



Submission by the International Observatory of Human Rights to the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression

Input on disinformation to inform the Special Rapporteur's annual thematic report to be presented to the Human Rights Council at its 47th session in June 2021

15 February 2021

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

1. The key challenges of disinformation are identifying and controlling false news. One of the biggest problems with online content is its fast pace, which makes identifying disinformation in real-time in order to prevent it from being spread, impossible.

‘Even when there is early evidence that a Fake Story is being circulated online, there is not much to do until there is certainty about it. Otherwise, e.g. removing a story or preventing people from sharing it based on early signals, could be perceived as an attempt of intervention and censorship.’¹

2. Disinformation impairs Freedom of Expression because people are unable to make informed decisions and have no access to true information.
3. Instant and global reach of Disinformation: Disinformation is shared among the population because it is profitable for the providers (websites want to have more traffic, so they can run and sell more ads and increase their financial gains) or to achieve political gains.
 - Social Media is set up for instant sharing and the goal is gaining as much user attention and user engagement as possible (CTR, Clickbait).
 - To achieve higher user attention and engagement algorithms are used to provide users with information that “aligns” with their views.
 - “Fake News” are ‘engineered to attract attention and trigger emotional reactions’² because their purpose is to go viral as quickly as possible.
 - As a result, the performance of content is more valued than its quality.

What legislative, administrative, policy, regulatory or other measures have Governments taken to counter disinformation online and offline? And what has been the impact of such measures on i) disinformation; ii) freedom of opinion and expression; and iii) other human rights?

¹ George Krasadakis, ‘Misinformation and Fake News: A Solution Powered by the Latest Digital Technologies’ (5 November 2020) Towards Data Science. Available at: <https://towardsdatascience.com/misinformation-and-fake-news-a-solution-powered-by-the-latest-digital-technologies-b585c73daddf>.

² George Krasadakis, ‘Misinformation and Fake News: A Solution Powered by the Latest Digital Technologies’ (5 November 2020) Towards Data Science. Available at: <https://towardsdatascience.com/misinformation-and-fake-news-a-solution-powered-by-the-latest-digital-technologies-b585c73daddf>.

4. States have typically pursued one of three models; Self-regulation or regulation by contract, interventionist regulation and finally co-regulation.³

Regulation by contract

5. Tech companies have conditional immunity from liability as long as they uphold their set terms of service. Regulation by contract relies on voluntary compliance on the side of the platforms as well as users of those platforms understanding their rights protected in the codes of conduct. These models are easily manipulated to circumvent the rule of law and it is often hard to hold platforms to account.⁴

Interventionist Regulation

6. Such frameworks are characterised by states accumulating disproportionate censorship powers. This may manifest through content-blocking, fines and even prison sentences.⁵ In interventionist models, the laws that underpin regulators powers are often overly broad. This problem is accentuated in countries in which the regulator lacks independence. Interventionist policies ensure the state is ultimately in charge of what speech is and is not acceptable and poses a number of problems.

Co-regulatory framework

7. A co-regulatory framework, such as that implemented in the EU's new Digital Services Act, differs from interventionist regulation but ultimately suffers from the same shortcomings. These frameworks involve tighter regulation, encouraged or supported, and while the state is not the ultimate arbiter, they tend to give state institutions too much scope over the regulation of online content.

A new framework?

8. NGO Article 19 has proposed a new framework, of which the International Observatory of Human Rights supports. This would see terms and services standardised to, at minimum, be compatible with the UN guiding principles. The state to address cases where they deem platforms have fallen short of these standards - effectively acting as a grievance procedure. Article 19 has suggested that, in practice the "creation of a new cause of action could be derived either from traditional tort law principles or, as noted above, the application of constitutional theory to the enforcement of contracts between private parties".⁶

What policies, procedures or other measures have digital tech companies introduced to address the problem of disinformation?

9. While platform's policies have varied in setting the parameters of when and why actions are taken, the procedures and measures taken to counter disinformation often converge.

