**UNESCO input to the Report on Disinformation of the**

**Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan**

**June 2021**

**QUESTION #1**

**What do you believe are the key challenges raised by disinformation? What measures would you recommend to address them?**

The topic of disinformation itself raises a major challenge in that it is very often coupled with a strong connotation that such content is automatically and intrinsically harmful. Missing in such a perspective is an explication of the object/s of harm (especially which human rights may potentially be harmed), and of the nuance provided by the Rabat Plan of Action on Hate Speech, as to when potential risk may become actual danger. Such assumptions can point the whole discourse on expression towards a position that can be one-sidedly restrictive. At the extreme, the focus can then become one whereby it is freedom of expression that is to blame for enabling such an egregious problem to surge into society, and therefore that there is an underlying problem (the right to free expression) that needs to be addressed in combating disinformation.

The alternative perspective, however, would be to recognise that free expression is not only a fundamental and protected human right, but it is also a primary means to combat disinformational expression (as elaborated below). Accordingly, disinformation’s potential harmful impact should be understood as encompassed potential damage to the basic right to freedom of expression, and likewise it then needs to be underlined that responses to disinformation should deal with specific problems without causing any undue limitation to the basic right as such.

Relevant to this starting point, and to a wider assessment of key challenges and responses, is the Cross-Regional Statement on “Infodemic” in the Context of COVID-19 endorsed in June 2020 by more than 130 United Nations member countries and official observers, which recommended that: “[counter disinformation] efforts are based, *inter alia*, on freedom of expression, freedom of the press and promotion of highest ethics and standards of the press, the protection of journalists and other media workers, as well as promoting information and media literacy, public trust in science, facts, independent media, state and international institutions.”[[1]](#footnote-2)

Extensive work has been done in these areas by a range of actors. One significant contribution in unpacking the challenges and responses is a study for the ITU/UNESCO Broadband Commission for Sustainable Development. The Commission established a Working Group on Freedom of Expression and Addressing Disinformation, co-chaired by UNESCO, which published in September 2020 a comprehensive 350-page report entitled *Balancing Act: Countering Digital Disinformation while respecting Freedom of Expression*.[[2]](#footnote-3) The Balancing Act report identifies 11 categories of responses, and assesses their strengths and weaknesses – ranging from legislative through to educational responses. This evaluation is based also on analysis of the challenges of disinformation per se.

The insights from the Balancing Act and from UNESCO’s other work in this area, reveal a foundational set of challenges related to the definition of disinformation:

* The way the problem is defined and understood leads to different responses. Understood as misinformation, the focus is put on users and the need to educate them. Understood as deliberate disinformation, the focus is drawn to the producers who instrumentalise untruth and how to disincentivise them. Understood as both, there is attention to dissemination, especially behaviours of sharing, orchestration and algorithmic amplification and recommendation. UNESCO’s use of the term “disinformation” as a chapeau covers false and misleading content as regards production, reception, transmission, and retransmission. It is also recognised that like the term “fake news”, the term “disinformation” can be instrumentalised to designate a vast range of bodies of content, even when these are truthful in character. This means that actors can misuse the label of fighting disinformation for diametrically opposed purpose. Therefore definitions, and their application, are an arena of potentially high contestation, and equally responses are likely to align accordingly.
* A further challenge in regard to definitions is that of the unit of analysis. While the focus is often on single untruths, there is also the matter of narratives built and evolved in a fabric of meaning – such as with conspiracy theories that can rest upon true facts, but false logic. This affects the methodology of what responses are appropriate (e.g. fact-checking can be relevant for single units of disinformation; media and information literacy is more relevant to understanding narrative. It is important that disinformation analysis examines more than individual items, even although the falsity of these (often in combination with some truths) can well constitute the building blocks of more sustained meanings which are overall misleading by standards of rigour, logic, scientific method and the body of evidence.
* A further definitional challenge is the narrowing of disinformation to “fake news”, which term detracts from the full range of genres and formats beyond news as such, where false and/or misleading content is found. A format-agnostic understanding is necessary, also to understand the semantic, cognitive and affective complementarities that are exploited by organised disinformation.
* In addition, there is often conflation of concepts such as hate speech and disinformation. While such expressions can be, and often are, combined in a toxic package, they are at least conceptually distinct. This can be seen in anti-vaxx disinformation which does not intrinsically incite hostility to particular groups. Therefore, while some responses to disinformation will also implicate responding to hatred, and vice versa, there are also distinctive dimensions to these potentially harmful types of expression. Not least in this regard is that international human rights standards tolerate a level of false expression without restriction (unless such expression harms other human rights – such as reputation, property ownership, safety, etc.), whereas restrictions on types of hate are well mandated.
* Finally, it is valuable to note that the interpretation of definitions of disinformation is subject to power, such as via companies’ payment and remit of fact-checkers, or the wording of regulation and law. This becomes particularly challenging in regard to the full spectrum of meaning between truth and falsity. A simple binary perspective means that the range of unknowns and of legitimately contested meanings is effectively ruled out of existence. An arbiter of truth is the logical outcome of such a power approach. A great deal of discourse does not lend itself to interpretation as being either true or false, and there is a risk that both legitimate debate and scientific uncertainties can be suppressed in the power play of implementing responses to “disinformation”.

