Information Integrity to Sustain Peace during Electoral Processes

Sustaining Peace during Electoral Processes (SELECT) Project

Contents

Acknowledgement & Disclaimers

Key terms and concepts

The vocabulary around the information integrity topic will vary between countries and regions. For the purposes of this document, the following definitions are proposed.

Information ecosystem is the "complex adaptive systems that include information infrastructure, tools, media, producers, consumers, curators, and sharers. They are complex organizations of dynamic social relationships through which information moves and transforms in flows."[1]

Information integrity is "the accuracy, consistency, and reliability of the information content, processes, and systems to maintain a healthy information ecosystem."[2] It requires citizen access to trustworthy, balanced and complete information relevant to their political perceptions and decision-making[3].

Information pollution is "verifiably false, misleading and manipulated content on- and offline, which is created, produced and disseminated intentionally or unintentionally, and which has the potential to cause harm"[4]—namely the presence of disinformation, misinformation and malinformation.

Hate speech is "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law" (United Nations, 1966), including "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. This is often rooted in, and generates, intolerance and hatred, and in certain contexts can be demeaning and divisive."[5]

Digital repression is "the use of information and communications technology to surveil, coerce, or manipulate individuals or groups in order to deter specific activities or beliefs that challenge the state."[6]

Propaganda is "the deliberate, systematic attempt to shape perceptions, manipulate cognitions and direct behaviour to achieve a response that furthers the desired intent of the propagandist."[7]

Political entities is a term used to refer to candidates, politicians and political parties.

Electoral violence is a form of political violence, designed to influence an electoral outcome, and thus the distribution of political power. It may be considered a tool to (re)direct the trajectory of the elections by actors who are dissatisfied with the expected outcome or hold issue with the perceived credibility of the process, taking place in the pre-election, during the election day(s) or post-election periods. It may also be an expression of protests against a process that was seen as unfair—be it by incumbents or challengers. Electoral violence may take the form of physical violence or other forms of aggression, such as coercion or intimidation.[i] Accordingly, certain forms of online abuse and digital repression are included.

---

[1] 2015 Why Information Matters, a Foundation for Resilience, Internews
[2] 2019 What is Information Integrity?, Yonder AI
[3] 2018 Protecting Information Integrity: National and International Policy Options, Club de Madrid
[4] 2022 Strategic Guidance on Information Integrity, UNDP
[5] 2020 United Nations Strategy and Plan of Action on Hate Speech Detailed Guidance on Implementation for United Nations Field Presences
[6] 2021 The Rise of Digital Repression: How Technology is Reshaping Power, Politics, and Resistance, Carnegie Endowment for International Peace, Steven Feldstein
[7] 2018 Propaganda & Persuasion, 6th Edition, Garth S. Jowett, Victoria O'Donnell

This document intends to provide an introductory overview of the topic. It shall propose parameters for study, define key terms and outline a provisional framework. Its audience shall be the global working group and the participants of the regional consultative sessions who shall come from a variety of organizations and backgrounds. This document should not be considered a policy document, and its contents are intended to be interrogated throughout the SELECT project. The form of the output from the SELECT project topics shall be detailed elsewhere.

Summary

The precise role information pollution plays in undermining electoral credibility and driving electoral violence remains unclear. It is clear, however, that information pollution is a critical factor in electoral processes and that involved actors can mitigate its potentially destabilizing effects through the adoption of a number of measures.

Election-related violence and challenges specifically related to information pollution in electoral processes have long pre-dated the advent of the Internet. However, widespread Internet access and the emergence of social media platforms herald a number of fundamental changes to paradigms of communication, transforming various aspects of societies, not least of them, elections.

Initially, there were high hopes for what the Internet could do to strengthen elections and democratic political processes more broadly. However, this has given way to profound concerns, with some now warning that social media will undermine the ability to conduct credible elections, and that they increase the potential for election-related violence. Despite this, it is agreed that new opportunities and challenges have emerged for the various actors involved in the election, be they the Election Management Body (EMB), politicians, the media, civil society, the newly risen technology companies and above all, the voters.

The purpose of this report is to gain a better understanding of the pertinent dynamics and to bolster the design of programming to support the information ecosystem around elections. In aid of this, UNDP sought information through a number of channels, in a review of the relevant literature, a series of regional consultations, expert meetings and a survey.

While there is increasing agreement, at least from the major platforms, on the importance to adhering to international human rights law in how content is policed, the application of such can be complex and unsatisfactory. While incitement to hatred may have clearer guidance and agreement, the broader information pollution space lacks clear foundations. Indeed, election processes in particular carry an expectation that the broadest set of voices be permitted freedom of expression, within appropriate parameters. Ultimately, such guardrails may provide a floor around which to operate; alone—at least as they currently stand—they are unlikely to resolve the broader concerns at hand. Furthermore, international human rights law can provide a more rigorous framework for considering how different rights and protections can coexist as well as some guidance away from provisions that are intended to be politically instrumentalized—particularly where the justification is shakily based on different sets of human rights protections.

The body of existing empirical research is often conflicted about the impact of information pollution upon election credibility and potential for election violence. What research there is typically focuses on a small handful of platforms and predominately focuses upon a Western context. Broadly, there is agreement that information pollution can contribute to affective polarization, which in turn can influence electoral outcomes, inflame tensions and contribute towards the instigation of election-related violence. Successful attempts to do so rely upon and exploit social cleavages and tensions and are conducted as an electoral strategy—in line with typical electoral-violence concerns. Accordingly, and as indicated by the research, context is vital for gauging the potential impact of information pollution and associated factors. Furthermore, the broader information ecosystem, including traditional media, should be reflected upon in order to truly understand the dynamics and vulnerabilities in a specific context.

The various sources all conclude there remains no single panacea to the ills that information pollution brings upon elections. Rather, there is a variety of information pollution programming around elections,

each with its own benefits and deficiencies. In order to support the design of a holistic information integrity strategy, this report suggests that programmes seek to address one or more of the following three concerns (1) prevention—to address the supply side of information pollution by preventing or deterring the creation of information pollution, (2) resilience—building public resilience to information pollution limiting the ability of users to be influenced or co-opted by information pollution and (3) countering—identifying and attempting to counter information pollution.

As practitioners consider where to target their efforts, they should often be guided by pre-existing principles, updated for the Internet-age. First and foremost, this means attempting to address the underlying societal tensions, at least to the extent possible within an election period. Prioritizing promoting, rather than stifling, freedom of expression is more likely to engender trust. Transparency remains a key approach to the credibility of election processes, as does the quality of the election administration. Meanwhile, activities to combat the ill-effects of information pollution, by engaging with the general public, are increasingly recognized.

Political actors are the most prominent producers or amplifiers of election information pollution, and accordingly, are important to consider when devising measures to improve the information ecosystem around an election, such as political party and candidate codes of conduct or peace pledges. When considering those most in need of protection, it is women and marginalized communities who are the most likely targets of information pollution, and accordingly deserve appropriate attention. Of course, the intermediaries—the media and now, technology companies—have a vital role to play. Nevertheless, for a truly effective strategy, the broadest set of stakeholders should be addressed in a context-appropriate fashion.

Both research and programming face significant challenges from the rapidly changing technological landscape, shifting audiences and legislative decisions. However, these shifts may also create new opportunities, in particular around transparency.

A number of recommendations were identified on the basis of the research exercise:
1. No single solution has been found to the myriad of information integrity concerns, nor is it expected one shall be.
2. Instead, a range of activities are required, which should be complimentary and tailored to the specific country context.
3. Underlying the efforts must be a successfully conducted election deserving credibility, without which the information integrity activities should not be expected to provide unearned legitimacy.
4. Responses should take a multi-stakeholder approach. Various actors, including political parties, civil society organizations, technology firms and governments, have different mandates and roles to play. Yet, all the actors are unlikely to realize their individual goals without collective and coordinated action.
5. Successful approaches need to be embedded in the unique context. To support this, a thorough assessment of the information environment is a critical starting point.
6. When considering the options before them, practitioners should elevate those centred on the promotion of human rights, as opposed to those likely to be restrictive and that may inadvertently undermine the credibility of the election process.
7. In particular, when considering the regulatory approaches, care should be given to ensuring advice is appropriate to the relevant context and rights while remaining wary of attempting to transplant legislation from one context to another.
8. When assessing the threat social media poses to elections and voter behaviour, the potential for polarization is particularly urgent—the extent of which depends upon various country-context factors, including drivers of polarization, levels of distrust towards institutions and partisan traditional media, critically when considering interventions are approaches that tackle the underlying societal tensions.

9. Engaging with political actors is particularly resonant in the context of an election competition, where they are often the producers or instigators of information pollution, but also the targets. Supporting the negotiation and implementation of codes of conduct that consider online activities by the different electoral stakeholders can help to diffuse tensions and limit the instrumentalization of polarization within campaign strategies.

10. Fact-checking is an important, but insufficient activity. While vital for promoting accountability, its ability to effect corrections in the minds of voters is challenged.

11. Building public resilience among audiences is a means to attend to the demand side of the challenge. However, we should also be cognizant of the challenges here, both with regards to overcoming bias and reasonableness of affecting such skills in such a broad cohort.

12. Corrosive narratives are often seen to undermine confidence in the public institutions related to the election process. For public institutions to be able to combat these attacks, they—and the organizations that support them—require the appropriate technical skills, toolkits and financial resources. Support is necessary to allow the widest range of State actors to operate within this new domain. These competences stretch from detecting and responding to information pollution to protecting the integrity of its systems.

13. Public trust in media and journalism has come under increasing strain, while the importance of the profession is vital to combatting information pollution. Activities that foster ethics in the field as well as support actors to expand their investigative capabilities can improve the electoral information ecosystem.

14. Social media has been increasingly maligned, and while perhaps deservedly so, this should not distract from the strengths it can have as a tool for engaging various stakeholders and communities, coordinating action and advocating for peace. Actors involved in elections should seek to support the positive role that social media and messaging services can play in promoting inclusive credible elections and preventing electoral violence.

15. The modern iteration of the information integrity field is still young, and as a community, the body of evidence of the efficacy of various interventions is still being built. As part of any activity, rigorous evaluation practices should be implemented.

16. Information pollution, and the drivers of it, do not restrict themselves to the period of electoral operations. Furthermore, many activities in the area require prolonged periods to be effective. Accordingly, activities need to be considered beyond a single election process but over multi-electoral cycles.

17. Information pollution is not limited to the Internet, and in the course of programming, practitioners should look beyond to explore the various ways that it migrates through societies.

18. Platforms are powerful actors in these efforts; however, they may require pressure to act appropriately, and the newer platforms may not have adequate policies or technologies. Digital companies may also require support to ensure adequate, timely and effective action to protect electoral integrity and remove posts inciting violence.

19. The technological landscape is expected to evolve in various and often unexpected ways—creating new threats and opportunities. This will call for continuous investment in counter-technologies, tactics and research.

Chapter 1 Introduction

SELECT project background

The overall objective of the Sustaining Peace during Electoral Process (SELECT) project is to build the capacity of both national electoral stakeholders and international partners to: (a) identify risk factors that may affect elections; (b) design programmes and activities specifically aimed at preventing and reducing the risk of violence and (c) implement operations related to the electoral process in a conflict-sensitive manner.

Against this background, the SELECT project has developed an inclusive research process to ensure a multi-regional lens that takes into consideration experiences and knowledge from a wide range of stakeholders. The research process will be applied to various research topics included in the SELECT project document whereby the topics identified have the potential to negatively contribute to or positively mitigate the potential for electoral violence. The aim of this topic-specific research process is to understand the main challenges in relation to the nexus between the topic and electoral violence and outline actionable solutions to be implemented in the second phase of the project. Any solutions presented are intended to be informative and not prescriptive, recognizing that each country context will be unique.

Each topic will be accompanied by a working group comprised of experts in the field and representatives of relevant organizations. The members of the working group shall share their experience and expertise, as well as support, with their networks. Participants of the working group and their organizations will be acknowledged for their contributions to the topic.

The outputs of this project will not constitute United Nations policy recommendations.

This report is dedicated to the first SELECT project research topic, which shall explore the prevention of election violence and its linkages to information integrity.

Framing of the subject

The advent of widespread Internet access and the evolution of social media platforms have created new networks connecting billions of humans across the world, changing the way people seek, share and are served information. In turn, new paradigms of communication have emerged and old trends have accelerated—transforming aspects of societies and economies. The flow of information through this evolving landscape is directed by a complex set of motivations and incentives, mediated through technology, socio-political conditions and user behaviour.

The Internet has become an inextricable part of modern political life. The benefits are potentially transformative, creating new and more egalitarian opportunities for involved parties to communicate and coordinate. The emergent social media platforms have been immensely disruptive to the conduct of elections and the ways by which political entities contest them—for better or worse. An early excitement has receded, replaced in the minds of many with concerns over the digital domain's possible harmful impacts on politics, and democracy more broadly, with some wary that the risks may outweigh the benefits (United Nations, 2021).

The Internet provides a variety of transformative tools, which can be used for good or ill. It offers new means for citizens to reach each other, congregate, collaborate and evolve their common narratives. Social media arguably lowers the barrier to entry into political debate and opens the door to candidacy for those who had previously been excluded. Furthermore, the Internet provides an environment by which citizens can more easily access information about elections and politics.

Several deleterious effects of information pollution upon elections have been posited, including unduly influencing voters, defamation of opposition candidates, undermining the EMBs (Kimari, 2020), erosion of the credibility of the election process, impositions upon the right to participate in political affairs, gendered disinformation, online gender-based violence, heightened polarization and fomenting election-related violence. Certainly, these and related problems were present prior to the Internet. Yet, concerns are that new technologies are reinforcing these challenges and providing a toolkit for new impediments. The ultimate fear is the digital pathways have unique vulnerabilities that undermine the political norms, societal cohesion and voters' free choice, all essential for democratic elections.

