAS-A-Lie-Can-Travel-Election-Disinformation-in-United-States-Brazil-France.pdf.

[160] IFES, "Video: Why Electoral Justice Matters," October 8, 2021, https://www.ifes.org/news/video-why-electoral-justice-matters.

[161] William T. Adler & Dhanaraj Thakur, *A Lie Can Travel: Election Disinformation in the United States, Brazil, and France* (Center for Democracy & Technology, 2021), https://cdt.org/wp-content/uploads/2021/12/2021-12-13-CDT-KAS-A-Lie-Can-Travel-Election-Disinformation-in-United-States-Brazil-France.pdf.

In the U.S., a social media influencer named Douglass Mackey "established an audience on Twitter with approximately 58,000 followers," and was found to have "conspired with others to use social media platforms, including Twitter, to disseminate fraudulent messages designed to encourage supporters of one of the presidential candidates (the "Candidate") to 'vote' via text message or social media, a legally invalid method of voting," ahead of the 2016, U.S. Presidential Election.[162] He was convicted of "the charge of Conspiracy Against Rights stemming from his scheme to deprive individuals of their constitutional right to vote," and could face a maximum 10 years in prison.[163]

- Defamation, such as that of individuals involved in electoral processes including candidates, party figures, election officials, and election workers.
- Campaign advertising, whether or not transparency is required on who pays for the ads and if it is possible to penalize false claims in ads.

In 2022, a ruling in Washington state found Facebook had "repeatedly violated campaign finance rules requiring platforms to release information about political advertisers on their sites," stating Meta "repeatedly broke the state's law requiring technology platforms make information about political ads available for public inspection in a 'timely manner.'"[164] The company was required to pay $24,660,000 in restitution, "the largest campaign finance penalty

[162] U.S. Department of Justice Office of Public Affairs, "Social Media Influencer Charged with Election Interference Stemming from Voter Disinformation Campaign," January 27, 2021, https://www.justice.gov/opa/pr/social-media-influencer-charged-election-interference-stemming-voter-disinformation-campaign; Misleading texts targeting voters have been identified in other contexts in 2023, such as in Poland, where voters receievd messages stating the ruling Law and Justice party were going to offer "provide funerals for pensioners for free." See: Vanessa Gera, "Polish Government Warns of Disinformation After Fake Messages Are Sent Out Before Election," October 12, 2023, https://apnews.com/article/poland-election-disinformation-430418b6b55c6ffc0a6a71e26bf9c51d.

[163] U.S. Department of Justice Office of Public Affairs, "Social Media Influencer Douglass Mackey Convicted of Election Interference in 2016 Presidential Race," March 31, 2023, https://www.justice.gov/usao-edny/pr/social-media-influencer-douglass-mackey-convicted-election-interference-2016.

[164] Naomi Nix, "Washington State Judge Rules Facebook Violated Campaign Finance Rules," September 2, 2022, https://www.washingtonpost.com/technology/2022/09/02/facebook-political-ads-details/.

anywhere in the country — ever."[165]

- Electoral justice systems should endeavor to set up mechanisms for and dedicate resources to monitoring disinformation on social media platforms to better inform approaches.

In 2022, the Electoral Tribunal of Panama (TE) and the Supreme Elections Tribunal (TSE) of Costa Rica teamed up via a "horizontal cooperation mission" to "explore possibilities to provide technical support to the TSE in preparation for the February 2022 national elections in Costa Rica in the area of monitoring political campaigns on social media, dissemination of fake news, disinformation and divisive and polarizing speech."[166]

- **International and citizen observers**
  - Electoral justices can advocate for engaging international observers in monitoring the fairness, transparency, and overall integrity of elections, and specifically the prevalence of mis- and disinformation and possible impacts on electoral processes within their jurisdiction.

- **Recourse for victims**
  - Electoral justice can ensure individuals involved in electoral processes who have been impacted by disinformation have clear paths for quickly seeking redress, as mitigating the impact of mis- and disinformation during election periods is often time-sensitive.
  - Electoral justices can ensure those who seek redress are aware of the protections they may have against possible retaliation.

- **Regulatory Frameworks for Reform**
  - It is important to note the role of an electoral justice in mitigating the harms of information pollution is entirely dependent on existing legislation in their jurisdiction and the level of independence and impartiality the justice system holds, which may pose limitations on how far justices can go on deciding appropriate penalties for perpetrators of mis- and disinformation or social

[165] Washington State Office of the Attorney General, "Judge Grants AG Ferguson's Request for Maximum $24.6m Penalty Against Facebook Parent Meta," October 26, 2022, https://www.atg.wa.gov/news/news-releases/judge-grants-ag-ferguson-s-request-maximum-246m-penalty-against-facebook-parent.

