Forum on Information & Democracy Working Group on Infodemics

GLOBAL PARTNERS DIGITAL

Global Partners Digital response

September 2020

About Global Partners Digital

Global Partners Digital is a social purpose company dedicated to fostering a digital environment underpinned by human rights.

Overview of Contribution

We would like to take this opportunity to share our contribution to the Forum on Information & Democracy Working Group on Infodemics. Global Partners Digital (GPD) recognises the pressing need to define a policy framework (set of recommendations) in order to combat infodemics and bring about systemic change while respecting human rights online, particularly the rights to freedom of expression and privacy.

In this contribution, we provide a framework that should underscore any rights-respecting approach to disinformation, misinformation and information chaos. We then unpack GPD's approach to disinformation laws and policies, examining both existing and proposed regulatory efforts to combat disinformation and misinformation. We further examine efforts undertaken by platforms. Through this analysis, we establish a clear need for policymakers and platforms to adopt a new model of dealing with disinformation and infodemics. Accordingly, we propose specific recommendations which address the four structural challenges identified in the call for contributions. We hope these recommendations are useful in the development of the working group's final output.

Framework for our Contribution

(i) International Human Rights Law & Standards

Our contribution and overall approach to disinformation, misinformation and information chaos is based on international human rights law, primarily the International Covenant of Civil and Political Rights (ICCPR). The most relevant human rights impacted by frameworks developed in response to these issues are the rights to freedom of expression and privacy. Article 19 of the ICCPR guarantees the right to freedom of expression, including the right to receive and impart information and ideas of all kinds regardless of frontiers. Article 17 of the ICCPR guarantees the right to privacy and provides that "no one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence". The rights to freedom of expression and privacy are also protected in other relevant treaties, such as Articles 13 and 16 of the Convention on the Rights of the Child.

Restrictions on the rights to freedom of expression and privacy are only permissible under international human rights law when they can be justified. In order to be justified, a restriction must meet a three-part test, namely that: (i) it is provided by law; (ii) it pursues a legitimate aim; and (iii) it is necessary and proportionate, which requires that the restriction be the least restrictive means required to achieve the purported aim.

It is important to remember that a state's obligation to ensure that these rights are not unjustifiably restricted exists both in relation to restrictions which stem from the actions of the state itself as well as those caused by third parties, such as private companies. As such, it makes no difference from the perspective of the individual affected whether any restrictions are imposed and enforced directly by the state (e.g. through creating criminal offences which are enforced by the police and the courts) or through third parties, particularly when the third party is acting in order to comply with legal obligations.

With respect to the actions of private companies specifically, the United Nations Guiding Principles on Business and Human Rights (UNGPs) makes clear that a state's international human rights obligations include establishing a legal and policy framework which enables and supports businesses to respect human rights. Principle 3 notes that this general obligation includes ensuring "that (...) laws and policies governing the creation and ongoing operation of business enterprises, such as corporate law, do not constrain but enable business respect for human rights".

Given the impact that online platforms have upon the enjoyment and exercise of the rights to freedom of expression and privacy, governments have a clear obligation to ensure that these rights are respected by these platforms. This includes ensuring that legislation and other measures do not constrain online platforms' ability to respect the right to freedom of expression or privacy themselves, nor should they directly or indirectly constitute a restriction on the enjoyment and exercise of those rights by those that use such platforms.

Given the limited existing interpretation and case-law of these frameworks as they apply to existing initiatives and regulatory efforts, we also make reference, as appropriate, to

Recommendation CM/Rec(2018)2 of the Council of Europe's Committee of Ministers to member States on the roles and responsibilities of internet intermediaries (Recommendation CM/Rec(2018)2),[1] and relevant commentary from the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (the UN Special Rapporteur). In addition, we acknowledge the 2017 Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda. This declaration was a joint initiative by the Special Rapporteurs on freedom of expression of the UN, OSCE, OAS and AU, to provide guidance on disinformation and regulatory responses designed to tackle the phenomenon. These guidelines and commentaries provide detail on the obligations of states with respect to the protection and promotion of human rights in the digital environment, with a particular focus on any legal frameworks that apply to internet intermediaries.

