5 Ecosystem Responses Aimed at Producers and Distributors

Chapter 5 of the report: Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression

Broadband Commission research report on ‘Freedom of Expression and Addressing Disinformation on the Internet’

Published in 2020 by International Telecommunication Union (ITU), Place des Nations, CH-1211 Geneva 20, Switzerland, and the United Nations Educational, Scientific and Cultural Organization, and United Nations Educational, Scientific and Cultural Organization (UNESCO), 7, Place de Fontenoy, 75352 Paris 07 SP, France

ISBN 978-92-3-100403-2

https://en.unesco.org/publications/balanceact

5.1 Legislative, pre-legislative, and policy responses

Authors: Trisha Meyer, Clara Hanot, Julie Posetti and Denis Teyssou

This chapter discusses legislative, pre-legislative and policy responses that originate from government actors (legislative, executive, judiciary) and which encompass regulatory intervention to tackle disinformation. These responses cover different types of regulatory action, ranging from inquiries and proposed laws through to legislation and law enforcement. They typically aim at using state power to shape the environment of the production, transmission and receiving of content, affecting either the entire circuit, or specific moments and actors within these.

Disinformation online is tackled from myriad perspectives, including through existing sets of legislation that are not specific to disinformation, but which nonetheless address some aspect of the phenomenon. This chapter cannot cover them comprehensively, but it is worth highlighting some of the means deployed in order to understand the wider legal and policy context in which disinformation-specific government responses develop. Thus the focus here is on legislation and policy strictly related to disinformation, unless it is clear that a legislative/policy measure has been expanded or repurposed to also tackle disinformation.

While institutional and individual self-regulatory approaches are major responses to disinformation, a number of State actors deem it necessary to have regulatory interventions as well. Some of these may be constraining, while others (less often) rewarding. The intention is to provide sufficient disincentives and (less often) incentives to change actors' behaviour. These responses are shaped by national/regional legal traditions, the strength of international legal and normative frameworks, and cultural sensitivities.

In the coercive dimensions of these kinds of interventions, it should be noted that laws applied to disinformation are often vague, which introduces a risk of over-blocking and censoring legitimate expression, including acts of journalism. A further issue is whether existing regulation on harmful expression (for example, on fraudulent claims to sell products) suffices, or whether new regulation is needed and how it can avoid undermining protections for legitimate freedom of expression. Related to this is whether there are effective legal provisions that, in tandem, also ensure that incitement of violent attacks on press freedom and journalism safety (including by disinformation purveyors) is prohibited.

In respect to which some regulatory interventions focus not on restraint, but rather on incentives, an issue is the extent to which there is transparency and equity as a fundamental principle of law. An example is whether there are open and fair systems for regulatory allocation of public funds towards fact-checking, counter-speech (see chapter 5.2 below), or news media, and which ensure that such spending is not abused for political purposes.

Methodology and scope

In order to identify relevant legislative, pre-legislative and policy responses related to disinformation this research has used three resources that cover a range of countries and approaches as a starting point: the Poynter "Guide to Anti-Misinformation Actions around the World" (Poynter, updated regularly[122]), the Library of Congress report *Initiatives to Counter Fake News in Selected Countries* (Library of Congress, 2019[123]) and the University of Oxford's *Report of Anti-Disinformation Initiatives* (Robinson et al., 2019).

In the analysis of these regulatory responses, the researchers have gone back to the primary sources (laws, policy documents, government press releases, websites, etc) to understand the government initiatives, to the full extent possible. If primary sources proved impossible to find, or where additional information was deemed necessary, secondary sources (news articles, academic reports, legal analyses, etc) were consulted. To be considered reliable, information gained through secondary sources needed to have been found on multiple websites. These secondary sources also led to the identification of additional disinformation-specific government responses.

Some countries propose or have passed legislation unique to disinformation. For others, the proposed amendments or legal basis for tackling disinformation are grounded in other sets of legislation, such as the penal code, civil law, electoral law or cybersecurity law. It is recognised that there are provisions pertaining to disinformation, false information, 'fake news', lies, rumours, etc. in far more sets of legislation than can be covered in one report. Cases have been included where disinformation-related (amendments to) legislation were recently proposed, passed or enforced, or a clear link to disinformation was made in the reporting, discussions and argumentation that led to the proposal, law or its enforcement. Fewer cases have seen countries engage in 'positive measures' as distinct from punitive, and these are discussed in chapter 5.2.

5.1.1 What and who do legislative responses monitor/target?

To understand the tensions and challenges of using legislative and policy responses for freedom of expression, it is worth recalling the right to freedom of opinion and expression, as found in Article 19 of the UN Declaration of Human Rights and echoed in the International Covenant on Civic and Political Rights:

> “*Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.*”

Regulatory measures seeking to constrain disinformation should be assessed in terms of the international standards that any restrictions to freedom of expression must be provided by law, be proven necessary to a legitimate purpose, and constitute the least restrictive means to pursue the aim. They should also be time-limited if justified as emergency response measures.

One way of reflecting on how speech is affected by law and policy in online environments, is by assessing responses targeting different *actors' behaviours*. Some responses seek to provide what could be understood as 'positive' measures - necessary conditions for executing the right to freedom of expression. Most measures, however, aim

[122] https://www.poynter.org/ifcn/anti-misinformation-actions/
[123] https://www.loc.gov/law/help/fake-news/index.php

to deter abusive forms of freedom of expression as defined in law, and thus produce what could be termed ‘negative’ measures.

Many of these measures are taken with the rationale of protecting citizens. On one side there are steps like data protection rules and media and information literacy policy to give people a level of basic protections and skills to participate in the online environment. At the same time, there are restrictions on expression that cause harm to others, such as incitement to hatred and violence (based on race, ethnicity, gender, or religion), defamation, Nazi propaganda (in specific settings), or harassment and threats of violence. These curbs on speech are justifiable in terms of international standards, although the Rabat Principles of the UN High Commissioner on Human Rights provide important nuance in this regard by setting a high threshold for restrictions.[124] Such constraining elements target inter alia three kinds of behaviours.

Firstly, the range of persons implicated in producing, enabling and distributing content deemed to be harmful are targeted for punishment when *they transgress speech restrictions*. A complication here is whether there can be unintended effects that violate legitimate expression which, even if false (such as in satire) or disturbing (such as in shocking), is not necessarily illegal under international standards. A second and fundamental issue is whether such measures, through design or application, are genuinely to protect the public, or rather to protect particular vested interests such as political incumbents. Additionally, there is the complication that this kind of intended constraint on speech is usually in the form of national-level restrictions that require cooperation from global internet communications companies which have become primary vectors for viral disinformation.

Secondly, competition and consumer protection rules, accompanied by sectoral rules, including phenomena such as laws on misleading advertising, provide the contours of acceptable *economic behaviour* on internet communication companies. However, as chapter 6.3 on de-monetisation responses explains, there is increased questioning within policy circles on whether current rules sufficiently deter economic profiteering from sensationalist and/or false content.

Thirdly, *technical behaviour* is steered through legally formulated cyber-policy seeking to deter use of the internet technologies for malicious intent, such as spam or coordinated information operations for disinformation purposes. Also noteworthy is increased collaboration on topics such as counter terrorism in order to share knowledge and practices among government and technical actors, within legal frameworks on terrorism.

Fourthly, regulatory interventions to channel behaviours of *political actors* include election and political campaign advertising rules.

On the side of enabling, rather than restrictive policy measures, there may be regulatory interventions to increase the availability of information as an alternative to disinformation. These can include enhanced transparency and proactive disclosure practices by *officials*, linked to access to information regimes. They may also include public funds to support *news media*, *fact-checking initiatives*, and *counter-disinformation campaigns by private or public entities*.

[124] Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. https://www.ohchr.org/EN/Issues/FreedomOpinion/Articles19-20/Pages/Index.aspx

Intergovernmental and State-based policy and legal responses to disinformation are cross-cutting and cover all types of actions. Based on the analysis above, four groups can be identified as targets of policy responses.

Firstly, *users considered to be fraudulent and abusive* are at the core of many regulatory responses from governments representing their rationale for action as not only the need to diminish incitements to hatred, violence, and defamation - but also more broadly and problematically, speech that is deemed to be 'false' that is perceived to be prejudicial to national security, international diplomacy, social order, and more (*see #8 in Table 3 below*). On the other hand, some governments also invest in support for *those presenting as ensuring information quality*: fact-checking, counter-disinformation, media and information literacy, and journalism initiatives in order to reliably inform users and empower them to detect disinformation (*see #1,2,3,9,10 Table 3 below*).

Secondly, government initiatives focusing on *internet communication companies* target their economic and technical behaviour. Based on the assumption that online platforms' algorithms enable the viral amplification of disinformation, many regulatory initiatives attempt to place greater obligations on these actors.

In lighter forms of government intervention, internet communications companies are requested to self-regulate and provide public insight into content moderation and political advertising practices and processes. In heavier forms of regulatory action, online platforms and internet intermediaries are required, formally or informally, to de-prioritise, block and take down certain types of content and websites and deregister particular users (*see #5,6,7 in Table 3 belows*).

To some extent, though not often directly targeted, the advertising industry can also be included in this category, as certain policy makers consider the online advertising business model to indirectly enable the financing of disinformation operations (*see #7 in Table 3 below*).

A third stakeholder in the scope of government responses targeting disinformation are *journalists and the news media*. Either by design or unintentionally, many regulatory responses catch journalists and news publishers in the criminalisation of publication and dissemination of false information, despite international protections for press freedom - indicating the need for caveats to shield journalists (*see #8 in Table 3*). In contrast, and as noted above, there are some interventions that have stimulated investment in independent journalism, as well as collaborations between news organisations and communities aimed to strengthen media and information literacy, and third party fact-checking initiatives (*see #1,2,10 in Table 3*), as part of recognising news media's potential role in countering disinformation.

Finally, some government responses target *political actors* (including political parties) themselves, by requiring them to meet new obligations for transparency in online political campaigning, such as the labelling of political advertising (*see #3,7 in Table 3* [125]) and/or by increasing fact-checking endeavours during election periods (*see #1,7 in Table 3*).

[125] See also chapters 4.1 and 7.1

5.1.2 Who do legislative, pre-legislative, and policy responses try to help?

State-based disinformation responses target all involved actors: end users (individuals, communities, audiences, etc), online platforms, advertisers, journalists and news organisations, politicians and political parties, and also domestic and foreign actors perceived to have malicious intent. These regulatory interventions seek to deter what they deem to be abusive forms of expression, with - in the focus of this study - relevance to disinformation, by means of policy and law. Their aim is presented as using 'negative' (i.e. constraining) measures to protect society and its right of access to information by constraining the presence of destructive and harmful disinformation. On the other hand, 'positive' (i.e. enabling) measures aim to affirm the right to freedom of expression by improving the ecosystem through programmes like Media and Information Literacy (MIL) and financial allocations to fact-checkers, media and/or counter-content. However, individual State-based interpretations of rights and responsibilities do not always align with the intent of international legal and normative frameworks designed to support freedom of expression.

'Negative' steps can restrict certain content or behaviour that authorities deem to be fraudulent or otherwise abusive in diverse ways. They focus primarily on moderating the public discourse, under the justification of minimising harm to others, to ensure public health, defence and security but also, at times, for political gain.

Interventions that restrict freedom of expression rights are a notoriously slippery slope, and thus international standards require that they must be provided for by law, be legitimate, proportionate, proven necessary, and the least restrictive means to pursue the stated objective. If they are introduced during emergency settings, they should also be limited by sunset clauses.

On the other hand, "positive" measures targeted at users are aimed, at least in part, at increasing Media and Information Literacy and empowering users via the content they access online. Similarly, they can empower and help enable the news media to investigate, verify, publish and disseminate public interest information.

With respect to the motivating factors, government actions primarily focus on encouraging other actors to tackle disinformation, but they also use the power of legal coercion against actors deemed to be active in the disinformation 'industry'. The theory of change underpinning these kinds of responses will depend on what and/or whom the targets are:

- For *users*, the assumption is that abusive speech can be curtailed through punitive measures, such as fines and arrests. Correlatively, change is expected through increasing the volume of, and access to, credible information, along with awareness-raising among citizens, and Media and Information Literacy programs designed to 'inoculate the herd' against disinformation, so that users are better able to understand and control their own content production/circulation/consumption.

- For *internet communication companies/PR and advertising industry*, the implied theory of change focuses on the role of law and policy in directly - or more often - indirectly reducing the economic and political incentive structures that fuel disinformation. This is also based on the assumption that the companies involved have an interest in thwarting actors who abuse the opportunities that the technology and contemporary business models create. In some cases, the aim is to control the information flows by ensuring that the companies make better use of technology such as AI to deal with issues at scale.

- For *journalists and news publishers*, similar to *users*, the working theory of change is that their publishing 'false' information and speech deemed to be 'abusive' (which, problematically, could capture robust critique as a product of independent journalism) can be curtailed through punitive measures, such as fines, censorship and arrests. The correlative assumption, one aligned with international human rights law, is that change can be effected through support for independent journalism, relying on the belief that the provision of factual and verifiable information shared in the public interest is a precondition for sustainable democracy and sustainable development.

