United Nations A/HRC/47/25

General Assembly

Distr.: General
13 April 2021

Original: English

Human Rights Council
Forty-seventh session
21 June–9 July 2021
Agenda item 3
Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development

Disinformation and freedom of opinion and expression

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

Summary

In the present report, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression examines the threats posed by disinformation to human rights, democratic institutions and development processes. While acknowledging the complexities and challenges posed by disinformation in the digital age, the Special Rapporteur finds that the responses by States and companies have been problematic, inadequate and detrimental to human rights. She calls for multidimensional and multi-stakeholder responses that are well grounded in the international human rights framework and urges companies to review their business model and States to recalibrate their responses to disinformation, enhancing the role of free, independent and diverse media, investing in media and digital literacy, empowering individuals and rebuilding public trust.

I. Introduction

1. More than 2,000 years ago, Octavian spun a vicious disinformation campaign to destroy his rival Mark Anthony and eventually become the first Roman emperor Augustus Caesar. Since those ancient times, information has been fabricated and manipulated to win wars, advance political ambitions, avenge grievances, hurt the vulnerable and make financial profit.

2. Disinformation is not a new phenomenon. What is new is that digital technology has enabled pathways for false or manipulated information to be created, disseminated and amplified by various actors for political, ideological or commercial motives at a scale, speed and reach never known before. Interacting with political, social and economic grievances in the real world, disinformation online can have serious consequences for democracy and human rights, as recent elections, the response to the coronavirus disease (COVID-19) pandemic and attacks on minority groups have shown. It is politically polarizing, hinders people from meaningfully exercising their human rights and destroys their trust in Governments and institutions.

3. Finding appropriate responses to disinformation is difficult, not least because the concept is undefined and open to abuse, and because the size and nature of the problem is contested in the absence of sufficient data and research. State responses have often been problematic and heavy handed and had a detrimental impact on human rights. Companies play a major role in spreading disinformation but their efforts to address the problem have been woefully inadequate.

4. At the core is a human rights challenge, aggravated by an information disorder. There is growing evidence that disinformation tends to thrive where human rights are constrained, where the public information regime is not robust and where media quality, diversity and independence is weak. Conversely, where freedom of opinion and expression is protected, civil society, journalists and others are able to challenge falsehoods and present alternative viewpoints. That makes international human rights a powerful and appropriate framework for addressing disinformation.

5. In the present report, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression recognizes the complexity of disinformation and outlines the conceptual and contextual challenges it poses. She sets out the relevant international legal standards, analyses the responses of States and companies to the phenomenon and then proposes recommendations for multifaceted responses grounded in international human rights law, transparency and accountability and in the multi-stakeholder engagement of States, companies and civil society. Building on the ground-breaking work done by her predecessor on the human rights responsibilities of digital platforms, the current Special Rapporteur calls for a review of the business models of the platforms and a recalibration of State responses to disinformation.

6. In drafting the report, the Special Rapporteur has benefited from the submissions of 119 civil society organizations and academic entities, three international organizations, three Member States and three companies, as well as several online consultations with civil society organizations and meetings with Member States, social media companies and experts.

7. The report does not purport to be comprehensive in its content or recommendations. It does not, for example, cover the issue of disinformation campaigns directed by State or State-sponsored actors towards the population of other States, as that is a complex subject requiring more consultations and reflection than would have been possible within the timeline of the present report.

8. The purpose of the present report is to open a dialogue with interested stakeholders, including Member States, companies and civil society, and contribute to ongoing discussions in various forums with a view to further refining and pursuing the conclusions and recommendations.

II. Conceptual and contextual challenges

A. Concept of disinformation

9. There is no universally accepted definition of disinformation. While the lack of agreement makes a global response challenging, the lack of consensus underlines the complex, intrinsically political and contested nature of the concept.

10. Part of the problem lies in the impossibility of drawing clear lines between fact and falsehood and between the absence and presence of intent to cause harm. False information can be instrumentalized by actors with diametrically opposite objectives. Truthful information can be labelled as "fake news" and delegitimized. Opinions, beliefs, uncertain knowledge and other forms of expression like parody and satire do not easily fall into a binary analysis of truth and falsity. Furthermore, false content that is spread online with the intent to cause harm (disinformation) can be picked up and shared by innocent third parties with no such intent (misinformation), the innocent vector boosting dissemination and adding credibility to the malicious campaigner. Intentionally or not, the harm occurs. Some forms of disinformation can amount to incitement to hatred, discrimination and violence, which are prohibited under international law.

11. The European Commission has described disinformation as verifiably false or misleading information that, cumulatively, is created, presented and disseminated for economic gain or to intentionally deceive the public and that may cause public harm.[1] The Broadband Commission for Sustainable Development, on the other hand, has approached disinformation as false or misleading content with potential consequences, irrespective of the underlying intention or behaviours producing and circulating messages.[2] National laws and regulations dealing with disinformation cover a varied combination of false or misleading information, the intention to cause harm or not and the nature of the harm caused or intended. Disinformation is often described in broad, ill-defined terms not in line with international legal standards.

12. Academics have developed a taxonomy of an information disorder in which "disinformation" is described as false information that is knowingly shared with the intention to cause harm, "misinformation" as the unintentional dissemination of false information and "malinformation" as genuine information shared with the intention to cause harm.[3] By setting out a holistic and interconnected picture of the problem, the information disorder framework encourages a multidimensional, varied and contextualized approach to disinformation.

13. Some academics have framed the phenomenon of disinformation as "viral deception" consisting of three vectors: manipulative actors, deceptive behaviour and harmful content.[4] The focus is on online behaviour rather than on the veracity of content. Some large social media platforms, including Facebook, refer to these vectors to inform their policies on responding to coordinated inauthentic behaviour.

14. Ultimately, the lack of clarity and agreement on what constitutes disinformation, including the frequent and interchangeable use of the term misinformation, reduces the effectiveness of responses.[5] It also leads to approaches that endanger the right to freedom of opinion and expression. It is vital to clarify the concepts of disinformation and misinformation within the framework of international human rights law.

---

[1] European Commission, Code of Practice on Disinformation (2018).

[2] Broadband Commission for Sustainable Development, *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression* (International Telecommunication Union and United Nations Educational, Scientific and Cultural Organization, September 2020), pp. 8, 18 and 25 ff.

[3] Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking* (Council of Europe, 2017), p. 5.

[4] Camille François, "Actors, behaviors, content: a disinformation ABC" (Transatlantic Working Group, September 2019).

[5] Submission from the United Nations Educational, Scientific and Cultural Organization.

15. For the purposes of the present report, disinformation is understood as false information that is disseminated intentionally to cause serious social harm and misinformation as the dissemination of false information unknowingly. The terms are not used interchangeably.

B. Actors and vectors

16. Disinformation spreads rapidly and widely through social media and messaging platforms, even in some of the most remote and fragile regions of the world,[6] although in overall terms it remains a small subset of all the information in circulation.[7] Digital technology has made it possible to share in new ways texts, images and videos, including "deep fakes" and "shallow fakes", that can give a distorted picture of reality.[8] False information is amplified by algorithms and business models that are designed to promote sensational content that keep users engaged on platforms. Disinformation thrives in an online environment that encourages amplification while reducing accessibility to plural and diverse sources of information.