(ii) Unpacking Disinformation, Misinformation and Infodemics from a Human Rights Perspective

In recent years, states have increasingly responded to the spread of disinformation and misinformation online through content based restrictions and regulation, which in some cases has imposed stricter moderation by platforms. At GPD, we have examined a myriad of disinformation laws and policies from a human rights perspective, which is informed by the framework referenced above. Our analysis involves making an assessment of the following six elements:

1. The precise scope of the law should be clear.
2. Speech or content should only be restricted where it is in pursuance of a legitimate aim. i.e. if it causes a particular harm to an individual's human rights, or a society's legitimate interest (such as the protection of democracy, national security or public health).
3. Any restrictions in the law should account for instances where the individual reasonably believed the information to be true.
4. Determinations of whether speech or content is disinformation should be made by an independent and impartial judicial authority.
5. Any responses or sanctions should be proportionate.
6. Intermediaries should only be liable for third party content where (a) an intermediary specifically intervenes in that content; or (b) an intermediary refuses to obey an order adopted in accordance with due process guarantees by an independent, impartial, authoritative oversight body (such as a court) to remove it, and they have the technical capacity to do so.

Unfortunately, nearly all laws and policies that we have examined fail to satisfy these elements, and they present a number of risks to individuals' right to freedom of expression. The outbreak of COVID-19 has only contributed to more problematic laws. We demonstrate this through the following case studies.

[1] Council of Europe, Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries, 7 March 2018.

United Kingdom: The UK government's Online Harms White Paper was released in 2019 and proposed placing legal obligations on online platforms to remove or restrict particular forms of illegal and harmful content, including disinformation. It defined disinformation as "information which is created or disseminated with the deliberate intent to mislead… to cause harm, or for personal, political or financial gain". But it is unclear how platforms would be able to determine intent - especially considering the scale of content platforms will have to review, which may lead to the over removal of legitimate material inadvertently flagged as disinformation.

Singapore: Article 7 of the Protection from Online Falsehoods and Manipulation Act 2019 law prohibits an individual from communicating a false statement of fact when such statement is likely to be prejudicial to state security, public health, public safety, public tranquility, public finances, amongst other reasons. Parts 3 and 4 of this law further enable any Minister to issue directions to control and prevent the communication of false statements of fact where he or she "is of the opinion that it is in the public interest" to do so. On the basis of this determination, a Minister can make further orders to block access to or restrict accounts. This is worrying because determinations are decided by partisan government officials, particularly those within the executive branch, as opposed to judicial authorities, which are less likely to be politically biased and simply charged with interpreting the law.

South Africa: South Africa passed legislation in 2020 to tackle the spread of COVID-19 related disinformation and misinformation. It criminalises the publication of any statement made "with the intention to deceive any other person" about COVID-19, the infection status of any person, or any measure taken by the government to address COVID-19. While the offence does not explicitly require the statement to be false, the "intention to deceive" suggests that the publication would need to be false or misleading. However, it would be difficult to determine what is false or misleading. The exact scope of what is considered to be "about COVID-19" or "any measure taken by the Government" is also unclear.

(iii) From Self Regulation and Content Regulation to Meta Regulation

Clearly, existing and proposed legislative responses to disinformation are flawed and can, themselves, pose serious risks to human rights, particularly the right to freedom of expression. We have not, in our capacity, come across a disinformation related law which closely aligns with relevant international human rights standards. Still, the need to address these issues continues to exist for policymakers around the world. It is spurred, in part, because digital platforms continue to fail to respond to disinformation and misinformation.

While some platforms were originally praised for their swift self-regulatory responses to COVID-19, the reality is that challenges which existed for platforms before the pandemic are now even more pronounced. Not only has there been an increased need for content moderation and enforcement, but the ability to provide human oversight of such processes has been drastically reduced, resulting in a dramatic increase in automated decision making. The scale of erroneous removals of content is likely to increase as platforms continue to resort to automated processes for content moderation. It has been widely observed that AI is at a very nascent stage

when it comes to analysing speech, and can only accurately identify a very small number of categories of speech which don't require an assessment of context or other nuances.[2] These issues are compounded by a lack of sufficient transparency provided by platforms as they relate to decisions surrounding content moderation and removal.

This failure on the part of platforms is substantiated through a number of studies. For example, Reuters Institute for the Study of Journalism recently found that 59% of posts on Twitter which were rated as false by fact-checkers remained up without warnings, whereas on Facebook 24% of false-rated content in a particular sample remained up without warning labels.[3] This is in spite of the unprecedented action taken by platforms against false and misleading content during COVID-19.