- For *politicians*, the theory of change implicit in related regulatory interventions is that political campaigning, which is largely unregulated online, can be governed by new or updated rules fit for the digital environment. The scrutiny during election periods, through political advertising transparency and increased fact-checking, is considered an incentive for political candidates not to use disinformation as a communication strategy.

The extent to which such perceptions of intervention cause and outcome effect are plausible is discussed in sub-section 5.1.6 below.

5.1.3 What are the outputs of legislative, pre-legislative, and policy responses?

The outputs of state-based responses are reports from inquiries, policy documents (and commissioned research supporting policy development), bills and legislation, and published judgments. In cases where the government takes action, the output would then also include the specific measure taken, such as a fine, an arrest, a campaign aimed to counter what the authority deems as disinformation, or an internet shutdown. In positive measures, there are allocations of resources and capacity-building steps such as for Media and Information Literacy, implementation of access to information regimes, strengthening news media, etc.

5.1.4 Who are the primary actors and who funds them?

Legislative, pre-legislative, and policy work is usually funded by the States, but in some cases - like the internet communications companies - the implementation costs are carried by private entities. Examples are compliance with required transparency of political advertising. This is in line with many commercial enterprises across a range of sectors that have to comply with legislation and policies designed to protect public interests and safety as part of the costs of doing business. At the same time, States may directly finance and execute their own counter-disinformation content campaigns, or media and information literacy programmes.

A multitude of government responses across the globe are covered in this chapter 5.1 as well as in 5.2, 117 responses across 61 countries and inter-governmental organisations. While the objective has been to demonstrate a range of experiences, omissions are inevitable. Most of these policy initiatives are very recent, and many might have been subject to change and review since the time of writing. In addition, inquiries might turn into legislative proposals, legislative proposals might not be adopted, new regulations might arise, amendments might be brought forth, etc. This mapping should therefore be regarded as an evolving tool. The table below also reflects general categories. It does not drill down to more granular identification of issues such as criminalisation of disinformation within the category of legislative responses.

This chapter contains the summary of the research findings. For an entry-by-entry analysis, please refer to Appendix A.[126]

The numbers at the top of the table below resonate with the range of disinformation responses as defined in the study's overall taxonomy, showing links between the legislative/policy responses and the other responses.

1. Monitoring/Fact-checking
2. Investigative
3. National and international counter-disinformation campaigns
4. Electoral-specific
5. Curatorial
6. Technical/algorithmic
7. Economic
8. Ethical and normative
9. Educational
10. Empowerment and credibility labelling

Policy type	Policy response		1. Monitoring/Fact-checking	2. Investigative	3. National and international counter-disinformation campaigns	4. Electoral-specific	5. Curatorial	6. Technical/algorithmic	7. Economic	8. Ethical and normative	9. Educational	10. Empowerment & credibility labelling
Inquiries, task forces and guidelines	**Actor**: *ASEAN, Australia, Belgium, Brazil, Canada, COE, Denmark, Estonia, European Union, India, Indonesia, International Grand Committee, Ireland, Italy, Japan, Mexico, Netherlands, New Zealand, OAS, South Africa, Republic of Korea, Spain, Sweden, Ukraine, UK, U.S.*											
	1	ASEAN's Ministers responsible for Information joint declaration	x		x					x	x	
	2	Australia's Electoral Assurance Taskforce				x	x	x	x		x	
	3	Australia's Parliament Joint Standing Committee on Electoral Matters: Democracy and Disinformation	x			x				x	x	
	4	Belgium's expert group and participatory platform	x								x	x
	5	Brazil's Superior Electoral Court	x		x	x	x					
	6	Canada's parliamentary committee report on 'Democracy under Threat'				x	x	x	x		x	
	7	Canada's Digital Citizen Initiative			x				x	x	x	x
	8	Canada's Critical Election Incident Public Protocol			x	x				x	x	

[126] See Appendix A

Policy type	Policy response		1. Monitoring/Fact-checking	2. Investigative	3. National and international counter-disinformation campaigns	4. Electoral-specific	5. Curatorial	6. Technical/algorithmic	7. Economic	8. Ethical and normative	9. Educational	10. Empowerment & credibility labelling
	9	COE's 'Information Disorder' study	x	x			x	x	x		x	x
	10	Denmark's Elections Action Plan	x	x		x						
	11	Estonia's Cyber Defence League						x				
	12	European Union Code of Practice and Action Plan on Disinformation	x		x	x	x	x	x		x	x
	13	India's social media platforms Code of Ethics				x	x	x			x	
Inquiries, task forces and guidelines	14	Indonesia's war room and 'Stop Hoax' campaigns	x	x	x	x	x			x	x	
	15	International Grand Committee on 'Disinformation and 'Fake News''				x	x	x	x			
	16	Ireland's Interdepartmental Group on 'Security of the Electoral Process and Disinformation'				x	x				x	
	17	Italy's 'Enough-with-the-Hoaxes' campaign and 'Red Button' portal	x	x		x					x	
	18	Japan's Platform Services Study Group				x	x					
	19	Mexico's National Electoral Institute	x			x	x				x	x
	20	Netherlands' 'Stay Critical' campaign and strategy	x	x		x					x	x
	21	New Zealand's parliamentary inquiry into 2016 and 2017 elections				x	x	x	x		x	
	22	OAS' Guide on freedom of expression and disinformation during elections	x	x		x	x	x	x		x	x
	23	South Africa's Political Party Advert Repository and digital disinformation complaints mechanism				x					x	x
	24	Republic of Korea's party task force	x	x								
	25	Spain's government hybrid threats unit	x	x		x						
	26	Sweden's investigation into development of psychological defence authority									x	x
	27	Ukraine's 'Learn to Discern' initiative									x	
	28	UK's House of Commons (Digital, Culture, Media and Sport Committee) inquiry into 'Disinformation and 'Fake News''	x			x	x	x	x		x	

Policy type	Policy response		1. Monitoring/Fact-checking	2. Investigative	3. National and international counter-disinformation campaigns	4. Electoral-specific	5. Curatorial	6. Technical/algorithmic	7. Economic	8. Ethical and normative	9. Educational	10. Empowerment & credibility labelling
	29	UK's House of Commons Foreign Affairs Committee inquiry into Global Media Freedom (sub theme on disinformation)	x	x	x					x		
	30	U.S.' Senate Select Committee on Intelligence inquiry into 'Russian Active Measures Campaigns and Interference in the 2016 US Election'				x						
Legislative proposals	**Actor**: *Argentina, Chile, France, Germany, India, Ireland, Israel, Nigeria, Philippines, Republic of Korea, Sri Lanka, UK, U.S.*											
	31	Argentina's Bill to create a Commission for the Verification of Fake News	x			x	x	x			x	
	32	Chile's proposal to End Mandate of Elected Politicians due to Disinformation				x						
	33	France's Online Hate Speech proposal					x	x				
	34	Germany's Network Enforcement Act update					x			x		
	35	India's proposed amendments to IT Intermediary Guidelines				x	x					
	36	Ireland's proposal to Regulate Transparency of Online Political Advertising				x	x					
	37	Israel's Proposed Electoral Law Amendments and 'Facebook Laws'				x	x					
	38	Nigeria's Protection from Internet Falsehood and Manipulation bill						x		x		
	39	The Philippines' Anti-False Content bill										
	40	Republic of Korea's law proposals				x	x			x		
	41	Sri Lanka's proposed penal code amendments								x		
	42	UK's Online Harms White Paper				x	x					
	43	U.S.' Tennessee State Legislature bill to register CNN and *The Washington Post* as "fake news" agents of the Democratic Party			x							x

Policy type	Policy response	1. Monitoring/Fact-checking	2. Investigative	3. National and international counter-disinformation campaigns	4. Electoral-specific	5. Curatorial	6. Technical/algorithmic	7. Economic	8. Ethical and normative	9. Educational	10. Empowerment & credibility labelling
Adopted legislation	**Actor:** *Argentina, Bangladesh, Belarus, Benin, Brazil, Burkina Faso, Cambodia, Cameroon, Canada, China, Côte d'Ivoire, Egypt, Ethiopia, France, Germany, Indonesia, Kazakhstan, Kenya, Malaysia, Myanmar, New Zealand, Oman, Pakistan, Philippines, Russian Federation, Singapore, Thailand, Vietnam*										
	44 Argentina's Political Party Financing Law				x			x		x	x
	45 Bangladesh's Digital Security Act	x				x	x		x		
	46 Belarus' Media Law						x		x		
	47 Benin's Digital Code								x		
	48 Brazil's Criminal Electoral Disinformation Law				x						
	49 Burkina Faso's Penal Code					x	x		x		
	50 Cambodia's Anti-Fake News Directives					x	x		x		
	51 Cameroon's Penal Code and Cyber Security and Cyber Criminality Law								x		
	52 Canada's Elections Modernisation Act				x	x					
	53 China's Anti-Rumour Laws					x	x		x		
	54 Côte d'Ivoire's Penal Code and Press Law								x		
	55 Egypt's Anti-Fake News Laws					x	x		x		
	56 Ethiopia's False Information Law								x		
	57 France's Fight against Manipulation of Information Law				x	x	x				
	58 Germany's Act to Improve Enforcement of the Law in Social Networks					x	x				
	59 Indonesia's Electronic Information and Transactions Law					x	x		x		
	60 Kazakhstan's Penal Code								x		
	61 Kenya's Computer Misuse and Cybercrimes Act								x		
	62 Malaysia's Anti-Fake News (Repeal) Act								x		
	63 Myanmar's Telecommunications Law and Penal Code								x		
	64 New Zealand's Electoral Amendment Act				x			x		x	
	65 Oman's Penal Code					x	x		x		
	66 Pakistan's Prevention of Electronic Crimes Act					x	x		x		
	67 The Philippines' Penal Code					x	x		x		

Policy type		Policy response	1. Monitoring/Fact-checking	2. Investigative	3. National and international counter-disinformation campaigns	4. Electoral-specific	5. Curatorial	6. Technical/algorithmic	7. Economic	8. Ethical and normative	9. Educational	10. Empowerment & credibility labelling
	68	Russian Federation's Fake News Amendments to Information Law and Code on Administrative Violations					x	x		x		
	69	Singapore's Protection from Online Falsehoods and Manipulation Act					x	x	x			
	70	Thailand's Computer Crime Act					x	x		x		
	71	Vietnam's Cyber Security Law					x	x		x		
Law enforcement and other state intervention		**Actor**: *Bahrain, Bangladesh, Benin, Cambodia, Cameroon, China, Côte d'Ivoire, Egypt, Germany, India, Indonesia, Kazakhstan, Latvia, Malaysia, Myanmar, Russian Federation, Singapore, Sri Lanka, Thailand, Ukraine*										
	72	Bahrain								x		
	73	Bangladesh				x		x				
	74	Benin								x		
	75	Cambodia								x		
	76	Cameroon								x		
	77	PR China					x	x		x		
	78	Côte d'Ivoire								x		
	79	Egypt						x		x		
	80	Germany					x					
	81	India						x				
	82	Indonesia					x	x		x		
	83	Kazakhstan					x	x		x		
	84	Latvia						x				
	85	Malaysia								x		
	86	Myanmar								x		
	87	Russian Federation						x				
	88	Singapore					x					
	89	Sri Lanka						x				
	90	Thailand								x		
	91	Ukraine						x				

Table 3. *Legislative, pre-legislative, and policy responses (mapped against study taxonomy)*

Inquiries, task forces and guidelines

With widespread misinformation and disinformation becoming a growing concern, several countries have set up dedicated task forces and inquiries to monitor and investigate disinformation campaigns. Such task forces have often been launched following disinformation campaigns perceived as a hybrid threat to the country's democratic integrity, or cyber-security. An additional aim of these governmental initiatives is educational, with many including a media and information literacy aspect (*see #9 in Table 3*), such as the Netherlands' 'Stay Critical' strategy (entry 20. in Appendix A) or Canada's Digital Citizen Initiative (entry 7. in Appendix A). In addition, 17 initiatives in this category include fact-checking (*see #1 in Table 3*). It can be highlighted that out of the 30 countries which have set up such inquiries or task forces, 21 have an electoral-specific focus (*see #4 in Table 3*), including a U.S. Senate Select Committee on Intelligence inquiry into interference in the 2016 U.S. Election (entry 30. in Appendix A), Australia's Electoral Assurance Taskforce (entry 2. in Appendix A) and Mexico's National Electoral Institute (entry 19. in Appendix A). Electoral-specific inquiries have the objective to investigate or prevent interference in legislative processes. Because online disinformation is a relatively new phenomenon, most of the initiatives identified are recent and still susceptible to evolution, including as regulatory initiatives.

