

# 8

# Challenges and Recommended Actions

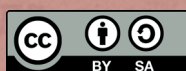
**Authors:** Julie Posetti and Kalina Bontcheva

## Chapter 8 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

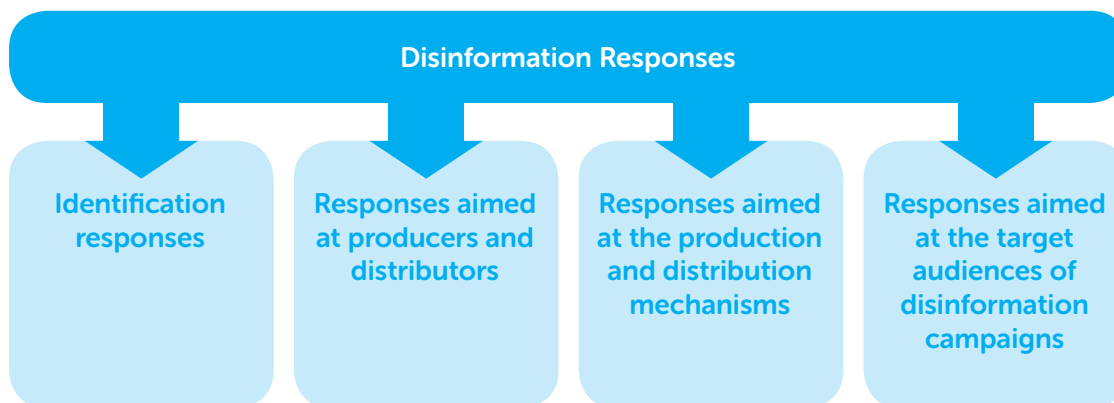


This research will be available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY SA 3.0 IGO) license. By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository

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

## 8.1 Recapping the typology

At this juncture it is appropriate to summarise the original typology of responses developed by the researchers for this report. Firstly, the types of responses were grouped under **four umbrella categories**:



Then, **11 separate modalities** of response were identified under these four umbrella categories:

1. **Identification responses** (aimed at identifying, debunking, and exposing disinformation)
  - Monitoring and fact-checking
  - Investigative
2. **Responses aimed at producers and distributors through altering the environment that governs and shapes their behaviour** (law and policy responses)
  - Legislative, pre-legislative, and policy responses
  - National and international counter disinformation campaigns
  - Electoral responses
3. **Responses aimed at the production and distribution mechanisms** (pertaining to the policies and practices of institutions mediating content)
  - Curatorial responses
  - Technical and algorithmic responses
  - Economic responses

#### 4. Responses aimed at the target audiences of disinformation campaigns (aimed at the potential 'victims' of disinformation)

- Normative and ethical
- Educational
- Empowerment and credibility labelling responses

These different categories of responses are synergistic and symbiotic in nature. They work separately but also interdependently to counter disinformation. For example, normative and ethical responses underpin many of the other response types, while monitoring, fact-checking, and investigative responses play an essential role in informing economic, curatorial and empowerment responses.

## 8.2 Thematic overview

The landscape mapping and research gap analysis presented in Chapter 3 demonstrates an abundance of studies on disinformation and counter-disinformation methods. However research pre-dating this report largely focuses on the developed West and Anglophone contexts.

In general, while there is a growing body of research, software, training and knowledge resource development focused on tackling disinformation, there is a comparative absence of that which focuses on disinformation in light of human rights, freedom of expression, and the burgeoning access to - and use of - broadband technology worldwide.

At the same time, there is a dearth of research into the impacts of exposure to disinformation and counter-disinformation content on those exposed to it. The research gap analysis also identifies a disconnect between academic research, journalistic investigations, and that commissioned by civil society and intergovernmental organisations. Additionally, it is observed that collaboration between these sectors is infrequent but potentially highly valuable. This report begins to fill the void for research into disinformation responses in the developing world, and it is a starting point for future research endeavours that emphasise linguistic, geographic and disciplinary diversity.

Through an analysis of identification responses (monitoring, fact checking and investigative) to disinformation around the world (Chapters 4.1 and 4.2), the study highlights, in particular, the value of cross-border, interdisciplinary and multi-stakeholder collaboration in initiatives designed to identify falsehoods and fabrications.

Major events, such as elections and public health emergencies, provide an opportunity for independent identification responses to reaffirm the value of facts, and to encourage public reflection of what content they treat as credible and what people decide to share. Identification responses are also important for monitoring the intersection of disinformation with hate speech - against women, minorities, migrants and other vulnerable citizens and communities, including children and the elderly who may be especially susceptible to disinformation.

As discussed, the huge volume, range of sources, and different types of disinformation make it hard to monitor, debunk, assess and raise awareness about the problem. The challenges are many and complex - the importance of working at scale, operating in multiple languages, and across myriad countries in real time being chief among them. The parallel need to achieve impact by limiting and stemming disinformation in measurable ways is also a key challenge, as is the difficulty of evaluating these efforts.