17. The global disinformation system is a highly lucrative business that is driven by commercial motives and that is becoming increasingly professionalized.[9] Technology companies are also purportedly allowing spreaders of misinformation to monetize their content, for instance by allowing junk news websites disseminating COVID-19-related conspiracies to post advertisements on their platforms.[10] Essentially, disinformation is a modern way in the digital era of making money by purposefully spreading lies.[11]

18. The use of new technologies for the production of disinformation or divisive content is exploited, for multiple motives (political, ideological or commercial), by multiple actors, including States, political parties, politicians and other powerful individuals or businesses supported by troll armies or public relations companies.[12] The false messages of these instigators are often conveyed, unwittingly or not, by traditional media,[13] celebrities or ordinary users and their peer-to-peer and friend-to-friend networks in a complex mix of exchanges between the online and offline worlds.

19. Ideologically driven non-State actors, including extremist or terrorist groups, also frequently engage in the dissemination of false news and narratives as part of their propaganda to radicalize and recruit members.[14] The security dimensions and the excessive responses by States to them add to human rights concerns.[15]

20. Notwithstanding the above, the growth of disinformation in recent times cannot be attributed solely to technology or malicious actors. It needs to be understood in the context of other factors, including: a struggling legacy media sector, challenged by digital

[6] Submission from Fondation Hirondelle.

[7] Submissions from the Reuters Institute for the Study of Journalism (University of Oxford) and the Center for Social Media and Politics (New York University).

[8] For a description of various methods of spreading disinformation, see, e.g., Kate Jones, "Online disinformation and political discourse: applying a human rights framework", Chatham House Research Paper (November 2019), pp. 11–12.

[9] Samantha Bradshaw, Hannah Bailey and Philip N. Howard, "Industrialized disinformation: 2020 global inventory of organized social media manipulation". (Computational Propaganda Project, University of Oxford, 2021), p. 21.

[10] See, e.g., United Kingdom of Great Britain and Northern Ireland, Parliament, "*Misinformation in the COVID-19 Infodemic*" (20 July 2020).

[11] Submission from Edith Cowan University.

[12] See, e.g., Dhanaraj Thakur and DeVan L. Hankerson, "Facts and their discontents: a research agenda for online disinformation, race and gender" (Center for Democracy and Technology, February 2021). See also the submission from PEN America.

[13] Media Matters for Democracy, "Disorder in the newsroom: the media's perceptions and response to the infodemic" (December 2020).

[14] See, e.g., United Nations Interregional Crime and Justice Research Institute, "Stop the virus of disinformation: the risk of malicious use of social media during COVID-19 and the technology options to fight it" (November 2020), pp. 15–17.

[15] Submission from Stanford University.

transformation and competition from online platforms and threatened by State pressure in some parts of the world; the absence of robust public information regimes; low levels of digital and media literacy among the general public; and the frustrations and grievances of a growing number of people, fuelled by decades of economic deprivation, market failures, political disenfranchisement and social inequalities, which make some individuals more susceptible to manipulation.[16]

21. Disinformation is not the cause but the consequence of societal crises and the breakdown of public trust in institutions. Strategies to address disinformation are unlikely to succeed without more attention being paid to these underlying factors.

C. Targets and victims

22. Although empirical research suggests that only a small proportion of people are exposed to disinformation,[17] the impacts on institutions, communities and individuals are real, broad and legitimate. Over 100 submissions received by the Special Rapporteur for the present report contain many concrete examples.[18] They suggest that much of the targeting is politically motivated against institutions and individuals in vulnerable situations and affects a wide range of human rights, including economic, social, cultural, civil and political rights.

23. There is clear evidence that robust public information regimes and independent journalism are strong antidotes to disinformation. Therefore, it is doubly disturbing that smear campaigns against journalists have become more pernicious on social media networks. Some political leaders have labelled the media as "the enemy of the people"[19] or financed entire "fake news industries" that drown out their reporting.[20] Such attacks both erode public trust in journalism and make journalists more fearful about doing their job.[21] According to one report, at least 34 journalists were jailed on charges of "fake news" in 2020, compared to only 1 in 2012.[22] Disinformation poses a threat not only to the safety of journalists but also to the media ecosystem in which they operate,[23] forcing the legacy media to divert precious resources from reporting to dispelling and debunking lies.

24. Disinformation has been used in several countries in highly visible ways to undermine the right to free and fair elections.[24] As an example, racially targeted disinformation campaigns were used to suppress votes from communities of colour in the three most recent major elections in the United States of America.[25] During the 2020 presidential election, then President Donald Trump and his surrogates repeatedly sought to erode confidence in the postal voting system and made baseless claims about election fraud on social media.[26] The detrimental impact of politically motivated disinformation has been felt on democratic institutions in many other countries too, chilling free speech, reducing the level of trust in the public sphere as a space for democratic deliberation, amplifying anti-democratic narratives, driving polarization and promoting authoritarian and populist agendas.[27]

---

16 Submissions from Article 19: International Centre against Censorship and the Internet Governance Project.

17 Submissions from the Center for Social Media and Politics and the Reuters Institute for the Study of Journalism.

18 These submissions will be made available at www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Report-on-disinformation.aspx.

19 Committee to Protect Journalists, "The Trump Administration and the media: attacks on press credibility endanger US democracy and global press freedom" (April 2020).

20 European Parliament, "Disinformation and propaganda: impact on the functioning of the rule of law in the EU and its Member States" (2019).

21 Media Matters for Democracy, "Disorder in the newsroom".

22 Submission from the Committee to Protect Journalists.

23 Ibid.

24 Submissions from the Center for Democracy and Technology and PEN America.

25 Young Mie Kim, "Voter suppression has gone digital", Brennan Center for Justice, 20 November 2018.

26 Submission from the International Center for Not-for-Profit Law.

27 See, e.g., submissions from the Association for Progressive Communications and Global Partners Digital.

25. Over the past year, the spread of disinformation and misinformation by non-State sources has posed significant challenges to the right to health and responses to the COVID-19 pandemic against a background of efforts by some Governments to withhold or falsify information.[28] As in the case of global health, in respect of climate change too scientific information has been discredited and environmental activists have been attacked through well-organized online disinformation campaigns.

26. Ideological and identity-based disinformation has fomented discrimination and hatred against minorities, migrants and other marginalized communities,[29] generating ethnic or religious tensions[30] that have culminated, at times, in violence offline, as happened in Ethiopia[31] and Myanmar.[32] Civil society organizations are calling for more research to be carried out to understand the full measure of disinformation on vulnerable and minority communities.[33]

27. Online gendered disinformation campaigns are increasingly being used to deter women from participating in the public sphere, mixing "old ingrained sexist attitudes with the anonymity and reach of social media in an effort to destroy women's reputations and push them out of public life".[34] Women journalists, politicians and gender equity advocates who speak out on feminist issues are particularly targeted.[35] There is also significant disinformation around the issue of sexual and reproductive health.[36]

28. Human rights defenders and civil society organizations, especially those representing marginalized and discriminated groups, are harshly attacked and subjected to verbal abuse and vilification by online disinformation campaigns.[37]

29. The negative impact of disinformation is undeniable and must be addressed. International human rights law provides a powerful antidote and a framework for formulating responses.