Legislative proposals

A majority of recent legislative proposals (8 out of 13 analysed) aim to tackle disinformation through curation and the prism of intermediary liability obligations for online platforms regarding misinformation/disinformation or hate speech (*see #5 in Table 3*). This is particularly the scope of France's Fight Against Online Hate Speech Law proposal (entry 37. in Appendix A), Ireland's Proposal to Regulate Transparency of Online Political Advertising (entry 39. in Appendix A) and Israel's Proposed Electoral Law Amendments and 'Facebook Laws' (entry 40. in Appendix A). Similar to inquiries and task forces, the legislative proposals sometimes have an electoral-specific focus (*see #4 in Table 3*), such as Chile's Proposal to End Mandate of Elected Politicians Due to Disinformation (entry 35. in Appendix A). Some other legislative proposals would criminalise the action of spreading disinformation (*see #8 in Table 3*). This can lead to a risk, highlighted on several occasions by human rights activists, of it being used against critical independent journalists.

Adopted legislation

According to this research, by March 2020, at least 28 countries had passed legislation related to disinformation, either updating existing regulations or passing new legislation. The scope of the established legislation varies from media and electoral laws to cybersecurity and penal codes. The regulations either target the perpetrators (particularly individuals and media entities) of what the authorities deem to be disinformation or shift the responsibility to the internet communication companies to moderate or remove specific content, such as the German Network Enforcement Act (entry 61. in Appendix A). In some cases, in particular where disinformation is defined broadly or where provisions are included in general penal codes, there is a major risk of censorship.

Law enforcement and other state intervention

By enforcement of existing or recently adopted laws, a number of State interventions have been justified on the grounds of limiting disinformation. Such actions can consist of fines, arrests or internet and website shutdowns. Enforcement targets individuals, and sometimes journalists and activists; foreign state media considered as disseminating disinformation (for example Latvia's shutdown of a website linked to another government (entry 88. in Appendix A)); or the internet communication companies deemed as responsible for the massive reach of disinformation (see Facebook fines in Germany (entry 84. in Appendix A)). A number of arrests have been pointed out by Human Rights organisations as arbitrary, and as harnessing disinformation to limit free speech. Internet shutdowns have also been observed to have been used by some governments under a professed rationale of preventing the spread of disinformation, despite such restrictions being blunt (over/under-inclusive) measures that limit access to the full range of information that a society would otherwise enjoy.

5.1.5 How are these responses evaluated?

Many of the impacts of disinformation can be hard to measure comprehensively, and the effectiveness of laws drafted to tackle disinformation are similarly difficult to evaluate. Nonetheless, one example is metrics of action taken by companies against online disinformation (flagging, review, filtering, blocking) which can be considered as criteria for evaluating the application of the law. For example, as a means of evaluating the implementation of the German Network Enforcement Act, platforms report every six months on the action taken on content flagged by users. Partially on this basis, the German government has now proposed updates to the law due assessing the reports having underreported the number of complaints received (Pollock, 2019). A second text revising the initial Network Enforcement Act was expected to be on the table in mid 2020, focusing on the complaint management of the platforms (German BMJV, 2020a; German BMJV, 2020b) (entries 37, 61 and 84. in Appendix A).

It has also become clear that certain laws are difficult to enforce in practice. For example, after the adoption of the French Fight Against Manipulation of Information Law (entry 60. in Appendix A), stakeholders and political candidates sought to demonstrate the limitations of this law. In addition, Twitter initially blocked an official communication campaign from the government to encourage people to vote, arguing it was complying with the law (LeFigaro, 2019). For many small countries worldwide, it is hard in practice to apply laws to international services which do not have significant business or physical presence within the national jurisdiction.

Governments, parliaments and courts can evaluate, and if necessary, revisit and amend existing legislation and policy. For example, the constitutionality of the 2018 Kenya Computer Misuse and Cybercrimes Act has been challenged in court and a judgment was expected in early 2020 (entry 64. in Appendix A). In 2018 Malaysia passed an Anti-Fake News Act. However, after a change of government, the law was repealed on the basis that existing laws (Penal Code, Sedition Act, Printing Presses and Publications Act, Communications and Multimedia Act) already tackle disinformation (entry 65. in Appendix A).

Non-State actors can exert pressure for policy change by publishing their own evaluations and positions on regulatory initiatives. Many civil society groups do in fact provide some evaluations, as do UN organisations such as UN Special Rapporteur on Freedom of Opinion and Expression.[127]

[127] https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/LegislationAndPolicy.aspx

5.1.6 Response case study: COVID-19 disinformation

The COVID-19 pandemic triggered a flurry of state-based actions to prevent and punish acts of potentially life-threatening disinformation (Posetti & Bontcheva, 2020a).[128] Around the world, parliaments, governments and regulators amended or passed laws or regulations enabling the prosecution of people for producing or circulating disinformation, with custodial sentences ranging up to five years (Quinn, 2020). These laws effectively criminalised acts of producing or sharing information deemed to be false, misleading and/or contradicting official government communications about COVID-19. Emergency decrees giving political leaders sweeping new powers were among these measures, along with the application of existing emergency acts to COVID-19 disinformation to enable arrests, fines and jail time for associated offences, such as in South Africa (South African Government, 2020). For example, in January 2020, the Malaysian Communications and Multimedia Commission (2020) detained four individuals suspected of spreading false news on the Coronavirus under Section 233 of the Communications and Multimedia Act.

These measures carried with them the risk of catching legitimate journalism in the net (UK Delegation to the OSCE, 2020). In some countries, producers of independent journalism were arrested and detained, or deported under these laws in the context of States responding to what they deemed to be false information (Simon, 2020; Eljechtimi, 2020). Freedom of expression rights were also affected more broadly due to the challenges of introducing emergency measures in ways that urgently address public health and safety threats, as well as cases of restricting access to official information. Limitations were often not justified, nor in line with the criteria of being legal, necessary, time-limited, and proportionate to the purpose.

Other kinds of policy responses have included support for news media as a bulwark against disinformation. In light of the negative impact of the crisis on the media sector (Tracy, 2020), along with recognition of the corresponding social value of maintaining news outlets, a number of countries took such action.

For example:

- Canada fast-tracked tax relief for media outlets, and put money into advertising specifically to be carried by news outlets (Canadian Heritage, 2020)

- State aid packages or tax exemptions to support news media and media employers were offered in Denmark, Belgium, Hungary and Italy (UNI Global Union, 2020).

- There were mounting calls (Aaron, 2020) for this kind of policy response, qualified by insistence on ensuring transparency, impartiality and independence of any such support mechanisms. Assistance for public service media was also being advocated (Public Media Alliance, 2020).

- A number of NGOs dedicated funds for COVID-19 coverage with state support (UK Government, 2020)

[128] See also the databases of freedom of expression abuses connected to COVID-19 disinformation responses (e.g. 'fake news laws') curated by the International Press Institute (IPI) https://ipi.media/covid19-media-freedom-monitoring/ and Index on Censorship https://www.indexoncensorship.org/disease-control/

5.1.7 Challenges and opportunities

The pace of technological change is a fundamental challenge, as every regulatory action can be quickly outpaced. Broad language can get around this challenge, but at the expense of allowing for interpretations for selective implementation and excessive scope, or for other actors to find loopholes to avoid compliance.

A further challenge is that while there are advantages to dealing with disinformation at the national level, where government initiatives are tailored for a specific political and social context, this does not apply at various supranational levels. This is particularly the case for measures targeting internet communications companies that operate globally. At the same time, it can be difficult for global actors to properly enforce divergent national regulation in the context of networked international information flows.

Some of the measures described in this chapter consist of updating existing legislation to diminish abuses of free expression, and to regulate elections, in order to limit the impact of disinformation on the ability of voters to make informed decisions. Where existing legislation includes protection of freedom of expression and democratic participation, updating or adapting these laws to ensure they can be applied to online disinformation may prevent rushed legislation that does not respect international human rights standards.

Under public and time pressure, legislation is often passed without sufficient debate or transparency, especially in the run-up to elections and in the context of major public health crises like the COVID-19 pandemic. It is noteworthy that some proposed and adopted legislation has been challenged in court, while other bills and acts have been amended or withdrawn in response to such developments.

Moreover, while some governments attempt in good faith to update the regulatory environment to tackle disinformation in the digital age, others have been seen to attempt to control citizens' speech by creating new illegal speech categories, or extending existing laws to penalise legitimate speech. The paradox to highlight here, is that governments that appear to be seeking to control speech for political gain try to legitimise their actions by referring to hate speech regulations and anti-disinformation laws. In other words, disinformation responses risk being used (or justified for use) for censoring legitimate expression - and clearing the field for official disinformation to spread unchecked.

This concern has been increasingly raised by human rights organisations around the world, pointing that such laws have led to abusive arrests of journalists and activists (Human Rights Watch, 2018b; Posetti & Bontcheva, 2020a; Posetti & Bontcheva, 2020b). However, while regulating against disinformation in tandem with safeguarding internationally enshrined rights to freedom of expression can be challenging, there are also opportunities to be noted. For example, when critical independent journalism is empowered as a response to disinformation, governments and private companies can be more effectively held accountable, and policy action can be evaluated and changed as appropriate.

Many legislative and policy responses push responsibility for action onto internet communication companies (especially the big global players), and hold them accountable for the widespread diffusion of disinformation online. But this is sometimes done with insufficient debate and transparency regarding the way measures are then implemented by the companies, and how inevitable risks might be mitigated. Private companies are increasingly required to implement government policy on disinformation, and in essence determine in their implementation the contours of acceptable and unacceptable speech, often with insufficient possibilities of redress for users.

An opportunity is to counter-balance restrictive approaches with enabling measures. Rather than create new expression-based crimes, or to restrict internet access, there are legislative and policy responses which help ensure that information rather than disinformation predominates online. In all cases, there is potential to mainstream assessments of impact on human rights, and on legitimate forms of expression in particular. This covers proposing, passing and implementing state-based responses to digital age manifestations of disinformation.

5.1.8 Recommendations for legislative, pre-legislative, and policy responses

Drawing on the research-based assessment of legislative, pre-legislative and policy responses to disinformation outlined above (and in the accompanying appendix) the following recommendations for action are presented for the consideration of individual States, which could:

- Review and adapt responses to disinformation with a view to conformity with international human rights standards (notably freedom of expression, including access to information, and privacy rights), and make provision for monitoring and evaluation.
- Develop mechanisms for independent oversight and evaluation of the efficacy of relevant legislation, policy and regulation.
- Develop mechanisms for independent oversight and evaluation of internet communication companies' practices in fulfilling legal mandates in tackling disinformation.
- Avoid criminalising disinformation to ensure that legitimate journalism and other public interest information are not caught in the nets of 'fake news' laws.
- Avoid internet shutdowns and social media restrictions as mechanisms to tackle disinformation.
- Ensure that any legislation responding to disinformation crises, like the COVID-19 disinfodemic, is necessary, proportionate, and time-limited.
- Support investment in strengthening independent media, including community and public service media, in the context of the economic impacts of the COVID-19 crisis threatening journalistic sustainability around the world.

5.2 National and international counter-disinformation campaigns

Authors: Trisha Meyer, Clara Hanot and Julie Posetti

This chapter highlights examples of State-based and intergovernmental initiatives aimed at the construction of counter-disinformation narratives, which provide factual information to refute the falsehoods embedded within disinformation narratives. It also discusses whether refutation is an effective disinformation response, based on the latest scientific studies on this topic. Government-run counter-disinformation campaigns have the potential to increase trust and transparency in authorities when they are transparent and serve to enhance dialogue with citizens. An inherent danger, however, is that these mechanisms constitute unidirectional strategic communications initiatives that serve incumbent political interests and also do not address some of the underlying causes of disinformation which would require policies beyond the informational level (such as economic development for marginalised groups or areas). By extension, some counter-disinformation initiatives can risk deepening partisan divides.

5.2.1 What and whom do these responses monitor/target?

Counter-disinformation initiatives launched by national and international authorities target both foreign and domestic disinformation campaigns. While some initiatives do not target a defined type of disinformation, others have a specific focus, such as the European Union External Action Service East Stratcom Task Force which primarily monitors disinformation that it assesses as coming from within countries outside the EU. Some debunking initiatives are actively set up for electoral periods, such as the website led by the Brazil Superior Electoral Court in the run up to the 2018 general elections. Disinformation related to health issues is also a concern that has prompted many dedicated counter-disinformation initiatives, particularly with the COVID-19 crisis.

5.2.2 Who do these responses try to help?

Many of these campaigns and initiatives focus on informing the general public about identified disinformation claims, such as in an electoral context, on a range of policy, natural disasters, and public health and safety concerns amongst others. In addition, the international outreach of such counter-disinformation campaigns can also be designed to preserve or improve the public perception of a country and its government (or regional bloc) on the international scene. These initiatives range from public diplomacy to propaganda. Some of this work provides analysis to military actors, such as the work conducted by NATO StratCom Center of Excellence, which both publishes reports and supports NATO's strategic communications capabilities.

The motivation behind counter-disinformation campaigns is based on refutation. The underlying assumption of those states launching anti-disinformation campaigns, is that debunking and providing accurate factual information to the public will mitigate the belief in, and influence of, non-factual information. There is also the intention to raise public scepticism based on the provenance of particular messages. A number of these initiatives also go beyond issues of factuality to present narratives and facts in a different light, often thereby hoping to exert geopolitical influence.