One particular challenge to note is the role of States and political actors as both vectors of, and respondents to, disinformation. However, this study is focused on describing and evaluating disinformation responses and thus the role of such actors as sources and amplifiers of disinformation has not been emphasised. Nevertheless, the prevalence of these sources as instigators and agents of disinformation underscores their obligation to respond to the crisis in transparent and accountable ways.

The importance of defending freedom of expression rights in tandem with responding to disinformation is also underlined through an emphasis on ensuring that fact-checking efforts are neither hampered by partnerships initiated by internet communications companies which may seek to limit the scope of fact-checking or use such initiatives as public relations cover, nor by State and political actors behaving as primary sources of disinformation and key amplifiers of it.

Another issue considered is the failure of verification and fact-checking efforts among some news publishers - either as a product of hyper-partisanship, state capture, 'platform capture', poor standards or under-resourcing - which can turn them into disinformation vectors, possibly eroding audience trust in the process. So, fact-checking and investigative capacity also needs to be developed - both inside journalism, and within civil society organisations that can help reinforce self-regulatory efforts and improve professional accountability.

This represents an opportunity to strengthen identification responses broadly. While some companies and NGOs have pledged some funding to fact-checking organisations, ongoing support throughout, and beyond, critical periods of elections and pandemics is needed. For example, verifying claims about migration and race, vaccinations and climate change will be increasingly important.

At the level of State-based responses to disinformation - including pre-legislative, legislative, policy-based, public education campaigns and electoral responses (chapters 5.1-5.3) - the initiatives are cross-cutting. Actors considered fraudulent or abusive are among the main targets of individual government's regulatory responses, the stated aim being to quell viral incitement to hatred and violence, safeguard against national security risks, deter the disruption of democratic elections, and avoid geopolitical fallout. These responses may also include investing in fact-checking initiatives, public interest journalism and media and information literacy projects. Additionally, they focus on internet communications companies, targeting their economic power and technical behaviour. Based on the assumption that the structure (peer-to-peer distribution) and the algorithms they use enable the amplification of disinformation, many regulatory initiatives attempt to encourage increased responsibility by these actors.

Interventions range from supporting self-regulation through to hard regulatory action that can result in takedown orders and content blocks. Professional disinformation purveyors, such as PR agencies specialising in viral disinformation may also be targeted, along with politicians themselves through new obligations regarding transparency in online political advertising and supporting fact-checking initiatives during elections.

The dramatic pace of technological change renders many attempts at regulation outdated before they are even applied. On the other hand, the associated desire to move quickly to curtail viral disinformation without appropriate debate, transparency and scrutiny entails significant risks for human rights, in particular freedom of expression, press freedom, access to information, and privacy. This can lead to undesirable consequences such as the effective criminalisation of journalism through so-called 'fake news' laws and other measures that may inadvertently or otherwise catch legitimate journalism in the net.

Chapters 6.1 - 6.3 deal with responses aimed at limiting the production and distribution of disinformation: curatorial responses, technical and algorithmic responses, and economic responses. Arguably the biggest challenge identified in connection with these responses is that while recognising the role that internet communications companies need to play in curtailing disinformation published on their platforms, there are potential issues with having (self-)regulatory power delegated to these private companies. This is especially the case where this reduces the accountability and judiciability of expression decisions at large that are the responsibility of States, and which should be in line with international human rights laws and standards. The risk is that censorship can effectively be privatised..

In democratic contexts, such delegation can be explicitly provided by regulations, in which case there can be public accountability for them. However, the companies are largely left to self-regulate content, for a range of political, economic and technological reasons. This underlines the urgent need for a robust appeals process and standardised transparency reporting on the way such decisions are taken - to both remove content and to leave it up in the case of targeted abuse and disinformation, for example. This issue is intensifying as the internet communications companies increasingly resort to automation and AI-based algorithms as cost-efficient means for controlling disinformation at scale and at a speed closer to real time. Since algorithms are subject to both potential implicit and explicit bias in their design and in the training data that is used to develop them (with particular implications for gender and racial equality), this is increasingly leading to significant problems, especially in circumstances where the companies have also limited users' ability to resort to a human appeals process. On the positive side, however, at a specific content level, technological responses are less susceptible to external pressure (e.g. from States) applied to individual human operators within a company to take particular action on a case of purported disinformation.

Demonetisation and advertising-linked responses to disinformation are a particular kind of technological response, which are focused specifically on reducing the creation and propagation of disinformation produced for profit. Again, similar to technological responses, the majority of demonetisation responses are largely in the hands of private actors, who are making inconsistent and opaque decisions. In this case, the problems are in the insufficient transparency provided by internet communications companies with regard to advertising, which in effect prevents independent scrutiny and oversight. The problem is acutely present across many platforms and countries not only in the realm of health disinformation (e.g. COVID-19; vaccinations) or issue-based advertisements, but also for political advertising. However, it has become a particular problem for Facebook, which has refused to eliminate micro-targeting of its users in political advertising, and resisted measures to subject direct political speech pertaining to politicians (and their parties and affiliates) to fact-checking, particularly in the United States (Suarez, 2020).