III. Applicable international legal framework

30. The Human Rights Council has affirmed that responses to the spread of disinformation and misinformation must be grounded in international human rights law, including the principles of lawfulness, legitimacy, necessity and proportionality.[38] While disinformation affects a wide range of human rights, the present report focuses on the freedom of opinion and expression in the light of the particular value that this right brings to efforts to counter disinformation.

31. Article 19 of the Universal Declaration of Human Rights and article 19 of the International Covenant on Civil and Political Rights guarantee the right to hold opinions without interference and to seek, receive and impart information and ideas of all kinds, regardless of frontiers and through any media. While freedom of opinion is absolute, freedom of expression may be restricted under certain circumstances. The State has a duty to refrain from interfering with that right and also an obligation to ensure that others, including businesses, do not interfere with it.

[28] A/HRC/44/49, paras. 45–47.
[29] A/HRC/46/57.
[30] Submission from the Ahmadiyya Muslim Lawyers Association.
[31] See www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26483&LangID=E.
[32] See www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=26808&LangID=E.
[33] See submissions from Access Now and the Center for Democracy and Technology.
[34] Nina Jankowicz, "How disinformation became a new threat to women", *Coda Story*, 11 December 2017.
[35] Submission from Media Matters for Democracy. See also the case of Maria Ressa, mentioned in AL PHL 12/2018.
[36] Submission from MSI Reproductive Choices.
[37] Submission from Digital Rights Foundation.
[38] Resolution 44/12.

32. The previous mandate holder has urged digital technology companies to apply international human rights standards in their business practices.[39] The exhortations have gained added urgency in the context of disinformation and misinformation.

A. Freedom of opinion

33. The right to freedom of opinion comprises two dimensions: an internal dimension closely connected to the right to privacy and freedom of thought and an external dimension related to freedom of expression. While the latter aspect is discussed frequently, the former has begun only recently to gain attention as a result of greater awareness and understanding of the manipulative techniques used by social platforms, State actors and others online to influence individuals in ways that could infringe their freedom of opinion.[40]

34. Article 19 of the Universal Declaration of Human Rights and of the International Covenant on Civil and Political Rights protects the right to hold opinions without interference. The right is absolute and permits no exception or restriction. Notwithstanding the absolute and broad nature of this right, in reality human beings are influenced constantly in their thought and opinion by others, and "the freedom to be subject to a wide range of influences is itself a dimension of our autonomy".[41] Therefore, in determining whether and how disinformation online might infringe freedom of opinion, the critical issue is the knowledge and consent of the rights holder.

35. The right to form one's opinion and to develop it by way of reasoning is an essential element of freedom of opinion. It is well established that the freedom of opinion includes the right not to express an opinion, as well as the right to change one's opinion whenever or for whatever reason a person so freely chooses.[42] In other words, involuntary disclosure is prohibited and mental autonomy is affirmed. Any effort to coerce the holding or not holding of any opinion is prohibited.

36. Punishment, harassment, intimidation and stigmatization for holding an opinion, including coercive, involuntary or non-consensual manipulation of the thinking process to develop an opinion, are violations of the right to opinion.[43] Coercive or manipulative action has been understood to include indoctrination, "brainwashing", influencing "the conscious or subconscious mind with psychoactive drugs or other means of manipulation".[44] The digital equivalent would be techniques that allow State and non-State actors to access and influence the thoughts and opinions of people without their knowledge or consent, such as content curation through powerful platform recommendations or microtargeting. Such techniques play a significant role in spreading disinformation and, as involuntary or non-consensual manipulation of thinking processes, contravene the right to freedom of opinion.[45]

[39] A/HRC/38/35, paras. 9–12; and A/74/486, paras. 40–55.

[40] Evelyn Aswad, "Losing the freedom to be human", *Columbia Human Rights Law Review*, vol. 52 (2020); Susie Alegre, "Rethinking freedom of thought for the 21st Century", *European Human Rights Law Review* (2017); and Kate Jones, "Protecting political discourse from online manipulation: the international human rights law framework", *European Human Rights Law Review* (2021).

[41] Kate Jones, "Online disinformation and political discourse: applying a human rights framework", p. 33.

[42] Human Rights Committee, general comment No. 34 (2011), para. 9.

[43] "Deliberate efforts to influence through non-consensual means violate this right when they rise to the level of either overwhelming mental autonomy or manipulating one's reasoning." See Evelyn Aswad, "Losing the freedom to be human", p. 329.

[44] Manfred Novak, *UN Covenant on Civil and Political Rights: CCPR Commentary*, 2nd ed. (Kehl am Rhein, N.P. Engel, 2005).

[45] Evelyn Aswad, "Losing the freedom to be human". See also a similar commentary on article 9 of the Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights) by Susie Alegre, "Rethinking freedom of thought for the 21st Century".

B. Freedom of expression

37. The right to freedom of expression is broad and inclusive, and encapsulates the freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers and through any media, offline or online.[46]

38. In the context of disinformation, two points are worth noting. Firstly, the right to freedom of expression applies to all kinds of information and ideas, including those that may shock, offend or disturb,[47] and irrespective of the truth or falsehood of the content.[48] Under international human rights law, people have the right to express ill-founded opinions and statements or indulge in parody or satire if they so wish.[49] Secondly, the free flow of information is a critical element of freedom of expression and places a positive obligation on States to proactively put information of public interest in the public domain, and promote plural and diverse sources of information, including media freedom. It can be a valuable tool for countering disinformation.

39. Freedom of expression may be restricted only in accordance with article 19 (3) of the International Covenant on Civil and Political Rights, which requires all restrictions to be provided by law and to be necessary for the legitimate aim of respecting the rights and reputations of others and for protecting national security, public order or public health or morals. In the light of the fundamental importance of this right to the enjoyment of all other human rights, the restrictions must be exceptional and narrowly construed.

40. The principle of legality requires the scope, meaning and effect of the law to be sufficiently clear, precise and public. Vague laws that confer excessive discretion can lead to arbitrary decision-making and are incompatible with article 19 (3) of the Covenant. Any limitation of disinformation must establish a close and concrete connection to the protection of one of the legitimate aims stated in article 19 (3). The prohibition of false information is not in itself a legitimate aim under international human rights law.

41. The directness of the causal relationship between the speech and the harm, and the severity and immediacy of the harm, are key considerations in assessing whether the restriction is necessary. The principle of necessity requires the restriction to be appropriate and proportionate to achieve the legitimate aim, using the least restrictive means to protect it. Criminal sanctions constitute serious interference with the freedom of expression and are disproportionate responses in all but the most egregious cases.

42. Given the fundamental importance of freedom of expression to democracy and the enjoyment of all other human rights and freedoms, international human rights law affords particularly strong protection to expressions on matters of public interest, including criticism of Governments and political leaders and speech by politicians and other public figures, and to media freedom.[50] This does not mean that disinformation in the context of political speech can never be restricted, but that any such restriction requires a high threshold of legality, legitimacy, necessity and proportionality. For instance, electoral laws may justifiably forbid the propagation of falsehoods relating to electoral integrity, but such a restriction must be narrowly construed, time-limited and tailored so as to avoid limiting political debate.