The need for action against disinformation that will be consistent with freedom of expression and other human rights is strongly recognised by the “Balancing Act” report noted above. The same report highlights the challenge of responses to disinformation that often risk violation of the right to freedom of expression. The report assesses that while the right to free expression is often blamed as being an enabler of disinformation, it is in fact the same right that is essential to combat disinformation. Examples include the case of journalistic exposes of both orchestrated disinformation and the problematic role of internet companies as vectors. Freedom of expression in the public sphere has enabled civil society groups to call out disinformation whether promoted knowingly by powerful political, state and corporate actors, as well as to challenge falsehoods that are innocently believed and shared by ordinary individuals. Without this point being strongly made, there is a danger that freedom of expression is treated as the problem, rather than as both a right and a solution. In turn, that approach would push in the direction of disproportionate and illegitimate restriction of freedom of expression. The default should always remain as unrestricted expression, with any curbs being exceptions and justifiable under the strict conditions of the ICCPR. To exceed these conditions is to play into the hands of disinformation which can thrive in the absence of freedom of expression.

Further in the Balancing Act report, it is recognised that a particular challenge lies in the generally aggressive character of disinformation. Because disinformation needs to displace truth (as well as recognised unknowns), it frequently operates in contestation mode. Its targets are truth-tellers – in particular those who are political actors, news media members, scientists, and public leaders. This observation is not to over-characterise disinformation as “information warfare” (although it may sometimes be part of such a geopolitical context). Instead, it is important to avoid a reductionism that lends itself to militaristic-style responses. The main point to recognise here is that disinformation does not exist as if it were simply a parallel and alternative universe of content, in relation to which autonomous subjects are positioned with full knowledge and capacity to assess and evaluate its offerings in relation to more traditional sources of credible and reliable content. Disinformation exploits unequal markets of information, and indeed seeks to enhance these. In brief, disinformation has an intrinsic tendency towards combat, to discredit, and hence also often to harass and intimidate. In this way, disinformational expression often seeks to suppress targets’ right to freedom of expression. Accordingly, disinformation not only threatens sustainable development in all dimensions (health, gender equality, climate change mitigation, etc), but also the right to freedom of expression itself.

The Balancing Act report registers two additional challenges. One is the fragmentation of responses. An illustration is that responses to COVID-19 disinformation too frequently involve silo-ised interventions, without recognition that many of the same disinformational actors and/or techniques are also evident in other spheres like elections. There is also fragmentation which means that restrictive responses are promoted (e.g. take-downs of social media content), without regard to enabling and empowering responses (e.g. Support for news media; Media and Information Literacy programmes). Finally, while disinformation exploits a range of platforms, there is limited shared of the information that is needed to effectively counter it.

A further and highly significant challenge noted in the Balancing Act report is that of monitoring both the problems and the attempted solutions. Without such information, it is evident that interventions are operating with sub-optimal knowledge. The lack of monitoring, however, is not just an oversight or omission. Many responses lack a theory of change in which monitoring would be an evident step, but even more they lack the data that could enable them to make more effective diagnosis and better evaluation and assessment of their interventions. The major problem here is the opacity of internet companies as the transmission engines of disinformation. Efforts to increase the supply of quality information (such as public service health announcements, the UN’s Verified initiative), or to “clean up the mess” at the receiver end (such as raising media and information literacy about emerging vaccination falsehoods), are confined to working in the dark about what is happening in the “black box” of transmission. Data that companies release are partial and self-serving, and the shortcomings are continuously revealed by journalists and NGOs who expose continuing problems notwithstanding the publicity about actions taken. It is not possible to evaluate the claims by the companies about their actions and the impact thereof, in the absence of disclosure that would allow for contextual evaluation. There is thus a dual challenge for responses to disinformation – a lack of monitoring, and a lack of transparency that could enhance monitoring.