The opening of the digital space has created new vectors for electoral violence. Norms of civility are weaker online, with harassment of women, youth, minorities and other marginalized communities particularly prevalent and fast-growing (UN WOMEN, 2021) (Izsák-Ndiaye, 2021). Citizens and politicians alike find themselves subject to such abuse though some research suggests that incivility is not inextricable from online life, with a more polite environment potentially cultivated through design, policies and incentives (Antoci, Delfino, Paglieri, Panebianco, & Sabatini, 2016). With regards to women, specific concerns exist that they may be disproportionately targeted, facing narratives that can be used to portray them as unfit for public office, as villains or dismissed in other ways. This targeting is further accentuated for women of minority groups (Sobieraj, 2020).

Elections are by their nature sovereign exercises, making interference by external actors in domestic political processes deleterious to the credibility of the processes. Without publicly accepted protections, a borderless digital ecosystem may create damaging concerns of new avenues for foreign interests to target individual citizens or influence the narrative of the election.

The framing of this study is guided by the features of election-related violence. The United Nations Policy Directive on Preventing and Mitigating Election-Related Violence describes electoral processes as the methods of managing and determining political competition with the outcomes deciding a multitude of critical issues, leading to a highly competitive environment where underlying societal tensions and grievances may be exacerbated and ultimately may lead to electoral violence—a form of political violence that may be physical or take other forms of aggression, including coercion and intimidation (United Nations Department of Political Affairs, 2016). This can happen spontaneously or be planned by political actors and their supporters, the latter who may be paid to commit attacks against candidates of political opponents or to create violent scenarios that ultimately favour the sponsoring party. It may be staged to look like random attacks, organized to blackmail political candidates, involve kidnapping and be accompanied by acts of coercion, intimidation and threats (Opongo & Murithi, 2022). Findings indicate that election-related violence is typically conducted between competing parties (Ginty & John, 2022). While the focus of the research is on election violence—violence that is related to the holding of an election delineated by actors, timing and motives (Birch, Daxecker, & Höglund, 2020)—we remain cognizant that the tensions and disputes that may emerge from an electoral process may fuel subsequent political violence and diminish the legitimacy of governments.

This paper attempts to reflect the disputes around the ways and degree by which the concerns raised above actually emanate from the online space. Certainly, anxieties around election-related violence and propaganda pre-date the Internet—in one form or another. Laying the blame for the social ills seen over the recent past solely at the feet of social media could prevent the identification of other factors driving division and political unrest (Bruns, 2019). However, the digital age heralds new tools and novel dynamics—such as virality, velocity, anonymity, homophily, automation and transnational reach (Persily, 2019)—which indicate that it is not just more of the same. Certainly, the lived experiences of many of those that have been spoken to for the purpose of this research make it clear that there are indeed urgent and critical issues to address.

At the same time, there are great challenges in resolving these concerns, especially regarding elections. A fundamental question is, if elections require a free and pluralistic public debate, then on what basis are platforms, or even governments, censoring discourse? In the electoral context, perhaps the most straightforward type of information to combat is the immutable facts that define the process, for example, the dates within the electoral calendar of polling or eligibility requirements, as would emanate from the Election Management Body (EMB) or legislation. While confusion around such facts must be combatted, there is a myriad of other, more complex information pollution concerns to address.

There is an understandable differentiation between hate speech and other forms of information pollution. Hate speech is widely recognized within international human rights law as being prohibited, and established tests exist to define such content—notably the Rabat Plan of Action threshold (United Nations High Commissioner for Human Rights, 2013). In practice, there remain complex categorization and automation issues (MacAvaney, et al., 2019) (Van Zuylen-Wood, 2019). However, while these concerns are extremely serious, they are in some ways less contentious. In principle, disinformation calls for a different approach.

There is no common definition, understanding or approach to 'disinformation' within international human rights law. There are several tensions involved in tracking disinformation—or more broadly, information integrity. Rights to freedom of expression; the freedom to hold, form and change opinions; freedom of information; and the freedom to participate in public affairs, while in contention, are various protections that contribute to the legal guidance. Freedom of expression is not limited to truths; in fact, it especially does not preclude offensive and disturbing ideas and information, irrespective of the truth or falsehood of the content (Report of the Secretary-General, 2022). Despite these protections, reasonable exceptions in particular circumstances are provided for (United Nations, 1966). Where restrictions are in place, they are expected to meet a high threshold of legality, legitimacy, necessity and proportionality (Khan, 2021). The above vital protections simultaneously signal the limitations of content moderation as a solution to information pollution and demonstrate a need to look more broadly for solutions that are based in promoting international human rights law protections.

This paper also recognizes that the aforementioned United Nations Policy Directive on Preventing and Mitigating Election-Related Violence does not attempt to directly address information integrity issues. While this paper seeks to provide material to support programmatic activities, it does not form United Nations policy.

UN Electoral Assistance Framework

UN electoral assistance is provided to Member States at their request or based on mandates from the UN Security Council or General Assembly only. The UN system-wide focal point for electoral assistance matters, the Under-Secretary-General for Political and Peacebuilding Affairs, decides on the parameters of such assistance, based on needs assessments led by the Electoral Assistance Division of the Department of Political and Peacebuilding Affairs (DPPA). Implementation is guided by UN electoral policies set by the Focal Point, in consultation with UN entities, including UNDP.

Chapter 2 Framework of analysis

Process

There is limited rigorous research that focuses purely on the relationship between election-related violence and information integrity centered on the online experience. This is especially true outside of the Western experience. A broader set of literature exists enquiring about how information integrity influences election processes and public behaviour, though they rarely provide uncontested conclusions. Considering some key dimensions of the election process, the overview below seeks to illustrate the current debates and the research underpinning different positions. Broadly, however, the current body of research describes digital media as posing benefits and risks to democracy. On the one hand, there is evidence that it can contribute to voter participation and mobilization. However, there are more mixed findings when it comes to political knowledge and concerns when it comes to trust.

The relationship between information pollution and elections is complex and at times counterintuitive. Some logical suppositions or popular narratives become confounded or diluted by evidence and empirical research. At the same time, the dynamic nature of the environment and data constraints contribute to difficulties in developing replicable studies, limiting the scope to test findings in new contexts.

It is important to keep in mind that investigation into this field of study continues at pace, and its findings are likely to evolve. It is also necessary to recognize the limitations of the current research, which is often centered on Western experiences and around the handful of companies willing to provide adequate datasets (Kubin & Sikorski, 2021).

Key research questions

The study will explore the following questions:

i.   What information integrity factors influence elections and the potential for election-related violence, and how does this vary based on the context or different conditions?

ii.   What responses have been devised to promote and maintain information integrity around elections, and what are key lessons learned?

iii.   How can these responses be applied elsewhere and what issues should practitioners consider in their planning, as appropriate to their specific context?

It is beyond the scope of this exercise to provide definitive answers. Rather, it aims to explore the questions to inform future programming. The report shall consider the existing research and marry this with insights from election practitioners across the world.

There are a variety of ways to explore the 'problem' at hand. One is to consider how information pollution consumption influences voters in the context of an election process. Another is to explore the production and supply side of the election information pollution. The role of platforms as channels and actors in this relationship must also be understood. Finally, in the context of an election process and the triggers of election-related violence, the question as to how trust in institutions, the EMB and the election process itself can be eroded or bolstered bears relevance.

Ultimately, this report posits that strong information integrity is vital to ensuring credible and peaceful elections. Its exploration of the subject is towards an understanding of what the risks are, how an information integrity environment may be achieved, and how this varies specific to the context.

Chapter 3 State of the Research

Impact on voter attitudes

The foundational phenomena being explored is the influence of information pollution on attitudes of voters. Various institutions are exploring this and related questions, with interesting and valuable conclusions.

Power to polarize

Research has raised doubts about the ability of information pollution to convert voters to political positions that are markedly different than their own. However, there is clearer consensus that online information pollution can contribute to polarization within electorates and populations, thus raising the prospect of election-related violence (Barrett, Hendrix, & Sims, 2021). Specifically, literature is concerned with affective polarization and the tendency for partisans to dislike and distrust out-groups.

It has been found that exposure to false information deepens partisan beliefs (Guess A. M., 2020), with disinformation campaigns that adhere to a widespread belief or hold some basis as reality being particularly successful (Moore, 2018). Conversely, it has been found that simply by being off social media, individual polarization can decrease (Allcott, Braghieri, Eichmeyer, & Gentzkow, 2020). Overall, it is clear information pollution can be a polarizing force in the minds of voters and that its potential impact may be tied to the pre-existing levels of polarization in the country. How Global South countries present has not been well researched; however, certain metrics may support assessments of susceptibility to information pollution, such as polarization indexes, media independence and trust in institutions.

A popular explanation for online polarization is 'filter bubbles' caused by platform algorithms (Pariser, 2011) that are described as creating and amplifying echo chambers where people hear only similar views, which in turn validate opinions and drive polarization. Recent evidence challenges the strength of this theory (Guess,

Lyhan, Lyons, & Reifler, 2018). This refined position holds that, while accepting that the 'echo chamber' theory is true for a few who conduct selective exposure, on average, users of social media are expected to experience more diversity than non-users (Newman, 2017). Thus, rather than being forced into echo chambers, this position views people who experience such as a self-selecting small minority of highly partisan individuals. Research on a handful of established Western democracies found only around 5 percent of Internet users exist in echo chambers, with the exception of the United States where rates of 10 percent or more exist (Fletcher, Robertson, & Neilsen). A reasonable concern is countries with high polarization will present akin, or worse, than the United States. Furthermore, communities with smaller linguistic pools may also present differently.

The outcome of personalization and recommender algorithms does not always lend itself to the echo chamber concern. One study conducted in the US demonstrated algorithmic personalization for people interested in false election narratives resulted in them being presented content more often challenging these false narratives (Bisbee, et al., 2022). Similarly surprising are findings that the forms of algorithmic selection offered by some search engines, social media and other digital platforms lead to slightly more diverse news exposure (Arguedas, Robertson, Fletcher, & Nielson, 2022). In particular, there is concern on what drives radical political content consumption on video-based platforms, with many pointing to the recommendation engine, while others contending it is a combination of user preferences, platform features and the supply-demand dynamics (Hosseinmardi, Ghasemian, Clauset, & Watts, 2021). However, what has been seen in the past may not hold, as more efficient algorithmic targeting develops, nor is there clarity on how these algorithms perform in different contexts.

Ironically, it has also been found that exposure to messaging by opposing political ideologies can entrench political views (Talamanca & Arfini,

2022) (Bail, et al., 2018). This implies that the greater diversity of political news exposure may exacerbate polarization, not diminish it. This may be reflected in findings that social media posts that reflect animosity towards opposing political views are significantly more likely to attract engagement (Rathje, Van Bavel, & Linden, 2021), which may incentivize such language and enhance the virality of such polarizing posts.

Ultimately the overriding concern is that the online space gives falsehood the advantage over the truth. One often-cited study used Twitter data to infer that information pollution diffuses "farther, faster, deeper, and more broadly" than non-polluted information, especially prevalent in polluted political information, and that it was humans, not bots, that were responsible for this spread (Vosoughi, Roy, & Arel, 2018). While the extent of this research is contested on the basis of more recent analysis (Juul & Ugander, 2021), as newer platforms with more successful engagement techniques arrive, these trends may deepen.

Another cited phenomena of political engagement on social media is the incivility of online discourse, which it turn drives mistrust on- and offline. This falls significantly more upon women, reporting greater experience of these risks than men (Microsoft, 2022). In the political context, the highly evidenced harassment of female political figures online is of particular concern, imposing a barrier to their participation in the electoral process. Such acts are not only harmful in themselves, but also have negative consequences upon the broader electoral process. This influences the broader participation of voters and would-be candidates who find this unappealing, given that exposure to opposing views may reinforce existing views, and that discontent increases partisan acceptance of misinformation (Weeks, 2015), contributing to polarization.

In some ways, irrespective of the actual impact, the mere concerns by some that others may be influenced by disinformation has the potential to undermine confidence in democracy (Nesbit, Mortenson, & Li, 2021).

News exposure

The degree to which people are exposed to information pollution may also influence its impact. In reality people typically consume little political news, with assessments finding that only a small amount of total online media consumption is spent on acquiring the news, and even a smaller fraction of this time is spent on fake news. For example, one study finds that for Americans, fake news comprises only 0.15 percent of the daily media diet, and news overall represents at most 14.2 percent of their daily media diet (Allen, Howland, Mobius, Rothschild, & Watts, 2020). In the months leading up to the 2016 US election, the average user is estimated to have seen between one and seven false stories online (Allcott & Gentzkow, 2017). Research into partisan WhatsApp groups in India found that little content was being shared that was hateful or misinformed (Chauchard & Garimella, 2022). However, not all people behave as the average, with a proportionately small group of people sharing a disproportionately large amount of extreme content (Mellon & Prosser, 2017).

In light of the limited exposure to online information pollution, consideration should be given to the powerful role traditional media outlets still hold in most demographics—particularly within the Global South. Assessments find that for many, it continues to outweigh social media as a source for news. While much attention is paid to the role of social media in spreading information pollution, it is only partially responsible for the spread, with traditional media also as a significant disseminator and shaper of the national information environment (Humprecht, 2018). This spread of information pollution by traditional media is in part driven by the newsworthiness of stories about false news and the need to repeat false news to address it (Boomgaarden, et al., 2020).

Irrespective of the actual impact, citizens often have a concern that others within their society are being unduly influenced by disinformation, and this belief alone has the potential to undermine confidence in democracy (Nesbit, Mortenson, & Li, 2021).

Polarization to election-related violence

Election-related violence is typically a strategic decision to aid electoral success. Broadly speaking, the opportunities to manipulate the information landscape revolve around polarizing election competition along pre-existing social cleavages. Considering that 'systemic, longstanding and unresolved grievances' are among the factors that lead a country to be particularly susceptible to election-related violence (United Nations Department of Political Affairs, 2016), it follows that information pollution can be used as a device to help steer voters towards election-related violence through an avenue of political polarization.

Furthermore, there are findings that the electoral process itself does tend toward intensifying underlying tensions into high intensity situations where the quality of disinformation and propaganda becomes immediately inflammatory, increasing the likelihood that long-term discrimination will turn into physical violence (Banaj & Bhat, 2018).