43. Disinformation is often used to foment hatred and violence. Article 20 (2) of the International Covenant on Civil and Political Rights provides that any advocacy of national,

[46] Human Rights Committee, general comment No. 34 (2011), para. 12, and Human Rights Council resolution 20/8.

[47] Human Rights Committee, general comment No. 34 (2011), para. 11. See also European Court of Human Rights, *Handyside v. the United Kingdom*, application No. 5493/72, judgment, 7 December 1976, para. 49.

[48] Human Rights Committee, general comment No. 34 (2011), paras. 47 and 49. See also European Court of Human Rights, *Salov v. Ukraine*, application No. 65518/01, judgment, 6 September 2005, para. 113: "Article 10 of the [European] Convention [on Human Rights, on freedom of expression] does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not be truthful."

[49] False statements can be restricted only if they also meet the restrictions and criteria set out in article 19 (3) of the International Covenant on Civil and Political Rights.

[50] Human Rights Committee, general comment No. 34 (2011), para. 38.

racial or religious hatred that constitutes incitement to discrimination, hostility or violence is to be prohibited by law. It does not call for criminalization, nor does it make any reference to untruthful information. The Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence is an authoritative road map for interpreting article 20 (2) and sets out six factors to determine the severity necessary to criminalize incitement: context; status of the speaker; intent; content and form of speech; reach of the speech; and likelihood of risk. It may in certain situations provide a relevant framework for addressing disinformation.[51]

44. Hate speech relating to racial or ethnic origin is prohibited under article 4 of the International Convention on the Elimination of All Forms of Racial Discrimination. The Human Rights Committee and the Committee on the Elimination of Racial Discrimination have both clarified that the prohibitions must be justified in strict conformity with article 19 of the International Covenant on Civil and Political Rights and that criminalization should be reserved only for the most serious cases.

45. The Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda sets out key principles drawn from international human rights law to guide States, companies and others.[52]

IV. State responses to disinformation: key concerns

46. State responses to disinformation can take various forms, ranging from measures to disrupt the Internet and legislation to censor, punish or restrict dissemination to the regulation of social media platforms. At one end of the spectrum, some States sponsor disinformation while ostensibly seeking to suppress it; on the opposite (and positive) end of the spectrum, States promote various measures to encourage the free flow of information, enhance media diversity and support media information and digital literacy as means of countering disinformation.

A. State-sponsored disinformation

47. State-sponsored disinformation can emanate from State institutions directly or from proxies targeting audiences within the State's own territory or abroad for political and strategic aims.[53] In the digital age, new techniques have significantly expanded the scale, speed and spread of such operations. When combined with the power, means and reach of a State, their impact can be devastating for human rights. Where States systematically and simultaneously suppress other sources while promoting their own false narratives, they are denying individuals the right to seek and receive information under article 19 (2) of the International Covenant on Civil and Political Rights.[54]

48. A notorious example of State-led disinformation involved the "Tatmadaw true news information team" in Myanmar, which posted online doctored and mislabelled photographs relating to the Rohingya crisis. In August 2018, Facebook blocked the accounts of the "team" for spreading hate speech.[55] Another example is the online practice of "red tagging" used by State agents in the Philippines to falsely brand activists, journalists and political opponents

[51] A/HRC/22/17/Add.4, annex, appendix.

[52] See www.osce.org/files/f/documents/6/8/302796.pdf. The Joint Declaration was adopted on 3 March 2017 by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression of the United Nations, the Representative on Freedom of the Media of the Organization for Security and Cooperation in Europe, the Special Rapporteur for Freedom of Expression of the Organization of American States and the Special Rapporteur on Freedom of Expression and Access to Information of the African Commission on Human and Peoples' Rights.

[53] The present report does not cover attempts by States to spread disinformation outside their borders.

[54] Mark Milanovic and Michael N. Schmitt, "Cyber attacks and cyber (mis)information operations during a pandemic", *Journal of National Security Law and Policy*, vol. 11 (2020).

[55] See https://about.fb.com/news/2018/08/removing-myanmar-officials/.

as leftists, communists, terrorists or subversives, thereby increasing the risk of them being arrested, attacked or killed.[56]

49. In recent years, in a number of countries, State-led disinformation campaigns have sought to influence elections and other political processes, control the narrative of public debates or curb protests against and criticisms of Governments. In the context of the COVID-19 pandemic, there have been various instances of State actors disseminating unverified claims about the origins of the virus responsible for COVID-19, denying the spread of the disease or providing false information on infection rates, fatality figures and health-care advice. Such disinformation has been detrimental to efforts to control the pandemic, endangering the rights to health and life, as well as people's trust in public information and State institutions.[57]

B. Internet shutdowns

50. During the past two years, the Internet has been shut down just before or during elections in several countries, including Belarus,[58] the Democratic Republic of the Congo,[59] Ethiopia[60] and Myanmar,[61] ostensibly to prevent the spread of disinformation online that could incite violence. Governments have also imposed Internet shutdowns during demonstrations or to silence dissent, for example in Bahrain and Venezuela (Bolivarian Republic of).[62] Ambiguous laws that permit broad government discretion to shut down or otherwise disrupt Internet connectivity and access to telecommunications have been documented in Tajikistan.[63] In contrast, a court in Indonesia ruled against the decision to shut down the Internet to prevent the spread of "fake news" during political unrest in Papua and West Papua on the grounds that it was excessive and unnecessary.[64]

51. The Human Rights Council has strongly condemned the use of Internet shutdowns that intentionally and arbitrarily prevent or disrupt access to information online.[65] Shutting down the Internet is an inherently disproportionate response, given the blanket nature of the act, which blocks multiple other uses of the Internet. As such, it violates the requirement of necessity and proportionality set out in international human rights law. It deprives individuals of all information and services online. It hinders voters from accessing information about elections, human rights defenders from documenting and sharing human rights concerns and journalists and the media from reporting on issues of public interest. By depriving people of information sources, Internet shutdowns do not curb disinformation but, rather, hamper fact-finding and are likely to encourage rumours. In many cases, they appear to be aimed at silencing minority voices and depriving them of access to vital information.[66]

C. Criminal laws

52. States have long had discreet laws to address the harm done by false information, for example in relation to defamation, consumer protection and financial fraud. More problematic is the use of criminal laws to punish the spread of loosely defined false information on issues of public interest. Some of these laws date back to colonial times and

[56] AL PHL 1/2021.
[57] A/HRC/44/49, para. 45.
[58] See www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=26164&LangID=E.
[59] See www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=24057&LangID=E.
[60] See www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26483&LangID=E.
[61] See www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=26431&LangID=E.
[62] A/HRC/35/22, para. 11.
[63] A/HRC/35/22/Add.2, para. 29.
[64] Moch. Fiqih Prawira Adjie, "Jokowi 'violates the law' for banning internet in Papua, court declares", *Jakarta Post*, 3 June 2020.
[65] Resolution 44/12.
[66] See the submissions from Access Now and the Centre for Law and Democracy.

