Response to the Call for Comments

United Nations Special Rapporteur on the Promotion and Protection of Freedom of Opinion and Expression - Thematic Report on Opportunities, Challenges, and Threats to Media in the Digital Age

January 24, 2022

Justitia

Justitia is Denmark's first judicial think-tank. Justitia aims to promote the rule of law, human rights and fundamental freedoms both within Denmark and abroad by educating and influencing policy experts, decision-makers, and the public. In so doing, Justitia offers legal insight and analysis on a range of contemporary issues.

Future of Free Speech Project

The Future of Free Speech is a collaboration between Justitia, Columbia University's Global Freedom of Expression and Aarhus University's Department of Political Science. We believe that a robust and resilient culture of free speech must be the foundation for the future of any free, democratic society. Even as rapid technological change brings new challenges and threats, free speech must continue to serve as an essential ideal and a fundamental right for all people, regardless of race, ethnicity, religion, nationality, sexual orientation, gender or social standing.

Introduction

News from free and diverse media sources is a fundamental tenet of liberal democracy, a central aspect of the right to freedom of information and expression and an enabler of public participation and dialogue. Today's media landscape is altering predominantly due to the rise of digital media. A handful of social media platforms act as the gatekeepers through which people access news and information in an easy and attractive way. Shulz (2019) argues that the growing number of users relying on social media for news is positively correlated to their discontent with mainstream media. However, as more and more individuals are given space on centralized platforms, the harms of free speech have been amplified since this process of "platformization" provide extremism, hatred, abuse and disinformation with new visibility. As a result, countries are placing increasing pressure on social media platforms to remove allegedly harmful content. For example, the infamous German Network Enforcement Act (NetzDG) discussed later in the report imposes a legal obligation on platforms to remove content such as insult, incitement and religious defamation within short time limits of 24 hours for 'manifestly illegal' content or risk a fine of up to 50 million EUR. At the same time, COVID-19 has brought to the forefront new challenges such as medical disinformation and also impacted media freedom and diversity globally by increasing pressure on traditional media houses to remain financially viable during the pandemic.

Please note that our response to this call emanates predominantly from Justitia's recent report entitled 'Framework of Frist Reference: Decoding a Human Rights Approach to Content Moderation in the Era of Platformization.'

Issue 1: Key trends, threats or challenges to the freedom, independence and diversity of media and safety of journalists & link to policies/practices of social media platforms

Trends/Threats/Challenges

The 2021 World Press Freedom Index found that journalism is under serious threat in almost three-quarters of the 180 ranked countries with only 12 countries having respectable press-freedom environments, the lowest number since 2013. Issues such as intimidation and criminal prosecution of journalists, and restrictive legislation passed to counter false information are some of the marked the situation of media freedom around the global, a reality which has worsened with COVID-19. At Justitia's Future of Free Speech Project, we tracked global restrictions to freedom of expression related to the pandemic. Between 1 February and 15 June 2020, we recorded at least 70 legislative and policy changes leading to various forms of censorship. We documented at least 703 arrests or detentions in 36 countries for, amongst others, allegedly spreading rumours or fake news related to the virus. Worryingly, many of the arrestees were journalists. Other measures have included the blockage of hundreds of news sites in Myanmar and the ban on printing of all newspapers in Iran. The Council of Europe Commissioner for Human Rights has stressed that while COVID-related disinformation must be combatted, some governments are 'using this imperative as a pretext to introduce disproportionate restrictions to press freedom; this is a counterproductive approach that must stop.'

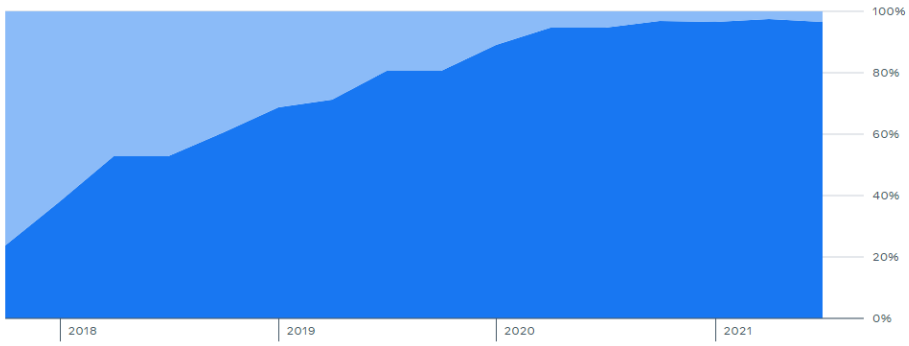
Link to Polices/Practices of Social Media Platforms