The *Balancing Act* report recommends actions that align all stakeholders to respecting international standards for freedom of expression and other human rights, including access to information and privacy. The report has proposed a 23-step tool ([page 14-17](https://www.broadbandcommission.org/Documents/working-groups/ExecSum_FoE_disinfo_report.pdf)) to assess disinformation responses from a human rights perspective and introduces a novel typology of 11 responses to disinformation, making holistic sense of the disinformation crisis on an international scale.

Other UNESCO-led initiatives and resources in 2020 have addressed key challenges specifically in response to Covid-19 disinformation. These illustrate the importance of responses that cover research, policy advice to diverse stakeholders, awareness-raising and capacity-building:

* In April 2000, UNESCO launched two policy briefs offering critical insights into the fast-growing COVID-19-related ‘disinfodemic’.[[3]](#footnote-4) Adapting the typology of the Broadband Commission report, while the report was under preparation, the policy briefs assess available responses holistically and in relation to impact on the rights to freedom of expression, access to information and privacy. The phenomenon of COVID-19 disinformation, its spread of social media, and responses to it were also addressed in UNESCO publications on journalism, press freedom and COVID-19, and on media independence.[[4]](#footnote-5)
* UNESCO, in partnership with the Innovation for Policy Foundation (i4Policy), launched the #DontGoViral online campaign in Africa to invite artists to amplify life-saving information to combat the COVID-19 pandemic among fans and followers, and engage with cultural and creative industries in awareness-raising.[[5]](#footnote-6) In three months, the campaign reached over 220 million people across the African Continent. The campaign addressed the urgent need to mitigate the spread of COVID-19 on the African Continent through culturally relevant and open-licensed information in local African languages. Partnerships were established in 2020 with both African media outlets and global outlets, notably the BBC World Service and France 24.
* UNESCO, with support from the European Union, launched a global project to strengthen the resilience of states in the face of the Covid-19 pandemic. The project is helping societies address the challenge of false information on the coronavirus. The “#CoronavirusFacts: Addressing the ‘Disinfodemic’ on COVID-19 in conflict-prone environments” project produced a series of relevant outputs, including resources hubs in several languages and a massive open online course (MOOC) taken by 9,000 journalists from 162 countries.[[6]](#footnote-7) A second MOOC for journalists in covering vaccines is currently under development, together with the World Health Organization (WHO) and the Knight Center for Journalism in the Americas at the University of Texas, as well as other major initiatives for fact-checkers in the Latin America and Caribbean region.[[7]](#footnote-8)
* In October 2020, UNESCO and the Republic of Korea launched a global competition HackingDisinfodemic inviting young people worldwide to design innovative solutions to misinformation and related online challenges, in partnership with WHO, the United Nations Population Fund, the UNESCO Mahatma Gandhi Institute of Education for Peace and Sustainable Development (MGIEP), and IBM.[[8]](#footnote-9) The hackathon was organized in the framework of Global MIL Week 2020 (24-31 October), under the theme of “Resisting Disinfodemic: Media and Information Literacy for Everyone and by Everyone.”[[9]](#footnote-10)

**QUESTION #2**

1. **What legislative, administrative, policy, regulatory or other measures have Governments taken to counter disinformation online and offline?**

The Balancing Act report [Chapter 5](https://en.unesco.org/sites/default/files/5_ecosystem_responses_aimed_at_producers_and_distributors_96_139_balancing_act_disinfo.pdf) includes Table 3 (pages 103-107) which identifies current legislative, pre-legislative, and policy responses (as of May 2020) across ten categories of response by national and intergovernmental actors, namely:

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

Chapter 5 shows that at least **28 countries** [as of April 2020] had passed legislation related to disinformation, either updating existing regulations or passing new legislation.

[Chapter](https://en.unesco.org/sites/default/files/5_ecosystem_responses_aimed_at_producers_and_distributors_96_139_balancing_act_disinfo.pdf) section 5.1.6 (page 110) highlights specific examples of national laws or regulations governing disinformation surrounding the COVID-19 pandemic.