have been found by domestic courts in Uganda,[67] Zambia[68] and Zimbabwe[69] to be unconstitutional and unjustified in modern democratic societies. In 2016, the regional court in West Africa found that the criminal offences of sedition, false news and criminal defamation in the Gambia infringed international law on expression and ordered them to be repealed.[70]

53. During the past decade, there has been a flurry of laws prohibiting "false news" of various forms on the Internet and social media platforms, with at least 17 States adopting legislation in the past year alone to address pandemic-related problematic information.[71]

54. Many of these "false news" laws fail to meet the three-pronged test of legality, necessity and legitimate aims set out in article 19 (3) of the International Covenant on Civil and Political Rights. They often do not define with sufficient precision what constitutes false information or what harm they seek to prevent, nor do they require the establishment of a concrete and strong nexus between the act committed and the harm caused. Words such as "false", "fake" or "biased" are used without elaboration and assertions based on a circular logic are made (for example, "a statement is false if it is false or misleading, whether wholly or in part, and whether on its own or in the context in which it appears").[72] In some cases, harm is defined in overly broad terms.[73] Unfettered discretion has been given to executive authorities without judicial oversight in some legislation, notably in Malaysia[74] and Singapore,[75] opening the possibility for abuse and arbitrary decision-making. Often, the prescribed punishment is excessively harsh and disproportionate, and can have a chilling effect on freedom of expression.[76]

55. The vague and overly broad nature of such laws allows Governments to use them against journalists, political opponents and human rights defenders. For instance, the ambiguous and broadly defined provisions in the Penal Code of Turkey and in anti-terrorism legislation criminalizing broad categories of speech, including expressions that "denigrate the Turkish nation" or "insult the President", have been used against many political activists and journalists in Turkey.[77] In Egypt, human rights defenders and journalists have been prosecuted for spreading "false news" after they published reports on the human rights situation in the country.[78] In Bangladesh, the detention of cartoonists, bloggers and journalists under the Digital Security Act have led to allegations of torture and death in custody.[79] The United Nations High Commissioner for Human Rights has expressed alarm at the sharp rise in the use of "false news" laws to clamp down on criticism of Governments in the wake of the COVID-19 pandemic in many countries in Asia.[80]

D. Social media regulation

56. At the heart of concerns about disinformation or misinformation online is the way in which such information can go viral, at great speed. States have responded to this challenge

[67] Uganda, Supreme Court, *Charles Onyango Obbo and another v. Attorney General*, constitutional appeal No. 2 of 2002, judgment, 10 February 2004.
[68] Zambia, High Court, *Chipenzi and others v. the People*, HPR/03/2014, 4 December 2014.
[69] Zimbabwe, Supreme Court, *Chavunduka v. Minister of Home Affairs*, case No. 2000 JOL 6540 (ZS), 22 May 2000.
[70] Community Court of Justice of the Economic Community of West African States, *Federation of African Journalists and four others v. the Gambia*, judgment No. ECW/CCJ/JUD/04/18.
[71] See https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/. See also the submission from the Centre for Law and Democracy.
[72] See the Protection from Online Falsehoods and Manipulation Act of Singapore, sect. 2 (2) (b).
[73] See, e.g., the Penal Code of Qatar, art. 136, and QAT 1/2020.
[74] OL MYS 5/2021.
[75] OL SGP 3/2019.
[76] OL BFA 2/2020.
[77] See, e.g., OL TUR 13/2020, AL TUR 18/2020 and AL TUR 20/2020.
[78] See, e.g., AL EGY 19/2020, AL EGY 15/2020, AL EGY 10/2020, UA EGY 6/2020 and UA EGY 1/2020.
[79] OL BGD 4/2018 and AL BGD 7/2020.
[80] See www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25920&LangID=E.

in different ways, some by adopting regulations that give the authorities a direct role in controlling content online, some by focusing not on content but on the processes by which disinformation or misinformation is handled by companies. Some laws or proposals raise privacy concerns, as instant messaging services may fall within their scope and require users to provide identification information, data to be traced or the use of filters.[81]

57. In recent years, several States have adopted laws that grant the authorities excessive discretionary powers to compel social media platforms to remove content that they deem illegal, including what they consider to be disinformation or "fake news". Failure to comply is sanctioned with significant fines and/or content blocking. This has been the case, for example, in Kenya,[82] Pakistan[83] and the Russian Federation.[84] In effect, such laws lead to the suppression of legitimate online expressions with limited or no due process or without prior court order and contrary to requirements of article 19 (3) of the International Covenant on Civil and Political Rights.[85] In Latin America, disinformation laws[86] that force platforms to decide whether to remove content without judicial orders are incompatible with article 13 of the American Convention on Human Rights.

58. The trend that sees States delegating to online platforms "speech police" functions that traditionally belong to the courts has continued. The risk with such laws is that intermediaries are likely to err on the side of caution and "over-remove" content for fear of being sanctioned. The German Network Enforcement Act, adopted in 2018 and recently amended, allows users to flag content that they believe is illegal under certain provisions of the Criminal Code and obliges platforms to remove the "violating content" within a short period of time or face heavy fines.[87] Although the German statute does not prohibit or penalize disinformation, it has been cited by other countries seeking to introduce unduly restrictive intermediary laws or social media regulations that would enable the removal of "fake news" without a judicial or even a quasi-judicial order.

E. Emerging trends

59. In countries where disinformation or misinformation is not explicitly banned, States have generally relied on social media platforms' terms of service to tackle disinformation. However, a new trend appears to be developing in the European Union, where the draft digital services act would, once adopted and enforced, require platforms and other intermediaries to adopt transparency and due process measures that could, among other things, help to address the problem of disinformation. Such regulatory proposals, which focus on transparency and due process obligations, rather than viewpoint- or content-based regulations, can make a positive contribution to the protection of human rights and greater public accountability of platforms. However, for the regulatory measures to work properly, the independence of the oversight body or regulator must be assured and scrupulously respected.[88]

60. The draft digital services act would also require large online platforms to conduct annual reviews of "significant systemic risks stemming from the functioning and use made of their service", including the "intentional manipulation of their service", which causes or could cause a negative effect on the protection of public health, minors, civil discourse, electoral processes and public security.[89] Appropriate mitigation measures would have to be adopted in response, subject to independent auditing. Very large online platforms would also be expected to comply with codes of conduct, including on disinformation, under independent regulatory oversight. Whether these due diligence measures will protect human

[81] See, e.g., BRA 6/2020; see also the submission from Derechos Digitales.
[82] OL KEN 10/2017.
[83] OL PAK 3/2020.
[84] OL RUS 4/2019.
[85] See https://transparency.facebook.com/content-restrictions.
[86] OL BRA 6/2020.
[87] OL DEU 1/2017.
[88] See, e.g., Eleonora Maria Mazzoli and Damian Tambini, "Prioritisation uncovered: the discoverability of public interest content online", Council of Europe study DGI(2020)19, pp. 40–43.
[89] See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en.

rights and address disinformation in the States members of the European Union will depend ultimately on how clearly and narrowly they are drafted into law and on the effectiveness and independence of the regulatory bodies.