5.2.3 What are the outputs of national and international counter-disinformation campaigns?

The work of counter-disinformation initiatives mainly consists of fact-checking activities and dissemination of what is officially considered as authoritative information. The verification is presented online and shared on social media in an attempt to reach the audience on the same platforms where they might encounter disinformation. The debunking can also be directly presented on social media channels, such as the Pakistani @FakeNews_Buster. Such monitoring work might also be presented in reports and extensive analysis to feed strategic communication efforts, such as the work done by the NATO StratCom Center of Excellence and the EEAS East Stratcom Task Force. Additionally, it might be shared with the news media for coverage.

5.2.4 Who are the primary actors and who funds these responses?

The initiatives presented below, collected through research up until May 2020, emanate from governments or international organisations and are thus publicly funded by authorities.

Counter-disinformation campaigns	**Actor:** *ASEAN, Brazil, Cambodia, Canada, China, Democratic Republic of Congo,* **EU/***EEAS, India, Indonesia, Malaysia, Mexico, NATO, Oman, Pakistan, Russian Federation, South Africa, Thailand, Tunisia, UK, UN, UNESCO, WHO*	
	1	Brazil's Superior Electoral Court
	2	Cambodia's TV programme
	3	Canada's programme of activities under the Digital Citizen Initiative
	4	China's Piyao government platform
	5	Democratic Republic of Congo's Ebola mis/disinformation response
	6	European Union EEAS East Stratcom Task Force
	7	India's Army information warfare branch
	8	India's Ministry of Information and Broadcasting FACT check module
	9	Indonesia's CEKHOAKS! debunking portal
	10	Malaysia's Sebenarnya.my debunking portal
	11	Mexico's Verificado Notimex website
	12	NATO StratCom Centre of Excellence
	13	Oman's government communications
	14	Pakistan's FakeNews_Buster' Twitter handle
	15	Russian Federation's Ministry of Foreign Affairs debunking page
	16	Thailand's Anti-Fake News Centre

	17	Tunisia’s Check News website
	18	UK Foreign and Commonwealth Office Counter Disinformation and Media Development programme
	19	WHO Coronarvirus Mythbusters campaign*
	20	UN Communications Response (COVID-19)*
	21	UNESCO coronavirus disinformation campaigns*
	22	ASEAN partnership to combat coronavirus disinformation*
	23	EU COVID-19 mythbusting campaign*
	24	South Africa’s COVID-19 landing page campaign*
	25	India’s WhatsApp coronavirus counter-disinformation* campaign
	26	UK Government’s COVID-19 disinformation rapid response unit*

*These initiatives are detailed in the coronavirus case study below

Table 4. *National and international counter-disinformation campaigns*

1. Brazil Superior Electoral Court (2018)

The Brazilian Superior Electoral Court (TSE) launched its own fact-checking and counter-disinformation website (Brazil Superior Electoral Court, 2018)[129] in the run-up to the general elections in October 2018. Reports of disinformation brought to its attention were passed on to the Public Prosecutor’s Office and the Federal Police for verification.

2. Cambodia TV Programme (2019)

In early 2019 the Cambodian Ministry of Information launched a weekly live TV programme on the National Television of Kampuchea to counter what it deems to be disinformation (Dara, 2019).

3. Canada Programme of Activities under the Digital Digital Citizens’ Initiative (2019)

In 2019, Canada funded a series of initiatives designed to raise awareness about the problem of disinformation and build capacity to combat the problem within broad publics (Canada Government, 2019c).

4. China Piyao Government Platform (2018-)

The Chinese government launched the Piyao (‘Refuting Rumours’) platform[130], hosted by the Central Cyberspace Affairs Commission in affiliation with the official Xinhua news agency, in August 2018. The platform encourages citizens to report disinformation and uses artificial intelligence to identify rumours. It also distributes state-approved news and counter-disinformation. The platform centralises the efforts of Chinese government agencies to refute what they deem to be disinformation (Qiu & Woo, 2018).

[129] http://www.tse.jus.br/hotsites/esclarecimentos-informacoes-falsas-eleicoes-2018/

[130] http://www.piyao.org.cn

5. Democratic Republic of Congo Ebola Mis/Disinformation Response (2018-)

In response to the spread of rumours and mis/disinformation about the Ebola virus in the Democratic Republic of Congo, health organisations (WHO, UNICEF, IFRC) collaborated to maintain a database of rumours spread within communities and via social media channels. Because disinformation can complicate the work of medical staff on the ground, the WHO provided fact-checking and risk communication advice for volunteers and frontline personnel in the context of the Ebola epidemic (WHO, 2018). The DRC Ministry of Health also recruited people to report mis/disinformation spread on WhatsApp. These monitoring efforts aim to develop the most appropriate strategy for responding and refuting in person, by radio and via WhatsApp (Spinney, 2019; Fidler, 2019).

6. European Union External Action Service East Stratcom Task Force (2015-)

In March 2015, the European Council tasked the High Representative in cooperation with EU institutions and Member States to submit an action plan on strategic communication. As part of the objective to better forecast, address and respond to disinformation activities by external actors, the task force was set up as part of the European External Action Service to address what it perceived as foreign disinformation campaigns.

For this objective, a small team within the EEAS was recruited to develop what it regarded as positive messages on the European Union in the Eastern Neighbourhood countries. It was also tasked to support the media environment in this region. Finally, the task force analysed the disinformation trend and exposed disinformation narratives, which it saw as emanating mainly from sources outside of the EU. The task force's work on disinformation can be found on their website (euvsdisinfo.eu)[131]. They also operate a Russian language website (eeas.europa.eu/ru/eu-information-russian_ru)[132] to communicate the EU's activities in the Eastern Neighbourhood (EU EEAS, 2018).

The EEAS Stratcom Task Force also operates a 'Rapid Alert System' between the EU Member States, launched in March 2019 as an element of the EU Code of Practice on Disinformation. The mechanism has been first put into use in the context of the Coronavirus crisis (Stolton, 2020).

7. India Army Information Warfare Branch (2019)

The Indian Defence Ministry approved the creation of an Information Warfare branch within the Army to counter what it deems to be disinformation and propaganda in March 2019 (Karanbir Gurung, 2019).

8. India Ministry of Information and Broadcasting FACT Check Module (2019)

Later, in November 2019 the Indian Government announced the creation of a FACT Check Module within the Ministry of Information and Broadcasting. The team will "work on the four principles of find, assess, create and target (FACT)" and will also report disinformation to the relevant government ministries (Mathur, 2019).

[131] http://euvsdisinfo.eu/

[132] https://eeas.europa.eu/ru/eu-information-russian_ru

9. Indonesia CEKHOAKS! Debunking Portal (2019-)

The Indonesian debunking portal 'CEKHOAKS!'[133] allows citizens to flag disinformation and hoaxes, as well as check which content has been debunked. This website is supported by the Indonesian ministry of Communication and Information Telecommunication, the Indonesian Anti-Slander Society, as well as other government agencies and other civil society organisations.

10. Malaysia Sebenarnya.my Debunking Portal (2017-)

The Malaysian Communications and Multimedia Commission set up a debunking portal 'sebenarnya.my'[134] in March 2017 and accompanying app in March 2018 in order to raise awareness and curb the spread of online disinformation (Buchanan, 2019).

11. Mexico Verificado Notimex Website (2019-)

In June 2019, Notimex, the news agency of the Mexican government, launched its own fact-checking and counter-disinformation website 'Verificado NTX'.[135]

12. NATO StratCom Center of Excellence (2014-)

Based in Riga, Latvia, the NATO Strategic Communication Center of Excellence is a NATO-accredited organisation, formed in 2014 by a memorandum of understanding between Estonia, Germany, Italy, Latvia, Lithuania, Poland, and the United Kingdom. It is independent of the NATO command structure and does not speak for NATO. The Netherlands and Finland joined in 2016, Sweden in 2017, Canada in 2018 and Slovakia in early 2019. France and Denmark were set to join in 2020. The centre analyses disinformation and provides support to NATO's strategic communications capabilities (NATO Stratcom COE, 2019).

13. Oman Government Communications (2018)

The Omani Centre for Government Communications provided training to help the media and communication departments within government institutions to monitor and refute disinformation (Al Busaidi, 2019).

14. Pakistan 'FakeNews_Buster' Twitter Handle (2018-)

The Pakistani Ministry of Information and Broadcasting launched a Twitter handle (@FakeNews_Buster)[136] to raise awareness and refute what it deems as disinformation in October 2018 (Dawn, 2018). A recurring tweet states that "[d]isseminating #FakeNews is not only unethical and illegal but it is also a disservice to the nation. It is the responsibility of everyone to reject such irresponsible behavior. Reject #FakeNews" (@FakeNews_Buster).

[133] https://stophoax.id

[134] https://sebenarnya.my/

[135] http://verificado.notimex.gob.mx

[136] https://twitter.com/FakeNews_Buster

15. The Debunking Page of the Ministry of Foreign Affairs of the Russian Federation (2017-)

The Ministry of Foreign Affairs of the Russian Federation has a dedicated webpage to raise awareness of published materials that contain information about the Russian Federation that is deemed to be false.[137] Following the passing of Amendments to the Information Law in 2019, the Russian Federation's media regulator Roskomnadzor was also expected to set up a "fake news database" (Zharov, 2019).

16. Thailand Anti-Fake News Center (2019-)

The Thai Digital Economy and Society Minister set up an intergovernmental 'Anti-Fake News Center' in October 2019 to monitor and refute disinformation, defined as "any viral online content that misleads people or damages the country's image" (Tanakasempipat, 2019b). In coordination with relevant authorities, correction notices are published through the centre's social media accounts, website (antifakenewscenter.com)[138] and the press. The Center has also issued arrest warrants (Bangkok Post, 2019).

17. Tunisia Check News Website (2019-)

A Tunisian fact-checking and debunking website (tunisiachecknews.com)[139] was launched in October 2019. The Tunisian High Independent Authority for Audiovisual Communication (HAICA) supervises the project and works in close collaboration with journalists from public media houses (national television, national radio and Agence Tunis Afrique Presse).

18. UK Foreign and Commonwealth Office Counter Disinformation and Media Development Programme (2016-2021)

In April 2018, in the context of disinformation around the Salisbury poisoning incident (Symonds, 2018), the Foreign and Commonwealth Office (FCO), together with the Ministry of Defence (MoD), Department for Culture, Media and Sport (DCMS) and the Cabinet Office, launched a programme on 'Counter Disinformation and Media Development'. This project is part of a broader set of 'Conflict, Stability and Security Fund programmes in the Eastern Europe, Central Asia and the Western Balkans region'.

The programme provides financial and mentoring support to organisations with the objective to "enhance the quality of public service and independent media (including in the Russian language) so that it is able to support social cohesion, uphold universal values and provide communities in countries across Eastern Europe with access to reliable information." By supporting civil society efforts to expose disinformation, it says that it expects to strengthen society's resilience in Europe.[140]

[137] https://www.mid.ru/en/nedostovernie-publikacii

[138] https://www.antifakenewscenter.com/

[139] https://tunisiachecknews.com/

[140] EU Disinfolab responsible for drafting this chapter 5.2 is grantee of the Foreign and Commonwealth Office Counter Disinformation and Media Development programme.

5.2.5 Response Case Study: COVID-19 Disinformation

Counter-disinformation campaigns have been strong elements of both State-based and intergovernmental responses to COVID-19 disinformation. They were rolled out quickly to mobilise online communities to help spread official public health information, as well as debunk content deemed to be false. Partnerships have been forged between various internet communications companies and authorities to provide interactive channels for official content. Measures in this category include campaigns and the creation of special units charged with producing content to counter disinformation.