³ Haw Ang, P. (2008). International Regulation of Internet Content: Possibilities and Limits. Access: <https://www.dhi.ac.uk/san/waysofbeing/data/governance-crone-ang-2008.pdf>

⁴ Article 19. (2018). Side-stepping rights: Regulating speech by contract. Access: <https://www.article19.org/wp-content/uploads/2018/06/Regulating-speech-by-contract-WEB.pdf>

⁵ Article 19. (2013). Freedom of expression and ICTs: Overview of international standards. Access: <https://www.article19.org/wp-content/uploads/2018/02/FoE-and-ICTs.pdf>

⁶ Article 19. (2018). Side-stepping rights: Regulating speech by contract. Access: <https://www.article19.org/wp-content/uploads/2018/06/Regulating-speech-by-contract-WEB.pdf>

10. In December 2020, UK Head of Public Policy at Facebook told the Sub-Committee on Online Harms and Disinformation that the tech platform “broadly think[s] about misinformation in two brackets. One is misinformation that can cause imminent harm and real-world danger, and other types of misinformation”, and that the procedure and measures used to address the two types vary.⁷ It has been described as a “two-pronged approach. One is about the kind of content we serve, then proactive interventions to direct people to those sources of information.”⁸
11. All the major tech companies can broadly think of misinformation along similar lines.
12. Common measures include: Removing misinformation, downranking them to make it harder to find, labelling content, blocking fake accounts, and promoting authoritative information.

Policies:

13. Policies - often encompassed in a platform’s terms and conditions, guidelines, and community standards - set out what content is and is not acceptable on their platform. A report looking into disinformation during COVID-19 states these policies: “often go beyond the requirements of the [UK’s] law, such as in the case of hate speech or graphically violent content”.⁹
14. However, Mehwish Ansari, Head of Digital for Article 19 has said: “as a general understanding of the way that dominant platforms have approached freedom of expression, their terms of service usually include lower standards for restrictions of freedom of expression than those permitted under international human rights law”¹⁰
15. The tech companies uphold that “their policies were their primary consideration when tackling misinformation, disinformation and other so-called ‘harmful but legal content online’.
16. Ambiguity over concepts such as ‘legal but harmful’, as well as the tech companies own policies being unclear and applied inconsistently, tend to have a chilling effect on human rights. Regulation of tech companies should primarily focus on standardising policies across different platforms, with the policies based on the UN Guiding Principles on Business and Human Rights and other sources of international human rights law. In turn, this will help to ensure the procedures and measures used by the platforms are consistent, both in their application and in their alignment with international standards.
17. States can then act in cases of grievance, potentially creating greater clarity about how, when and the process for which users seek remedies.

Facebook:

18. *Advertising:* In response to the coronavirus advert Facebook further “refined” their advertising policy. Expanding the policy to include “adverts that might fall short of that test of “Is it fake?”, but might be scaremongering” or discouraging users to take the vaccine.¹¹
19. *Correct the Record:* Facebook has implemented a “Correct The Record Tool”, long lobbied for by Global Advocacy group Avaaz and which activists and politicians have previously touted as one of the

⁷ <https://committees.parliament.uk/oralevidence/1446/pdf/>

⁸ Ibid

⁹ <https://committees.parliament.uk/publications/1954/documents/19089/default/>

¹⁰ <https://committees.parliament.uk/oralevidence/1368/pdf/>

¹¹ Ibid

most effective measures to counter disinformation.¹² This tool's effect is two-fold *a)* By informing via notification that they have engaged with verified misinformation (expanded during COVID to users who shared, clicked on, or liked a verified piece of misinformation information); *b)* By linking users to authoritative information e.g. If you search a COVID-related term, a pop-up would direct you to an NHS or WHO source of information.¹³¹⁴

Twitter:

20. Twitter has also begun applying warning labels to content that has been independently fact-checked and debunked. However, Facebook has gone further in its "Correct the Record" tool, as it retroactively provides authoritative information so some people who have previously encountered disinformation.
21. Twitter's Global Policy Director, Nick Pickles however rejected academic research around the use of "correct the record" tools, saying: "a number of studies around correct the record are not peer reviewed"¹⁵

Google:

22. Google has particularly emphasised its role in funding quality journalism, through projects such as the global Journalist Emergency Relief Fund and making a \$1 million donation to the International Center for Journalists.

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

23. Article 19 has noted that "On content removal as an example, while the mechanisms put in place by dominant social media platforms to remove content generally feature some procedural safeguards, none contains all the appropriate safeguards, so they fall short of international standards on freedom of expression and due process in some way."
24. The UK Digital, Culture, Media and Sports Committee highlight that "The Competition and Markets Authority's recent market study final report into the Online Platforms and Digital Advertising, concluded that weak competition in digital advertising caused by players such as Facebook and Google "undermines the ability of newspapers and others to produce valuable content, to the detriment of broader society".¹⁶

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

25. A major problem with current regulation among states is how to remedy situations where users feel their legitimate exercise of free speech has been infringed upon.
26. A huge hurdle lies in the lack of transparency in the process. There is simply not enough reporting on behalf of the companies on what content has been removed, under what basis (infringing terms of services, at the request of a government), the number of complaints received about wrongful removal

¹² <https://time.com/5540995/correct-the-record-polling-fake-news/>

¹³ <https://committees.parliament.uk/oralevidence/1446/pdf/>

¹⁴ <https://committees.parliament.uk/publications/1954/documents/19089/default/>

¹⁵ <https://committees.parliament.uk/oralevidence/459/html/>

¹⁶ <https://committees.parliament.uk/publications/1954/documents/19089/default/>

of content and the outcome of such complaints.