Finally, this study has examined responses aimed at the target audiences of disinformation campaigns: normative and ethical responses, educational responses, and empowerment and credibility labelling responses (chapters 7.1 - 7.3). Responses in the first category - normative and ethical - can be comparatively simple and affordable to implement (although harder to assess in terms of impact), and they can work as counternarratives

that appeal to individuals' moral compasses, or appear to be in alignment with cultural values like anti-racism or anti-misogyny. A problem arises, however, when moral compasses and societal norms are not linked to the principles of access to information, freedom of expression, press freedom, and privacy - as enshrined in international human rights law.

For example, one of the most significant risks associated with State-based responses to disinformation is posed by the kind of legal and regulatory approaches that go against the international norms of freedom of expression (including its corollary press freedom) and privacy. And one of the biggest opportunities exists for internet communications companies to rise above concerns of profit, power and reputation management in order to take genuine action to deal with political disinformation that threatens democracy, in parallel with health disinformation that threatens lives. Ultimately, though, legitimate normative and ethical responses to disinformation can be de-legitimised by the individuals, organisations and States who disagree with the intention behind them, in the same way that credible journalism can be misrepresented as 'fake news' by those seeking to avoid being held to account. That is why these responses need to work in tandem with creative Media and Information Literacy (MIL) interventions, and those that are designed to empower social media users and news consumers.

In the case of educational responses, interventions target a broad spectrum of age groups - from school children to university students, through to journalists, and older citizens, although achieving a comprehensive scale of reach is far from being accomplished in many countries. One of the biggest challenges and opportunities is making Global Citizenship Education (GCED) and Media and Information Literacy (MIL) accessible to children around the world. This is a mission that requires significant investment in interventions (best developed in consultation with multiple stakeholders, including the internet communications companies), and appropriate education for teachers in training as well as for those already working in the field. Another important challenge is to ensure that MIL and GCED interventions are designed holistically and include reference to issues such as hate speech, digital safety and security literacy, as well as identity, human rights and the political economy of media and internet companies' business models. Educational interventions also face the challenge of integrating understanding of the role of journalists and human rights defenders, especially the women and marginalised communities among them, who are frequently targeted in disinformation campaigns. In the bigger picture, it is also important to ensure that a society's emphasis is upon solving the root causes of the disinformation problem, rather than simply building resilience to it as if this was a stand-alone solution.

Lastly, the analysis of empowerment and credibility labelling responses presented here highlighted the important challenges connected with the need to tackle diversity, exclusion and impartiality with reference to who determines what is trustworthy and credible and how it is signalled. There are particular concerns about the implications of these systems for media plurality given the ways in which they can include or exclude media outlets (particularly if they depend upon access to particular technologies or skill sets), and the potential for them to be weaponised in the context of 'fake news' laws. The longer term monitoring, evaluation and updating of these systems was also identified as an area of concern.

Many of the responses to disinformation described in this study are still relatively new and have not yet been broadly adopted. In some cases, this is because certain technologies are still under development or adaptation, because they have not found broad traction, or because there are legal or ethical impediments. For example, when credibility and labelling approaches are not widely used, this clearly impacts on their effectiveness,

and limits the understanding of their potential. This illustrates Collingridge's dilemma (Collingridge, 1980), which essentially posits that the social consequences of technology often cannot be predicted until the technology has already been developed, at which point it is often too late, or at least much more difficult to change. Here, through a convergence of interdisciplinary expertise and pragmatic experience that is focused on a human rights-based assessment of responses to disinformation, an attempt has been made to respond to Collingridge's dilemma at a policy development level.

## 8.3 Applying the typology to the disinfodemic

In two UNESCO-published policy briefs about responses to COVID-19 disinformation, the typology of responses to disinformation detailed in this report was used to assess the applicability of the formulation to a specific disinformation crisis (Posetti & Bontcheva, 2020a; 2020b). In the context of the COVID-19 disinformation crisis, framed as a 'disinfodemic' by these reports, common and intersecting issues were analysed, and the underlying assumptions, challenges and opportunities of each of the response types were systematically dissected, providing an instructive framework for this more general and comprehensive study.

This process demonstrated the possibility of applying the framework to other specific disinformation crises, such as those associated with elections, anti-vaccination campaigns and climate change.

## 8.4 Cross-cutting assessment

The responses assessed in this study rest on underlying assumptions, some of which may be open to question and call out for scrutiny. They may be implicit rather than explicit in some cases, and in others they may serve to undermine the intended outcomes of the interventions. Some assumptions may be blind to issues of human rights, while others may incorrectly assume that there would not be unintended effects that harm these rights.

Many of the response modalities presented here seek to strengthen and increase the visibility of genuine public interest information (such as independent journalism, and legitimate public health information, or election safeguarding efforts). Others are aimed at quashing disinformation (or at least downgrading its prominence or pseudo-authoritative character), and there are also instances of responses designed to exert political control or resist regulation.