Examples of these response types deployed to counter COVID-19 disinformation include:

- World Health Organisation mythbusting: In a press conference, a World Health Organisation official declared that "[w]e need a vaccine against disinformation" (WHO, 2020). After the outbreak of the COVID-19 epidemic, WHO set up an official 'Myth Buster's' page[141] to provide reliable information on the disease, as well as an 'EPI-WIN' website.[142] (*This initiative is 19. in the table above*)

- The UN Secretary General launched a UN Communications Response initiative "to flood the internet with facts and science", while countering the growing scourge of misinformation, which he describes as "a poison that is putting even more lives at risk" (UN News, 2020; UN Department of Global Communications, 2020). In May, the initiative was rolled out as "Verified", with the aim being to create a cadre of "digital first responders" to increase the volume and reach of trusted, accurate information surrounding the crisis.[143] (*This initiative is 20. in the table above*)

- UNESCO published two policy briefs deciphering and dissecting the 'disinfodemic' (Posetti & Bontcheva, 2020a; Posetti & Bontcheva, 2020b) which formed part of a broader campaign to counter disinformation and influence policy development at the individual State level. It also produced content in local languages under the rubric of "misinformation shredder".[144] (*This initiative is 21. in the table above*). UNESCO also operated a global campaign called FACTS during the commemorations of World Press Freedom Day on 3 May 2020, audio content was produced in numerous languages for radio stations worldwide, and subsequently launched a further initiative titled Don't Go Viral.[145]

- The UN Global Pulse teams in New York, Kampala and Indonesia are building situational awareness around the outbreak, emergence, and spread of 'infodemics' that can drive efforts across all pillars of the UN, and analytics that identify successful efforts to increase the reach and impact of correct public health information.[146] To this end, they are creating and scaling analytics tools, methodologies, and frameworks to support UN entities to better understand the operational contexts in which they counter the negative effects of COVID-19 in Africa. Based on scientific methodologies, direct support to WHO Africa focuses on providing analytical support and products based on the following methodologies: 1) Short term qualitative and quantitative analysis of digital signals

[141] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters
[142] https://www.epi-win.com/advice-and-information/myth-busters
[143] https://www.shareverified.com/en; https://news.un.org/en/story/2020/05/1064622
[144] https://en.unesco.org/news/faq-covid-19-and-misinformation-shredder-african-local-languages
[145] https://en.unesco.org/commemorations/worldpressfreedomday/facts-campaign; https://en.unesco.org/covid19/communicationinformationresponse/dontgoviral; https://en.unesco.org/covid19/communicationinformationresponse/audioresources
[146] https://www.unglobalpulse.org/project/understanding-the-covid-19-pandemic-in-real-time/

based on rumours and misinformation provided by field offices; 2) Continuous monitoring based on an adaptive taxonomy which allows identification of rapidly evolving 'infodemics' as well as quantitative evaluation of temporal evolution of particular topics. This includes predictive analytics of rumours and concepts along the lines of size, geographic and channel reach; 3) Sentiment and emotion analysis around particular concepts, including the appearance and escalation of hate speech. This will allow the teams to develop a framework for optimizing the messaging provided by WHO and partners to counter the disinformation.

- The Foreign Ministers of ASEAN and the People's Republic of China met to coordinate their action against COVID-19. In particular, the ministers agreed to strengthen their cooperation in risk communication "to ensure that people are rightly and thoroughly informed on COVID-19 and are not being misled by misinformation and 'fake news' pertaining to COVID-19" (ASEAN, 2020). It has not been precisely described how this cooperation would work in practice. (*This initiative is 22. in the table above*)

- The European Parliament has published guidance on dealing with COVID-19 myths.[147] (*This initiative is 23. in the table above*)

- The South African government has regulated that all internet sites operating within zaDNA top-level domain name must have a landing page with a visible link to www.sacoronavirus.co.za (national COVID-19 site).[148] (*This initiative is 24. in the table above*)

- The Indian Government launched a WhatsApp chatbot designed to counter COVID-19 related disinformation (Chaturvedi 2020). (*This initiative is 25. in the table above*)

- The UK Department for Digital, Culture, Media and Sport (DCMS) set up a dedicated unit to monitor and respond to disinformation on the pandemic, with regular engagement with the internet communications companies (Sabbagh, 2020). This initiative included a 'rapid response unit' which is designed to "stem the spread of falsehoods and rumours which could cost lives" (UK Parliament, Sub Committee on Online Harms and Disinformation, 2020). To complement this effort, the Department of International Development (DFID) supported an initiative to limit the spread of disinformation related to the disease, particularly in South East Asia and Africa. The programme focused on verifying information with help from partner media organisations, such as BBC and sharing reliable news, with help from several selected influencers (UK Department for International Development, 2020). (*This initiative is 26. in the table above*)

5.2.6 How are national and international counter-disinformation responses evaluated?

As many of the initiatives presented in this chapter are quite recent, there is little evidence of meaningful evaluation. At the same time, it appears that the initiatives have also not explicitly embedded monitoring and evaluation activities in their plans which would entail assessment of their intended (and unintended) impact and effectiveness. As they

[147] https://www.europarl.europa.eu/news/en/headlines/society/20200326STO75917/disinformation-how-to-recognise-and-tackle-covid-19-myths

[148] https://sacoronavirus.co.za

are publicly funded by governments, it is to be presumed that their effectiveness may be assessed internally by governmental services, or externally by civil society and media organisations seeking accountability and transparency. In the context of international operations, the initiatives are evaluated by the Member States that support them. For instance, the April 2018 Foreign Affairs Council (EU Foreign Affairs Council, 2018) "commended the work conducted by East StratCom Task Force" in the context of what it saw as the need to strengthen the resilience of the EU and its neighbours.

5.2.7 Challenges and opportunities

Counter-disinformation campaigns can appear to the target audiences as legitimate and convincing if the institutions initiating them are trusted. Such debunking strategies can also remain within the boundaries of freedom of expression, by refuting content that is not banned as being "false". Where such campaigns are factually grounded and subject to scrutiny, it can be presumed that they are more effective than covert efforts and/or those which are narrative-driven to the point of being propagandistic.

The refutation of 'disinformation' can also be dismissed by critical and disengaged audiences as a public relations exercise for government bodies, rather than a neutral fact-checking exercise. This can, in turn, fuel scepticism and conspiracy theories about State intervention and entrench distrust in State actors, especially those with a history of censorship and propaganda. This is compounded by the risk of governments promulgating their own 'alternative facts' as an exercise in seeding disinformation. Where the same actors themselves might be implicated in the adoption of disinformation tactics, this could be a factor that causes their work of debunking falsehoods to boomerang.

In communications, the 'Barbara Streisand effect' is a widely known theory, according to which the attempt to hide or censor a piece of information can rebound with the opposite unintended consequence of this information going viral in the Digital Age (Masnick, 2003). It is named after the singer Barbara Streisand for her attempt to remove an aerial picture of her property in Malibu which had the opposite effect of drawing more attention to it. This assumption could be tested in relation to the debunking of disinformation as well, including when governmental initiatives are involved.

In cognitive science, this unintended impact is also presented as the "backfire effect", according to which the refutation of information can reinforce the reader's belief in it (Cook et al., 2014; Nyhan & Reifler, 2006). In an analysis of the psychological efficacy of messages countering disinformation, some researchers recommend that when there is a need to repeat a lie to debunk it, it is best to limit the description of it (Sally Chan et al., 2017).

More recent research, however, has not found evidence that retractions that repeat false claims and identify them as false result in increased belief in disinformation. On the contrary, providing a detailed alternative explanation was found to be more effective (Ecker et al., 2019). Some research suggests that debunking should use the modality of a 'truth sandwich' as described by linguist George Lakoff (Hanly, 2018), where the false information is enveloped by true information, and is not given first or last prominence in the narrative.[149] However, further research is needed into differences between governmental debunking and independent debunking.

[149] See detailed discussion of the literature in Chapter 3

One limitation to point out is that refutation only works on identified false claims. Disinformation takes different forms which do not necessarily consist of straightforward false claims, but can involve a decontextualised or misleading application of information to frame an issue, and is often merged with strong emotive resonance.

On the opportunity-side, campaigns led by public authorities can mobilise significant resources - both financial and human - to monitor and fact-check content, and circulate the results. The public character of such initiatives can also lead to public engagement and debate, such as through parliamentary or other oversight mechanisms.

5.2.8 Recommendations for national and international counter-disinformation campaigns

Individual states could:

- Engage more closely with civil society organisations, news organisations, and academic experts to aid development of well-informed campaigns responding to different types of disinformation.
- Consider campaigns designed to raise awareness of the value of critical, independent journalism and journalists in protecting societies from disinformation.
- Invest in research that measures the efficacy of counter-disinformation campaigns.

Researchers could:

- Conduct audience research to test responses to a variety of national and intergovernmental campaign types (e.g. online/offline, interactive, audio-visual) among different groups (e.g. children and young people, older citizens, socio-economically diverse communities, those with diverse political beliefs, those who are identified as susceptible to being influenced by and/or sharing disinformation).

Internet communications companies could:

- Expand financial support for, and heighten the visibility of, intergovernmental anti-disinformation campaigns beyond crises like the COVID-19 pandemic.

5.3 Electoral-specific responses

Authors: Denis Teyssou, Julie Posetti and Kalina Bontcheva

This chapter deals specifically with electoral responses designed to protect voters and the integrity and credibility of elections, through measures that detect, track, and counter disinformation that spreads during election campaigns. Such disinformation threatens democratic processes more generally within a growing number of countries around the world (UNESCO, 2019).

Here, the spotlight is on initiatives launched, either by news media or NGOs, and sometimes by electoral bodies themselves. Their aim is to prevent jeopardising elections and undermining democracy, while preserving universal standards of election integrity applicable to all countries throughout the electoral cycle - from the lead up to elections, during election campaigns, during ballots, and in the aftermath (Norris et al., 2019).

State-based legal and policy responses are detailed in chapter 5.1, while chapter 5.2 tackles counter-disinformation campaigns from States and intergovernmental actors.

Internet Age realities complicate pre-internet normative standards such as those set out in the *Handbook on the Legal, Technical, and Human Rights Aspects of Elections* (United Nations, 1994):

> *Use of the media for campaign purposes should be responsible in terms of content, such that no party makes statements which are false, slanderous, or racist, or which constitute incitement to violence. Nor should unrealistic or disingenuous promises be made, nor false expectations be fostered by partisan use of the mass media.*

5.3.1 What and whom do electoral disinformation responses target?

The challenges to trust in parts of the news media, combined with the proliferation of user-friendly digital tools that make it easier to create synthetic media that mimics credible journalism, increase the spread of disinformation during election periods (Ireton & Posetti 2018; Norris et al., 2019)[150]. While some falsehoods and myths that spread via orchestrated campaigns are mistaken as factual, the main damage might actually be the systematic erosion of citizens' capacity to even recognise truth. The effect would be to reduce elections to popularity contests which have no need of verified information, eroding the modality of informed voters making rational political choices as a core concept of democratic life.

The kind of disinformation that impersonates legitimate news content is often debunked within a short period of time. However, the purpose behind it is not necessarily to

[150] See also the discussion of trust in chapter 7.1

create a belief based on falsehoods, but rather to "…undermine established beliefs and convictions…to destabilize, to throw suspicion upon powers and counterpowers alike, to make us distrust our sources, to sow confusion." (Eco, 2014). While this observation applies to disinformation across a range of issues (e.g. vaccination, climate change, migration), it can have very direct significance during elections.

For instance, in the context of the 2016 UK EU membership referendum known as the 'Brexit' vote, some researchers argued that voters' exposure to disinformation on social media played a major role in the results (Parkinson 2016; Read, 2016; Dewey 2016). Others, however, pointed out the complexity and polarisation of the political situation as bigger factors (e.g. Benkler et al., 2018; Allcott & Gentzkow, 2017; Guess et al., 2018b), while some highlighted the role of biased coverage in the UK press (Davis, 2019; Freedman, 2016). One Foreign Policy assessment noted the failure of journalistic accountability to professional standards of truth-telling (Barnett, 2016):

> *"Mainstream media failed spectacularly (…) most of UK national press indulged in little more than a catalogue of distortions, half-truths and outright lies: a ferocious propaganda campaign in which facts and sober analysis were sacrificed to the ideologically driven objectives of editors and their proprietors…[Their] rampant Euroscepticism also had an agenda-setting role for broadcasters."*

Another factor here was the failure of objectivity norms within journalism due to the misapprehension that both sides of the debate (i.e. those campaigning for the UK to leave the EU, and those campaigning for it to stay) needed to be given equal weighting, rather than be assessed on the basis of the evidence in the public interest. However, it should be acknowledged that the news media had to navigate a 'pro-leave' political communications strategy designed "to destabilise the discourse while controlling [their] own message based on emotional appeals to voters", which when mixed with disinformation had a powerful impact on democratic deliberations (Beckett, 2016).

Another example highlighting the need to counter election-related disinformation on Facebook and other social media sites was the 2016 U.S. Presidential election. While scholars have emphasised the pre-existing polarisation of American politics, the significance of orchestrated disinformation campaigns (e.g. the Cambridge Analytica scandal) is recognised as a factor in the wider equation (Benkler et al., 2018; Allcott & Gentzkow, 2017; Guess et al., 2018).

Another key concern to be addressed is disinformation associated with political advertising, and its potential to dishonestly influence voters. Such content can be distributed as messages on social networks, within closed chat apps, and in the form of memes, videos, and images to persuade, mobilise, or suppress voters and votes (Wood & Ravel, 2018). Such advertising is designed to affect people's political opinions and voter turnout or suppression. The advertiser pays to produce those effects and can distribute such adverts through microtargeting on social media and search. Some political adverts look like organic content or native advertising, and are also less traceable and thus not easily amenable to counter-narratives. It should also be noted that in many countries the standards applied to political advertising on social media websites are also generally lower than broadcast licensing allows.

Political advertising spending was surging ahead of the 2020 U.S. presidential election, with one digital marketing firm forecasting that the total campaign advertising spend would jump 63% from the 2016 election, to $6.89 billion (eMarketer, 2020). According to this report, the highly partisan political environment was driving more Americans to donate to their preferred candidates than in previous elections, which in turn was funneling more money into political advertising. While television was predicted to account for the largest share of political advertising (66 percent of the total), digital advertising – with Facebook being the primary platform – was expected to grow more than 200 percent from the previous presidential election, according to the same source. Facebook's ability to offer reach, as well as contentious voter targeting capabilities (Harding-McGill & Daly, 2020), along with its ease of use make it particularly appealing to political advertisers.