27. Regardless of the regulatory model used, a lack of transparency in the process almost always ensures that companies and governments cannot be held to account. Article 19 has argued that this means “there is no easy answer to the question of better or worse models. It goes back to the more fundamental approach that ARTICLE 19 proposed, which looks to hold companies to the UN Guiding Principles on Business and Human Rights and centring the international human rights framework as they develop their community standards on terms of service, and for the state to provide an effective remedy for free speech violations by private actors.”

28. Basing policies around the UN Guiding Principles and existing international law will also help to standardise policies across platforms. This provides greater clarity on behalf of the users as what constitutes free speech and allows states to act as a grievance procedure.

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.

Education and Awareness

29. An effective long-term measure is educating users of social platforms on detecting disinformation and verifying content. In doing so, the freedom of speech is not impacted in any way, while the negative impact of disinformation is minimized/neutralised. On the hand, users do not believe false information as easily if they come across it, and on the other hand, they are less likely to circulate it themselves. Awareness raising and training of critical thinking should be made compulsory in schools, and outside-of-school training programs including content literacy tools should be provided. To this effect, media literacy and public awareness campaigns have been implemented in Australia, Canada, Denmark, the Netherlands or Nigeria.¹⁷ For example the school project *Faktana, kiitos* (“Facts, please!”) in Finland has brought 124 journalists to meet around 7,200 school students between September and December 2017.¹⁸ States should be obligated to carry out media literacy campaigns and report on their implementation.

Advertising

30. An effective measure to stop the deliberate spreading of disinformation is to cut off the ‘flow of digital advertising funds to shady operators peddling disinformation’,¹⁹ provided the motivation for creating the disinformation is purely financial. If a website or content creator loses the monetary benefit from advertisements and promotions, they are likely to be losing their incentive to create and share the disinformation.

‘Professor Ang Peng Hwa of the Wee Kim Wee School of Communication and Information at Nanyang Technological University says the effectiveness of this has already been shown in the United States, when Facebook, Apple, Twitter, YouTube and Spotify removed podcasts, pages and channels of fake news peddler Infowars.’²⁰

¹⁷ See Poynter, ‘A guide to anti-misinformation actions around the world’ (2019). Available at: <https://www.poynter.org/ifcn/anti-misinformation-actions/#germany%20https://www.loc.gov/law/help/fake-news/index.php>.

¹⁸ IPI, ‘New Finnish Project brings journalists to schools to teach media literacy’ (2018). Available at: <https://ipi.media/new-finnish-project-brings-journalists-to-schools-to-teach-media-literacy/>.

¹⁹ Yasmine Yahya, ‘3 key strategies to fight fake news’ (30 September 2018) The Straits Times, available at: <https://www.straitstimes.com/politics/3-key-strategies-to-fight-fake-news>.

²⁰ Yasmine Yahya, ‘3 key strategies to fight fake news’ (30 September 2018) The Straits Times, available at: <https://www.straitstimes.com/politics/3-key-strategies-to-fight-fake-news>.

31. Therefore, Governments have to issue clear regulations demonetizing false content.

Alternative Information

32. Another strategy to tackle disinformation is simply to create more truthful information. By making alternative information available alongside misleading information, users are exposed to different opinions and the effects of disinformation can be limited. 'Instead of killing the story', the disinformation is swamped with the truth.²¹ Academic research suggests that this approach could make a real difference in readers' perceptions.²² Facebook has already tested this method by providing alternative information with a story through its function *Related Articles*.²³ This function does not restrict freedom of speech, but increases it, and does not 'imply any editorial judgement' about the truthfulness of the stories shared.²⁴

Transparency

33. A similar approach, which does not restrict the freedom of speech, is to increase the transparency of information. This measure does not go after wrong, misleading information either, but instead attempts to increase credibility of information through transparency. A successful example is the *Transparent Journalism Tool* in Spain, which introduces 'radical transparency in the editorial process'.²⁵ Publications by the online news site Público are shared with the reasons for covering a certain topic, the number of people working on the content, validated sources, consulted documents, among other things. This not only increases the trustworthiness of articles shared with this editorial information, but it also simultaneously decreases the credibility of articles that are published without this additional information. Greater transparency is also needed in regard to the reporting and accountability in the use of algorithms and artificial intelligence on Social Media.