61. Another emerging trend is draft laws or proposals that seek to restrict the amplification of illegal or otherwise harmful content, for example in Brazil,[90] France[91] and the United States.[92] Such measures could help to reduce the impact of disinformation if the laws are drafted in clear and precise language so that they do not infringe on free speech and seek only to suppress undue reach.

62. The permanent suspension of former United States President Trump from Twitter, Facebook, YouTube, Reddit and other platforms following the events of 6 January 2021 at the United States Capitol have also prompted regulatory proposals to sanction companies for removing lawful content. In Poland, a new proposed law would require social media companies to put back content deemed lawful by a body largely controlled by the Government.[93] In Brazil, terms of service that allow for viewpoint-based removal of content could lead to the suspension of services or other sanctions. Such "must-carry" obligations or related obligations characterized by the "duty of impartiality" could be detrimental if they are not in line with international human rights law, which, while strongly upholding freedom of expression, prohibits hate speech.[94]

V. Company responses: key concerns

63. Companies do not have the same human rights obligations as States. They are, however, expected to respect human rights in their activities and operations in line with the Guiding Principles on Business and Human Rights. At a minimum, they should conduct regular human rights impact assessments of their products, operations and policies and implement due diligence processes with a view to identifying, preventing or mitigating any actual or potential adverse impacts on human rights. They should also put in place a remediation process for users. The Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda and the reports of the previous mandate holder also provide important guidance for companies on human rights standards applicable to their policies and content-moderation practices.

64. In response to the challenges raised by disinformation and misinformation, the largest United States-based social media platforms[95] have adopted a range of policies and tools. They generally ban what they consider to be "false news" and various deceptive practices that undermine authenticity and integrity on their platforms. Some have adopted specific policies on COVID-19-related misinformation[96] and on civic integrity.[97] Enforcement can take various forms, from the application of labels and the issuance of warnings to the removal of content and closure of accounts. Problematic information is made less visible or its reach is reduced.[98] Companies also cite efforts to promote authoritative content.[99] Facebook has established a third-party fact-checking programme.[100] Most recently, Twitter announced a

[90] OL BRA 6/2020.
[91] OL FRA 5/2018.
[92] Submission from Access Now.
[93] See, e.g., Richard Wingfield, "Poland: draft law on the protection of freedom of speech on online social networking sites", 8 February 2021.
[94] See, e.g., the finding of a court in Rome in *Facebook v. CasaPound*, case No. 80961/19, 29 April 2020.
[95] Owing to the lack of easily accessible information, the present section refers mainly to Facebook, YouTube and Twitter and, to some extent, to TikTok.
[96] Twitter and YouTube, for example.
[97] See https://help.twitter.com/en/rules-and-policies/election-integrity-policy.
[98] See, e.g., Tessa Lyons, "Hard questions: what's Facebook's strategy for stopping false news?", 23 May 2018. See also Twitter's enforcement options, available from https://help.twitter.com/en/rules-and-policies/enforcement-options.
[99] See, e.g., www.youtube.com/howyoutubeworks/product-features/news-information/.
[100] See www.facebook.com/journalismproject/programs/third-party-fact-checking.

new community-based approach to fighting misinformation.[101] Beyond social media platforms, messaging services such as WhatsApp have limited the ability of their users to forward messages to an unlimited number of people, particularly at sensitive times, like elections.[102]

65. While these measures are generally positive, they are an insufficient response to the challenges posed by disinformation. Reactive content moderation efforts are simply not enough to make a meaningful difference in the absence of a serious review of the business model that underpins much of the drivers of disinformation and misinformation.[103] Moreover, content moderation efforts continue to display the same long-standing problems of inconsistent application of companies' terms of service, inadequate redress mechanisms and a lack of transparency and access to data that hampers an objective assessment of the effectiveness of the measures that have been adopted. Furthermore, although the platforms are global businesses, they do not appear to apply their policies consistently across all geographical areas or to uphold human rights in all jurisdictions to the same extent.

A. Advertisement-driven business model

66. Algorithms, targeted advertising and the data harvesting practices of the largest social media companies are largely credited with driving users towards "extremist" content and conspiracy theories that undermine the right to form an opinion and freedom of expression.[104] There is a real concern that the systematic collection of data about users' activities online and targeted advertising may violate their right to freedom of opinion under article 19 (1) of the International Covenant on Civil and Political Rights. The lack of transparency with which companies automatically curate content online also points towards an unacceptable level of intrusion into individuals' right to form their ideas free from manipulation and right to privacy. By designing their products with highly personalized content to encourage addictive engagement, companies further promote a system that significantly undermines people's agency and choice in relation to their information diet.[105] Finally, there is evidence to suggest that the recording of people's private thoughts as expressed through online searches and other online activities could be used against them by commercial actors or Governments in a discriminatory manner.[106] There is also a concern that collecting data on ethnicity or political affiliation with no limitations or safeguards in countries with a history of political violence could be dangerous.

67. Despite these concerns, it is not clear whether social media platforms have sought to review their business model as part of their human rights due diligence efforts. Nor is there sufficient publicly available information to enable users, researchers and activists to understand the way in which algorithms promote certain kinds of content.

68. The main area where the largest companies appear to have adopted some measures is in relation to political advertising. The extent of election-related disinformation has belatedly prompted social media companies to create advertisement archives or libraries that enable some scrutiny of political advertising on their platforms. The information available, however, is often too limited and checks on political advertisers are optional.[107] Moreover, users are exposed to political ads by default rather than through an opt-in mechanism. More generally, it is unclear what criteria or objectives are used for the purposes of targeted advertising and whether they are compatible with human rights standards.

[101] Keith Coleman, "Introducing Birdwatch, a community-based approach to misinformation", 25 January 2021.
[102] See, e.g., the submission from Les femmes, la force du changement.
[103] Submission from the Center for Social Media and Politics.
[104] See, e.g., the submission from Vodafone.
[105] A/HRC/44/49, para. 60.
[106] Evelyn Aswad, "Losing the freedom to be human", p. 363.
[107] Submission from Privacy International.

69. Greater transparency is needed on the issue, including through information about the targeting, actual reach and amount spent on advertisements.[108] In the case of political advertising, this should take place alongside reforms by States to electoral laws to ensure that online political advertising and digital campaigns do not undermine the integrity of elections and democratic processes.

B. Application of rules

70. Social media companies rely on a range of content policies to tackle disinformation or misinformation online. The applicable rules can be hard to find, strewn as they often are across various parts of the companies' websites in community standards, policies, leadership statements, newsrooms, product information pages and business help centres.[109] A common concern is that the definitions are often overly broad: they do not always clearly spell out what kind of harm and what likelihood of harm will lead to content removal, labelling or other sanction. The application of the coordinated inauthentic behaviour policy may also have an adverse effect on freedom of expression, particularly on legitimate campaigning activities.[110] The lack of clarity in the companies' definitions hinders consistency in the implementation of companies' rules and policies.