These factors combined have given rise to 'sock puppet farms' operated by disinformation agents that span State-linked propaganda units, profiteers, and public relations firms that have begun specialising in creating orchestrated disinformation networks using a host of tactics beyond advertising. These are known as 'black PR firms' (Silverman et al., 2020; Bell & Howard, 2020). There is mounting concern about the role such disinformation purveyors might play in electoral contexts. The danger is that these networks, which also specialise in 'astroturfing', are designed to mimic authentic citizens and organic political movements and therefore generate a veneer of legitimacy which can make their content go more viral than recognisable political advertising.

Another example of deceptive online identities and behaviour emerged in the 2019 UK general election when the name of the Twitter account for the Conservative Party's campaign headquarters (@CCHQPress) was changed to @FactCheckUK, and the accompanying avatar was changed to resemble that of a fact-checking organisation during a televised leaders' debate. Each tweet posted during the debate began with the word "FACT". After the debate, the account name and avatar were changed back. The ultimately victorious Conservative Party defended the act, while Twitter accused the party of misleading the public, a view echoed by the independent fact-checker FullFact (Perraudin, 2019). This weaponisation of fact-checking for political gain during an election campaign underscored the value of such services as tools of trust, while also triggering significant concerns within the fact-checking and journalism communities.

Journalistic actors have responded to these forms of election-related disinformation with investigative reporting and forensic analysis of the data (Ressa 2016; Silverman et al., 2020)[151]. Fact-checking organisations have built on these traditions with electoral specific projects (see below).

State-based responses have involved calls for tighter regulation of political advertising, propaganda networks, and voter targeting in some contexts (Dobber et al., 2019; Kelly, 2020b), but advocated for looser regulation in others (@TeamTrump 2019). The distinction in approaches can be explained in part by the potential for political loss or gain for ruling political parties.

Besides journalists, the other major respondents to electoral disinformation are the internet communications companies themselves. During 2020, a public disagreement erupted between Twitter and Facebook over divergent approaches to fact-checking and identifying disinformation associated with the U.S. President's claims about electoral processes (Smith, 2020b). Fact-checking and flagging his claims as misleading fell within both companies' guidelines on the issue of monitoring and checking electoral disinformation. In May 2020, Twitter took the unprecedented step of attaching a warning

[151] https://www.theguardian.com/news/series/cambridge-analytica-files

label to the relevant tweets (NPR, 2020). This was a move which Facebook CEO Mark Zuckerberg strongly disagreed with, arguing that private companies should not be "arbiters of truth" (Halon, 2020). However, as discussed in chapter 7.1, avoiding being an arbiter of truth does not exclude taking any action against the promotion of clear falsehoods (Kaye, 2020b). In response to Twitter's decision to implement its policies regarding electoral disinformation responses, the U.S. President immediately announced (on Twitter) that he would move to "strongly regulate" or "shutdown" social media companies via an Executive Order (Smith & Shabad, 2020). Civil society organisations focused on freedom of expression condemned the threat, and others said the resulting Executive Order could not be implemented without a change in the law (Article 19, 2020b).

5.3.2 Who do electoral disinformation responses try to help?

Electoral responses are aimed at protecting voters from exposure to disinformation and reducing the likelihood of it influencing their political views and voting intentions in ways that would not have been the case without its impact.

In the context of polls, political advertising (including highly personalised, and individually targeted messaging) has been employed extensively by political parties and candidates with the purpose of influencing voters. For instance, during the 2016 UK EU membership referendum, the #VoteLeave campaign used targeted adverts containing disinformation regarding the weekly cost of Britain's EU membership and Turkey being set to join the EU (Cadwalladr, 2018). In a number of instances, the disinformation was disguised in statistics, which raised complex issues of calculations of costs and benefits. Given that political contests invariably involve selective use of statistics, there are grey areas about when legitimate campaigning blurs into acts of definitive disinformation, although invented statistics are clearly not acts of information. The term disinformation (along with misinformation and 'fake news') themselves can be weaponised to brand particular reality claims as being beyond the pale of accuracy and honesty. These challenges underline both complexity for, and the significance of, responses to electoral disinformation.

Another way in which electoral responses help voters deal with disinformation is to expose the actors behind the problem. For example, many voters do not always know that a major route for targeting them with disinformation is through automated accounts (bots and cyborgs). However, there are well researched cases during the 2016 U.S. presidential elections, the 2016 Philippines presidential election, the UK EU membership referendum, and the 2017 French presidential elections. Political bots, in particular, have been shown as trying to influence voter opinion, e.g. attack political leaders or journalists, although some evidence seems to indicate that bots in certain cases do not change voter intent (Howard et al., 2018b). Nevertheless, it is the case that during elections a large number of (coordinated) bots and sockpuppet accounts were used for spreading disinformation and political rumours (Ressa 2016; Phillips & Ball, 2017; Howard et al., 2018; Gorrell et al., 2018; Howard & Kollanyi, 2016). Exposing such phenomena is part of electoral responses that can sensitise voters to covert disinformation operations and 'dirty tricks' designed to subvert the norms and rules of fair campaigning in a poll.

5.3.3 What output do electoral-specific responses publish?

Outputs of electoral responses to disinformation can include a range of real-time detection, debunks, counter-content, as well as retrospective assessments. They can also entail campaigns linked to voter education, and regulations about electoral conduct.

Electoral responses are treated as a stand-alone response category in this report due to the major impact that disinformation has on democratic processes and citizens' rights during elections. However, this category of responses, due to its very nature, typically involves a combination of monitoring and fact-checking, regulatory, curatorial, technical, educational and other responses, which are separately categorised in the typology presented in this report. They are cross-referenced as applicable in this chapter. Therefore, the outputs from electoral responses essentially equal a subset of the combined outputs produced by these other categories of responses (e.g. election-specific fact-checks, election ad archives).

5.3.4 Who are the primary actors behind electoral-specific responses and who funds them?

a. Political fact-checking in the U.S.

The history of political fact-checking in the U.S. is strongly tied to the coverage of presidential elections, and to the amount of falsehoods spreading during breaking news events. To date, much detailed analysis of the practice of political fact-checking has been focused on the U.S., where the practice is said to have originated (Birks, 2019). In fact, the U.S. fact-checking movement emerged in response to the news media's perceived failure to adequately call out campaign trail falsehoods (Spivak, 2010).

The first independent fact-checking organisation was Spinsanity[152], which was founded in 2001 (Graves, 2013). It was active during the 2004 presidential campaign, producing more than 400 articles. Next, just before the 2004 U.S. presidential election, FactCheck.org was launched as "a nonpartisan, non-profit consumer advocate for voters" which aimed to equally monitor the major political parties, talk shows, TV advertisements, official websites, press releases and media conference transcripts[153]. Another prominent initiative was The Fact Checker, launched by *The Washington Post* prior to the 2008 election (Kessler, 2017). It pioneered a rating system based on one to four Pinnochios.

Another major development in political fact-checking for the 2008 election was the creation of Politifact, the largest independent fact-checking outlet in the United States (Aspray & Cortada, 2019; Drobnic Holan, 2018). They became noteworthy for the quality of their fact-checking and their special Truth-O-Meter rating system (a scale ranging from True, Mostly True, Half True, Mostly False, False, up to Pants on Fire[154]). This Truth-O-Meter became an iconic feature of Politifact (Adair, 2018) which received a Pulitzer Prize in 2009 for its coverage of the 2008 presidential campaign. FactCheck.org was also a nominee (Graves, 2013).

In 2010, Politifact expanded its fact-checking by licensing its brand and methodology to U.S. state-based media partners. Three years later, Politifact launched Punditfact to check the accuracy of claims by pundits, columnists, bloggers, political analysts, the hosts and guests of talk shows, and other members of the media (Hollyfield, 2013).

At present, FactCheck.org, Politifact, the *Washington Post*'s Fact Checker and Snopes are considered to be the most important political fact-checking outfits in the U.S. (Graves 2013; Aspray & Cortada, 2019).

[152] http://www.spinsanity.org/about/
[153] https://www.factcheck.org/spindetectors/about/
[154] https://www.politifact.com/truth-o-meter/

They are facing formidable challenges since, as researchers have argued, there are two media ecosystems in the U.S.: one, "the insular right-wing media ecosystem", which shows "all the characteristics of an echo chamber that radicalizes its inhabitants, destabilizes their ability to tell truth from fiction, and undermines their confidence in institutions"; and another, representing the majority of the news media, that is "closer to the model of the networked public sphere" (Benkler et al., 2018). In this dual media ecosystem, fact-checking websites are perceived as systematically biased by the "insular right-wing" and are generally not trusted or believed by this group (Ibid).

b. Political fact-checking in Europe

Fact-checking as a response to political disinformation started in Europe with a blog launched by UK's Channel 4 News in 2005, to cover a parliamentary election (Graves & Cherubini, 2016). It was followed by similar French press blogs: Désintox from Libération in 2008, and Les Décodeurs from *Le Monde* in 2009. Both were inspired by Politifact and FactCheck.org with the aim of fact-checking politicians and public figures, as well as election campaigns. The British charity FullFact.org began in 2009, with the intent to "fight bad information". That year, it was also joined by the BBC's Reality Check (Birks, 2019). In the Netherlands, the fact-checking project Nieuwscheckers began the same year, within the Journalism and New Media school of Leiden University. Fact-checking has expanded rapidly in Europe, with particular reference to elections. From the 34 permanent outlets active in 20 European countries in 2016, fact-checking at the beginning of 2020 involved some 66 active outlets in 33 countries in the region, according data from Duke's University Reporters's Lab.

Presidential or general elections have often been a catalyst for the extension of the fact-checking movement: either by running a 'real life' experiment, or triggering the establishment of more permanent operations. For instance, French journalists really started fact-checking during the 2012 Presidential election campaign (Bigot, 2019).

In Austria, Spain and Italy, fact-checking went mainstream via TV broadcasting. Austria's public service broadcaster ORF began Faktencheck in 2013 to fact-check politicians on live TV shows in the run-up to the general elections. The same year, in Spain, El Objetivo broadcast a prime time program on fact-checking on La Sexta TV to fact-check politicians amid the Spanish financial crisis. A couple of years later, a similar program made by Pagella Politica was broadcast in Italy on the national channel, TV RAI2.

In 2018, a report from the European Commission's High Level Expert Group on disinformation suggested several strategies in order to overcome disinformation and protect EU elections, as well as elections in Member States, such as enhancing transparency in political advertising, developing tools to empower users, and promoting media literacy. Later that year, the European Commission announced measures for securing free and fair elections to the European Parliament in May 2019 (European Commission, 2018b). Those measures include recommendations for State members to create a national network of relevant authorities to detect and respond to cybersecurity and disinformation threats, greater transparency in online advertising and targeting, and tightening rules on European political party funding.

c. Political fact-checking in the rest of the world

In the Asia-Pacific, Duke University's Reporters' Lab's database registers 47 fact-checking organisations. As elaborated below, regarding elections, disinformation was particularly prevalent in India, Indonesia, the Philippines and in Republic of Korea, while being substantially lower in comparison in Japan, Singapore, and Australia. In other parts of the world, recent fact-checking initiatives have developed more to debunk disinformation than to verify political claims and discourses, so they are addressed in chapter 4.1.

All political parties in India began using social media in the 2014 election campaigns, with an emphasis on targeting first-time voters (Kaur & Nair, 2018). More recently, WhatsApp has evolved into India's main channel of disinformation (Kajimoto & Stanley, 2019).

A large part of the disinformation debunked in India is political, either pertaining to local disputes or about tensions with neighbouring Pakistan. It is noteworthy that in the legislative assembly election campaign in Delhi in February 2020, members of one political party spread two viral deepfakes videos on WhatsApp with messages targeting a political rival (Christopher, 2020).

In Indonesia, disinformation is often present during important elections, exploiting religious and ethnic fault lines (Kajimoto & Stanley, 2019). The main actors are called 'political buzzers'. They aim to promote their electoral stance, while undermining their rivals' campaigns with hate speech, hyper-partisan discourse, or religious and ethnic-based disinformation.

The Philippines is another country suffering the proliferation of disinformation in online political discourse, especially since the run-up to the 2016 presidential election. The heightened 'indecency' and incivility in political discourse since that period is frequently blamed on so called 'patriotic trolls' and orchestrated online networks (Ressa 2016; Ong & Cabañes, 2018). It has been argued that some fact-checking efforts undertaken by news organisations and NGOs in the Philippines fail to address the underlying causes of disinformation because they do not address the "professionalized and institutionalized work structures and financial incentives that normalize and reward 'paid troll' work" (Ong & Cabañes, 2018). Click farms and the practice of astroturfing, especially on Facebook, have been regularly reported since 2016 in the Philippines.