[Chapter](https://en.unesco.org/sites/default/files/5_ecosystem_responses_aimed_at_producers_and_distributors_96_139_balancing_act_disinfo.pdf) section 5.2.4 includes Table 4 (page 114-118) which lists national examples of counter-disinformation campaigns.

Legal measures to protect election integrity are discussed in the report specifically in Chapter 5.1 and Appendix A (p.322), under two dedicated sections - one on legislative proposals and another on adopted legislation.

The report illustrates how electoral commissions and government committees have sought to provide reliable information on candidates and parties, as well as work with the internet communications companies towards the promotion of such information. Another important kind of regulatory response targets transparency and integrity of online adverts during election periods.[[10]](#footnote-11) The response of Internet shutdowns to disinformation is described as generally being a disproportionate restriction on freedom of expression and access to information, and even counter-productive to electoral credibility.[[11]](#footnote-12) Some governments have enforced these in the run up to polls saying they are seeking to protect citizens from electoral disinformation and propaganda.

Broadly speaking, the scope of established legislation varies from media and electoral laws to cybersecurity and penal codes. 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. In some cases, regulations avoid content interference by focussing more on process issues, including requirements for clear policy and opportunities for redress by users.

In some cases, in particular where disinformation is defined broadly or where provisions are included in general penal codes, there is a significant risk of censorship, according to the report.

Ensuring that laws and policies aimed at countering disinformation respect freedom of expression requires greater understanding among judicial operators of international standards related to freedom of expression. Among resources available include guidelines for the role of judicial operators in the protection and promotion of the right to freedom of expression during the COVID-19 pandemic, including in addressing laws aimed at curbing COVID-19 disinformation.[[12]](#footnote-13)

Among the panoply of government responses, there are also several policy responses that specifically include support to news media as defence against information.

(The information above is derived from the [Balancing Act report.](https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf)

**What has been the impact of such measures on i) disinformation; ii) freedom of opinion and expression; and iii) other human rights?**

One major challenge to understanding the full impacts of such measures on disinformation and on freedom of expression and other human rights is a lack of data and information. Such gaps in knowledge could be improved through greater transparency from internet platforms about the scale of disinformation on their platforms and the effectiveness of counter measures.

While the Balancing Act report does not attempt to quantify or evaluate impacts, it outlines the theory of change underpinning Member State responses, including based on what and/or whom the targets are. This is vital if impact were to be assessed in terms of the intention of the measures being taken.

In the view of the report, the theories of changes that underly the actions of Member States to curtail disinformation differ by target group:

-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

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

-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 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.

Although comprehensive impact has been hard to assess, it is seems that a number of responses have been insufficient to reduce disinformation, which assumes a “whack-a-mole” character, and which also requires resources which can cost money. The scale at which the Internet companies have been able to develop without having to invest in information quality, makes it predictable that their measures (assuming they are prepared to pay for them) will have limits. Indeed, journalists continuously expose the continuing character of disinformation (and other potentially harmful content) notwithstanding companies’ claims to have treated a vast volume. What is also clear is that ways to address disinformation algorithmically have caught non-disinformational content in the net, thereby diminishing the availability of such content. Meanwhile, it appears that curational logics of elevating and recommending that content which attracts engagement (and enragement), work at cross purposes to moderational efforts. In other words, the effectiveness of corporate responses to disinformation is evidently limited.

The impact of governmental efforts, directly and indirectly, on combating or otherwise reducing disinformational content, is not clearly tracked. However, as shown in the Balancing Act report, there are numerous cases of governmental efforts both spreading disinformation, as well as harming those who are part of the solution to disinformation – such as independent and critical journalists. Thus, a number of arrests on the grounds of spreading disinformation have been pointed out by Human Rights organizations as arbitrary, and as harnessing disinformation to limit free speech. There are clear risks to privacy, when such actions are based on authorities accessing data without just cause. 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.

(The information provided above is excerpted from the [Balancing Act report).](https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf)

**c.   What measures have been taken to address any negative impact on human rights?**

In addition to legislative responses, the [Balancing Act report](https://www.broadbandcommission.org/Documents/working-groups/FoE_Disinfo_Report.pdf) analyses the use of algorithms by internet companies (e.g. social media platforms) in order to scale the identification and evaluation of disinformation.