We highlight two interrelated aspects to this issue. (i) The increasing removal rate of allegedly hateful content, the change of platform policies on misinformation/disinformation post-COVID-19 (and the impact the two have had on media diversity); (ii) The encroaching role of states vis-à-vis the functioning of social media platforms.

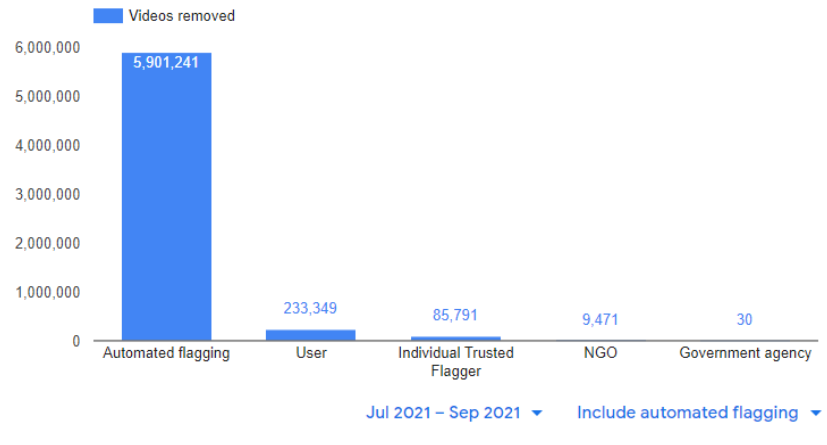
Issue 1: There is a Private social media companies, which are note not bound by International Human Rights Law, have become the ultimate arbiters of harm, truth and the practical limits of the fundamental rights to freedom of expression. For example, in relation to Facebook, the first graph (as extracted from Facebook's Community Standards Enforcement Reports) demonstrates the massive rise in removal of the broadly conceptualized notion of hate speech between 2018 and 2021. The second graph demonstrates a respective rise in relation to proactive removal due to advances in the use of Artificial Intelligence (AI).

PROACTIVE RATE

Of the violating content we actioned for hate speech, how much did we find before people reported it?



The increased use of AI for content removal can also be seen with [YouTube:]



Technology handling content such as hate speech is still at its 'infancy'. The algorithms developed to achieve this automation are habitually customized for content type, such as pictures, videos, audio and text. The results of enhanced moderation of contentious areas of speech such as 'hate speech' and the use of AI have contributed to a deterioration of media diversity due to the impact such technologies may have on reporting on contentious issues. For example, YouTube removed 6,000 videos documenting the Syrian conflict. It shut down Qasioun News Agency, an independent media group reporting on war crimes in Syria. Several videos were flagged as inappropriate by an automatic system designed to identify extremist content. Other hash matching technologies, such as PhotoDNA, also seem to operate in 'context blindness' which could be the reason for the removal of those videos. In sum, as also noted by the OSCE Representative on Freedom of the Media, the use of AI could seriously jeopardize our human rights, in particular the freedom of expression and media pluralism.

We have also witnessed the removal of fake news/misinformation/disinformation. Platforms such as Facebook and Instagram generally downrank such content. However, following the onset of the pandemic, there was an increasing trend towards removal. For example, Facebook has introduced a detailed section on COVID-19 misinformation in 2020 which provides that, amongst others, false vaccine claims are to be removed. Instagram followed suit, noting that some COVID-19 related content that could cause harm could be removed.

Issue 2: We are witnessing a global trend of states cracking down on internet intermediaries and social media platforms by imposing obligations to quickly remove content broadly defined as, for example, hate speech or disinformation. In 2019 and 2020 respectively, Justitia demonstrated how the NetzDG has been replicated in over 20 countries around the world, most of which rank as 'not free' or 'partly free' by Freedom House. The countries discussed in the reports require online platforms to remove vague categories of content that include 'false information' (Kyrgyzstan, Nigeria and Morocco), 'blasphemy'/'religious insult' (Indonesia, Austria, Turkey), 'hate speech' (Austria, Cambodia), incitement to generate anarchy (Cambodia) and 'personal and privacy rights' (Turkey). The chilling effect of such legislation is even higher as compared to Germany, because unlike Germany, many of these states do not have the same robust protection of the rule of law, and often lack institutional safeguards (such as independent courts) to enforce constitutional protections of freedom of expression.