In Republic of Korea, almost all newspapers and broadcasters launched fact-checking initiatives during the 2017 presidential election (Kajimoto & Stanley, 2019). They aimed to tackle the spread of disinformation, including a collaborative endeavour with academia, SNU Factcheck[155], launched by Seoul National University to enable a fact-checking platform used by 26 news outlets to cross-check disputed information. Other examples in the wider region include the FactCheck Center[156] and MyGoPen, who tackle disinformation on the popular messaging service LINE.

d. Collaborative media responses on elections

Due to the sheer volume of online disinformation and candidate statements in need of fact-checking during elections, a number of media organisations have begun pooling their resources into well-coordinated, collaborative initiatives. Some are national and others are international in nature. The rest of this section discusses some prominent examples.

[155] http://factcheck.snu.ac.kr/

[156] https://tfc-taiwan.org.tw/

Country-based Collaborative Responses

Electionland was the first **U.S**. joint endeavour in 2016, launched by Propublica with Google News Lab, WNYC, First Draft, Gannett's U.S. Today Network, Univision News, and the CUNY Graduate School of Journalism, to monitor disinformation in social media around the 2016 election day. The project involved 600 journalism students and over 400 reporters located across the U.S. (Bilton, 2016; Wardle, 2017b).

In **Europe**, CrossCheck France (funded by Google News Lab) was among the first collaborative journalism projects on debunking false stories, comments, images and videos about candidates, political parties and all other election-related issues that circulated online during the French presidential election campaign in 2017. It involved more than 100 journalists from 30 French and international media organisations, with some academics and technology companies. In total, 67 debunks were published on CrossCheck's own website, as well as on the websites of the newsroom partners (Smyrnaios et al., 2017). The pioneering collaboration in debunks attracted 336,000 visitors (95% French) (Ibid).

Prior to the UK's 2017 general election, the non-profit First Draft established CrossCheck UK[157], with a dedicated workspace for British journalists providing alerts, facilitating collaborative reporting and investigating suspicious online content and behaviour. In terms of the response categories presented in this report, CrossCheck UK is an investigative response, as the focus is on the disinformation narratives and context rather than on labelling individual claims as true or false. The funding sources for this version of CrossCheck are unclear.

At that time, one of the major challenges faced by such media-focused investigative disinformation responses was the need to establish shared methodology, knowledge and tools[158]. This is where First Draft's contribution was instrumental, alongside the development of innovative tools developed specifically to support collaborative content verification and fact-checking (Mezaris et al., 2019).

First Draft's CrossCheck collaborative methodology was also adopted by the Spanish Comprobado initiative (in collaboration with Maldita.es) to fight disinformation during the country's 2019 general election. In addition to political fact-checking and investigation of disinformation, a new challenge that was addressed was disinformation on private messaging apps (WhatsApp in particular). Comprobado implemented strict quality controls on verification by requiring the approval of at least three of the 16 project members. Based on lessons learned in previous initiatives, Comprobado carefully selected what viral content should be debunked, and how, so as to avoid giving oxygen to disinformation.

In 2018, **collaborative election-focused verification initiatives started spreading worldwide**. One example is the Mexican Verificado 2018[159], led by Animal Politico, Newsweek in Spanish, Pop Up Newsroom and AJ+ Spanish. It aimed to debunk 'fake news' and verify the political discourse during the Mexican 2018 election campaign. It was ground-breaking in scale, as it involved more than 60 media, civil society organisations and universities - all aiming to help citizens decide who to vote for based on confirmed, accurate information. Each report labeled with the Verificado 2018 hashtag was reviewed and supported by the whole network of partners. Verificado 2018 was funded by the

[157] https://firstdraftnews.org/project/crosscheck-uk/
[158] https://firstdraftnews.org/verification-toolbox/
[159] https://verificado.mx/metodologia/

Facebook Journalism Project, the Google Digital News Initiative, and Twitter, as well as the organisation Mexicans Against Corruption and Impunity, and foundations such as Open Society and Oxfam. The initiative won the 2018 U.S. Online Journalism Association Award for Excellence in Collaboration and Partnerships.

The Verificado 2018 initiative was also ground-breaking in terms of the support provided by the internet communications companies. Key to its success were: 1. The companies' provision of access to data about the most shared stories or search engine queries and 2. curatorial measures to promote the verified information. Typically, however, news media and independent fact-checkers lack such comprehensive support and data access from the platforms, which complicates their work significantly.

The Verificado 2018 project is also notable in that it adopted and adapted a collaborative platform originally created by by Verificado19S[160] to manage collaborative citizen response and rescue operations.

In **Latin America** a prominent example is Comprova[161], a partnership of 24 Brazilian media organisations, established for the 2018 elections but still ongoing. The project is coordinated by Abraji (Associação Brasileira de Jornalismo Investigativo) with First Draft. As with many other collaborative fact-checking projects, the Google News Initiative and the Facebook Journalism Project provide financial and technical support to Comprova. Projor, a non-profit organisation focused on issues concerning Brazil's media, was also an early supporter. During the elections, Comprova monitored and verified the veracity of viral information shared by unofficial sources on social media and messaging applications (mainly WhatsApp) (Tardáguila & Benevenuto et al., 2018). WhatsApp monitoring relied on crowdsourced suggestions for content to be verified, leading to 67,000 pieces of information being submitted. This clearly demonstrates the huge volume of potentially problematic political content circulating through these closed messaging apps, and the impossible task that rigorous fact-checking and verifying such content presents.

Another prominent example, this time from Argentina, is Reverso[162]. A massive collaborative project, it was promoted and coordinated by the fact-checker Chequeado, AFP Factual, First Draft and Pop-Up Newsroom, in which more than 100 media and technology companies came together during the 2019 Argentinian presidential election campaign. In order to achieve maximum reach, Reverso debunked (180 articles and 30 videos produced over the six month campaign) were published simultaneously by all partners. The team monitored Facebook, Instagram and Twitter; private messaging apps (mainly WhatsApp); and platforms, such as YouTube and Chequeo Colectivo[163] (a crowdsourcing platform from Chequeado). From an innovation point-of-view and through collaboration with BlackVox[164] the Reverso team also managed to verify fake audio files[165] of candidates shared on WhatsApp[166].

[160] https://verificado19s.org/sobre-v19s/

[161] https://projetocomprova.com.br/about/

[162] https://reversoar.com/

[163] https://chequeado.com/colectivo/

[164] https://blackvox.com.ar/

[165] https://www.clarin.com/politica/reverso-creo-nuevo-metodo-conicet-verificar-audios-virales-whatsapp_0_1638FLWL.html

[166] https://www.poynter.org/fact-checking/2019/meet-forensia-a-software-ready-to-debunk-fake-whatsapp-audio-files/

The last prominent Latin American example is Uruguay's Verificado.UY[167] project, in which over 30 partners monitored and debunked disinformation during the Uruguayan presidential elections in October 2019. Training, financial and technological support were provided by First Draft. It focused on two types of verification: rumours spreading on social networks, and statements of politicians and candidates.

In **Australia** and **Asia** respectively, examples include CrossCheck Australia[168] (managed by First Draft), which monitored the 2019 Australian federal election and the collaborative Checkpoint project in India which was operated ahead of national elections there in 2019 (see Chapter 6.1 for details).

In **Africa**, First Draft worked in partnership with the International Centre for Investigative Reporting in Nigeria and 16 newsrooms to establish CrossCheck Nigeria[169] in the run up to the February 2019 Nigerian elections. Support for the project was provided by the Open Society Foundation. It built on knowledge and technology from previous First Draft collaborative initiatives, including Comprova in Brazil and CrossCheck in France. A key feature of this work is the 'CrossCheck' methodology which involves journalists from different newsrooms checking on each other's work to ensure that the principles of transparency, accuracy and impartiality are adhered to.

Another example is the South African Real411[170] ('411' being internet slang for information) project. What is particularly notable is that unlike the previous media- and First Draft-driven examples, Real411 was launched by an NGO (Media Monitoring Africa) and it also involved the South African Electoral Commission. Similar to the other initiatives, it offers an online platform for citizens to report instances of alleged disinformation. This platform, however, incorporates a governmental aspect response, as it is connected to the Directorate of Electoral Offences. Complaints are considered by a panel of experts including media law, and social and digital media representatives. They make recommendations for possible further action for the consideration of the Electoral Commission, including referring the matter for criminal or civil legal action; requesting social media platforms to remove the offensive material; and issuing media statements to alert the public and correct the disinformation. The Real411 site contains a database of all complaints received and their progress. To help distinguish between official and fake adverts, political parties contesting the 8 May, 2019 general elections were asked to upload all official advertising material used by the party to an online political advert repository at www.padre.org.za. This initiative has also been adapted to deal with COVID-19 disinformation.

International Collaborative Responses

EU-wide collaborative responses: FactCheckEU.info[171] was established by the International Fact-Checking Network (IFCN), bringing together 19 European media outlets from 13 countries (the European signatories of IFCN's Code of Principles) to counter disinformation in the European Union ahead of the European Parliament elections in May 2019 (Darmanin, 2019). The core focus was on providing debunks on disinformation or facts about Europe to reduce misperceptions (e.g. islamophobia, immigration). Citizens could submit claims for verification through a web Q&A form which were then picked

[167] https://verificado.uy/

[168] https://firstdraftnews.org/project/crosscheck-australia/

[169] https://crosschecknigeria.org/about/faqs

[170] https://www.real411.org/

[171] https://factcheckeu.info/en/

up by one of the partners. For maximum coverage, the articles were published in their original languages and translated into English.

This initiative was entirely independent of EU institutions and other governmental actors. The platform was built by the newspaper *Libération* and the web agency Datagif with an innovation grant (U.S.$50,000) from the Poynter Institute. Other costs — primarily the salary of a full-time project coordinator for six months and the costs of translating content — were covered through financial support from Google (€44,000), the Open Society Initiative for Europe (€40,000), and the IFCN (€10,000).

Two other collaborative fact-checking initiatives were launched in parallel: the EU-funded Disinformation Observatory (SOMA, reviewed in chapter 4.1)[172], along with CrossCheck Europe by First Draft[173].

At the time of writing, First Draft's CrossCheck initiative was expanding as a global network of reporters and researchers that collaboratively investigates online content during elections and beyond. Building on the previous campaigns in the U.S., France, Brazil, Nigeria, Spain, Australia and the EU, it seeks to demonstrate that news organisations can work together on a global scale, to produce more effective, efficient and responsible reporting against disinformation.

e. Responses by the internet communications companies

Ahead of the 2020 U.S. presidential election, fears were mounting about exacerbated polarisation, foreign interference, and the rise of new forms of digital content manipulation such as so-called deepfakes: synthetic videos or audio files created through machine learning (see chapter 6.2). Against this background, in 2019 Facebook's vice-president of Global Affairs and Communications, former UK Deputy Prime Minister Nick Clegg, said that "Facebook made mistakes in 2016" but he added that the company had spent the three years since "building its defenses to stop that happening again" (Clegg, 2019). He then enumerated the actions taken by Facebook to crack down on 'inauthentic' accounts – qualified by him as the main source of 'fake news' and malicious content – such as "bringing in independent fact-checkers to verify content" (see chapter 4.1 for an analysis of Facebook's third-party fact-checking network and 7.1 for an assessment of the ethical issues involved) and "recruiting an army of people – now 30,000 – and investing hugely in artificial intelligence systems to take down harmful content".

With respect to false or misleading political advertising, Facebook has been extensively criticised for its policy. The U.S. Sen. Elisabeth Warren accused Facebook of turning its platform "into a disinformation-for-profit machine" and followed up to make a point by publishing a fake advertisement saying: "Breaking news: Mark Zuckerberg and Facebook just endorsed Donald Trump for re-election"[174]. It was reported in October 2019 that Facebook had changed the rules from preventing any advertisements with "false and misleading" content, defined as "deceptive, false, or misleading content, including deceptive claims, offers, or methods," to include a narrower definition prohibiting "ads that include claims debunked by third-party fact checkers or, in certain circumstances, claims debunked by organizations with particular expertise." (Legum, 2019).

[172] https://www.disinfobservatory.org/the-observatory/

[173] https://firstdraftnews.org/project/crosscheck-europe/

[174] https://twitter.com/ewarren/status/1183019880867680256

Facebook further limits its fact-checking of politicians and political parties through guidelines for third party fact-checking partners that state: “posts and ads from politicians are generally not subjected to fact-checking” (Facebook, 2019b.) The guidelines align to “Facebook’s fundamental belief in free expression, respect for the democratic process, and the belief that, especially in mature democracies with a free press, political speech is the most scrutinized speech.” (See chapters 4.1 and 7.1 for further discussion of these issues). The guidelines indicate: “If a claim is made directly by a politician on their Page, or in an ad or on their website, it is considered direct speech and ineligible for our third party fact checking program — even if the substance of that claim has been debunked elsewhere.”

In contrast with Facebook’s comparatively hands-off approach to political disinformation, Twitter CEO Jack Dorsey announced that the platform he founded would stop running all political advertisements commencing November 22, 2019. He said that this reflected concerns that “paying to increase the reach of political speech has significant ramifications that today’s democratic infrastructure may not be prepared to handle”[175]. As discussed earlier in this chapter, Twitter and Facebook engaged in a public disagreement in mid 2020 over fact-checking and debunking the content published by political leaders, in an incident triggered by Twitter’s decision, for the first time, to flag misleading tweets from the U.S. president connected to electoral processes.