While the nature of the disinformation problem and its impacts may be assessed differently around the world, and by different actors, all of the interventions presented here are designed to effect change.<sup>486</sup> This is why they have implicit in them a ‘theory of change’. What they seek to change and why varies, and the reasons for action may be diverse. For instance, while news organisations fear the impact of disinformation on the value of their journalism (due to online attacks designed to discredit critical reporting, for example) and the business model implications of eroding trust, internet communications companies do not necessarily see disinformation as a problem of economics but rather as a public relations issue and a potential regulatory problem. Governments may wish to regulate for various reasons, one being because they are not satisfied with the companies’ responses, or because they see an opportunity to chill critical independent journalism through purported counter-disinformation efforts such as ‘fake news’ laws. Even though the ‘theory of change’ behind these interventions is not usually elaborated, the strengths and weaknesses of the particular theory being relied upon are fundamental to the efficacy of the interventions, as well as any unintended effects.

Application of the typology developed as part of this research to specific disinformation crises, such as the COVID-19 ‘disinfodemic’ case study referenced throughout this study, reveals the shortcomings of simply repurposing existing counter-disinformation responses (like those designed to protect elections, for example) to new types and manifestations of disinformation. In the case of COVID-19 disinfodemic responses, pre-existing intervention models (e.g. those applied to climate change denialism and the anti-vaccination movement) were assumed to be an adequate base for responses to the hyper-viral and extremely deadly pandemic-related disinformation. But in many cases, these were not fit for purpose due to the global scale, speed, and range of impacts associated with the pandemic which generated such immense confusion and uncertainty. Far more concerted and complementary interventions, across a wider range of actors, were needed in the face of the ‘disinfodemic’.

An underlying assumption in many initiatives rolled out in response to disinformation is that they in effect operate in terms of hunches about what is needed, and how an intervention is expected to work. This is because they operate in the absence of empirical evidence. Few actors dealing directly with disinformation appear to make provision for independent oversight or long term impact assessment, including monitoring and evaluation for unintended effects. Key among these risks is a long-term undermining of the right to freedom of expression, including press freedom, access to information, and privacy protections.

However, a further issue is that accountability for some of the responses is not always obvious or transparent. It is also apparent that many responses are not cognisant of international standards in terms of limitations to freedom of expression rights, in particular with regard to necessity and proportionality. Such overreach infringes the legitimate right to freedom of expression, and especially press freedom which is a precondition for the supply of information that can help overcome the challenge of disinformation.

These accountability issues were exacerbated in the context of the COVID-19 pandemic which demanded swiftly conceived responses rolled out under emergency conditions, in order to deal with an unprecedented global public health threat with massive social and economic ramifications amplified by the disinfodemic.

<sup>486</sup> Note: many of the insights and text that follows have been prepublished in the UNESCO Policy Series “Disinfodemic” <https://en.unesco.org/covid19/disinfodemic>



## 8.5 Taking stock of challenges and opportunities

- Time frames: Some responses - like new regulations - are geared towards immediate results, others such as user empowerment are more medium-term. Then, there are measures like developing critical Media and Information Literacy (MIL), which take longer to embed but which may have enduring outcomes. Others - like support measures for journalistic coverage designed to counter disinformation - are more time-specific. It is worth noting that different problems and opportunities operate within different time-frames.
- Complementarities: The 11 types of responses to disinformation modeled here are in many ways complementary to each other. They can be recognised as a holistic package of interventions. For example, in many cases, journalists have exposed online disinformation that had remained undetected (or unrecognised) by the internet communication companies enabling its transmission. In the bigger picture of responses, actions by these companies need to receive attention. This is because the use of power and policy, and the attention to audiences, are the categories of responses that cannot alone 'fix' the disinformation problem - they need to work in tandem with actions taken by the industry to stop transmission of disinformation.
- Contradictions: There are cases where one type of response can work against another. An example would be an imbalance whereby there is over-emphasis on having top-down regulation, while at the same time neglecting the need for bottom-up empowerment. Another example would be the resistance of internet communications companies to removing content associated with disinformation-laden attacks on journalists on the grounds of 'free speech'. This highlights a tension whereby 'free speech' can be justified as a reason to avoid responsibility for responding swiftly and decisively to disinformation that actively undermines press freedom (a corollary of freedom of expression) and journalism safety. In other words, preserving 'free speech' without observing and preserving press freedom rights (which include the need to protect journalists) is not a sustainable approach to mitigating overreach in responses to disinformation.
- Another tension would be the act of catching journalists in nets set for disinformation agents through the criminalisation of the publication or distribution of false information (e.g. via 'fake news' laws), precisely when journalism is needed to counter disinformation. It can also be noted that counter-disinformation content needs to coexist with, not compete with, nor be at the expense of, independent journalism. The different interventions therefore need to be aligned, rather than going in separate directions.
- Gender: There is gender-blindness in many of the responses to disinformation, which risks missing the subtle differences in how false content often targets people, as well as overlooking differences in the way people respond to the content concerned. It is also important to note that established patterns of behaviour by disinformation agents include gendered attacks online (ranging from abuse and threats of sexual violence to digital security and privacy breaches). There is also the issue of women and girls' access to information, which is often

restricted in certain contexts, and threatened by the presence of domestic violence, potentially limiting their access to counter-disinformation efforts.