This report highlights a study by Ranking Digital Rights that looked into the issue of transparency in relation to algorithm-based recommendation engines. The study reviewed five internet companies including Apple (iOS), Google (Search, YouTube, Android), Facebook (Facebook), Microsoft (Bing, OneDrive) and Twitter, and found governance gaps and weak human rights due diligence. It notes that “none of the five U.S.-based platforms evaluated make explicit public commitments to protect human rights as they develop and use algorithmic systems” and that “companies operating major global platforms do not provide evidence that they are conducting risk assessments that enable them to understand and mitigate human rights harms associated with how their use of algorithmic systems and targeted advertising-based business models affect internet users”. Only one U.S. company (Microsoft) disclosed that it conducts impact assessments on its development and use of algorithmic systems. None of the eight companies in the study disclosed whether they conduct risk assessments on how their targeted advertising policies and practices affect users’ freedom of expression and information rights, or their right to privacy or to non-discrimination.

Although most major internet companies now produce transparency reports on levels of content or account takedowns as well as investigatory reports on outcomes in countering particular disinformation campaigns, these reports do not include in-depth transparency on the implications of their use of algorithms, machine learning and other forms of automated decision-making in regard to human rights. Nor do they explain on what criteria these methods are considered effective interventions. There is scant information on the sharing and circulation of disinformational content prior to its moderation, and nor on the implication of recommendation and content-elevation algorithms as a factor in this.

The absence of deeper transparency on usage of algorithmic systems, or on implementation of human rights due diligence prevents effective external evaluation of their effectiveness in countering disinformation or their impact on freedom of expression and other rights. Transparency reports provide aggregate figures on enforcement around for example, false accounts, but do not provide detail.

It is evident that most internet companies make explicit decisions in their policies around how they understand freedom of expression, which typically fails to reference international human rights law as a guiding standard. These policies have cascading implications into the design and application of algorithms and other automated systems, and in decision-making around what is escalated to human review. Despite some improvements in overall human rights due diligence within policies by internet, search and messaging companies, important gaps still remain. These issues have elicited criticism for failure to systematically invest in impact assessments that thoroughly engage with civil society and other stakeholders as the companies enter new markets/societies with existing products. Similarly, the companies are criticised for not evaluating emerging risks in existing markets. This is compounded by the lack of transparency which complicates external oversight on platforms and their algorithms, including access to better evaluation data on successful identification as well as identified false positives and false negatives. Additionally, the companies are criticised for not engaging in “abusability testing”, where platforms invest resources into seeing how their platforms can be abused to harm consumers.

While recognising the role that internet communications companies need to play in curtailing disinformation published on their platforms, there are potential issues with having regulatory power informally delegated by States to these private sector companies, according to the Balancing Act report. 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 standards. This can amount to privatised censorship. Where delegation is explicitly provided by regulations, there can be public accountability for these regulations in democracies which respect the rule of law and issues of necessity and proportionality.

Tricky terrain is entered into when regulations criminalise disinformation, particularly when these are vague and/or disproportionate in terms of international human rights standards. However, consumer regulation about data protection and the ability to appeal decisions, as well as regulation for transparency companies report on how decisions are taken, could be less complex from a freedom of expression point of view.

(The information provided above is excerpted from the Balancing Act report).

**QUESTION #3**

1. **What policies, procedures or other measure have digital tech companies introduced to address the problem of disinformation?**

Further information is presented notably in [Chapter 6](https://en.unesco.org/sites/default/files/6_responses_within_production_and_distribution_140-201_balancing_act_disinfo.pdf) of the Balancing Act report. Below are selected examples addressed in the report that may be of interest to the Rapporteur.

**Connecting users to reliable information**

Some internet communications companies (e.g. Facebook, YouTube, Instagram, WhatsApp, Twitter, Linkedin) themselves have taken action to connect their users to reliable information about the pandemic by linking any query on the coronavirus to the World Health Organization (WHO) main hub and their WHO mythbusters page, or to the local government’s ministry of health.

**Supporting journalism and partnering with fact-checking organizations**

Facebook merits attention as the only large-scale international “third party verification” programme among the internet communications companies, which was launched shortly after the 2016 U.S. presidential election. The ‘Balancing Act’ report analyses it on pages 72-78 (Chapter 4).