More broadly, there has been work to understand how speech can be a driver of inter-group violence, including within an electoral context. While some researchers have posited a causal relationship, there is limited evidence, in part due to it requiring time to influence subjects and the difficulty of controlling for other influences (Buerger, 2021). While much study has been conducted looking at the role of television, radio or SMS, rather than the Internet, the findings suggest a relationship between the use of these channels to air messages that incite (United Nations Human Rights Council, 2011) (Deane & Ismail, 2008) or harmful rumors (Osborn, 2008) and the occurrence of election violence. Debate exists; however some research finds that the removal of hate speech online has benefits to offline violence (Durán, Müller, & Schwarz, 2022).

The threshold for mobilizing people to commit election violence is typically high. While some of the research raised above posits limited potency for using information pollution to influence the broad population, its ability to deepen polarization in certain segments of society can help to increase the propensity for election-related violence. Even the mobilization of relatively small groups of adherents is sufficient to instigate an outsized disruption of the election process.

Partisan violence lends itself to increasing approval by supporters, while driving away non-supporters, further exacerbating partisan polarization (Daxecker & Prasad, Voting for Violence: Examining Support for Partisan Violence in India, 2022).

Contextual factors

The effects of information pollution vary between different demographics and contexts. Ultimately, for an accurate understanding of the influencing factors in a specific country, a case-specific landscape assessment is probably required.

History, society, culture and politics all play a part in understanding disinformation in a particular context, with an analysis of how social differentiation, such as race, gender and class, shape dynamics of disinformation. Institutional power and economic, social, cultural and technological structures further shape disinformation dynamics (Kuo & Marwick, 2021).

Since digital communication practices both exist within a particular socio-political context and shape that socio-political context, the relationship of violence is situated between the technological and social. Where the broader environment contains animosity towards particular out-groups, in-group users are predisposed to believe and share information pollution about out-groups (Banaj & Bhat, 2018).

Some countries appear more resilient to online disinformation than others. There are a variety of indicators that signal this resilience. These may include: the level of existing polarization, the types of political communication, trust in the news media, the presence of public service media outlets, fragmentation of audiences, the size of the advertising market and the level of social media usage. In practice, it is assumed that these criteria are more varied and context specific and will require an assessment to uncover.

Age has been demonstrated to be a significant factor in the sharing or acceptance of information pollution, as has political orientation. With age, older adults present this behaviour despite their awareness of misinformation and cynicism about news (Munyaka, Hargittai, & Redmiles, 2022). Gender, education and other identity features are considered to be less relevant, at least in some contexts (Rampersad & Althiyabi, 2019).

The underlying nature of relationships in a community may also influence behaviour, with trust in content and onward sharing potentially guided by ideological, family and communal ties. Behaviour is further influenced, for example, by the propensity to share and believe content guided by trust in the source (Banaj & Bhat, 2018).

Key actors

Election Management Body

Public trust in the conduct of elections is the cornerstone to the peaceful acceptance of the outcome, and a key guardian of this trust is the EMB (Elklit & Reynolds, 2002). The confidence held by election stakeholders in the EMB as an effective and impartial entity can underpin much of the resilience a process has to mitigate against the emergence of disputes and election violence. Many of the well-trod principles of effective electoral management continue to be relevant, such as impartiality in action, professionalism and, perhaps most applicable here, transparency (Wall, 2006) (Kerr & Lührmann, 2017).

Building the effectiveness of the EMBs as credible and capable institutions has long been a core tenant of electoral assistance and the prevention of election-related violence. However, the impact of new tools to undermine the integrity of electoral processes raises questions about the challenges EMBs may face in the future and how they should best respond. Similarly, there are clearly limits to what an EMB can achieve in an area that is outside of its

traditional remit and which often has roots extending beyond the election process.

The formal responsibilities for the EMB will also depend upon their legal remit. Tasks such as monitoring the campaign and enforcing rules around advertising content and spending are complicated by the online realm. However, in order to defend the credibility of the institution and election, commissions may choose to proactively take action to improve the quality of the information environment.

There are increasingly troubling reports of election administrators being the targets of information pollution and harassment in the online and physical space (The Bridging Divides Initiative, 2022). There are concerns that increasingly public personal information and online data privacy issues provide more opportunities for attacks against public officials and techniques such as doxing (Zakrzewski, Election workers brace for a torrent of threats: 'I KNOW WHERE YOU SLEEP', 2022)(Zakrzewski, 2022).

Operating in the online domain, however, introduces a number of operational and financial challenges that are hard to overcome. Professional firms who work on social media analysis and attributing influence operations can be eye-wateringly expensive. The tools that exist are relatively immature and require a set of technical and analytical capabilities, which EMBs are not well-versed in. The platforms themselves often impose barriers to what State authorities can access by way of information. Taken together, the aforementioned obstacles call for investments in various capabilities and new approaches by international assistance providers.

Political entities

While information pollution concerns are often viewed through the platform or citizen lens, some argue that this is inadequate, especially outside of Western contexts (Abhishek, 2021). For election-related information integrity concerns, a supply-side approach in which political actors are the key producers is a valuable prism to examine the impact of information pollution (Daxecker & Prasad,

Poisoning Your Own Well - Misinformation and Voter Polarization in India, 2022).

Political entities and election-related violence

There is a wealth of analysis suggesting that political actors play a central role in the incitement of election-related violence. Incidents are typically conducted between parties, with incumbents the main perpetrators (Ginty & John, 2022).

Ultimately, for incumbent governments, the decision to foment election-related violence is driven by the fear of losing authority (Hafner-Burton, Hyde, & Jablonski, 2013). The act has varying purposes. In the pre-election period, it can be used to change the electoral competition in their favour, for example by depressing turnout or mobilizing supporters. In the post-election phase, it may be used against public demonstrations or to punish winners (Bekoe & Burchard, 2017).

Some scholars argue electoral violence should not only be viewed as a strategy wielded by incumbents or the opposition: individual motivations may differ from the political groups' and leadership's goals, leading electoral violence to be fuelled by individual revenge dynamics and grievances or local power competitions (Hafner-Burton, Hyde, & Jablonski, 2014). Additionally, countries experiencing armed conflict often see armed groups as perpetrators of electoral violence to achieve their own objectives, intertwining it with other forms of political violence (Daxecker & Jung, 2018).

Political entities and information pollution

Election periods are typically rife with political information pollution, often conducted or inspired by the political contestants—in particular opposition groups or embattled incumbents. The influence a political entity has over their supporters translates to their ability to convince them of the supposed veracity of information pollution narratives (Siddiqui, 2018).

The instrumentalization of information has always been part of electoral campaigns, being used strategically to further prospects—with potential advantage (Kurvers, et al., 2021). However, researchers believe political parties and governments are escalating their capacity to use social media for information pollution. Numerous cases have been identified where they are outsourcing activities to the private sector. Bots or enlisted influencers are being used to bolster efforts. There are various election-specific examples of candidates or parties using social media to voice their disinformation or instances of fake accounts to artificially amplify their messages. (Bradshaw, Bailey, & Howard, 2020)

The use of social media manipulation strategies has increased during presidential elections in many countries. This is propelled by the contracting of global data mining players to collect and analyse data on voters and electoral patterns and then use it to target advertising and messaging to influence decisions. Targeted disinformation campaigns in many countries have also aimed to sway the electoral outcome, undermining credibility and confidence in electoral institutions, and fueling social tensions and violence during elections (Mutahi, 2022).

A particular challenge represents the fact that many of the actors at risk of inciting election-related violence are often responsible for setting the rules around the electoral process, campaign and, to some extent, information pollution. Thus, any regulatory process should be inclusive and transparent to relieve concerns of conflicts of interest, grounded in human rights protections. Related, legislative approaches should look to bolster confidence by insulating regulators from political interference, as well as providing effective routes for appeals and redress (Report of the Secretary-General, 2022).

Challenges in moderating political entities

There is a difficult balance to be struck when considering the appropriateness of certain rhetoric by political actors around the election. While there is an understandable desire for politicians to be limited to sharing factual information, in practice this is a complex criteria

to enforce. Furthermore, freedom of expression and human rights protections to impart information and ideas are not limited to 'correct' statements, as the right also protects information and ideas that may shock, offend and disturb. Prohibitions on disinformation may therefore border violation of international human rights standards, while, at the same time, this does not justify the dissemination of knowingly or recklessly false statements by official or State actors (UN, OSCE, OAS, & ACHPR, 2017). Broadly, there is an expectation that rights that exist 'in real life', should persist online.

Certainly, a different standard should apply to speech that infringes upon human rights or qualifies as hate speech or incitement to violence. Experience indicates that voter propensity towards election violence is low, requiring political elites to invest significantly to mobilize supporters to engage in violence. Among the most effective types of messages that can lead to election-related violence are those which stoke fear in their supporters, often targeting minority groups. The intersection between election-related violence and hate speech is particularly concerning (Siddiqui, 2018).

Key platforms specifically limit content moderation of materials posted by politicians. Given that in most cases election-related violence is incited by politicians, this weakens another line of defence—though perhaps justifiably so. Any restriction on political discourse should respect that election processes require freedom of expression and a plurality of voices. The decisions about which messages constitute harm can become complex in a process that is, at its heart, a contest between political rivals and ideologies seeking to win the support of the population. Unlike in other contexts, such as during the COVID-19 pandemic, there are rarely authoritative official institutions to set out the relevant facts.

Traditional media

The mass media plays a vital role in shaping the degree of trust enjoyed by an electoral process. However, media institutions are currently experiencing a decline in their own public trust,

with few countries reporting more than 50 percent of people trusting most of the news most of the time. Furthermore, increasing proportions of news consumers say that they actively avoid the news. How citizens consume news varies from country to country, though online news outlets are increasingly overtaking traditional media. Younger users are migrating from websites to get their news from apps. (Newman, Fletcher, Robertson, Eddy, & Nielsen, 2022). Research indicates the expected correlation between exposure to 'fake news' and lower trust in media institutions. It also reveals that greater exposure to information pollution may lead to greater trust in political institutions depending upon voter alignment with political entities in power and the specific media environment (Ognyanova, 2020).

The decline in trust in media has various causes, including the perception that media hold political or elite biases or that media outlets are subject to interference by politicians and businessmen. There is an overarching need for independent, transparent and open press reporting on electoral processes to permit better scrutiny and accountability of elections and their results. Accordingly, it is prudent to consider the media environment in the country to understand the tools at hand.

Independent public service broadcasters remain well trusted in those countries where they have been appropriately established. It has been contended that the structure of public broadcasting, as opposed to commercial broadcasting, poses some limits on its ability to counter disinformation but also provides protections from attacks, for example fiscal and structural resilience (Bennett & Livingston, 2020).

Journalists operate in a difficult and dangerous space. There are various reports of them being the target of harassment and even election-related violence, just as the practice of journalism is being degraded. However, the role of independent journalism has possibly never been more vital to the conduct of an election, as a means for providing scrutiny, transparency and public education (America, 2021).

Not all mainstream media adhere to journalistic standards. While media are rarely the instigators of violence, they can produce a political environment that is conducive for polarization and violence. In the context of armed conflicts, media frequently become polarized, acting as propagandists for conflict actors (Hoglund, 2008). Accordingly, it is prudent to consider the media environment in the country to understand the tools at hand and the need for media reform.

Foreign actors

While foreign intervention in elections has taken various forms over the past decades, the evolution of the information environment has created unprecedented vectors for them to seek influence. There is broad agreement that such intrusions should not be tolerated. However, indications are that condemnation and impacts fall along partisan lines, with those who stand to lose from the action becoming more outraged and losing faith in the democratic process (Tomz & Weeks, 2020). However, here also there is debate over the level of influence that these actors can meaningfully wield, at least independent of domestic political cooperation.

There are scenarios where the foreign influence is channelled through State broadcasters, in which case the attribution is relatively simple, and mechanisms to better inform the public of the risk have some promise (Nassetta & Gross, 2020). Various actors have been working to better support the identification of such operations and to devise response strategies.

Agents

Threat actors have various means of deploying their information pollution. Of course, political entities typically have sufficient standing within their communities to personally instigate harmful messages—and in some cases it is part of their political platform—which their supporters will organically disseminate. However, in other cases they will enlist others to initiate and or amplify content.

The presence of bots has been a concern, with the power to amplify opinions—including

information pollution—by increasing engagement of posts, as seen in past elections (Bradshaw & Howard, Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation, 2017) ((Wardle & Derakhshan, 2017), which has been seen to influence the agenda of media outlets (Vargo, Guo, & Amazeen, 2018).

Alternatively, humans can be engaged to disseminate and amplify partisan messages and information pollution narratives in a coordinated fashion, be it through the engagement of 'influencers' or the use of networks of humans (troll farms) directing account activity. A trend has been observed where the use of bots is giving way to human agents aided by technology (Bradshaw, Bailey, & Howard, 2020).

As discussed before, private strategic communications firms, enlisted to spread propaganda on behalf of candidates and political parties, are growing and deploying many of the tools described above.

Platforms

Platforms are both the conduit for spreading online information pollution, as well as a potential actor in remedying such. A confluence of factors has led to finger-pointing at platforms for, at best, not being sufficiently motivated to manage harms, and at worst, putting profit over the well-being of their users (Haugen, 2021). Some contend that in order to boost engagement, personalization algorithms are designed to promote controversial content, regardless of the veracity, and the platforms are not motivated to appropriately moderate disinformation. Platforms themselves have on occasion contested this narrative, with some arguing that they delivered personalization without promoting sensational content since a shortsighted quest for 'clicks' would undermine their longer-term profitability and reputation (Clegg, 2021).

Platforms vary in their capability to address disinformation (Allcott, Gentzkow, & Yu, Trends in the diffusion of misinformation on social media, 2019) each with different policies, resourcing and technical features. While some platforms are more associated with information

pollution than others, this may simply be a function of greater market share and relatively more transparency in their operations.

While most of the well-known platforms have signed up, at least, nominally, to the goals of protecting human rights and imposing a responsible content moderation policy, other platforms are emerging that do not share similar concerns, and research into them is yet to develop, for example, Rumble or Truth Social.