- Age demographics, particularly regarding children and older people in response to the disinfodemic are also under-considered in many of the responses.

## 8.6 Overview assessment

Disinformation thrives in the absence of verifiable, trustworthy information. Equally, it can also flourish amid high volumes of content when people may find it difficult to distinguish credible information from disinformation, between what is a verified fact and what is not. It exploits people's need for sense-making of complex developments, as well as their fears, hopes and identities. This is why a multi-faceted approach is needed - one that also goes beyond the realm of communications and contested content, to include practical steps like social inclusion and solidarity, the reinforcement of ethics and values at the personal and community levels, and the embedding of peace-building principles within online communities. Any coherent strategy to fight the realm of information pollution also needs to recognise the value of securing a holistic and analytical approach to the problem.

In this wider context, it is evident that freedom of expression, access to information and independent journalism - supported by open and affordable internet access - are not only fundamental human rights, but also essential parts of the arsenal against chronic disinformation - whether connected to a pandemic, elections or climate change.

It should be noted that the fight against disinformation is not a call to suppress the pluralism of information and opinion, nor to suppress vibrant policy debate. It is a fight for facts, because without evidence-based information for every person, access to reliable, credible, independently verifiable information that supports democracy and helps avert worsening the impacts of crises like pandemics will not be possible.

## 8.7 Disinformation Responses: Freedom of Expression Assessment Framework

This 23-step assessment tool is designed to assist UNESCO Member States to formulate legislative, regulatory and policy responses to counter disinformation at the same time as respecting freedom of expression, access to information and privacy rights. The tool could be applied to proposed legislation and policy in development to assess - step by step - appropriateness in reference to international human rights laws and norms.

1. Have responses been the subject of multi-stakeholder engagement and input (especially with civil society organisations, specialist researchers, and press freedom experts) prior to formulation and implementation? In the case of legislative responses, has there been appropriate opportunity for deliberation prior to adoption, and can there be independent review?

2. Do the responses clearly and transparently identify the specific problems to be addressed (such as individual recklessness or fraudulent activity; the functioning of internet communications companies and media organisations; practices by officials or foreign actors that impact negatively on e.g. public health and safety, electoral integrity and climate change mitigation, etc.)?
3. Do responses include an impact assessment as regards consequences for international [human rights frameworks](#) that support freedom of expression, press freedom, access to information or privacy?
4. Do the responses impinge on or limit freedom of expression, privacy and access to information rights? If so, and the circumstances triggering the response are considered appropriate for such intervention (e.g. the COVID-19 pandemic), is the interference with such rights narrowly-defined, necessary, proportionate and time limited?
5. Does a given response restrict or risk acts of journalism such as reporting, publishing, and confidentiality of source communications, and does it limit the right of access to public interest information? Responses in this category could include: 'fake news' laws; restrictions on freedom of movement and access to information in general, and as applied to a given topic (e.g. health statistics, public expenditures); [communications interception](#) and targeted or mass surveillance; data retention and handover. If these measures do impinge on these journalistic functions or on accountability of duty-bearers to rights-holders in general, refer to point 4. above.
6. If a given response does limit any of the rights outlined in 4., does it provide exemptions for acts of journalism?
7. Are responses (e.g. educational, normative, legal, etc.) considered together and holistically in terms of their different roles, complementarities and possible contradictions?
8. Are responses primarily restrictive (e.g. legal limits on electoral disinformation), or there is an appropriate balance with enabling and empowering measures (e.g. increased voter education and Media and Information Literacy)?
9. While the impact of disinformation and misinformation can be equally serious, do the responses recognise the difference in motivation between those actors involved in deliberate falsehood (disinformation) and those implicated in unwitting falsehood (misinformation), and are actions tailored accordingly?
10. Do the responses conflate or equate disinformation content with hate speech content (even though international standards justify strong interventions to limit the latter, while falsehoods are not per se excluded from freedom of expression)?
11. Are journalists, political actors and human rights defenders able to receive effective judicial protection from disinformation and/or hateful content which incites hostility, violence and discrimination, and is aimed at intimidating them?
12. Do legal responses come with guidance and training for implementation by law enforcement, prosecutors and judges, concerning the need to protect the core right of freedom of expression and the implications of restricting this right?