Apart from founding partner Google News Initiative, First Draft has also obtained grants and donations from many philanthropic foundations as well as support from the Facebook Journalism Project and Twitter.

WhatsApp, Facebook, Google, and Twitter have pledged some funding to fact-checking organizations. Ongoing support throughout and beyond critical periods of elections and pandemics is needed. Verifying claims about vaccinations and climate change is particularly significant going ahead.

**Banning “coordinated inauthentic behavior”**

Facebook banned in 2019 a number of dark PR firms for attempting to influence elections, or for what it calls “coordinated inauthentic behavior” in various countries. However, these kinds of activity are still widespread, and new companies promoting such services can be easily set up.

**Encouraging research on disinformation, but with restricted data**

While some companies have encouraged research into disinformation, there is reluctance to make their data available for this purpose. For example, Facebook has announced $2m for research into “Misinformation and Polarisation” – although with the proviso that “No data (Facebook, Messenger, Instagram, WhatsApp, etc.) will be provided to award recipients”.

Twitter has explained its data disclosure policy in an article by its Head of Site Integrity, and Facebook has been criticised by researchers for delays in providing data access but has recently released a larger data set in line with its commitments. Another issue is the restriction of access to a limited number of researchers, who are also frequently the recipients of large grants from these companies.

**Changing advertising policies**

Twitter CEO Jack Dorsey announced that the platform would stop running all political advertisements commencing November 22, 2019.

**Content moderation and curation:**

With regard to curatorial including algorithmic responses [[Chapter 6](https://en.unesco.org/sites/default/files/6_responses_within_production_and_distribution_140-201_balancing_act_disinfo.pdf)], the Broadband Commission report encourages internet communications companies to:

* Work together in a human rights frame, to deal with cross-platform disinformation, in order to improve technological abilities to detect and curtail 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 organizations that practice critical, ethical independent journalism.
* 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 internationally enshrined 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.

(The information provided above is excerpted from the Balancing Act report).

1. **To what extent do you find these measures to be fair, transparent and effective in protecting human rights, particularly freedom of opinion and expression?**

After assessing the responses to disinformation, the *Balancing Act* report recognises shortfalls in the protection of human rights such as freedom of expression – as well as privacy. This is basis on which the report provides specific recommendations for the consideration of individual States:

* 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.

* Invest in monitoring, measuring and assessing the effectiveness of electoral

responses to disinformation.

* Engage more closely with civil society organizations, news organizations, 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.
* 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.

(The information provided above is excerpted from the Balancing Act report)

  **c.            What procedures exist to address grievances and provide remedies for users, monitor the action of the companies, and how effective are they?**

The ‘Balancing Act’ report notes that 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. At the same time, as indicated above, there are severe limitations to the transparency of the companies, which do not enable effective monitoring. (As a recent case in point, a New York Times journalist reported that use of Facebook’s Crowdtangle data portal showed that extremist content was receiving the most engagement. The company responded to the effect that a more important metric was “reach” rather than “engagement” and that mainstream media content had far greater reach than extremist content. But this data was reportedly not available through Crowdtangle).

Redress actions can be taken on the basis of existing law and community standards. Yet, a major limitation in the compliance of social media companies with national regulation needs to be noted, as they operate globally and do not necessarily fall into the legal frameworks of the jurisdictions where they operate. There operations are also geographically very uneven in terms of resourcing to deal with disinformation in different languages and different countries.

The report recommends that internet communication companies:

• Provide detailed and frequent public transparency reports, including specific

information on the viewing and spread of disinformation, suspension of accounts

spreading disinformation, removals and other steps against disinformation,

including demonetisation, as these responses can have significant human rights

and freedom of expression implications.

• Establish robust third party/external review mechanisms for content moderation

and ensure the ability to appeal decisions, including machine-driven ones. This

includes the need to review decisions not to remove content, as well as decisions

to delete it.

• Ensure that curatorial responses encourage users to access journalism from

independent and professional news organizations or others publishing critical,

evidence based public interest information (e.g. independent researchers and bona

fide civil society organizations).

• Increase their efforts against orchestrated disinformation-laced attacks on

journalists by excluding users who are part of such assaults on press freedom and

act as obstacles to efforts to counter disinformation.

• Take steps to ensure appropriate support for content moderators, including training, commensurate wages for work done, and provision for psychological health.