Content moderation and safety tools

Platforms are defined, in part, by their content moderation approach, as well as their features, design, commercialization decisions and algorithms. Content moderation policies and decisions help shape the culture and experience of the platform and compliance with the law, underpinning their safety and attractiveness to users. Despite this, platforms are frequently cited as being opaque about their content moderation policies.

Correctly moderating content is challenging, especially with the massive scale and velocity of information around an election overwhelming human moderation capacity. Unfortunately, while platforms have long applied artificial intelligence to conduct content moderation, there are serious limitations to such approaches, such as: inaccurate decisions, bias towards particular populations and potentially censoring political ideas (Gorwa, Binns, & Katzenbach, 2020). Ultimately, content moderation is unlikely to remove all offending material while maintaining all appropriate content. Instead success may be more realistically defined by the degree to which it is correct (Douek, Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability, 2021).

The effectiveness of content moderation will differ by medium of content, complicated by the move from text to audio and video. Encryption and ephemeral content pose serious hurdles. While particular content moderation measures exist to challenge end-to-end encryption, they are reserved for certain illegal content, such as child sexual abuse material or violent extremist materials.

There are concerns about how platforms vary their service and site safety features between countries. Companies will prioritize their attention and resources providing some countries a better class of service than others (Zakrzewski, De Vynck, Masih, & Mahtani, 2021). For example, Facebook only opens Elections Operations Centers to address information integrity in some countries (Elliott, 2021).

A common difficulty platforms face is moderating content in various languages, with their local contexts, nuances and dialects. Automated measures have limits, while human moderation is also challenging, especially for less used languages (Facebook Oversight Board, 2021) (Stecklow, 2018) (Fatafta, 2021). For many, this demonstrates an underinvestment of resources by platforms and the limitations of artificial intelligence.

Policies and legislation

Platforms and other electoral information businesses should remain committed to ensuring that their actions respect human rights (Human Rights Council, 2011). While the platforms have increasingly agreed to apply international human rights law to their content moderation policies, it may be insufficient in tackling the thorniest questions (Douek, The Limits of International Law in Content Moderation, 2021). For example, there are questions regarding whether international human rights law provides any basis for censure of information pollution systematically conducted by foreign parties (Ohlin, 2021). Also, its restrictions around coordinated behaviour—foreign or domestic— also are considered absent. At the same time, the attempts to align with international human rights law can shield platforms, to some degree at least, against States who attempt to impose repressive content policy.

Despite typically having special policy frameworks for electoral content, platforms face dilemmas as they seek to achieve censure of unacceptable behaviour, support to credible elections and protection of freedom of expression. Platforms and election practitioners are unlikely to always agree on where this balance sits and the actions taken.

Governments are increasingly pursuing forms of platform regulation. These include requirements to prevent or remove otherwise illegal content, including hate speech. Some regulation imposes a vaguer criterion covering legal but still harmful content. By and large, platforms already restrict some legal yet undesirable content, arguably to create a hospitable online environment, support monetization and adhere to the spirit of the company. However, there is the inevitable tension between the self-regulation a corporation seeks to apply and what is culturally and legally acceptable in a sovereign country. It is unsurprising that governments have sought to assert greater control over the social media platforms; there are concerns over how these controls may be instrumentalized by illiberal or authoritarian governments.

Much of the legislation planned or in-place contains massive potential penalties for non-compliance. Some fear a chilling effect as platforms steer towards over-censorship to be sure to remain within the law.
Policy makers and legislators have various ways to approach the regulation of content moderation. Some approaches focus on the handling of individual content decisions and determining what is ineligible. Content-specific regulation faces the various challenges described above, including the overwhelming volume of content to parse, the complexity of designing moderation rules and the challenge of making consistent and correct content moderation decisions. Regulating for more process around moderation has been increasingly welcomed, for example, granting people the right to appeal content moderation decisions and notifications of actions taken to their content. However, a new line of thinking considers how to regulate activity to address the broader systems in play and direct effort upstream of individual cases, for example, producing annual content moderation plans and compliance reports or separating internal functions to limit problematic incentives (Douek, 2022).

Election campaign advertising has shifted substantially from traditional to digital channels over the last decade, raising concerns about transparency and the effectiveness of campaign financing regulations. In some ways, political advertising becomes simultaneously more individualized, tailored and opaque. The approach increased the potential for polluted information being diffused among voters on a large scale, without oversight or interventions against politicians' claims (Council of Europe, 2017). However, while some platforms seek to enhance transparency by publicizing archives of ads, the degree of access and detail in these vary substantially. Simultaneously, where some major platforms have banned political advertising, this may prompt a shift to alternative platforms with less regulation and transparency (Brennen & Perault, 2021).

Influence operations in pre-Internet traditional media can build support for political parties and lead to increased political violence in conflict environments or where there is civil unrest (Bateman, Hickok, Courhesne, Thange, & Shapiro, 2022). If the information ecosystem is to support healthy elections, all elements of media—traditional and online—must be considered and addressed. Just as the degree to which social media is a concern varies within and between countries, the broader media environment carries the same concerns.

Future considerations

The landscape around information integrity concerns is rapidly shifting, which is expected to frustrate the work of practitioners, while creating new opportunities. The intermediaries, threats and the citizenry will all change in fundamental ways.

Legislation targeting large platforms has been or is being introduced in various parts of the world, and while the Digital Services Act in the European Union is maybe the most discussed, there are initiatives around the world already reshaping the Internet. However, various legislative efforts in the United States and challenges to the Supreme Court may have particularly far-reaching impacts. Two of the many planks of such legislation are content moderation directions and processes, and requirements for more transparency and data provision. One long-standing frustration of practitioners and researchers alike has been the difficulty on accessing data from the platforms,

which legislation may redress, at least in some jurisdictions.

Where the legislative road ultimately leads is unclear, with some instances likely to pose potential misuse for political or authoritarian goals, while others will strive to protect freedom of expression and digital rights.

The fragmentation of the social media landscape is likely, with newcomers already eroding the market share of incumbents. As each platform is unique in its own approach, culture and technical features, previous approaches and assumptions will need constant review and renewal. Continuing from current trends, it may also be the case that the popularity of platforms and their impacts will vary from country to country. As each new platform emerges, it may also take time for them to develop the same content moderation tools and features that have been developed by incumbents.

Another trend that may potentially further accelerate is the role of different parts of the technology 'stack' (the various technologies and companies that are required for platforms to operate) in content moderation, which may lead to broader and more indiscriminate removal of material. Consequentially, new technology ecosystems may emerge to insulate potentially offending platforms with different conventions around content moderation principles. The shift away from text, the continued rise of end-to-end encryption and the increasing effectiveness of artificial intelligence recommender engines are likely just the changes we currently experience.

More profound changes may also be felt. For example, the current centralized design of platform architectures, governed by a single entity, may give way to more distributed services governed by a myriad of hosts or even governed by the users themselves. Such approaches may dilute the power of 'big tech' and change the content moderation landscape profoundly. Effectively, the above changes describe a future where no controlling authority is possible, as is one where authority is fractured across innumerable bodies—which would have radical implications upon content moderation, accountability and economic models.

The intersection between the information pollution and other related information exploits, primarily hacking of election technology to discredit election systems or to manipulatively release sensitive information about the administration or individuals, may become more frequent.

The rapid development of machine learning and artificial intelligence will undoubtedly influence the space, though how exactly remains hard to predict. In the near term, fear exists that it will support the cheap and simple creation of persuasive and personalized mass information pollution that is seemingly indistinguishable from authentic content—be it textual, video or otherwise. It may prove more effective at espousing narratives of incitement to violence than traditional actors. At the same time, similar tools can be directed to better identify concerning content or activity. All the while, these technologies will be directed to making social media platforms more engaging. It remains unknown if, on balance, these developments will support information pollution or their opponents or if it will remain a cat-and-mouse game of competing capabilities.

Looking further ahead, a greater adoption of virtual reality and augmented reality will expose users to information pollution in increasingly intimate ways and further shift electoral activities into the cyber domain. Threat actors will invariably adjust to a new environment, exploiting technological advances and regulatory opportunities to seek to drive more persuasive and pervasive information pollution and avoid detection. We already see the intersection between artificial intelligence and better synthetic videos and audio creating increasingly powerful—and easy to produce—disinformation. There is also an underlying recognition that future generations will be digital natives, which will change the way they are able to assess and assimilate information compared to those—from some countries at least—accustomed to stronger media governance. They may be—in part depending upon investments in digital literacy—better digital citizens.

As practitioners consider responses, they need to accept that the ground is moving under their feet. In order to maintain effectiveness,

approaches need to be continuously challenged and investments are required to innovate new approaches and technology. Just as aggressors will continuously innovate to achieve their ends and leverage new opportunities, so must election practitioners as they seek to protect the information ecosystem.

Some of the dynamics listed above also point towards increasingly complex content moderation landscapes, which would indicate the importance of investments in public resilience measures.

Chapter 4 Regional analysis

The foundation of the research project was a series of workshops, consultations and discussions that were held with practitioners and experts in the fields of information integrity, conflict prevention and elections, as well as a survey that wa!s distributed online. Some of the discussions centred around regions, while others were with experts on a specific topic. Several themes were identified, some of which were common across the geographic regions engaged, while others differed depending upon the context or type of interlocutor.

Around 200 persons were engaged in the research process, covering 40 countries, as well as individuals from regional organizations. Predominantly, the participants were from the electoral community, emanating from the Global South and working on specific country contexts. The activity also included several academics, international organizations and representatives of social media platforms.

A widely discussed topic was the efficacy of fact-checking exercises—be they by civil society organizations or the EMB. While it was agreed that fact-checking is an important, useful and even brave endeavour, there were concerns, which reflect those raised in the relevant literature. First, most participants stated that they believed that the corrections did not travel as far or as fast as the falsehood that it was intended to address. This hypothesis was contested by at least one participant who stated the metric typically used—namely 'likes'—was misleading and that impressions were a more accurate measure and provided examples where this was evidenced. These two viewpoints are in line with the general state of research, with recent studies providing at least some credence to the latter position. While this would imply a greater confidence in the abilities of fact-checking, the participants also reflected upon the need to understand what techniques would support the most effective distribution of corrections to the public.

However, a more fundamental concern was raised, which is that once a belief has been set, it is harder to reverse, irrespective of the facts. This is also a concern held by literature and fact-checking organizations themselves. Swift responses to polluted narratives was viewed as key, with the highest probability of debunking success depending on immediate reactive measures taken.

The limitations to correcting beliefs led to the most common refrain from all the regional sessions—the need for strengthening civic education, both with regard to digital literacy and democracy. Goals included encouraging greater civility in public debate and increasing the understanding of the electoral process. The approach of 'pre-bunking' expected false narratives was also raised by some as a way to strengthen public resilience to information pollution. However, a contrary view was held by some experts, namely that it was not reasonable to put the burden of combatting information pollution upon individual citizens.

Participants described complex and varying information ecosystems; however they generally acknowledged the role of traditional media alongside the Internet. Some described a lifecycle of information pollution where harmful narratives form on the Internet before jumping to traditional media where they were more widely disseminated, while others presented a more circular process by which information would migrate back and forth. The key differentiator appeared to be the level of connectivity in each country. Where countries had limited connectivity, online information swiftly jumped to the more traditional media—with radio being a key medium.

In various countries, participants complained that media has been captured by partisans. They believed that this was in part due to the changing media landscape making older financial models unviable, leaving them vulnerable to deep-pocket interests.

The most prominent attribution for information pollution was to political entities or their proxies such as partisan media organizations. Accordingly, many placed an emphasis on the means to constrain these actors from engaging in such techniques. Codes of conduct were suggested as a potentially effective method, be they binding or voluntary. One factor that was cited to contribute to the functioning of such a code of conduct was the inclusion of the political parties within its design. However, concerns were raised that the efficacy of codes of conduct would be limited to electoral processes only, with the proliferation of information pollution resurging outside of election periods during which these codes of conduct are nullified.

Codes of conduct with provisions on social media, online campaigning and digital advertisement have been valuable tools to engage with various stakeholders during elections and establish accountability standards. There are good examples of this in Myanmar and Nigeria where the online environment was improved by political parties' commitment to codes of conduct that regulate online content and practices (Mutahi, 2022). A small number of countries in Eastern Europe expressed concerns about foreign influence operations; however in other regions, this was not raised.

The role of platforms received some criticism, though they were rarely the most significant concern. Overall, participants were concerned that platforms were not incentivized to combat threats to information integrity or that the engagement created by polarized debate was financially beneficial. On a more practical level, some participants complained about difficulties engaging with platforms or of inconsistent support to EMBs, depending upon the region—for example complaining about not being able to be registered as government organizations. Among the platforms, Twitter was largely seen as cooperative and communicative, while Facebook received criticism for only prioritizing information pollution during the election duration and being unresponsive outside of these periods. There was greater agreement on the need for the platforms to provide more transparency, with government intervention as required. There was a sharp minority position among some experts who viewed

personalization and financial incentives to be prime drivers of the concerns and requiring concerted advocacy—primarily by pressure from international institutions or domestic regulation. They believed governments should press upon platforms a duty of care for content or activity that "cause[s] significant physical or psychological harm to individuals" as well as conduct due diligence obligations for platforms' content moderation.

Various participants highlighted messenger platforms, particularly WhatsApp and Telegram, as being a concerning vector for disinformation, with end-to-end encryption making it harder for any party to meaningfully monitor the content shared.

The EMBs involved, and some of the advisors engaged, raised the limitations they had working in the space. They expressed fundamental questions about their role in the area of work, primarily due to it being outside of their core competencies and established toolsets.

Despite the various issues raised, when the role of government regulation was brought up, there was near uniform resistance to greater State control over online content—as it relates to disinformation. Participants shared concerns that it would be used for political advantage or to suppress dissent. As one participant put it, open societies have more trusted governments and confidence in their citizens to fight fake news; however, more repressive States will take advantage of this. One variant of this was participants from countries who believed they are targets of foreign disinformation—here stronger measures were called for, including the imposition of sanctions against perpetrators. Another perspective shared by one of the experts questioned the need for new legislation, but rather advocated for the use of existing tools, for example defamation or libel laws.