71. This problem is exacerbated by the over-reliance on automated filters that are unable to capture nuance or understand context. The limits of technology and the political nature of determining what constitutes disinformation combined with the lack of transparency on content moderation decisions increase the risk that permissible content will be removed. It underlines the need for human involvement in content removal decisions, particularly when there is a risk of real-world injury or violence.[111]

C. Remedies

72. Companies continue to fail to provide adequate remedies for wrongful actions taken on the basis of disinformation or misinformation. Appeals mechanisms for wrongful decisions are crucial to offset the significant risks inherent in large social media companies using imperfect filters to remove content. Appeals do not appear, however, to be available for enforcement actions taken by companies such as labelling and demotions. Nor do they appear to be available to challenge decisions taken on the basis of coordinated harm or inauthentic behaviour policies. Moreover, it is unclear whether appeals mechanisms are available in a range of languages.

73. In addition to internal complaints mechanisms, proposals for third-party oversight bodies can be a valuable means of strengthening remedies. The Facebook Oversight Board, an external complaints' mechanism, is a novel experiment. While it is too early to assess its effectiveness, it should be evaluated in due course through a transparent, multi-stakeholder participatory process, as it could yield valuable lessons for the sector. It is also necessary to consider industry-wide, multi-stakeholder measures such as the establishment of social media councils, particularly for smaller players. Such multi-stakeholder bodies could provide policy recommendations, including on disinformation and misinformation, and consider appeals from the decisions made by participating companies.

[108] Kate Jones, "Online disinformation and political discourse", p. 54. See also Nathalie Maréchal, Rebecca MacKinnon and Jessica Dheere, "Getting to the source of infodemics: it's the business model" (New America, May 2020), pp. 55–56.

[109] See, e.g., the Facebook Oversight Board decision in case 2020-006-FB-FBR.

[110] See www.accessnow.org/rights-groups-to-facebook-on-tunisias-disappeared-accounts-were-still-waiting-for-answers/.

[111] Submission from Global Partners Digital.

D. Geographical disparities

74. A significant problem in companies' approach to content moderation is the apparent disparity with which they implement their policies in different parts of the world.[112] While the United States has benefited from special election or voting information centres or advertisement libraries, most countries do not seem to receive the same level of investment.[113] Even in the United States, Spanish-language resources are less well funded than English-speaking ones. Some companies are only gradually developing temporary country-specific election misinformation policies in places where they operate,[114] but they seem to be a minority. Moreover, the ability to flag content appears to be fully available only in the United States, much less so in Latin American countries.[115]

75. Fact-checking services are more limited in many parts of the world.[116] Information reported by whistle-blowers raised concerns about the deliberate inaction of companies in less affluent markets. If confirmed, the disparities would show a very different quality of content moderation, undermining civic space in developing countries.[117]

76. The largest companies, based in the United States and influenced by its politics and public opinion there, appear to be driven by United States and European priorities. They do not invest sufficient resources in understanding the local factors that feed disinformation online in other parts of the world, especially developing countries.[118] A thorough understanding of the local political, social and economic context, language proficiency and close cooperation with civil society in countries where disinformation is more prevalent are necessary.

E. Political pressure

77. Companies do not appear to apply their terms of service or community standards consistently to public figures. Nor do they appear to have developed clear policies to protect political content from being censored as "false news" or to address content produced by public figures that advocates incitement to violence.

78. The inconsistent application of community standards led to criticism of the large platforms during the United States Presidential elections in 2020. In February 2021, Facebook banned accounts linked to the military following the coup d'état in Myanmar but has not committed itself to doing the same in other situations. When platforms were pressured by the authorities to close the accounts of journalists and human rights defenders covering the farmers' protests in India earlier this year, they appear to have complied.[119] In April 2020, it was reported that Facebook had agreed to significantly increase compliance with the request of the Government of Viet Nam to censor "anti-State" content and closed the accounts of many human rights defenders after the authorities took down its servers, which slowed down the platform and made it inoperable for seven weeks.[120]

79. In the absence of clear policies, companies are vulnerable to pressure to clamp down on legitimate political speech and facilitate State-instigated disinformation. Companies should base their community standards on international human rights standards, including those relating to the right to freedom of expression, under which the speech of politicians, political parties and other public figures and entities would benefit from a high degree of protection. International human rights law permits the removal of content posted by a public

[112] See, e.g., the submission from Intervozes.
[113] Submission from Privacy International.
[114] See the submission from Facebook on its strategy in Myanmar.
[115] Submission from Derechos Digitales.
[116] See Mahsa Alimardani and Mona Elswah, "Trust, religion and politics: coronavirus misinformation in Iran", in *2020 Misinfodemic Report: COVID-19 in Emerging Economies* (Meedan, 2020).
[117] See www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo.
[118] Submission from the Centre for Law and Democracy.
[119] IND 2/2021.
[120] See www.reuters.com/article/us-vietnam-facebook-exclusive-idUSKCN2232JX.

figure only on very narrow grounds, including incitement to violence and hatred. By grounding their terms of service firmly in international standards, companies would be better placed to resist pressure to remove legitimate speech since they would be upholding the very instruments and principles to which States have formally committed themselves.

F. Transparency, accountability and access to data

80. Lack of transparency and access to data continue to be the major failings of companies across almost all the concerns in relation to disinformation and misinformation. They prevent independent scrutiny and affect accountability and trust. Opacity is disempowering for users and denies agency.

81. Most of the largest social media companies produce transparency reports twice a year, but they do not share more precise and meaningful information about action taken to address disinformation or misinformation.[121] The Facebook transparency report, for example, only provides information about the removal of fake accounts but not about content. Similarly, no information is provided about the number of items of content being labelled, nor any appeals against decisions to remove content or suspend accounts in relation to disinformation policies. There is no data available about user engagement with disinformation or misinformation, including numbers of shares, views, reach and number of complaints or requests for removal. Greater information is also needed about the reliability and accuracy of the artificial intelligence systems deployed to identify and remove content.[122] The overall lack of transparency regarding companies' content moderation systems and processes makes it impossible to assess the effectiveness of the measures adopted by the companies and their impact on human rights. This is compounded by the lack of sufficient access to data by researchers, academics or civil society to enable them to make more independent and objective assessments.

82. Greater transparency is also needed on agreements between companies and Governments, especially when the agreements involve giving more prominence to government messaging or to the removal of content or other restrictions on speech.

VI. Conclusions and recommendations

83. **In a report devoted to disinformation, it is easy – but dangerous – to lose sight of the value that digital technology offers to democracy, sustainable development and human rights, or the vital importance of the right to freedom of opinion and expression in that equation. That is why attempts to combat disinformation by undermining human rights are short-sighted and counterproductive. The right to freedom of opinion and expression is not part of the problem, it is the objective and the means for combating disinformation. The COVID-19 pandemic has starkly exposed both the imperative of upholding the right and the challenges of confronting disinformation and misinformation.**

84. **Disinformation is a complex, multifaceted phenomenon with serious consequences. It destroys people's trust in democratic institutions. It thrives where public information regimes are weak and independent investigative journalism is constrained. It disempowers individuals, robbing them of their autonomy to search, receive and share information and form opinions. In the platform world, individuals are regarded as users, not as rights holders with agency.**

85. **Disinformation is problematic, but so too are the responses of States and companies. Laws and policies are often being made with sub-optimal knowledge of online harm, without adequate data, research or public consultations. States have resorted to disproportionate measures such as Internet shutdowns and vague and**

[121] See, e.g., Working Group on Infodemics, *Policy Framework* (Forum on Information and Democracy, November 2020), pp. 17 ff., See also the submission from the Association for Progressive Communications.