Technical Solutions

34. Various technical solutions are available to tackle the problem of disinformation and to give prominence to trustworthy information. For example, search engines can list their results based on quality rather than popularity. Social Media Platforms can make it a requirement for users to read through a link, before they are able to share it or limit the amount of how often a message can be forwarded.²⁶ 'A good solution would require advanced digital technologies and protocols for assessing content (e.g. to efficiently spot and fact-check the less reliable content)'.²⁷

²¹ Alberto Alemmano, 'How to Counter Fake News? A Taxonomy of Anti-fake News Approaches' (2018) Cambridge University Press, available at: <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/how-to-counter-fake-news-a-taxonomy-of-antifake-news-approaches/EA53D30745F601834218DDD7DB90950A>.

²² Bode, L and Vrada, EK, "In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media" (2015) 4(65) *Journal of Communication* 619–638 available at <https://academic.oup.com/joc/article-abstract/65/4/619/4082315?redirectedFrom=fulltext>.

²³ Sara Su, 'New Test With Related Articles' (25 April 2017). Available at: <https://about.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/>.

²⁴ Alberto Alemmano, 'How to Counter Fake News? A Taxonomy of Anti-fake News Approaches' (2018) Cambridge University Press, available at: <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/how-to-counter-fake-news-a-taxonomy-of-antifake-news-approaches/EA53D30745F601834218DDD7DB90950A>.

²⁵ IPI, 'Spanish project bets on 'radical transparency' to restore reader trust' (2018). Available at: <https://ipi.media/spanish-project-bets-on-radical-transparency-to-restore-reader-trust/>.

²⁶ Alex Hern, 'WhatsApp to impose new limit on forwarding to fight fake news' (07 April 2020) The Guardian. Available at <https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news>.

²⁷ George Krasadakis, 'Misinformation and Fake News: A Solution Powered by the Latest Digital Technologies' (5 November 2020) Towards Data Science. Available at: <https://towardsdatascience.com/misinformation-and-fake-news-a-solution-powered-by-the-latest-digital-technologies-b585c73dadfd>.

Fact-Checking and Individual Verification

35. Fact-checking might not be the best way to stop the spreading of disinformation immediately, since this is generally too slow to stop the instant spreading online. Moreover, many say that it is actually counter-productive: 'As evidence suggests, to categorise a piece of news as fake and thereby give it greater publicity, gives the news piece a boost and spreads its reach even further'.²⁸ Yet, fact-checking online content is a good tool to gain an overview and to be able to monitor the extent of the problem. In this way, fact-checking and identifying authentic content can serve as a retrospective large-scale analysis rather than a real-time classification of content. An example for this action is the [European Digital Media Observatory](#). Additionally, providing fact-checking services and platforms that identify fraudulent content, helps users to investigate themselves whether sources are legitimate or not. For example, the [SocialTruth](#) project provides easy access to various verification services for individual users to prevent the spread of disinformation.²⁹

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.

36. Measures like censorship, internet shutdowns,³⁰ blocking websites,³¹ as well as arrests and imprisonment based on publishing of alleged disinformation³² gravely infringe the freedom of opinion and expression. Laws prohibiting "fake news" are often very vague,³³ and do not comply with the requirements for limitations of human rights. A report from UNESCO 2020 notes that restrictive responses to disinformation, like 'measures that intentionally or unintentionally criminalise critical journalism, such as so called "fake news" laws' often violate international standards requiring proportionality and necessity.³⁴

37. Similarly, *Dunja Mijatović*, Council of Europe Commissioner for Human Rights, stressed that 'measures to combat must never prevent journalists and media actors from carrying out their work or lead to content being unduly blocked on the Internet'.³⁵ Prison sentences for journalists reporting on the pandemic, for example in Hungary and Russia, the obligation on online news outlets to 'prevent the dissemination of "harmful information"' in Azerbaijan, or the removing from content and blocking of

²⁸ Alberto Alemmano, 'How to Counter Fake News? A Taxonomy of Anti-fake News Approaches' (2018) Cambridge University Press, available at: <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/how-to-counter-fake-news-a-taxonomy-of-antifake-news-approaches/EA53D30745F601834218DDD7DB90950A>.