Online violence against women was a serious concern expressed in all regions. All regions have experienced an increase in political participation by women, with an associated increase in online harassment; hate speech; and false, circulating narratives targeting women politicians and female political supporters. In the Arab States, social media platforms were both

seen as a positive and negative force for political participation: while it enables women to launch campaigns outside of traditional, patriarchal structures, it also leaves them vulnerable to online attacks that dissuade women from political exposure. Experts describe a complex relationship between gender, harassment and online violence, which manifests online and offline. It was noted that harassment of women online was not solely the purview of male assailants, but women also participated.

At the same time, discussions also highlighted a number of benefits that the online space had brought. Despite the concerns raised regarding online harassment, it was reminded that on balance the Internet has brought an unprecedented means for women and marginalized groups to engage in the electoral process. More broadly, it was pointed out that platforms provided a means to challenge power structures.

participation. However, these should not be understood as prescriptive, and context-specific approaches are essential for programming.

Chapter 5 Thematic priorities areas

Building and sustaining a strong information environment requires intensive and tailored effort. There are a range of options that can be applied, a list that will continue to grow given the pace of innovation and evolving landscape.

Judging which options are most effective is difficult. The empirical knowledge base on the efficacy of combatting influence operations is limited, except for fact-checking efforts—especially those efforts made by platforms themselves (Courchesne, 2021). However, we have attempted to compare the existing research with the experiences shared with practitioners to identify promising options.

As a guiding principle in the design of any plan, activities should avoid undue or onerous freedom of expression restrictions; rather, it should favour those creating an enabling environment for freedom of expression (United Nations, 2019) (United Nations, 2019)

When considering the specific threat of election-related violence, it is worth noting the types of interventions that have been shown to work and then aligning these with the concerned actors in country. The key lessons here are that State actors are more susceptible to 'attitude-transforming' techniques, such as mediation and agreements, towards the goal of creating frameworks to pressure actors to non-violent competition. However, capacity-building efforts are more appropriate for non-State actors (Birch & Muchlinski, 2018).

A healthy strategy should include a variety of responses attempting to address three broad categories of interventions. These categories provide a framework to support the design of a holistic set of activities.

The three categories of intervention are the following, categorizing the programmatic activities developed:

Countering – Probably the most established pillar is to identify and attempt to counter information pollution. Fact-checking is the most established of such interventions, though others such as strategic communications and media monitoring are other variations of the same theme.

Resilience – Building public resilience to information pollution is increasingly in vogue as a programmatic area. It seeks to limit the ability of users to be influenced or co-opted by information pollution.

Prevention – Actions can address the supply side of information pollution. By preventing or deterring the creation of information pollution, the impact is of course nullified. Activities that most align with this are legislative or voluntary efforts to prevent various actors from producing or sharing information pollution, with political actors and platforms as the key actors within the election space.

These will map against the electoral cycle and will vary depending upon the category. Resilience activities will typically require a longer period of time to become effective, and hence broader resilience efforts would be best started early, continuing throughout the electoral cycle, becoming more election focused in the run-up to the process. Preventative efforts may come to the fore in the run-up to the election; however, much work will often be required in advance of the election to establish the frameworks and conduct trainings. Finally countering information pollution will be more active in response to specific election milestones—most likely peaking around polling and results processes. However once more, the efforts to establish the structures and empower responders will need to take place in advance.

Overall, information integrity programming must take place throughout the election cycle—commencing well in advance of the election itself—and the intensity will vary depending upon the type of activity.

Cross-cutting are a number of activities that have the ability to determine the success of interventions, such as coordination efforts, institutional strengthening, monitoring and research, and the protection of digital rights.

Countering

The most established branch of programmatic activities is the tactical response to detected information pollution. Over time, there has been a recognition that there are contradictions in its outcomes and limitations in its efficacy. Unfortunately, some disinformation and hateful messages are believed to spread faster than facts on social media depending upon their virality, posing additional challenges in countering hate speech through peaceful messages and reliable information (Juul & Ugander, 2021) (Vosoughi, Roy, & Aral, 2018).
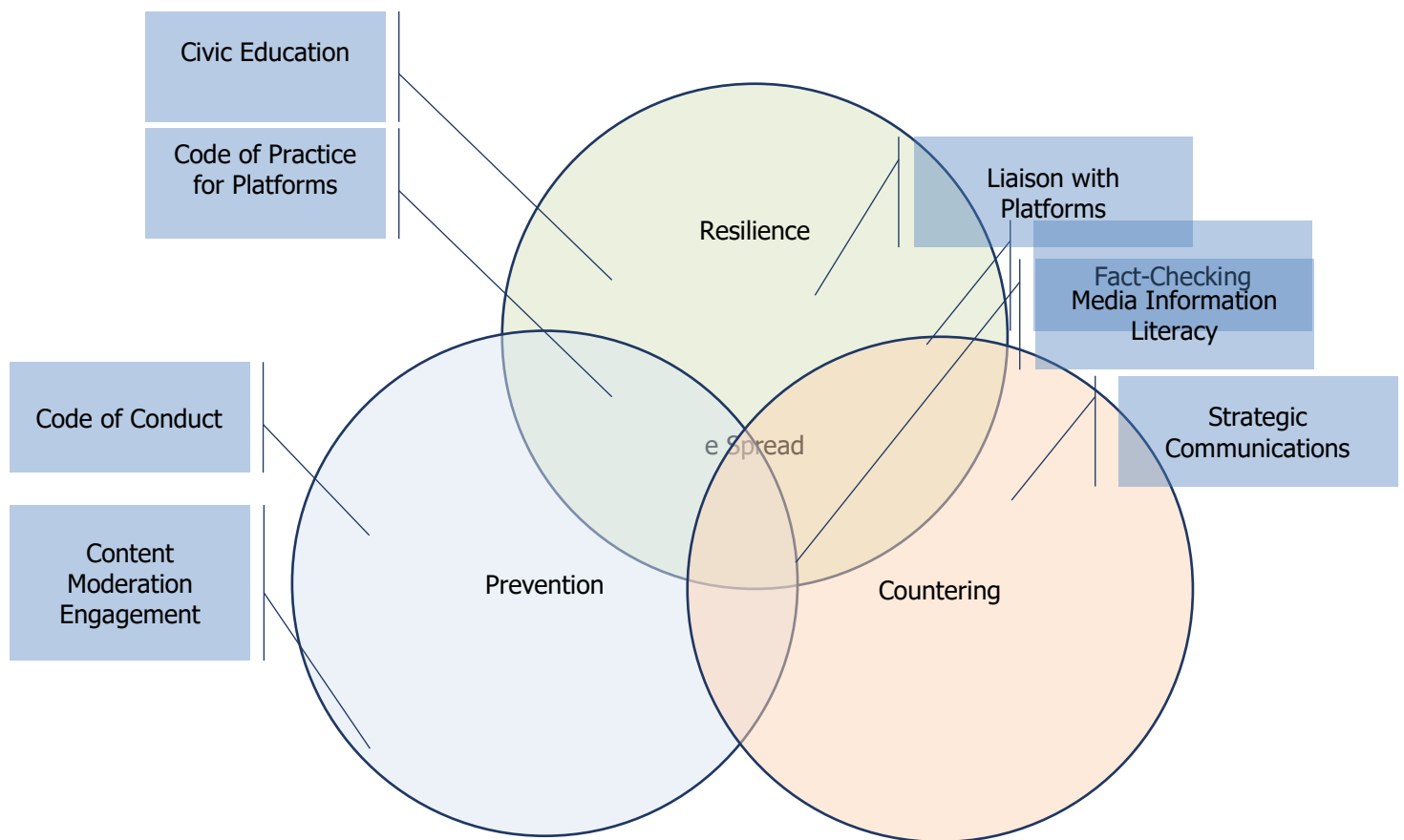

Figure 1: The three categories of intervention

And yet, it remains a critical aspect of the programmatic menu, and if appropriately delivered, is highly effective at tackling information pollution in an election. Fact-checking, as covered below, has been the main activity; however, there are variations such as State media monitoring or components like strategic communications that can also be developed.

Fact-checking

A commonly used technique to address beliefs from information pollution is to identify false narratives and communicate correct information. While 'debunking' or exposing a person to correct information may reduce misperceptions, it is limited in its ability to change people's opinions on candidates (Nyhan, Porter, Reifler, & Woods, 2020)(Nyhan, Porter, Reifler, & Woods, 2020)In the face of corrections, the debunking effect is weakened when an audience supports the initial information pollution (Chan, Jones, Jamieson, & Albarracin, 2017).

Some posit a 'backfire effect' where corrections actually increase misperceptions (Nyhan & Reifler, When Corrections Fail: The Persistence of Political Misperceptions, 2010)though more recent findings contest these effects (Wood & Porter, 2019). Another related constraint is the so called 'Streisand effect', where the attempt to censor information leads to increased exposure of news compared to no action being taken (Jansen & Martin, 2015). The above highlights some of the limitations of fact-checking, identified by academic researchers and practitioners alike (Tompkins, 2020).

Broadly, fact-checking, however, has a positive impact (Walker, Cohen, Holbert, & Morag, 2020). How the debunking of fact-checking is communicated has a significant impact upon how effective it is. Of course, ensuring wide distribution is vital, as are the source of the information pollution and the speed of the correction (Walker & Tukachinsky, A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It?, 2019). That is, if it provides the corrected details rather than just labeling and if it does not repeat the original falsehood. Where

information pollution is being addressed, it is the trustworthiness of the source, rather than their expertise, which has the most impact on the permeance of the incorrect beliefs (Ecker & Luke, 2021).

Of course, these are not simple hurdles to overcome. In many cases, pulling together approved rebuttals can be a time-consuming process especially for complex and technical issues or where there are various approvals. The original sources are often not motivated to issue corrections, and universally trustworthy channels may be rare.

IFCN accreditation is a critical milestone for fact-checking institutions. While hard to achieve, it opens up a number of opportunities, including funding from platforms and access to data. However, it is also the case that a report by a fact-checking organization will not necessarily lead to action by platforms.

For fact-checkers, and those working with them, there are also protection concerns that should be considered within the establishment of a programme. Various fact-checking organizations across the world report operating in fear of government reprisals or public harassment. In some cases, the threats come with the association that fact-checking parties have with social media firms (Stencel, 2020).

Resilience

Ultimately, no intervention will completely prevent the existence of information pollution. However, where the citizenry is effectively equipped to critically navigate the information ecosystem, the dangers of information pollution will be dampened. Building such resilience rests upon both the orientation of citizens in media literacy and the structures to permit free and plural public debate.

Civil education

Media and information literacy has been raised by many as the new silver bullet, a status once held by fact-checking. However, here also, practitioners should remain circumspect to the impact that can have. Clearly, a more critical and digitally sophisticated populace would

naturally be expected to be better able to spot and filter information pollution. But there is already widespread concern among citizens in much of the Americas, Africa, the Middle East and Europe, with many countries demonstrating that over half their Internet users are concerned about false information online (Knuutila, 2022). That said, fear of information pollution is not the same as being resistant to it. There appears to be limits to the efficacy of digital literacy in preventing people from sharing fake information, even if it does increase their ability to identify such items (Sirlin, 2021). The design of such programmes also appears critical, as with one country-level example of election-related in-person media literacy. The study found their intervention had only limited ability for participants to identify information pollution, while motivated reasoning led to supporters of the incumbent becoming less able to identify pro-attitudinal content (Badrinathan, 2021).

It is also worth considering the pace of digital change in this context. While education for school-aged children is valuable in addressing various digital harms, it may be limited in the election content for no other reason than by the time school-aged children largely age to the point that they are politically engaged, some of the lessons may simply be redundant. While of course life-long education would also be valuable, it is not possible to implement as widely.

Another approach for fostering citizen resilience is to design peace and anti-violence messaging campaigns. Traditional versions of these have been demonstrated to have an impact in preventing electoral violence (Collier & Vicente, 2014) (UNESCO, 2020). Various actors are exploring how these campaigns can be effectively deployed online.

Social media has been instrumental in preventing, reducing and responding to electoral conflict and violence. For example, different actors have used social media to monitor tensions and violence as part of early-warning and response mechanisms. While social media can be a vector for inciting violence, it can also be used as a channel for spreading peaceful messages. It has provided a platform for political discussions, ushered in online activism

and supported organizing and mobilization around political questions and demands. It has been used by various election stakeholders, including EMBs, the judiciary, security agencies, traditional media, civil society and political parties, to share information, dispel misinformation, fact-check, spread peace messages, conduct civic education and promote the participation of young people in elections (Mutahi, 2022).

Pre-bunking

A more targeted approach to building public resilience to information pollution is 'pre-bunking' or 'inoculation'. Here, users are pre-warned of possible false narratives in an effort to limit their influence if encountered during the process (Cook, Lewandowsky, & Ecker, 2017), which has seen encouraging research. Different routes can be used to deliver such, including social media posts, video or even games (Roozenbeek, Van Der Linden, & Nygren).

Within an election, there are generally some obvious narratives that deserve pre-emptive action. To identify these, the first call would be to look at the electoral cycle to identify risk points, for example, the correctness of the voter register, the accuracy of the election results and disincentives to voting, among others. At the same time, it is not possible to identify or address all the narratives—and they will inevitably be unique from country to country—so triaging of the most damaging and likely threats is required.

Some of the same concerns and tactics that exist with debunking messaging remain valid, for example, taking care not to accidentally create or reinforce false narratives, structuring messages to be most effective and having strong coordination with media organizations (Garcia & Shane, 2021). Others hold concerns about the extent that inoculation is viable in 'Global South' countries since it does not address the specific political economy of propaganda specific to these contexts (Abhishek, 2021).

Prevention

Underlying prevention activities are working with—or against—the actors who are responsible for the creation of harmful content to stop it from being made, released or distributed in the first place.

Legislative measures

Member States are increasingly exploring legislation or binding regulations related to information integrity—specific to electoral events or with more general reach. The application of such approaches can be quite contentious, with possible concerns that they are unduly contravening freedom of expression, are imprecise or do not make a reasonable connection between the expression and a harm. Concerns exist that they might be misused by governments against critics and political adversaries (Khan, 2021).