Many of the restrictions on contentious speech under the aforementioned legislation are difficult to reconcile with international human rights standards. For example, consider the manner in which such laws have been abused in countries is well depicted by the case of journalist Yavesew Shimelis, a prominent government critic. In March 2020, he posted on Facebook that, in anticipation of COVID-19's impact, the government had ordered the preparation of 200,000 burial places. His Facebook profile was suspended and the police detained him. In April, he was charged under Ethiopia's new 'Hate Speech and Disinformation Prevention and Suppression Proclamation No.1185/2020.' His trial commenced 15th May 2020. Since his release, Shimelis is no longer as outspoken.

Issue 2: Legislative/Administrative/Policy or other Measures to promote media independence and Pluralism and their Impact

Measures

We refer to measures taken at a regional level in Europe:

The Council of Europe has developed a Platform to report on serious threats to the safety of journalists and media freedom in Europe in order to reinforce the Council of Europe's response to the threats and member states' accountability.

The European Union has started consultations on the proposed European Media Freedom Act. The European Media Freedom Act despite being long overdue, could be a promising development. It is necessary to underline that the European Union, through its forthcoming Digital Services Act (DSA) might, in fact, shrinking civic space and potentially hampering media diversity. As noted by the European Federation of Journalists:

'The service providers would be expected under the DSA to search and delete *any* type of potentially illegal content under EU and national law. Given the plurality of and divergences among national laws regulating freedom of expression, it is expected that companies play safe and ban a wider range of content than what would be strictly necessary and proportionate. This undemocratic system of corporate censorship needs to be prevented in the future legislation.'

The DSA is currently at the EU Parliament, and we would urge that the institution should take into account the freedom of information and expression and place greater emphasis on its protection. At the moment, the Act imposes strict obligations on platforms at the risk of monitoring and fines to remove 'illegal content' without adequate assessment of the free speech implications/shrinking civic space impact. The current Act will lead to the furtherance of the 'regulatory race to the bottom', over-blocking, and removal of legitimate and unharmful content.

Recommendation of Changes/Additional Measures

Our key recommendation in relation to the DSA, and the process of content regulation on is that content moderation should be within the framework of International Human Rights Law. This has extensively been discussed in our 'Framework of First Reference' report where we recommend that to ensure a sustainable future for media pluralism, considering that platforms have become the central vehicle for news sharing, adequate safeguards to online free speech need to be secured especially in platforms' content moderation processes. Our recommendations need to be read in light of contemporary developments such as the deterioration of media pluralism and the increasing censorship/silencing/persecution of journalists as referred to above. To this end, we propose:

- Content moderation of contentious areas of speech, namely hate speech and disinformation occur within the framework of International Human Rights Law, with removal of such content being legitimate, necessary and proportional and in line with Article 19 of the International Covenant on Civil and Political Rights.

- The Rabat Plan of Action test should be implemented in order to assess whether content should be removed.
- Only disinformation entailing real and immediate harm should be subject to removal. For other categories of disinformation, platforms may resort to lesser restrictive forms of moderation such as downranking content, flagging content, providing users access to reliable/official sources of medical information, among others.
- AI based content filtering algorithms should be designed and deployed with 'humans-in-the-loop' as these algorithms are susceptible to bias and may be unable to appreciate the context of a post or the nuance of language.
- A voluntary pledge by platforms where they commit to adopting content moderation practices that are compliant with international human rights law standards that ensure greater transparency and consistency.
- The creation of a Free Speech Framework Agreement to be administered by the Office of the UN High Commissioner for Human Rights (OHCHR) under the auspices of the Special Rapporteur on Freedom of Opinion and Expression. to ensure compliance with the voluntary pledge.

We remain at your disposal for any clarifications/further information you may require.

Yours Sincerely,
Jacob Mchangama
Executive Director, Justitia