13. Is the response able to be transparently assessed, and is there a process to systematically monitor and evaluate the freedom of expression impacts?
14. Are the responses the subject of oversight and accountability measures, including review and accountability systems (such as reports to the public, parliamentarians, specific stakeholders)?
15. Is a given response able to be appealed or rolled-back if it is found that any benefits are outweighed by negative impacts on freedom of expression, access to information and privacy rights (which are themselves antidotes to disinformation)?
16. Are measures relating to internet communications companies developed with due regard to multi-stakeholder engagement and in the interests of promoting transparency and accountability, while avoiding privatisation of censorship?
17. Is there assessment (informed by expert advice) of both the potential and the limits of technological responses which deal with disinformation (while keeping freedom of expression and privacy intact)? Are there unrealistic expectations concerning the role of technology?
18. Are civil society actors (including NGOs, researchers, and the news media) engaged as autonomous partners in regard to combatting disinformation?
19. Do responses support the production, supply and circulation of information - including local and multilingual information - as a credible alternative to disinformation? Examples could be subsidies for investigative journalism into disinformation, support for community radio and minority-language media.
20. Do the responses include support for institutions (e.g. public service messaging and announcements; schools) to enable counter-disinformation work? This could include interventions such as investment in projects and programmes specifically designed to help 'inoculate' broad communities against disinformation through media and information literacy programmes.
21. Do the responses maximise the openness and availability of data held by state authorities, with due regard to personal privacy protections, as part of the right to information and official action aimed at pre-empting rumour and enabling research and reportage that is rooted in facts?
22. Are the responses gender-sensitive and mindful of particular vulnerabilities (e.g. youth, the elderly) relevant to disinformation exposure, distribution and impacts?
23. If the response measures are introduced to respond to an urgent problem, or designed for short term impact (e.g. time sensitive interventions connected to elections) are they accompanied by initiatives, programmes or campaigns designed to effect and embed change in the medium to long term?

## 8.8 Comprehensive recommendations for action

The recommendations below build upon the chapter-specific recommendations on particular types of disinformation. They aggregate key points from the chapters in order to set out a full list of options for each individual stakeholder group in regard to the range of disinformation types. This gives an easy-to-use overview of the holistic range of actions which each stakeholder group can consider undertaking in order to optimise the effectiveness and freedom of expression dimensions of their responses. At the same time, partnerships within and across each stakeholder group are recognised as essential for success.

### **Cross-cutting recommendations aimed at all actors:**

- Encourage the strengthening of the range of diverse responses to disinformation, and ensure that these are all in line with international human rights standards.
- Facilitate and encourage coordinated, global multi-stakeholder cooperation and exchange of good practice across continents and states, towards effective implementation of holistic measures for tackling online disinformation.
- Encourage donors to invest specifically in countermeasures to disinformation that strengthen Media and Information Literacy, freedom of expression, independent journalism and media development.
- Increase official transparency and proactive disclosure of official information and data, and monitor this performance in line with the right to information and SDG indicator 16.10.2 that assesses the adoption and implementation of constitutional, statutory and/or policy guarantees for public access to information.
- Promote privacy-preserving, equitable access to key data from internet communications companies, to enable independent analysis into the incidence, spread and impact of online disinformation on citizens around the world, and especially in the context of elections, public health, and natural disasters.
- Invest in independent research into the fast-moving nature and scale of disinformation responses, as well as the need to address the challenges of studying new and rapidly evolving social platforms, including those received or perceived mainly as entertainment and social spaces (e.g. TikTok).

### **The Broadband Commission could:**

- Continue monitoring, measuring and assessing the impacts of responses to disinformation against human rights frameworks, including use of the assessment framework presented above.
- Encourage members who are 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 member companies to consider swift and decisive responses to political and electoral disinformation, as has happened in the field of COVID-19 related disinformation, with due regard to the difference between these two subject fields.
- Encourage its members to integrate this study into their activities, and to bring it to the attention of their stakeholders.

**Intergovernmental and other international organisations, as appropriate, could:**

- Increase technical assistance to Member States at their request in order to help develop regulatory frameworks and policies, in line with international freedom of expression and privacy standards, to address online disinformation. This could involve encouraging the uptake of the 23-step disinformation response assessment framework developed for this study.
- Particularly in the case of UNESCO with its mandate on freedom of expression, step up the work being done on disinformation in partnership with other UN organisations and the range of actors engaged in this space.
- Invest in researching, monitoring, measuring and assessing the impacts of responses to disinformation against human rights frameworks, including using the assessment framework presented here.
- Work together with States and NGOs towards Media and Information Literacy initiatives targeting potentially-vulnerable groups.
- Consider convening multilingual conferences, knowledge sharing, and workshops focused on Media and Information Literacy as a response to disinformation.
- Increase work in Media and Information Literacy and training of journalists as significant responses to disinformation.
- Increase support to media institutions in developing countries, including through UNESCO's International Programme for the Development of Communications (IPDC) to enable them to continue producing public interest journalism, and equipping them to combat disinformation.
- Support gender sensitive responses to disinformation.
- Encourage donors to invest specifically in countermeasures to disinformation that strengthen independent fact checking, Media and Information Literacy, freedom of expression, independent journalism and media development.