The design of such policies should take into careful consideration human rights commitments, the broader democratic environment in the country and the nature of regulatory bodies and other institutions in place. The regulation should be carefully tailored and the result of a truly inclusive consultation, ultimately complying with the requirements of legality, necessity and proportionality under human rights law (Report of the Secretary-General, 2022).

Broadly, in many circumstances, it may be more appropriate to identify other mechanisms to curtail the actions of various actors, such as codes of practice or codes of conduct. At the same time, existing laws that censor defamation and harassment might suffice to tackle information pollution without the need for expansive new powers.

Engagement with political stakeholders and other information pollution actors

If the root of much election-related information pollution is political actors, then finding ways to dissuade them from engaging in such activities is logical, despite being challenging. Of course, they are not alone as creators or propagators of harmful content. Media entities, influencers and others are just some of the other potential sources.

Various methods may be used with these actors in order to attempt to dissuade them from engaging in information pollution. These may include, for example, codes of conduct, trainings, State monitoring and mediation. At the same time, programme designers should be realistic about the various competing incentives upon actors' behaviour. Alongside these efforts may be monitoring activities to motivate compliance.

Electoral Management Bodies, political parties, candidates, citizens, journalists and other stakeholders have negotiated and agreed to a code of conduct during elections in many countries. While most of these cover general rules of behaviour by the actors, they have also incorporated social media elements. For instance, election stakeholders in Myanmar (2015, 2020), Georgia (2020) and Kosovo (2021) have committed to declarations/codes of conduct that regulate their social media behaviour ahead of elections.

Codes of conduct have proved particularly useful in enabling political actors to reaffirm their commitments to fair play in elections. Although investment in codes of conduct is promising, self-regulation may only have limited effects, especially if there are no robust enforcement mechanisms.

Furthermore, the effectiveness of the codes of conduct is watered down if some of the individuals and groups who can meaningfully contribute to the code's implementation and whose actions could exacerbate conflicts do not sign it. It has been particularly useful in some contexts for the monitoring committees of codes of conduct to have a presence on social media and communicate directly with the public about their monitoring. Furthermore, it is important to build local and individual buy-in for codes of conduct (Mutahi, 2022).

Platform engagement

While each platform has its own features, user culture, monetization model and content rules,

they each have significant levels of control over what content is published on them and how this is distributed. However, they demonstrate varying levels of investment in addressing information pollution, as well as, at best, mixed outcomes. Be it by working with platforms or conducting advocacy towards them, it is possible to make a meaningful impact upon the information ecosystem. In-country election practitioners are often well placed to identify risks that need to be navigated, communicate features that should be deployed and identify areas of collaboration.

One increasingly common tool of note is the establishment of codes of practice—which may be a more desirable approach to governing platform behaviour than legislation. Social media codes of conduct regulate discussions on such issues as disinformation affecting elections and trolling, which if left unchecked undermine the public's trust of the electoral process and its legitimacy. These may cover the activities of platforms alone or also incorporate political parties. For the 2021 Dutch legislative elections, political parties and Internet platforms including Facebook, Google, Snapchat and TikTok agreed on voluntary rules in a code of conduct, the first of its kind in the European Union. The companies made transparency commitments regarding online political advertisements during election campaigns. Another country that has launched a social media code of conduct is Kosovo.

Media codes of conduct have come to now encompass social media as another way of regulating online speech during election periods. Technological innovation is also underway by third parties. For example, users may also be able to take actions to protect themselves, using the tools available; for example, services are being built to prevent online harassment, where platforms permit them (Chou, 2021).

Chapter 6 Considerations & recommendations

The study of information integrity is a nascent field, with too few definitive answers. Looking at the specific question that this document explores—how to prevent information integrity threats contributing to election violence—it is clear from the complexities outlined that no single solution exists. However, it is hoped that this document, and the associated materials, provide some insights to help programming design and implementation by election practitioners.

There are alignments, contradictions and gaps between the research and the practical experience collated for this report. As noted by senior experts spoken to, the research available is expected to explode in the coming years, which may bridge some divides or provide more clarity on why there are these discrepancies—though there are likely no easy answers. Furthermore, the constantly evolving technological, social and political landscape will invariably create new questions or confound old answers.

Election violence has been a long-standing threat to democracy; however, the information age has crafted new tools that stand to potentially spur division and incite election violence. As electoral practitioners, this new dimension of complexity complicates an already challenging task and calls for new ways of thinking and skills. However, rather than to merely consider it a burden, it is also vital to explore how the digital age may fundamentally enhance the work of practitioners to better support peaceful, credible and inclusive electoral processes.

On the basis of the research efforts, below are several recommendations. These recommendations are built to guide election practitioners as they consider how to design programmes in the election and information integrity field.

No silver bullet

There is no reasonable way to control exactly what content exists on the Internet or even an individual platform, and even if there were, there is no consensus on what should be disallowed in the name of information pollution.

Some electoral information integrity challenges are ingrained within political culture, electoral processes and questions around how to apply fundamental human rights. International conventions provide some direction on the types of behaviour that should be addressed with censure, and voters certainly deserve reliable information, yet there are a host of grey areas. The intrinsic nature of electoral competition creates a complex set of dynamics whose harms cannot be fully neutralized, only ameliorated.

While the toolkit to protect the electoral information ecosystem continues to grow and mature, each activity or approach comes with its own limitations and trade-offs. At the same time, the challenges are evolving, thus eroding the capacity of previous approaches.

Multi-pronged responses

The absence of a silver bullet demands a multi-pronged approach. These should be designed to ensure complementarity between the various endeavours and that they are tailored to the specific local context. As discussed in more detail within separate recommendations, responses should look to consider means to build public resilience, attempt to limit the creation of information pollution and respond to information pollution. Furthermore, the approaches should be designed with consideration of the broader information integrity activities and concerns that exist outside of electoral periods.

Given the lack of evidence on what measures are particularly effective, it further makes sense

to avoid putting all eggs in one basket. However, not all these activities can or should be delivered by one actor.

Credible elections as a must

While the new information landscape and threats have created a range of novel and exciting programming options, the core task of successfully running an election that is credible remains vital. Support to the professional conduct of an election is made only more necessary by the new fragility of the information ecosystem. Established routes to build trust in the electoral process should underpin and overlap with work focused on the information ecosystem. Related to this is the heightened value of transparency. If information has become the core currency online, and if genuine and plausible information is not published, then there is an opportunity for another actor to fill the vacuum with false and harmful messages.

Coordinated action

No single entity can resolve the myriad of information integrity challenges present in an election, and certainly not an EMB alone. The diverse range of organizations working on these issues is encouraging; however, how they work together is likely key to their collective success. Different models of partnerships are being developed. Election 'war rooms' provide for a joined-up crisis response. Social media councils are envisaged to allow coordination on advocacy. Information integrity coalitions bring together government and non-governmental actors including the private sector, political parties and ordinary citizens.

An effective programme of work requires a multi-stakeholder approach and the ability to creatively craft solutions. There are choices and actions that can be taken by the various actors in an election process, including citizens, civil society, State authorities, private sector platforms, traditional media and, perhaps most importantly, political figures. Together, they can build a strong information ecosystem to aid peaceful and credible elections. Multi-stakeholder dialogue and collaboration on social

media and preventing electoral violence should therefore be promoted and involve national authorities, social media companies, political parties and civil society.

Grounded in context

Each election is defined by a host of factors, creating a unique set of risks and opportunities. A starting point for any exercise should be an assessment of the information environment, directed by the particular political, security and social concerns in the country. Given the rapidly evolving dynamics, a continuous review should be in place as much as possible.

Centered in human rights

Cut broadly, interventions fall into two camps. The first is to address concerns by imposing controls over the information ecosystem. While potentially essential in some cases, such activities may be to the detriment of freedom of expression, the right to participate in public affairs and other essential human rights—intentionally or as an unfortunate side effect. Such efforts may ultimately undermine trust and threaten the credibility of the election.

Rather, interventions rooted in the protection and promotion of human rights should be prioritized, for example those which seek to promote transparency, access to information, media freedoms and public education. Where restrictions are to be established, they are expected to meet a high threshold of legality, legitimacy, necessity and proportionality.

A practical component of this, on the platform side, should be an adherence to human rights due diligence and regular transparent impact assessment. These may be guided by the approaches outlined in the UN Guiding Principles on Business and Human Rights (United Nations Human Rights Council, 2011).

Caution around restrictive regulation

While elections are clearly a vital part of any democratic society, actions taken to secure them should also consider the broader democratic landscape, human rights

environment, and how these relate to digital rights.

In practical terms, this calls for caution when providing advocacy or advice on legislative or regulatory activities. While there are certainly flaws with the approaches taken by platforms, risks—perceived or otherwise—should be balanced against an increase in State authority. While there may be legislative approaches that are adopted in some countries or regions, it is not necessarily wise to transplant them to another context. Certainly, given the novel nature of much legislation, little evidence has been collected on how they materially impact an election even in the contexts they were designed for.

Rightly, in most democracies, campaign regulations have long been in place to provide at least some semblance of a level playing field and to impose some guardrails on speech around the political process. How these translate onto the online space remains a work in progress, and one that is not solely in the control of the national authorities, with platforms often in the driver's seat.

Combat the causes of polarization

The research and conclusions drawn above point towards the threat that information pollution poses with regards to polarizing voters and politics, and how these can threaten credible and peaceful elections. We identify negative synergies between information pollution and incitement to election-related violence. They both rely on the existence of underlying cleavages, which they can inflame and exploit. Hence, practitioners should prioritize activities that aim to ease societal tensions, address excesses of division and foster confidence in the election process.

Focus on political actors

While there are various threat actors, by far the most potent in an election context are domestic political actors. These dominate as the sources of information pollution, driven by a motivation to incite election-related violence or discredit an election.

However, the means to constrain political actors may be limited. Using regulatory tools or codes of conduct can have a positive influence, supporting the negotiation and implementation of codes of conduct on online campaigning and the use of social media by the different electoral stakeholders can contribute to diffuse tensions. Monitoring and sanction mechanisms can help motivate compliance.

Support to the appropriate actors involved in advocating or guiding social media regulation, as well as on advocacy and monitoring of codes of conduct and platform codes of practice, is important. Broader programming is also required to respond more effectively to alarming rhetoric in a timely and persuasive fashion, before it has the opportunity to deepen divisions and drive election-related violence.

Fact-checking as a foundation, not a solution

For some time, fact-checking was the first and only line of defence against most forms of electoral information pollution. However, evidence and the experiences of many in the field indicate that its effectiveness is less definitive than hoped. It faces the challenge that it is hard to change people's minds once they have formed ideas that correspond to pre-existing beliefs. While fact-checking is a vital tool for accountability, and indeed, has some value in mitigating against information pollution, it should instead be considered the last tool—with additional attention being paid to 'up-stream' resilience.

Building public resilience

Much of the attention given to building public resilience to information pollution is warranted. Increasing the public's ability to critically assess content and identify information pollution can help to nullify its impact. There are a variety of methods that are being considered here, for example: digital, media and information literacy, and pre-bunking.

However, its effectiveness should not be overstated. As explored before, the receptiveness to disinformation may be less

related to the knowledge that it is false, than that the themes align with the recipient's political bias. Thus, enhanced media literacy is valuable but cannot resolve some of the most dangerous scenarios.

This also carries a number of difficult operational challenges, such as the complexity of reaching broad audiences, difficulties to identify most at-risk individuals and the need to commence well in advance of an election process.

Defending the credibility of public institutions

Building trusted and capable institutions represents one of the more traditional approaches to preventing election-related violence and remains vital. If anything, the more fluid the information landscape the more important it is that technical mistakes in the operation of the election are avoided, and that when they do take place, they are well communicated to the public.

Specifically in the case of information integrity threats, it is clear that the EMB cannot assume all burdens in combatting these threats. However, while the mandates of EMBs and other relevant State institutions will vary between countries, some programming aimed at promoting information integrity remains inescapable. At a minimum, they will be required to protect their own institutions from attacks or defamation if they are to support the credibility of the election, and to do so, they require the tools—and financial resources—to defend themselves.

The myriad of new challenges that election EMBs will need to address require a different technical toolkit, which in turn will generate additional requirements for providers of election assistance. Beyond the EMB, other domestic regulatory authorities, as well as civil society and the media, can also be supported to operate in this new domain, for example with advice on online monitoring of hateful messages, investigations, counter-messaging and conflict-sensitive coverage of the elections.

The integrity of the information technology infrastructure and data of institutions and public

officials is also vital, in particular as they provide an intersection with information pollution. Election Management Bodies in particular hold significantly sensitive systems such as their voter register and results tabulation systems. Hacked data can be leaked as part of misinformation campaigns. Furthermore, there are cases of public officials being harassed and having their online information misused. Accordingly, the cybersecurity and cyber hygiene practiced can have a tangible impact upon exposure to information pollution—and this is another space in which support may be provided.

How and whether these activities and tools can be funded in a sustainable fashion remains a question. This is an even more acute concern when we remember the purveyors on information pollution are increasingly funnelling capital into their activities.

Supporting independent media

Journalism and the traditional media outlets have long been a vital component of a healthy information ecosystem. For various reasons, some related to the new digital platforms, the practice of journalism is under increasing attack and suffers diminished trust.

As with other electoral components, codes of conduct agreed between media outlets can be used, to some extent, to incentivize professional coverage and build public trust. In countries where there is concern of media capture by partisan interests, this likely requires strong monitoring to be effective.

Activities that support impartial and effective journalism are increasingly vital. Specific to information pollution, assisting appropriate institutions or individuals in the practice of fact-checking, verifying sources and navigating the new dynamics can be valuable. Supporting journalistic capabilities in using the new data sources to better implement their activities, such as open-source investigation techniques, have the potential to provide new capabilities to their field. Unfortunately, also the added task of exploring means to support journalists in protecting themselves from abuse has become unavoidable.

Social media for good

Much has been made of the detrimental impacts social media can have. But ultimately, it is a tool and can be wielded for ill or for good. There is great potential benefit in using social media to enhance the electoral process. The Internet offers the ability to reach previously unengaged demographics and to create meaningful connections with them. For example, social media can empower disenfranchised communities such as women and youth in societies where they are traditionally marginalized.