²⁹ European Commission, 'SocialTruth: fighting fake news with trust at the times of Covid-19' (2020). Available at: <https://ec.europa.eu/digital-single-market/en/news/socialtruth-fighting-fake-news-trust-times-covid-19>

³⁰ As have been implemented for example in India, Sri Lanka or Cox Bazar in Bangladesh.

³¹ As has been applied for example in Hungary, Romania, Algeria, Thailand and the Philippines, see Jenna Hand, 'Fake News laws, privacy & free speech on trial: Government overreach in the infodemic?' (2020). Available at: <https://firstdraftnews.org/latest/fake-news-laws-privacy-free-speech-on-trial-government-overreach-in-the-infodemic/>. As well as in Russia, see IPI, 'Rush to pass 'fake news' laws during Covid-19 intensifying global media freedom challenges' (2020). Available at: <https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/>.

³² This happened for example in Bahrain, Bangladesh, Belarus, Benin, Burkina Faso, Cameroon, Côte d'ivoire, Egypt, Indonesia, Kazakhstan, Myanmar, Rwanda, Saudi Arabia and Thailand, see Poynter, 'A guide to anti-misinformation actions around the world' (2019). Available at: <https://www.poynter.org/ifcn/anti-misinformation-actions/#germany%20;%20https://www.loc.gov/law/help/fake-news/index.php>. As well as in Morocco and Algeria, see Kacha, 'In a post-COVID-19 world, "fake news" laws, a new blow to freedom of expression in Algeria and Morocco/Western Sahara?' (2020), Amnesty International. Available at: <https://www.amnesty.org/en/latest/news/2020/05/in-a-post-covid19-world-fake-news-laws-a-new-blow-to-freedom-of-expression-in-algeria-and-morocco-western-sahara/>.

³³ For example in Brazil, Croatia or France.

³⁴ UNESCO, 'Disinfodemic. Dissecting responses to COVID-19 disinformation' (2020) Policy Brief 2. Available at: https://en.unesco.org/sites/default/files/disinfodemic_dissecting_responses_covid19_disinformation.pdf.

³⁵ Council of Europe, 'Press freedom must not be undermined by measures to counter disinformation about COVID-19' (2020). Available at: <https://www.coe.int/en/web/commissioner/-/press-freedom-must-not-be-undermined-by-measures-to-counter-disinformation-about-covid-19>.

websites without being able to appeal against this decision in Romania, are further examples of measures that violate human rights.³⁶

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

1, Working Definition

First and foremost, a working definition of disinformation has to be established. We therefore encourage the Special Rapporteur to issue the following definition of disinformation:

Disinformation is understood as verifiably false, misleading or manipulated information that is created, presented and disseminated for economic or political gain or to intentionally deceive the public, and may cause threats to democracy, public health, the environment, public security or public harm.

To ensure the maximum of freedom of speech, disinformation should not include inadvertent errors, satire, parody, or political commentary.

2: Guidelines

Alongside a definition of disinformation, the Special Rapporteur is encouraged to issue concrete guidelines for state measures that are in line with the freedom of expression, based on the Joint Declaration on ‘Fake News’, Disinformation and Propaganda from 3 March 2017. These guidelines should recommend positive measures to regulate online content, such as education, raising awareness, regulating advertisement, mandating alternative information and transparency, providing fact-checking services and other incentives.

Negative measures for disinformation, such as criminal sanctions, fines, or the removal of content, should not be suggested, but expressly opposed, with the exception of content inciting criminal behaviour or constitution defamation, cyberbullying and harassment in accordance with Art 20 (2) ICCPR. To ensure that the definition of disinformation is not misused by autocratic regimes to shut down political opposition, it is important to ensure that – if negative measures such as removing content are implemented – the power of deciding whether information is “false, misleading or manipulated” does not lie with the government, but with the judiciary or an independent Online Media Council, consisting of multiple institutions.

Moreover, the guidelines should include recommendations regarding working together with corporations that provide online platforms. The General Principles regarding the liability of Intermediaries from the Joint Declaration of 2017 should be reiterated, while encouraging a direct and active collaboration with social media companies. Clear Guidance is needed on the criteria for content moderation, and on procedures of removing content.

Research Contact

Louise Pyne-Jones, Head of Research, International Observatory of Human Rights
+44 (0)7799 000 613 | louise@observatoryihr.org
www.observatoryihr.org

³⁶ Ibid.