[166] Katherine Batista-Sánchez, "Electoral Tribunals of Panama and Costa Rica Work Together to Combat Fake News and Disinformation in Elections," September 22, 2021, https://www.idea.int/news-media/news/electoral-tribunals-panama-and-costa-rica-work-together-combat-fake-news-and.

media companies who facilitate this content.[167] Furthermore, the way election-related cases are handled by judicial systems and the mandates of courts vary. Advice provided must be considered contextually, as its applications may vary across jurisdictions. It is therefore crucial that adequate legal frameworks are put in place globally to mitigate the impacts posed by election-related information pollution.

- Electoral disinformation manifests on different products which are enforced by distinct teams within social media companies. "This points to a key concern with regard to the current industry responses to viral deception: while disinformation actors exploit the whole information ecosystem in campaigns that leverage different products and platforms, technology companies' responses are mostly siloed within individual platforms (if not siloed by individual products).
- In order to guide regulatory and industry remedies, a "Disinformation ABC" (Actors, Behaviors, Content) was developed by Graphika and Berkman Klein Center for Internet & Society at Harvard University.[168]
    - This "ABC" also seeks to reconcile approaches throughout applicable disciplines (e.g., cybersecurity, consumer protection, content moderation) and stakeholders.
    - While public debate and media coverage of election disinformation has been largely concerned with actors (i.e. the perpetrators), the technology industry has invested in better regulating behavior (targeting coordinated and inauthentic behavior) while governments have been most preoccupied with content (what is acceptable to post on social media).
    - This concise "ABC" framework doesn't aim to propose one definition or framework to rule them all, but rather seeks to lay out three key vectors characteristic of viral deception in order to guide regulatory and industry remedies. Manipulative actors, deceptive behaviors, harmful content: each vector presents different characteristics, difficulties, and implications. Unfortunately, they are also often intertwined in disinformation campaigns, suggesting that effective and long-term approaches will need to address these different vectors with appropriate remedies.

---

[167] *Electoral Justice: An Overview of the International IDEA Handbook* (International IDEA, 2010), https://www.idea.int/sites/default/files/publications/chapters/electoral-justice-handbook/electoral-justice-handbook-overview.pdf; Countering Disinformation, "Legal and Regulatory Responses to Disinformation," accessed September 14, 2023, https://counteringdisinformation.org/topics/legal/6-enforcement.

[168] *Actors, Behaviors, Content: A Disinformation ABC Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses* (Transatlantic Working Group, 2019), https://www.ivir.nl/publicaties/download/ABC_Framework_2019_Sept_2019.pdf.

UPCOMING ELECTIONS IN 2024[169]

Algeria	Presidency	TBD
Australia	Federal Election, Territory Elections, State Election	TBD
Austria	Legislative Election, state elections	TBD
Belarus	Parliamentary Election	TBD
Belgium	Federal Elections, Chamber of Representatives	June
Cambodia	Senate	TBD
Canada	State Elections	TBD
Chad	Presidency	TBD
Comoros	Presidency	TBD
Croatia	Parliamentary Election	July
Dominican Republic	General Election	May
Egypt	Presidency	February
El Salvador	Presidency, legislative assembly	February
European Union	European Parliament	June
Finland	Presidency	January
Georgia	Parliamentary Election	TBD
Germany	State Elections	TBD
Ghana	General Election	December
Iceland	Presidency	June
India	General election, state legislative assemblies,	TBD

[169] IFES ElectionGuide, "Elections," accessed September 14, 2023, https://www.electionguide.org/elections/type/custom/?country_id=&election_institution_type_id=&year=.

	urban local bodies	
Indonesia	Presidency, Regional Representative Council, House of Representatives	February
Ireland	Local Elections	TBD
Lithuania	Parliamentary Election	TBD
Mali	Presidency	February
Mauritania	Presidency	TBD
Mauritius	General Election	TBD
Mexico	General Election	June
Moldova	Presidency	November
North Macedonia	Parliamentary Election	TBD
Palau	General Election	TBD
Panama	Presidency, National Assembly	May
Portugal	Regional Election	January
Romania	Legislative Election, Local Elections, Presidency	November
Russia	Presidency	March
Rwanda	Presidency	TBD
Senegal	Presidency	February
Slovakia	Presidency	TBD
South Africa	General Election	TBD
South Korea	Legislative Election	April
South Sudan	General Election	TBD
Spain	Regional Elections	TBD
Sri Lanka	Presidency	TBD

Taiwan	Presidency, Legislative Election	January
Tunisia	Presidency	TBD
Turkey	Local Elections	TBD
United Kingdom	General Election, Local Elections, Northern Ireland Assembly Election	TBD
United States	Presidency, House, Senate, Gubernatorial Elections	November
Uruguay	General Election	October