**Individual states could:**

- Actively reject the practice of disinformation peddling, including making a commitment not to engage in public opinion manipulation either directly or indirectly - for example, via 'influence operations' produced by third party operators such as 'dark propaganda' public relations (PR) firms.
- Review and adapt their responses to disinformation, using the 23-step framework for assessing law and policy developed as an output of this study, with a view to conformity with international human rights standards (notably freedom of

expression, including access to information, and privacy rights), and at the same time making provision for monitoring and evaluation of their responses.

- Increase transparency and proactive disclosure of official information and data, and monitor this performance in line with the right to information and SDG indicator 16.10.2 that assesses the adoption and implementation of constitutional, statutory and/or policy guarantees for public access to information.
- Promote affordable connectivity for all in line with UNESCO's concept of 'Internet Universality and the four ROAM principles (Rights, Openness, Accessibility and Multi-stakeholder participation).
- Support transparent social media councils and/or national ombuds facilities in order to help give users recourse to independent arbitration and appeals against moderation steps perceived to be unfair.
- Support investment in strengthening independent media, including community and public service media, as the economic impacts of the COVID-19 crisis threaten journalistic sustainability around the world.
- Earmark funding and support for Media and Information Literacy focused on combatting disinformation, especially through educational interventions targeting children, young people, older citizens, and potentially-vulnerable groups.
- Work with internet communications companies to establish privacy-preserving, secure data exchanges and facilitate access to social media data for journalists, and academic researchers, and NGO-based researchers where appropriate, to enable thorough investigations and preservation of historically-important data (especially as associated with elections, pandemics and other important flashpoints).
- Avoid criminalising disinformation to ensure legitimate journalism and other public interest information is not caught in the nets of 'fake news' laws, etc.
- Ensure that any legislation or regulation responding to disinformation crises, like the COVID-19 'disinfodemic', is necessary, proportionate and time-limited.
- Develop mechanisms for independent oversight and evaluation of the freedom of expression implications and efficacy of 'fake news' legislation, along with other relevant national policies and normative initiatives.
- Ensure gender sensitivity in their strategies and public responses to disinformation
- Encourage the uptake of the recommendations below for political parties and actors in reference to elections and campaigning.

**Electoral regulatory bodies and national authorities could:**

- Strengthen legal measures concerning privacy protection, freedom of expression and political advertising in order to better protect against electoral disinformation.
- Improve transparency of election advertising by political parties, candidates, and affiliated organisations through requiring comprehensive and openly available ad 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.
- Educate and empower citizens to detect and report disinformation during elections.
- Develop voter literacy through linking civics literacy with digital citizenship education and Media and Information Literacy.
- Work with journalists and researchers in fact-checking and investigations around electoral disinformation networks and producers of 'dark propaganda'.

#### **Political parties and other political actors could:**

- Speak out about the dangers of political actors as sources and amplifiers of disinformation and work to improve the quality of the information ecosystem and increase trust in democratic institutions.
- Refrain from using disinformation tactics in political campaigning, including the use of covert tools of public opinion manipulation and 'dark propaganda' PR firms.
- Consider following in the footsteps of political parties in recent elections where the contestants pledged to avoid disinformation.<sup>487</sup>
- Commit to transparency and accountability regarding scrutiny by critical journalistic actors and other mechanisms supporting open societies, and condemn threats against journalists including the use of disinformation as a weapon against the news media.
- Submit their online political adverts to independent fact-checking processes.

#### **Law enforcement agencies and the judiciary could:**

- Ensure that law enforcement officers are aware of freedom of expression and privacy rights, including protections afforded to journalists who publish verifiable information in the public interest, and avoid arbitrary actions in connection with any laws criminalising disinformation.
- For judges and other judicial actors: Pay special attention when reviewing laws and cases related to addressing measures to fight disinformation, such as criminalisation, in order to help guarantee that international standards on freedom of expression and privacy are fully respected within those measures.

<sup>487</sup> In Uruguay, political parties in 2019 agreed a pact to refrain from disinformation; to avoid actions or expressions that use aggravating tones against adversaries; and to set up a consultation mechanism when threats or challenges arise to fulfilment of their agreement. <https://www.undp.org/content/dam/uruguay/docs/GD/undp-uy-pacto-etico-definformacion.pdf/> In Germany, political parties committed to avoiding social media 'bots' and microtargeting. <https://www.bpb.de/gesellschaft/digitales/digitale-desinformation/290568/relevanz-und-regulierung-von-social-bots>