Social media has long been used as a means to coordinate advocacy and defend human rights. The opportunity to do so should be protected, especially in the case of contested elections. Using these channels for pro-peace or counter-messaging to better anticipate election issues and diffuse tensions can help the peacefulness in the offline world. Social media can be used to increase trust and knowledge about elections, promote political participation, and support early warning, mapping of hotspots and tracking of violence.

Monitoring and research at the core

In an area of work that seeks to address challenges that have merely emerged in the recent past, little is fully understood. In particular, there is a lack of know-how pertaining to the effectiveness of different interventions and how exactly to make them most impactful. As described in this document, researchers are still grappling with fundamental questions. With an understanding that the community is still trying to work out how to 'fix' the problem, and that the nature of the problem will vary widely from context to context, it is vital that projects include a significant monitoring component. Furthermore, where possible, there is great value in considering how projects can also contribute to the broader research efforts through rigorous data collection and analysis.

Breaking out of the electoral cycle

Information pollution, and the drivers of it, do not restrict themselves to the period of electoral operations. Political discourse continues in perpetuity, often preparing the ground for the various contests that proceed an election. Furthermore, many activities in the area require prolonged periods to be effective, in particular those that are intended to influence human knowledge and behaviour. Accordingly, activities need to be considered beyond a single election process and over multi-electoral cycles with full acknowledgement of the costs this entails.

Beyond the Internet

The information ecosystem is more than just the online space. 'Content' moves between mediums, transforming along its path. Its route and behaviour will differ from country to country, in large part based upon the types of traditional media (radio, television etc.), citizens' access to different media and the culture of the society. Additionally, the historic and contemporary role of traditional media in the creation and spread of inflammatory content may well be a more important concern to the peaceful nature of the election than that of new technologies. Without a comprehensive understanding of the entirety of the information landscape, it is plausible that responses will inadequately address the main drivers of information pollution or fail to stem problems at their source.

Platforms' potential and frustrations

For many reasons, stakeholders often look to social media platforms as the solution to the information pollution issue. However, while they have an important role to play, they are unlikely to unilaterally resolve the challenge of information pollution. Platforms can, when motivated, do much to support information integrity around an electoral process. However, they do not apply these resources in a uniform fashion, and how they devise their engagement strategy towards a certain country may be at odds with the actual concerns on the ground. Accordingly, it is wise for various stakeholders and practitioners in country to advocate together for the attention and programming they believe are required.

Digital companies may require support and scrutiny to ensure adequate algorithm management along with timely and effective action to protect electoral integrity and remove posts inciting violence. Not all platforms are equal in their ability to support information integrity around elections. The underlying technical features, audiences and profitability will all make a difference. The rate at which platforms grow, and in some cases their ideologies, make them immature in terms of content moderation or other programmes. As part of the assessment of the information landscape, understanding which platforms are a concern is as important as is then deciding where best to target attention.

Plan for a strange and unusual future

The rapid evolution of the Internet and associated technologies assure an unpredictable future. The likely radical changes over the coming years and decades will provide new opportunities and challenges to the conduct of credible elections and the programmatic activities. The fluid nature of the subject will have profound impacts upon actors involved in upholding information integrity. Solutions that rely on specific platforms or data sources may find themselves irrelevant as audiences migrate to various new sites. There will be a need to fund continuing innovation, development of technologies and constant research to understand the current scope of concern.

Bibliography

Abhishek, A. (2021). Overlooking the political economy in the research on propaganda. *Harvard Kennedy School (HKS) Misinformation Review*. doi:https://doi.org/10.37016/mr-2020-61

Affairs, U. N. (2016). *Policy Directive: Preventing and Mitigating Election-related Violence.*

Allcott, H., & Gentzkow, M. (2017). *Journal of Economic Perspectives, 31*(2), 2011 - 236. doi:https://doi.org/10.1257/jep.31.2.211

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 629-76. doi:10.1257/aer.20190658

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research and Politics*. doi:https://doi.org/10.1177/2053168019848554

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosytem. *Science Advances*. doi:https://doi.org/10.1126/sciadv.aay3539

America, P. (2021). *Hard News: Journalists and the Threat of Disinformation.* Retrieved from https://pen.org/wp-content/uploads/2022/04/PEN-America-Complete-Survey-Results_Hard-News.pdf

Antoci, A., Delfino, A., Paglieri, F., Panebianco, F., & Sabatini, F. (2016). Civility vs. Incivility in Online Social Interactions: An Evolutionary Approach. *PLOS One*. doi:10.1371/journal.pone.0164286

Arguedas, A. R., Robertson, C. T., Fletcher, R., & Nielson, R. K. (2022). *Echo Chambers, Filter Bubbles, and Polarisation: a Literature Review.* Reuters Institute for the Study of Journalism, University of Oxford. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-01/Echo_Chambers_Filter_Bubbles_and_Polarisation_A_Literature_Review.pdf

Badrinathan, S. (2021). Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. *American Political Science Review, 115*(4), 1325-1341. doi:doi:10.1017/S0003055421000459

Bail, C. A., Argyle, L. P., Brown, T. W., Volfovsky, A., Bumpus, J. P., Chen, H., . . . Merhout, F. (2018). Exposure to opposing views on social media can increase political polarization. *PNAS, 115*(37), 9216-9221. doi:https://doi.org/10.1073/pnas.1804840115

Banaj, S., & Bhat, R. (2018). *WhatsApp Vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India.* LSE, Department of Media and Communications.

Barrett, P. M., Hendrix, J., & Sims, J. G. (2021). *Fueling the Fire: How Social Media Intensifies U.S. Political Polarization — And What Can Be Done About It.* NYU Stern Center for Business and Human Rights. Retrieved from https://bhr.stern.nyu.edu/polarization-report-page

Bateman, J., Hickok, E., Courhesne, L., Thange, I., & Shapiro, J. N. (2022). *Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research*.

Bekoe, D. A., & Burchard, S. M. (2017). The Contradictions of Pre-election Violence: The Effects of Violence on Voter Turnout in Sub-Saharan Africa. *African Studies Review, Volume 60*, 73 - 92. doi:https://doi.org/10.1017/asr.2017.50

Bennett, W., & Livingston, S. (2020). The Role of Public Broadcasting. In W. Bennett, & S. Livingston, *The Disinformation Age* (pp. 211-258). Cambridge: Cambridge University Press.

Birch, S., & Muchlinski, D. (2018). Electoral violence prevention: What works? *Democratization, 25*(3), 385-403. doi:10.1080/13510347.2017.1365841

Birch, S., Daxecker, U., & Höglund, K. (2020). Electoral violence: An Introduction. *Journal of Peace Research*, 3-14. doi:https://doi.org/10.1177/0022343319889657

Bisbee, J., Brown, M. A., Lai, A., Bonneau, R., Tucker, J., & Nagler, J. (2022). Election Fraud, YouTube, and Public Perceptionof the Legitimacy of President Biden. *Journal of Online Trust and Safety*. doi:10.54501/jots.v1i3.60

Boomgaarden, H., Strömbäck, J., Lindgren, E., Vliegenthart, R., Damstra, A., & Tsfati, Y. (2020). Causes and consequences of mainstream media dissemination of fake news: literature review and

synthesis. *Annals of the International Communication Association*, 157-173. doi:https://doi.org/10.1080/23808985.2020.1759443

Bradshaw, S., & Howard, P. N. (2017). *Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation.* University of Oxford. Retrieved from http://governance40.com/wp-content/uploads/2018/11/Troops-Trolls-and-Troublemakers.pdf

Bradshaw, S., Bailey, H., & Howard, P. N. (2020). *Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation.* Oxford Internet Institute. Retrieved from https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/02/CyberTroop-Report20-Draft9.pdf

Brennen, J., & Perault, M. (2021, March 19). *How to increase transparency for political ads on social media*. Retrieved from Brookings: https://www.brookings.edu/blog/techtank/2021/03/19/how-to-increase-transparency-for-political-ads-on-social-media/

Bruns, A. (2019). *Are filter bubbles real?* UK: John Wiley & Sons. doi:orcid.org/0000-0002-3943-133X

Buerger, C. (2021). *Speech as a Driver of Intergroup Violence: A Literature Review.* Dangerous Speech. Retrieved from https://dangerousspeech.org/speech-as-a-driver-of-intergroup-violence-a-literature-review/

Chan, M.-p. S., Jones, C. R., Jamieson, K. H., & Albarracin, D. (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*. doi:https://doi.org/10.1177/0956797617714579

Chauchard, S., & Garimella, K. (2022). What Circulates on PartisanWhatsAppin India? Insights from an Unusual Dataset. *Journal of Quantitative Description: Digital Media 2*, 1-42. doi:https://doi.org/10.51685/jqd.2022.006

Chou, T. (2021). *Here's how to fix online harassment. No, seriously*. Retrieved from Wired: https://www.wired.co.uk/article/social-media-harassment-open-apis

Clegg, N. (2021). *You and the Algorithm: It Takes Two to Tango*. Retrieved from Medium: https://nickclegg.medium.com/you-and-the-algorithm-it-takes-two-to-tango-7722b19aa1c2

Collier, P., & Vicente, P. C. (2014, February). Votes and Violence: Evidence from a field experiment in Nigeria. *The Economic Journal, 124*(574), pp. F327–F355. doi:https://doi.org/10.1111/ecoj.12109

Cook, J., Lewandowsky, S., & Ecker, U. K. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLoS One*. doi:https://doi.org/10.1371/journal.pone.0175799

Council of Europe. (2017). *Internet and Electoral Campaigns: Study on the use of internet in electoral campaigns.* Retrieved from https://rm.coe.int/use-of-internet-in-electoral-campaigns-/16807c0e24

Courchesne, L. I. (2021). Review of social science research on the impact of countermeasures against influence operations. *Harvard Kennedy School (HKS) Misinformation Review*.

Daxecker, U., & Jung, A. (2018). Mixing votes with Violence: Election Violence around the World. *SAIS Review of International Affairs, 38*(1), 53-64. doi:10.1353/sais.2018.0005

Daxecker, U., & Prasad, N. (2022). Poisoning Your Own Well - Misinformation and Voter Polarization in India. *European Research Council (ERC)*.

Daxecker, U., & Prasad, N. (2022). *Poisoning Your Own Well: Misinformation and Voter Polarization in India.* University of Amsterdam.

Daxecker, U., & Prasad, N. (2022). Poisoning Your Own Well: Misinformation and Voter Polarization in India.

Daxecker, U., & Prasad, N. (2022). Voting for Violence: Examining Support for Partisan Violence in India.

Deane, J., & Ismail, J. A. (2008). The 2007 General Election in Kenya and Its Aftermath: The Role of Local Language Media. *The International Journal of Press/Politics*. doi:https://doi.org/10.1177/1940161208319510

Douek, E. (2020). Australia's 'Abhorrent Violent Material' Law: Shouting 'Nerd Harder' and Drowning Out Speech. *Australian Law Journal*. Retrieved from https://ssrn.com/abstract=3443220

Douek, E. (2021). Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability. *Columbia Law Review, 121*(3). doi:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3679607

Douek, E. (2021). The Limits of International Law in Content Moderation. *UC Irvine Journal of International, Transnational, and Comparative Law*. Retrieved from https://scholarship.law.uci.edu/ucijil/vol6/iss1/4

Douek, E. (2022, December). Content Moderation as Systems Thinking. *Harvard Law Review, 136*(2). doi:https://harvardlawreview.org/wp-content/uploads/2022/11/136-Harv.-L.-Rev.-526.pdf

Durán, R. J., Müller, K., & Schwarz, C. (2022). The Effect of Content Moderation on Online and Offline Hate: Evidence from Germany's NetzDG. *SSRN*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4230296

Ecker, U. K., & Luke, A. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*(49), 631-633. doi:https://doi.org/10.3758/s13421-020-01129-y

Elklit, J., & Reynolds, A. (2002). The Impact of Election Administration on The Legitimacy of Emerging Democracies: A New Comparative Politics Research Agenda. *The Journal of Commonwealth & comparative politics*. doi:10.1080/713999584

Elliott, V. (2021, November 22). Facebook may be dangerously unprepared for Libya's upcoming presidential election, documents show. Retrieved from https://restofworld.org/2021/facebook-libya-election/

Facebook Oversight Board. (2021). *Annual Report*. Retrieved from https://oversightboard.com/attachment/425761232707664/

Fatafta, M. (2021, November 18). *Facebook is bad at moderating in English. In Arabic, it's a disaster*. Retrieved from Rest of the World: https://restofworld.org/2021/facebook-is-bad-at-moderating-in-english-in-arabic-its-a-disaster/

Fletcher, R., Robertson, R. T., & Neilsen, R. K. (n.d.). How Many People Live in Politically Partisan Online News Echo Chambers in Different Countries? *Journal of Quantitative Description: Digital Media*, 2021. doi:https://journalqd.org/article/view/2585/2076

Fukuyama, F., Richman, B., Goel, A., Melemed, D., Katz, R. R., & Schaake, M. (n.d.). *Middleware for Dominant Digital Platforms: A Technological Solution to a Threat to Democracy.* Stanford Cyber Policy Centre - Freeman Spogli Institute. Retrieved from https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/cpc-middleware_ff_v2.pdf

Garcia, L., & Shane, T. (2021). *A guide to prebunking: a promising way to inoculate against misinformation.* Retrieved from First Draft News: https://firstdraftnews.org/articles/a-guide-to-prebunking-a-promising-way-to-inoculate-against-misinformation/

Ginty, R. M., & John, A. W.-S. (2022). Peacemaking and Election Violence. In I. von Borzyskowski, & R. Saunders, *Contemporary Peacemaking* (pp. 307-331). Palgrave Macmillan. doi:https://doi.org/10.1007/978-3-030-82962-9_16

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and politicla challenges in the automation of platform governance. *Big Data and Society*. doi:http://dx.doi.org/10.1177/2053951719897945

Guess, A., Lyhan, B., Lyons, B., & Reifler, J. (2018). *Avoiding the Echo Chamber about Echo Chambers.* Knight Foundation. Retrieved from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/133/original/Topos_KF_White-Paper_Nyhan_V1.pdf

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances, 5*(1). doi:https://www.science.org/doi/10.1126/sciadv.aau4586

Guess, L. L. (2020). "Fake news" may have limited effects beyond increasing beliefs in false claims. *Harvard Kennedy School (HKS) Misinformation Review*.