**QUESTION #4    Please share information on measures that you believe have been especially effective to protect the right to freedom of opinion and expression while addressing disinformation on social media platforms.**

The Balancing Act **recommendations** are based on steps considered to be effective as regards identifying the roles of individual states, political actors, electoral regulatory bodies and law and judiciary agencies, and social media platforms. Accordingly, some of the measures listed are set out below.

States could :

* Review and adapt their responses to disinformation, using the 23-step tool for assessing law and policy developed as an output of the Commission’s report, with a view to conformity with international human rights standards (notably freedom of expression, including access to information, as well as privacy rights), and at the same time making provision for monitoring and evaluation of their responses.
* 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.
* 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.

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’ public relations firms.

Electoral regulatory bodies and national authorities could:

* Improve transparency of all election advertising by political parties, candidates and affiliated organizations through requiring comprehensive and openly available advertising databases and disclosure of spending by political parties and support groups.
* Work with journalists and researchers in fact-checking and investigations around electoral disinformation networks and producers of ‘dark propaganda’.

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.
* Judges and other judicial actors may 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.

**QUESTION #5 Please share information on measures to address disinformation that you believe have aggravated or led to human rights violations, in particular the right to freedom of opinion and expression.**

The Balancing Act report finds that there are cases where one type of counter-disinformation response can work against another. An example is an over-emphasis on responses through top-down regulation at the expense of bottom-up empowerment.

Further, there is the phenomenon 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). This works directly against the role of independent, critical journalism as a counter to disinformation.

A similar example exists in cases of internet communications companies not removing disinformation-laden attacks on journalists on the grounds of ‘free speech’. In this way, a very particular understanding of expression undermines press freedom and journalism safety, and therefore journalism’s service against disinformation.

The report advocates that different interventions need to be aligned, rather than going in separate directions.

**QUESTION #6 Please share any suggestions or recommendation you may have for the Special Rapporteur on how to protect and promote the right to freedom of opinion and expression while addressing disinformation.**

UNESCO’s model of Internet Universality calls for **multi-stakeholder consultation** on digital issues, including such cooperation in the fight against disinformation. This is vital to avoid unilateral “capture” of disinformational responses by companies or governments. The problem and the responses require co-ownership and balancing out of the limits associated with single constituencies, even although there are different roles and obligations of the different actors.

Further, as the Balancing Act report advocates, a multi-faceted approach is needed and must including **addressing socio-economic drivers of disinformation**: through rebuilding the social contract and public trust in democratic institutions, promotion of social cohesion, particularly in highly polarised societies, and addressing business models that thrive on paid disinformation content such as advertising that crosses the line, through to fraudulent content masquerading as legitimate news or factually-grounded opinion.

The ralso eport urges that each actor seeking interventions against disinformation include systematic **monitoring and evaluations** within their response activities. These should cover effectiveness, as well as impacts on the right to freedom of expression and access to information, including on the right to privacy. The findings also underline the need for increased transparency and proactive disclosure across all 11 kinds of responses to disinformation. This aligns with the spirit of Sustainable Development Goal target 16.10 which calls for public access to information and fundamental freedoms.

UNESCO encourages internet providers, platforms and donors to invest further in **independent fact-checking, critical professional journalism, media development and Media and Information Literacy (MIL)**, especially through educational interventions targeting children, young people, older citizens, and vulnerable groups.

Stakeholders should 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 word, and especially in the context of elections, public health, and natural disasters.

Increased **transparency and proactive disclosure** across all responses to disinformation aligns with the spirit of Sustainable Development Goal target 16.10 which calls for public access to information and fundamental freedoms.

In addition to promoting all the above, the UN Special Rapporteur can play a role in encouraging system-wide approaches within the United Nations. As disinformation poses an obstacle to the mandates of each agency, fund and programme, so the generic elements of the problem and possible solutions have relevance across the system.

For the UN to provide effective and meaningful support to Member States on this complex issue, the Rapporteur could propose the formation of a UN intra-agency group to share information on disinformation and the importance of enhanced transparency from internet platforms to address it. Such an interagency group, composed of a range of UN entities, could provide a coordinated response and common voice to engage with relevant counterparts such as internet companies, civil society, international organizations.