[122] See, e.g., the submission from the Association for Progressive Communications.

overly broad laws to criminalize, block, censor and chill online speech and shrink civic space. These measures are not only incompatible with international human rights law but also contribute to amplifying misperceptions, fostering fear and entrenching public mistrust of institutions.

86. **Company responses have been reactive, inadequate and opaque. The large platforms are focused on improving content moderation while ignoring human rights concerns about their business models, lack of transparency and the inadequate due process rights of users.**

87. **The fundamental challenge for States, companies and the media is to restore public trust in the integrity of the information order. Tackling disinformation requires multidimensional, multi-stakeholder responses that are well grounded in the full range of human rights and the proactive engagement of States, companies, international organizations, civil society and the media. The need for multi-stakeholder dialogue and partnerships cannot be overstated.**

88. **States are the primary duty bearers with obligations to respect, protect and fulfil human rights. In keeping with their obligation to respect human rights, States should not make, sponsor, encourage or disseminate statements that they know or should reasonably know to be false, or authorize Internet shutdowns as a means of combating disinformation. They should refrain from restricting freedom of expression online or offline except in accordance with the requirements of articles 19 (3) and 20 (2) of the International Covenant on Civil and Political Rights, strictly and narrowly construed.**

89. **Criminal law should be used only in very exceptional and most egregious circumstances of incitement to violence, hatred or discrimination. Criminal libel laws are a legacy of the colonial past and have no place in modern democratic societies. They should be repealed.**

90. **States have the duty to ensure that companies respect human rights. They should not compel companies to remove or block content that is legitimate under international human rights law, nor require them to make determinations on the legality of content under national laws that should be done by the courts. Companies should be transparent about such requests from States and refrain from making deals behind closed doors.**

91. **State regulation of social media should focus on enforcing transparency, due process rights for users and due diligence on human rights by companies, and on ensuring that the independence and remit of the regulators are clearly defined, guaranteed and limited by law.**

92. **Data protection is key to reorienting the advertisement-driven business model of the digital economy, which drives the information disorder and related human rights abuses. States should adopt strong data protection laws and update electoral and other relevant laws to limit the pervasive tracking and targeting of individuals and their activities online.**

93. **Diverse and reliable information is an obvious antidote to disinformation and misinformation. States should fulfil their duty to ensure the right to information, firstly, by increasing their own transparency and by proactively disclosing official data online and offline and, secondly, by reaffirming their commitment to media freedom, diversity and independence. Ensuring the safety of journalists online and offline and ending impunity for threats, intimidation, harassment, attacks and killings of journalists, including women journalists, bloggers, cartoonists and human rights defenders is key to restoring confidence in the public sphere as a safe place for democratic deliberations.**

94. **Media information and digital literacy empowers people and builds their resilience against disinformation and misinformation, as noted recently by the General Assembly.**[123] **It should become part of the national school curriculum and engage the young and old alike. Along with digital literacy, more attention must be given to digital**

[123] Resolution 75/267.

inclusion so that people in developing countries who are now totally dependent on social media platforms and messaging applications for connectivity (through zero rating) can have meaningful, free, open, interoperable, reliable and secure access to the Internet.

95. **Companies are obliged to respect human rights under international human rights law. Although digital platforms are private actors, they have a far-reaching impact on human rights in the public space. As such, they are accountable not only to their users but to society at large. There is growing concern about the market dominance of the largest companies, as well as about the harmful effects of their current business models. Companies should proactively respond to these concerns, going beyond improving content moderation to reviewing their business models, acknowledging the agency and autonomy of users as rights holders and empowering them by increasing transparency, control and choice and by ensuring due process.**

96. **In line with the Guiding Principles on Business and Human Rights, social media companies should review their business models and ensure that their business operations, data collection and data processing practices are compliant with international human rights standards, including article 19 of the International Covenant on Civil and Political Rights, as well as data protection principles and relevant national consumer protection standards. They should also conduct human rights impact assessments of their products, particularly of the role of algorithms and ranking systems in amplifying disinformation or misinformation. Such assessments should be conducted regularly and ahead of and following significant events such as national elections or major crises like the COVID-19 pandemic.**

97. **Companies should review their advertising models to ensure that they do not adversely impact diversity of opinions and ideas and are clear on the criteria used for targeted advertising. They should provide meaningful information about advertisers in online advertisement repositories and give users the choice to opt in to be exposed to advertising.**

98. **Companies should adopt clear, narrowly defined content and advertising policies on disinformation and misinformation that are in line with international human rights law and after consultation with all relevant stakeholders. Furthermore, they should adopt clear policies relating to public figures that are consistent with international human rights standards and apply them consistently across geographical areas. They should ensure that all policies are easily accessible and understandable by users and are enforced consistently, taking into account the particular contexts in which they are applied.**

99. **Companies should provide clear and meaningful information about the parameters of their algorithms or recommender systems and ensure that those systems enable users to receive a diversity of viewpoints by default while also enabling them to choose the variables that shape their online experience.**

100. **Companies should publish comprehensive, detailed and contextualized transparency reports, including separate reports to address exceptional circumstances such as the COVID-19 pandemic, that include a breakdown of the actions taken against disinformation- or misinformation-related content and appeals against those actions, including number of shares, views, reach, complaints and requests for removal.**

101. **Users must have proper recourse. Companies should establish internal appeals mechanisms for a broader range of content moderation decisions and types of content, such as coordinated inauthentic behaviour. They should also explore the creation of external oversight mechanisms such as social media councils.**

102. **In the age of the Me Too movement, both States and companies should confront gender disinformation online as a priority and also give special attention to its consequences in the real world. Companies should introduce appropriate policies, remedies and mechanisms that are tailored from a gender perspective across all aspects of the platform experience and that are designed in consultation with those affected by this pernicious behaviour. States should also integrate fully gendered perspectives into**

their policies and programmes to address disinformation and misinformation, including in media, information and digital literacy programmes.

103. **As actors on a global stage, companies should invest more resources to develop their understanding of local contexts that drive disinformation and misinformation and address the disparities in their knowledge, languages, policies and services in relation to developing countries, minorities and other vulnerable groups, drawing on the perspectives of local civil society and groups targeted by disinformation and misinformation.**

104. **A major challenge to addressing disinformation and misinformation is knowledge gaps arising from the lack of access to data, especially in relation to developing countries. More attention should be given to making data available for research, policymaking, monitoring and evaluation. The concept of "differential privacy"[124] could provide a way forward for access to mega data for research purposes while respecting the human rights and safety concerns of users.**

105. **Last but not least, the United Nations human rights system and, in particular, the Human Rights Council, has a major role to play in ensuring that all efforts to address disinformation and misinformation are grounded firmly in international human rights law, including respect for freedom of opinion and expression. The Council should consider holding regular multi-stakeholder consultations with States, companies, civil society organizations and relevant international and regional actors and establishing initiatives on the subject of safeguarding and promoting human rights in the digital space.**

[124] Working Group on Infodemics, *Policy Framework*, p. 125.