Hafner-Burton, E. M., Hyde, S. D., & Jablonski, R. S. (2013). When Do Governments Resort to Election Violence? *British Journal of Political Science*. doi:https://doi.org/10.1017/S0007123412000671

Hafner-Burton, E. M., Hyde, S. D., & Jablonski, R. S. (2014, January). When Do Governments Resort to Election Violence? *British Journal of Political Science, 44*(1), 149-179. doi:https://doi.org/10.1017/S0007123412000671

Haugen, F. (2021, October 4). Whistleblower: Facebook is misleading the public on progress against hate speech, violence, misinformation. (S. Pelley, Interviewer) Retrieved from https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-misinformation-public-60-minutes-2021-10-03/

Hoglund, K. (2008). Violence in war-to-democracies transitions. In K. A. Jarstad, & T. D. Sisk, *From War to Democracy.*

Hosseinmardi, H., Ghasemian, A., Clauset, A., & Watts, D. J. (2021). Examining the consumption of radical content on YouTube. *PNAS*. doi:https://www.pnas.org/doi/10.1073/pnas.2101967118

Human Rights Council. (2011). *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework.* United Nations. Retrieved from https://documents-dds-ny.un.org/doc/UNDOC/GEN/G11/121/90/PDF/G1112190.pdf?OpenElement

Humprecht E, E. F. (2020). Resilience to Online Disinformation: A Framework for Cross-National Comparative Research. *The International Journal of Press/Politics*, 25(3):493-516.

Humprecht, E. (2018). Where 'fake news' flourishes: a comparison across four Western democracies. *Information, Communication & Society*. doi:https://doi.org/10.1080/1369118X.2018.1474241

Humprecht, E., Esser, F., & Aelst, P. V. (2020). Resilience to Online Disinformation: A Framework for Cross-National Comparative Research. *The International Journal of Press/Politics*. doi:https://doi.org/10.1177/1940161219900126

Izsák-Ndiaye, R. (2021). *If I Disappear - Global Report on Protecting Young People in Civic Space.* United Nations, Office of the Secretary-General's Envoy on Youth. Retrieved from https://www.un.org/youthenvoy/wp-content/uploads/2021/06/Global-Report-on-Protecting.-Young-People-in-Civic-Space.pdf

Jansen, S. C., & Martin, W. (2015). The Streisand Effect and Censorship Backfire. *International Journal of Communication*, 656-671. Retrieved from researchgate.net/publication/273947761_The_Streisand_Effect_and_Censorship_Backfire

Juul, J. J., & Ugander, J. (n.d.). Comparing information diffusion mechanisms by matching on cascade size. *PNAS*. doi:https://www.pnas.org/doi/10.1073/pnas.2100786118

Juul, J. L., & Ugander, J. (2021, November 8). Comparing information diffusion mechanisms by matching on cascade size. (D. Watts, Ed.) *PNAS*. doi:https://doi.org/10.1073/pnas.2100786118

Kerr, N., & Lührmann, A. (2017). Public trust in manipulated elections: The role of election administration and media freedom. *Electoral Studies*, 50-67. doi:https://doi.org/10.1016/j.electstud.2017.08.003

Khan, I. (2021). *Disinformation and freedom of opinion and expression.* Human Rights Council. Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. Retrieved from https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf?OpenElement

Khan, I. (2021). *DISINFORMATION AND FREEDOM OF OPINION AND EXPRESSION :REPORT OF THE SPECIAL RAPPORTEUR ON THE PROMOTION AND PROTECTION OF THE RIGHT TO FREEDOM OF OPINION AND EXPRESSION.* United Nations. United Nations. Retrieved from https://documents.un.org/prod/ods.nsf/xpSearchResultsM.xsp

Kimari, P. M. (2020). Fake News and the 2017 Kenyan Elections. *South African Journal for Communication Theory and Research*, 46:4, 31-49.

Knuutila, A. N.-M. (2022). Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries. *Harvard Kennedy School (HKS) Misinformation Review*.

Kubin, E., & Sikorski, C. v. (2021). The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association*. doi:https://doi.org/10.1080/23808985.2021.1976070

Kuo, R., & Marwick, A. (2021). Critical disinformation studies: History, power, and politics. *Harvard Kennedy School (HKS) Misinformation Review*. doi:https://doi.org/10.37016/mr-2020-76

Kurvers, R. H., Hertz, U., Karpus, J., Balode, M. P., Jayles, B., Binmore, K., & Bahrami, B. (2021). Strategic disinformation outperforms honesty in competition for social influence. *iScience*. doi:https://doi.org/10.1016/j.isci.2021.103505

Lee, E. (2022). Virtual Governments. doi:http://dx.doi.org/10.2139/ssrn.4028568

MacAvaney, S., Yao, H.-R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019, August 20). Hate speech detection: Challenges and solutions. *PLOS One*. doi:https://doi.org/10.1371/journal.pone.0221152

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research and Politics*. doi:https://doi.org/10.1177/2053168017720008

Microsoft. (2022). *Civility, Safety & Safety Global Report.* Retrieved from https://www.microsoft.com/en-us/online-safety/digital-civility?activetab=dci_reports%3aprimaryr3

Moore, M. (2018). Democracy Hacked: Political Turmoil and Information. London: Oneworld Publications.

Munyaka, I. N., Hargittai, E., & Redmiles, E. M. (2022). The Misinformation Paradox: Older Adults are Cynical about News Media, but Engage with It Anyway. *Journal of Online Trust and Safety*. doi:10.54501/jots.v1i4.62

Mutahi, P. (2022). *Role of social media in preventing electoral violence: Lessons learned and best practices. Presented to the European External Action Service.*

Nassetta, J., & Gross, K. (2020). State media warning labels can counteract the effects of foreign misinformation. *Misinformation Review*. doi:https://doi.org/10.37016/mr-2020-45

Nesbit, E. C., Mortenson, C., & Li, Q. (2021, March). The presumed influence of election misinformation on others reduces our own satisfaction with democracy. *Harvard Kennedy School Misinformation Review*. doi:https://doi.org/10.37016/mr-2020-59

Newman, N. (2017). *Overview and Key Findings: Digital News Report.* Retrieved from Reuters Institute for the Study of Journalism: http://www.digitalnewsreport.org/survey/2017/overview-key-findings-2017/

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Digital News Report 2022.* Reuters Institute. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-06/Digital_News-Report_2022.pdf

Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*. doi:https://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., Porter, E., Reifler, J., & Woods, T. (2020). Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior*. doi:https://doi.org/10.1007/s11109-019-09528-x

Ognyanova, K. L. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School (HKS) Misinformation Review*.

Ohlin, J. D. (2021). Election Interference: The Real Harm and the Only Solution. *Defending Democracies: Combating Foreign Election Interference in a Digital Age*. doi:https://dx.doi.org/10.2139/ssrn.3276940

Opongo, E. O., & Murithi, T. (2022). *Elections, Violence and Transitional Justice in Africa.* Routledge.

Osborn, M. (2008). Fuelling the Flames: Rumour and Politics in Kibera. *Journal of Eastern African Studies, 2*(2), 315-327. doi:https://www.tandfonline.com/doi/abs/10.1080/17531050802094836

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you. London: Viking/Penguin Press.* London: Viking/Penguin Press.

Persily, N. (2019). The Internet's Challenge to Democracy: Framing the Problem and Assessing Reforms. *Kofi Annan Commission on Elections and Democracy in the Digital Age*. doi:https://pacscenter.stanford.edu/publication/the-internets-challenge-to-democracy-framing-the-problem-and-assessing-reforms/

Rampersad, G., & Althiyabi, T. (2019). Fake news: Acceptance by demographics and culture on social media. *Journal of Information Technology & Politics, 17*(1). doi:https://doi.org/10.1080/19331681.2019.1686676

Rathje, S., Van Bavel, J. J., & Linden, S. v. (2021). Out-group animosity drives engagement on social media. *PNAS, 118*(26). doi:https://doi.org/10.1073/pnas.2024292118

Report of the Secretary-General. (2022). *Countering disinformation for the promotion and protection of human rights and fundamental freedoms.* United Nations. Retrieved from

https://undocs.org/Home/Mobile?FinalSymbol=A%2F77%2F287&Language=E&DeviceType=Desktop&LangRequested=False

Roozenbeek, J., Van Der Linden, S., & Nygren, T. (n.d.). Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School (HKS) Misinformation Review*. doi:https://doi.org/10.37016//mr-2020-008

Siddiqui, N. (2018). Who do you believe? Political parties and conspiracy theories in Pakistan. *Party Politics*. doi:https://doi.org/10.1177/1354068817749777

Sirlin, N. E. (2021). Digital literacy is associated with more discerning accuracy judgments but not sharing intentions. *Harvard Kennedy School (HKS) Misinformation Review*.

Sobieraj, S. (2020). *Credible Threat: Attacks Against Women Online and the Future of Democracy.* doi:https://doi.org/10.1093/oso/9780190089283.001.0001

Stecklow, S. (2018, August 15). *Inside Facebook's Myanmar operation - Hatebook*. Retrieved from Reuters Special Report: https://www.reuters.com/investigates/special-report/myanmar-facebook-hate

Stencel, M. (2020, May 6). *Abuse and threats come with the territory for many of the world's fact-checkers*. Retrieved from Poynter.org: https://www.poynter.org/fact-checking/2020/abuse-and-threats-come-with-the-territory-for-many-of-the-worlds-fact-checkers/

Talamanca, G. F., & Arfini, S. (2022). Through the Newsfeed Glass: Rethinking Filter Bubbles and Echo Chambers. *Philosophy & Technology*. Retrieved from https://link.springer.com/article/10.1007/s13347-021-00494-z

The Bridging Divides Initiative. (2022). *Tracking Threats and Harrassment Against Local Officials.* Princeton University. Retrieved from https://bridgingdivides.princeton.edu/sites/g/files/toruqf246/files/documents/Threats%20and%20Harassment%20Report.pdf

Tompkins, A. (2020, 10 12). *Is fact-checking effective? A critical review of what works – and what doesn't*. Retrieved from DW Akademie.

Tomz, M., & Weeks, J. L. (2020). Public Opinion and Foreign Electoral Intervention. *American Political Science Review*. doi:https://doi.org/10.1017/S0003055420000064

UN WOMEN. (2021). Guidance Note: Preventing Violence Against Women in Politics. doi:https://www.unwomen.org/en/digital-library/publications/2021/07/guidance-note-preventing-violence-against-women-in-politics

UN, OSCE, OAS, & ACHPR. (2017). *Joint Declaration on Freedom of Expression and "Fake News".* doi:https://www.osce.org/files/f/documents/6/8/302796.pdf

UNESCO. (2020). *UNESCO and IEBC Launch Communication Campaign Against Disinformation and Electoral Violence*. doi:https://en.unesco.org/news/unesco-and-iebc-launch-communication-campaign-against-disinformation-and-electoral-violence

United Nations. (1966). *International Covenant on Civil and Political Rights.*

United Nations. (1966). *International Covenant on Civil and Political Rights.* General Assembly resolution 2200A (XXI). Retrieved from https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights#article-19

United Nations. (2019). *Freedom of Expression and Elections in the Digital Age.* United Nations Special Rapporteur on the Promotion and Protection of the Right to Freedom of Expression. Retrieved from https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/ElectionsReportDigitalAge.pdf

United Nations. (2021). *Our Common Agenda.* Retrieved from https://www.un.org/en/content/common-agenda-report

United Nations Department of Political Affairs. (2016). *Policy Directive: Preventing and Mitigating Election-related Violence.* Retrieved from https://dppa.un.org/sites/default/files/ead_pd_preventing_mitigating_election-related_violence_20160601_e.pdf

United Nations High Commissioner for Human Rights. (2013). *Report of the United Nations High Commissioner for Human Rights on the expert workshops on the prohibition of incitement to*

national, racial or religious hatred. Human Rights Council. Retrieved from https://documents-dds-ny.un.org/doc/UNDOC/GEN/G13/101/48/PDF/G1310148.pdf?OpenElement

United Nations Human Rights Council. (2011). *Report of the High Commissioner forHuman Rights on the situation of human rights in Côte d'Ivoire.* Retrieved from https://digitallibrary.un.org/record/706235?ln=en

Van Zuylen-Wood, S. (2019, February 26). "Men are Scum" : Inside Facebook's War on Hate Speech. *Vanity Fair*. Retrieved from https://www.vanityfair.com/news/2019/02/men-are-scum-inside-facebook-war-on-hate-speech

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, 2028-2049. doi:https://doi.org/10.1177/1461444817712086

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*, 1146–1151. doi:https://doi.org/10.1126/science.aap9559

Vosoughi, S., Roy, D., & Arel, S. (2018). The spread of true and false news online. *Nature*, p. 1146. doi:10.1126/science.aap9559

Walker, N., & Tukachinsky, R. (2019). A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It? *Communication Research*. doi:https://doi.org/10.1177/0093650219854600

Walker, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*. doi:https://doi.org/10.1080/10584609.2019.1668894

Wall, A. (2006). Electoral management design: The international IDEA handbook.

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making.* Council of Europe. Retrieved from https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

Weeks, B. E. (2015). Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*.

Wood, T., & Porter, E. (2019). The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence. *Political Behaviour*. doi:https://doi.org/10.1007/s11109-018-9443-y

Zakrzewski, C. (2022, November 8). *Election workers brace for a torrent of threats: 'I KNOW WHERE YOU SLEEP'.* Retrieved from Washington Post: https://www.washingtonpost.com/technology/2022/11/08/election-workers-online-threats/

Zakrzewski, C., De Vynck, G., Masih, N., & Mahtani, S. (2021, October 24). How Facebook neglected the rest of the world, fueling hate speech and violence in India. *The Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/