**Relevant UNESCO documents and reports for consideration by the Special Rapporteur**

*Balancing Act: Countering Digital Disinformation while respecting Freedom of Expression.* <https://en.unesco.org/publications/balanceact>

*Combating the disinfodemic: Working for truth in the time of COVID-19.* <https://en.unesco.org/covid19/disinfodemic>

*Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training.* <https://en.unesco.org/fightfakenews>

*Social Media and Elections:* [https://unesdoc.unesco.org/ark:/48223/pf0000370634](https://unesdoc.unesco.org/ark%3A/48223/pf0000370634)

*World Trends in Freedom of Expression and Media Development*. <https://en.unesco.org/world-media-trends>

* World Trends in Freedom of Expression and Media Development: Global Report 2017/2018*:* [https://unesdoc.unesco.org/ark:/48223/pf0000261065](https://unesdoc.unesco.org/ark%3A/48223/pf0000261065)
* Elections and media in digital times. [https://unesdoc.unesco.org/ark:/48223/pf0000371486](https://unesdoc.unesco.org/ark%3A/48223/pf0000371486)
* Journalism, press freedom and COVID-19. [https://unesdoc.unesco.org/ark:/48223/pf0000373573?posInSet=1&queryId=0216815c-9a38-457c-8e20-b224c31b03e5](https://unesdoc.unesco.org/ark%3A/48223/pf0000373573?posInSet=1&queryId=0216815c-9a38-457c-8e20-b224c31b03e5)
* Reporting Facts: Free from Fear or Favour*.* [*https://unesdoc.unesco.org/ark:/48223/pf0000375061*](https://unesdoc.unesco.org/ark%3A/48223/pf0000375061)
1. <https://onu.delegfrance.org/IMG/pdf/cross-regional_statement_on_infodemic_final_with_all_endorsements.pdf> [↑](#footnote-ref-2)
2. <https://en.unesco.org/publications/balanceact> [↑](#footnote-ref-3)
3. <https://en.unesco.org/covid19/disinfodemic> [↑](#footnote-ref-4)
4. [https://unesdoc.unesco.org/ark:/48223/pf0000373573?posInSet=1&queryId=0216815c-9a38-457c-8e20-b224c31b03e5](https://unesdoc.unesco.org/ark%3A/48223/pf0000373573?posInSet=1&queryId=0216815c-9a38-457c-8e20-b224c31b03e5); [https://unesdoc.unesco.org/ark:/48223/pf0000375061](https://unesdoc.unesco.org/ark%3A/48223/pf0000375061) [↑](#footnote-ref-5)
5. <https://en.unesco.org/news/dontgoviral-unesco-and-i4policy-launch-campaign-crowdsource-local-content-combat-infodemic> [↑](#footnote-ref-6)
6. <https://en.unesco.org/covid19/disinfodemic/coronavirusfacts>; <https://journalismcourses.org/coursecat/journalism-in-a-pandemic/> [↑](#footnote-ref-7)
7. <https://en.unesco.org/news/unesco-partners-knight-center-chair-and-who-webinar-journalists-covering-covid-19-vaccines>; <https://journalismcourses.org/course/desinformacion-y-fact-checking-en-tiempos-de-covid-19-en-america-latina-y-el-caribe/>; <https://portalcheck.org/> [↑](#footnote-ref-8)
8. <https://en.unesco.org/news/call-applications-hackingdisinfodemic> [↑](#footnote-ref-9)
9. <https://en.unesco.org/commemorations/globalmilweek/2020/howtocelebrate> [↑](#footnote-ref-10)
10. [https://unesdoc.unesco.org/ark:/48223/pf0000370634](https://unesdoc.unesco.org/ark%3A/48223/pf0000370634); https://unesdoc.unesco.org/ark:/48223/pf0000371486 [↑](#footnote-ref-11)
11. [https://unesdoc.unesco.org/ark:/48223/pf0000261065\_eng](https://unesdoc.unesco.org/ark%3A/48223/pf0000261065_eng); <https://en.unesco.org/sites/default/files/declaration_accra.2018-05-03.pdf>; https://en.unesco.org/sites/default/files/wpfdaddisdecl3\_may.pdf [↑](#footnote-ref-12)
12. [https://unesdoc.unesco.org/ark:/48223/pf0000374208?posInSet=4&queryId=0fb1857b-3274-42dd-8e1a-24bf6065926e](https://unesdoc.unesco.org/ark%3A/48223/pf0000374208?posInSet=4&queryId=0fb1857b-3274-42dd-8e1a-24bf6065926e) [↑](#footnote-ref-13)