In December 2019, Google announced a commitment to “a wide range of efforts to help protect campaigns, surface authoritative election news, and protect elections from foreign interference”. (Spencer, 2019). Google said it wanted to “improve voters’ confidence in the political adverts they may see on our ad platforms”. The company announced changes including limiting election adverts and audience microtargeting to age, gender, and general location (postal code level). They also clarified their advertising policies by explicitly prohibiting “deep fakes”, misleading claims about the census process, and adverts or destinations making demonstrably false claims that could significantly undermine participation or trust in an electoral or democratic process.

As part of its election advertising transparency, Google says it provides both in-ad disclosures and an online transparency report[176] (only available for Europe, UK, India and the U.S.) that shows the actual content of the advertisement themselves, who paid for them, how much they spent, how many people saw them, and how they were targeted. “We expect that the number of political ads on which we take action will be very limited—but we will continue to do so for clear violations,” the company said. However, Google faced criticism ahead of the U.S. election in 2020 when it refused to remove advertisements from a group accused of voter suppression for falsely claiming that there is a material difference between absentee voting and voting by mail. Facebook, however, agreed to remove similar advertisements from the same group (Stanley-Becker 2020).

f. Regulatory responses to electoral disinformation

Electoral commissions or dedicated government units can also play a key role in fighting electoral disinformation through targeted responses. Examples include actions taken by the Australian Electoral Commission in 2019, including authorisation of electoral communications (AEC, 2019a), and the Spanish ‘hybrid threats’ government unit which

[175] https://twitter.com/i/events/1189643849385177088

[176] https://transparencyreport.google.com/political-ads/home?hl=en

focuses on cyber security, monitoring, and at times refuting of disinformation (Abellán, 2019).

Naturally, election integrity can be protected through legislative measures. These are discussed specifically in Chapter 5.1 and Annex A, under two dedicated sections - one on legislative proposals and another on adopted legislation.

Electoral commissions and government committees can also provide reliable information on candidates and parties, as well as work with the internet communications companies towards the promotion of such information. For example, the Canadian government created the Critical Election Incident Public Protocol in 2019[177] as a mechanism to notify citizens of election integrity threats as well as inform candidates, organizations or election officials who have been targets of attacks. Another example is the Indonesian Ministry of Communication and Information Technology, which in 2019 organised a 'war room' to detect and disable negative and violating content (Board, 2019). An example of a cooperation response is the approach of the Mexican National Electoral Institute (INE), who signed a cooperation agreement with Facebook, Twitter and Google to limit the spread of electoral disinformation and disseminate practical election information during their 2018 and 2019 elections.

Another important kind of regulatory response targets transparency and integrity of online adverts during election periods. For example, in 2019 the Irish government introduced a legislative proposal to regulate the transparency of online paid political advertising within election periods (Irish Department of the Taoiseach, 2019). A complementary approach is to encourage or to legislate that political parties need to log their online advertising in a public database. In 2019, the South African Electoral Commission for example created the Political Party Advert Repository (padre.org.za) for this purpose.

Responses have also enrolled citizens in helping them discover, report, and act upon electoral disinformation. One example, as already discussed above, is the real411.org portal created in co-operation with the Electoral Commission of South Africa. Another is the Italian government's 'red button' portal, where citizens could report disinformation to a special cyber police unit. The police unit would investigate the content, help citizens report disinformation to the internet communication companies, and in case of defamatory or otherwise illegal content, file a lawsuit (la Cour, 2019).

Another kind of response has been internet shutdowns, although these are widely regarded as disproportionate and even counter-productive to electoral credibility. Some governments have enforced these in the run up to polls saying they are seeking to protect citizens from electoral disinformation and propaganda (Al Jazeera, 2018; Paul, 2018).

There are also some examples of international responses. The European Union adopted an Action Plan on Disinformation, ahead of the 2019 European elections, which aimed to build capacities and cooperation within the EU and among its Member States (European Commission and High Representative, 2018). The European External Action Service also runs a website aiming to provide counter-narratives to disinformation. Another example is the guide to guarantee freedom of expression regarding deliberate disinformation in electoral contexts by the Organization of American States (OAS, 2019). It provides recommendations to a wide variety of actors: legislative branch, judiciary, executive branch, electoral authorities, Internet communication companies, political parties,

[177] https://www.canada.ca/en/democratic-institutions/services/protecting-democracy/critical-election-incident-public-protocol.html

telecommunications companies, media and journalists, fact checkers, companies that trade data for advertising purposes, universities and research centres (OAS, 2019).

5.3.5 How are electoral responses evaluated?

Since many of the electoral-specific responses are actually covered within other response type categories (e.g. fact-checking, curatorial) used to specifically target election-oriented disinformation, the methods and findings from their respective evaluations, as outlined elsewhere in this report, apply fully here.

Regarding transparency in political adverts, Facebook says that "ads about social issues, elections or politics are held to a higher transparency standard on its platform". It adds: "All inactive and active adverts run by politicians on Facebook will be housed in the publicly available, searchable ad library[178] for up to seven years", thereby enabling assessment.

Nevertheless, an Institute for Strategic Dialogue (ISD, 2019) study on the European elections concluded that the Facebook Ad Library "is full of shortcomings". Its classification of adverts is "often haphazard. For example it was accused of having originally wrongly labelled heating engineers in Italy and the Dungeons and Dragons computer game in Germany as 'political' content', while adverts from the far-right German party AfD were missing from the adverts library. In a blog post,[179] Mozilla also complained that Facebook's advertising archive application programming interface was "inadequate", meeting only two of experts' five minimum standards.

The same study (ISD, 2019) argued that internet communications companies are "simultaneously putting freedom of speech at risk, with over-zealous and misguided censorship, while being unable to keep up with many malign campaigns and tactics," the latter also representing threats to freedom of expression. ISD also reported counter-productive measures in Germany, for example, where Twitter's attempts to enable speedy reporting of disinformation appeared to have been gamed by far-right networks, leading to the removal or suspension of the accounts of anti alt-right activists and Jewish-interest newspapers, as well as the victims of harassment, rather than those of the perpetrators.

5.3.6 Challenges and opportunities

Recent research outlines an evolution of disinformation tactics. The already mentioned ISD foresees that "populist parties, far-right cyber militias and religious groups are adapting the tactics more notoriously used by States." The London-based Institute for Strategic Dialogue sees an evolution "away from so-called 'fake news' towards an aggressive 'narrative competition', with the promotion of a 'culture war' dynamic around issues like migration, Muslims in Europe, family vs. progressive values and, increasingly, climate policy" (ISD, 2019). The result is that the connections between political parties and online content are often blurred or fully opaque, the identities of the actors behind messages can be concealed, and there is a lack of transparency around the mechanisms of 'reach'.

In the 2019 EU elections campaign, non-profit activist network Avaaz (Avaaz, 2019) used a crowdsourcing platform[180] to identify new tactics of far-right networks across the

[178] https://www.facebook.com/ads/library/
[179] https://blog.mozilla.org/blog/2019/04/29/facebooks-ad-archive-api-is-inadequate/
[180] https://fake-watch.eu/

European Union that were adopting the following practices: using fake and duplicate accounts to amplify disinformation spread; abnormal coordination behaviour from specific alternative outlets to share identical content and hate speech; recycling followers with misleading page-name changes; clickbait; and boosting political or divisive agendas through popular entertainment pages. The challenge is to enable such monitoring and exposure during elections, and at scale.

In a 2019 report on the UK General Election (Election Monitoring, 2019), eight organisations highlighted the lack of transparency about the collection and processing of voter data by political parties, and the lack of transparency of political advertising and targeted messaging, including exaggerated and misleading claims. More concerns noted included "opaque funding arrangements" to push "paid content" to voters, bot-like activity in discussions around political parties and policies, spamming of disinformation and conspiracy theories by hyper-partisan actors on Facebook, online harassment of key political figures and journalists, and even the creation of biased polling organisations. The eight signatories, including ISD, Full Fact and the Computational Propaganda Project from Oxford University, called for electoral reform to counter those digital threats to democracy.

The internet communication companies have faced calls to address the challenge of surfacing and promoting reliable information sources during elections, especially as against issues such as deceiving voters (e.g. voter suppression) or undermining trust in the election process (Goldzweig, 2020). As part of their responses to COVID-19 disinformation, the companies have already demonstrated that they have the technical capabilities to do so (Posetti & Bontcheva, 2020a; Posetti & Bontcheva, 2020b) and their challenge is to adapt these to promote reliable information from authoritative sources during elections, such as electoral bodies and/or independent bodies monitoring election integrity.

Another challenge that needs addressing is the funding model for fact-checking and verification organisations, and the sometimes limited transparency associated with these efforts. For example, if an internet communications company controls the fact-checking standards applied to official fact-checks on its sites conducted by third party fact-checkers during elections, and refuses to fund certain content being fact-checked or to apply the results, this may affect the efficacy of fact-checking and how independent and trustworthy such fact-checking efforts are regarded.

Similarly, if fact-checking non-profits and research institutes investigating disinformation content and networks proliferating on social media during elections are directly funded by such companies, what are the implications for their independence, and what safeguards are put in place to ensure funders do not apply undue pressure to these organisations?

These considerations are especially important in light of the great challenge unfolding for internet communications companies to balance their dual responsibilities to uphold freedom of expression rights, while simultaneously consistently flagging, curtailing and blocking disinformation and misinformation during election periods, while facing mounting pressure from powerful political actors to be treated as exceptions to the rules.

Taken together, all these examples highlight the ongoing significant challenges surrounding election disinformation and voter targeting and manipulation. With multiple national elections happening globally on an annual basis, and hundreds of regional and state elections, this presents both major ongoing challenges to the internet communications companies and governments worldwide. But it also brings significant opportunities and impetus to efforts by independent fact-checkers, journalists, media, civil society, researchers, and national and international organisations to continue - and even expand - their key roles in monitoring, uncovering, countering, curtailing, and evaluating the impact of disinformation.

5.3.7 Recommendations for electoral-specific responses

Given the challenges and opportunities identified above, and the considerable potential harms of disinformation accompanying elections, the following policy recommendations can be made.

Governments and international organisations could:

- Invest in monitoring, measuring and assessing the effectiveness of electoral responses to disinformation.
- Work with internet communications companies to ensure the responses that they initiate are appropriately transparent and measurable, as well as implemented on a truly global scale.
- Encourage internet communications companies to apply the same swift and decisive responses to electoral disinformation as they have to disinformation related to COVID-19.
- Coordinate an initiative to support privacy-preserving, equitable access to key data from internet communications companies, in order to enable independent research on a geographically representative scale into the incidence, spread, and impact of online disinformation on citizens during elections.
- Facilitate and encourage global multistakeholder cooperation and exchange of best practice across continents and States, towards effective implementation of holistic measures for tackling online disinformation during elections.

Internet communications companies could:

- Recognise the significant damage potentially caused by political disinformation, specifically in the run-up to elections (including disinformation in online advertising) and engage in a multi-stakeholder dialogue on the policies and methods they adopt specifically during election periods. These could include temporary restrictions on pre-election political advertising; additional transparency information for political adverts placed during election periods; election-specific policies for promoting reliable information sources; and deployment of additional content moderation and fact-checking resources.
- To deal with cross-platform electoral disinformation, collaborate on the setting of broad industry-wide norms for dealing with electoral disinformation that support democracy and aid self-regulation.
- Collaborate on improving their ability to detect and curtail election disinformation, as cross-platform methods of manipulation are often practiced during elections.
- Apply the lessons learned from responding with urgency to the COVID-19 'disinfodemic' and apply those lessons to the management of political and electoral disinformation.

- Contribute significantly towards funds for fully independent research into manifestations and impact of election disinformation, as well as independent evaluation of the effectiveness of the companies' own disinformation responses, with such initiatives to be managed by arms-length independent funding boards.

- Work together, and under the guidance of the UN Special Rapporteur for the Right to Opinion and Freedom of Expression, along with other independent international experts, to develop a consistent policy approach for dealing with disinformation agents who hold powerful political office while using their sites.

Electoral regulatory bodies and national authorities could:

- Strengthen legislation that helps protect citizens against electoral disinformation (e.g. data protection, freedom of expression, electoral advertising transparency).

- Improve transparency of all election advertising by political parties and candidates through requiring comprehensive and openly available advertising databases and disclosure of spending by political parties and support groups.

- Establish effective cooperation with internet communication companies on monitoring and addressing threats to election integrity.

- Seek to establish and promote multi-stakeholder responses including especially civil society.

- Help educate and empower citizens to detect and report disinformation during elections.

- Improve citizens' knowledge and engagement with electoral processes through civics education and voter literacy initiatives.

- Co-operate with news organisations and specialist researchers in surfacing disinformation and probing disinformation networks.

Media and independent fact-checking organisations could:

- Consider expanding fact-checking during elections to live broadcasts and webcasts, to enable greater reach and impact.

- Carry out research into assessing the efficacy of the different approaches to debunking and containment of disinformation during elections, including responses implemented by regulatory bodies and the internet communication companies.