### Internet communications companies could:

- Intensify multi-stakeholder engagement and transparency about their policies in general and application thereof, including their responses to disinformation.
- Implement their responses on a global scale, rather than being limited to certain countries, and ensure coverage in all significant languages.
- Provide more financial support to: independent fact-checking networks, independent journalism (especially those focused on investigations targeting disinformation content and networks, and also to local news organisations which are particularly fragile), and independently-provided/delivered Media and Information Literacy initiatives.
- Avoid interventions that appear designed primarily as public relations or brand management exercises, make contributions with 'no strings attached', and improve transparency related to such funding.
- Support independently managed funds for research into cases of disinformation, its impact, and responses to it, including independent evaluations of the effectiveness of companies' own disinformation responses. Ensure a diversity of funding recipients along with transparency regarding the research methods and findings.
- Work together, through a human rights frame, to deal with cross-platform disinformation, in order to improve technological abilities to detect and curtail problems of false and misleading content more effectively, and share data about this.
- Develop curatorial responses to ensure that users can easily access journalism as verifiable information shared in the public interest, prioritising news organisations that practice critical, ethical independent journalism.
- Work to boost the visibility of credible news content and financially compensate news producers whose content benefits their businesses, especially as many news organisations removed paywalls and other barriers to content access during the COVID-19 pandemic as a counter-disinformation measure.
- Avoid overreliance on automation for content moderation, recognise the need to expand human review capacity and remedies for redress, and transparently monitor these matters.
- Ensure appropriate pay, training and psychological support for the people working in content moderation.
- Recognise that if health disinformation and misinformation can be quickly dealt with in a pandemic on the basis that it poses a serious risk to public health, action is also needed against political disinformation - especially at the intersection of hate speech – when it too can be life-threatening. The same applies to disinformation related to climate change.
- Recognise that press freedom and journalism safety are critical components of the right of freedom of expression, meaning that online violence targeting journalists (a frequent feature of disinformation campaigns) cannot be tolerated.

- Apply fact-checking to all political content (including advertising, fact-based opinion and direct speech) published by politicians, political parties, their affiliates and other political actors.
- Produce detailed and frequent public transparency reports, including specific information on identification of the origins, scale, views, flow and types of disinformation, removals of disinformation, demonetisation of disinformation content, and suspension of accounts spreading disinformation, as well as provide information on other curational steps such as labelling and appeals.

#### **The media sector could:**

- Redouble their efforts as professional frontline responders to disinformation, through increased investment in fact-checking, debunking, disinformation investigations, and ensuring robust lines of questioning about responses to disinformation, as well as by enhancing accountability and transparency with regard to political actors, states, institutions, and the corporate sector.
- Report on the human rights implications of responses to disinformation, including those impacting on freedom of expression and access to information, as well as privacy rights.
- Consider mythbusting and investigative collaborations into disinformation with other news organisations and audiences, including internationally. Partnerships with researchers and civil society organisations can also be successful.
- Focus innovation efforts on countering disinformation through accessible and engaging story formats, such as infographics and podcasts along with collaborative, data-driven investigations.
- Ensure that experiences in a range of developing countries are not overlooked in coverage of disinformation and responses to it.
- Ensure preparedness of staff for safety risks associated with reporting on disinformation, e.g. increased security threats, online abuse, physical attacks, and ensure gender sensitivity in responding to these dangers.
- Undertake coverage of the issues of transparency, accountability and independence of institutions and individuals engaged in fact-checking and/or evaluation of the credibility of sources of information.

#### **Civil society could:**

- Reinforce the call for responses to disinformation to conform to international human rights standards.
- Partner with journalists, news organisations and researchers on investigative and monitoring projects about disinformation and responses to it.
- Strengthen the roll-out of Media and Information Literacy projects, and of programmes that support independent journalism.

- Consider programmes targeting children as well as older citizens who are under-served by Media and Information Literacy campaigns, and therefore more susceptible to exploitation by disinformation agents.
- Produce counter-content and campaign against disinformation.

**Researchers could:**

- Strengthen their scientific enquiry agendas to focus on disinformation, the responses to it, and the impacts of these responses.
- Study under-researched formats such as interactive gaming where disinformation and countermeasures may effectively target young people.
- Undertake Participatory Action Research projects that respond to critical incidents connected to disinformation, and can also provide urgent knowledge.
- Collaborate with journalists, news organisations, and civil society groups on projects that help surface and combat disinformation, along with monitoring and assessment exercises focused on responses to it.
- Study cross-platform disinformation campaigns to get a more rounded, holistic perspective on the problem and responses to it.
- Pursue independent, longitudinal, quantitative and qualitative monitoring and evaluation of disinformation responses implemented by the internet communications companies.
- Develop new technological tools to assist journalists and other verification professionals in detecting and analysing disinformation, ensuring also that such tools put freedom-of-expression, privacy, algorithmic transparency and accessibility at their core.
- Use the typology and assessment framework developed through this study to analyse various types of disinformation responses as they emerge and evolve, and assess their efficacy and impacts with specific reference to freedom of expression challenges.
- Invest in studies that address the impacts of disinformation on consumers, including behavioural science investigations that build knowledge about what motivates people to share and/or give credence to disinformation.
- Prioritise studies targeting users' behaviour in relation to engagement with, and redistribution of, credible, verified information such as that produced by independent news publishers and journalists.
- Ensure female experts are visible as a way of addressing gender inequalities in international debates on disinformation.