In light of these circumstances, entire new frameworks and more nuanced measures are necessary to now tackle the infodemic and the dissemination of false or misleading information online. While many have advocated for digital literacy and fact-checking initiatives to solve the current infodemic, we need more concrete proposals. A shift from content moderation to meta regulation needs to occur to truly mitigate the harms of another infodemic. In the section below we provide and expand on these potential recommendations.

Substantive Recommendations

We present here a list of proposed recommendations which address the four structural challenges identified in the call for contributions. Please note that many of these recommendations touch upon multiple structural challenges.

Recommendation 1: States should avoid content based restrictions on disinformation, particularly through criminal laws, which should only be used in the most severe circumstances where there is an intention to cause some clear, objective public harm.

This recommendation speaks to the reality of most content based restrictions on disinformation containing ambiguous definitions of what constitutes "fake news" or disinformation. These vague and highly subjective terms—such as "unfounded", "biased", "false", and "fake"—do not adequately describe the content that is prohibited. As a result, they provide the authorities with broad remit to censor the expression of unpopular, controversial or minority opinions, as well as criticism of the government and politicians. Such ambiguity may also incentivise self-censorship due to fears of prosecution or other penalties. These content based restrictions are thus inappropriate tools for tackling disinformation and infodemics.

[2] See, for example, Center for Democracy & Technology, "Mixed Messages? The Limits of Automated Social Media Content Analysis", 28 November 2017, available at: https://cdt.org/insight/mixedmessages-the-limits-ofautomatedsocial-media-content-analysis.
[3] S. Brennen et al., "Types, Sources, and Claims of COVID-19 Misinformation", available at: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation

Recommendation 2: Platforms should not be expected to make determinations on the legality of content under national law. It should be up to platforms to decide what terms of service and content moderation policies they apply to content that is legal (even if harmful), including disinformation, however it may be appropriate to require those online platforms which do develop such terms of service and content moderation policies to ensure that those terms of service are clearly understood, and enforced fairly and consistently. Assessments of whether this is the case should involve an independent entity.

Whether terms of service include disinformation or misinformation should be left to the online platforms to determine. However, if they do introduce terms of service or community standards relating to these types of content (which most large platforms have to date) meta regulation could require platforms to ensure that terms are clear to users, and that they have mechanisms in place to ensure that those terms of service are enforced as fairly and consistently as possible. This should not simply be left to platforms, but also evaluated and assessed by an independent entity which would look at all systems in place dealing with content-based harm. This would provide increased transparency around the entire system - including the setting of terms of service, enforcement system design, moderator training, and efficacy of take-downs or other forms of moderation.

Recommendation 3: States should require certain platforms to submit transparency reports or relevant information on their advertising, targeting practices, and algorithmic decision making, particularly as they relate to political advertising and public health crises. However, the scope of platforms to be included must be proportionate and devised as narrowly as possible. Platforms should only be included where there is clear evidence of harm being caused or facilitated by their services.

While we strongly recommend that certain platforms or companies be required to submit transparency reports, we caution against responsibilities being applied without there being any evidence of harm being caused or facilitated by their services. Their inclusion in scope would be neither proportionate nor appropriate without such evidence. We are particularly concerned about responsibilities being applied to smaller platforms where such requirements may be onerous financial burdens. Moreover, appropriate transparency reporting requirements should not incentivise companies to act in a way that creates risks to individuals' human rights. Even the possibility of heavy or disproportionate sanctions may skew incentives and exacerbate risks to freedom of expression.

Recommendation 4: States should consider measures that facilitate appropriate data sharing by platforms to designated third parties. We recommend the data trust model as one solution, with researchers given the ability to independently assess and report findings.

Further data sharing is needed to make empirically based judgements. Without adequate information from platforms, policymakers and researchers will be unable to properly assess compliance with international human rights standards, or be able to craft effective solutions or appropriate limitations on systems designed for automated amplification and audience targeting.

Recommendation 5: States need to develop and effectively enforce data protection legislation which tackles the issues of micro-targeting and surveillance of users. Policymakers should examine existing frameworks, such as the GDPR, as useful models for potential legislation.

While platforms vow to fight COVID-19 related misinformation, reports indicate that Facebook continues to enable advertisers to target users based on undetermined data points and potentially dangerous characteristics, including "interest in pseudoscience".[4] The "interest in pseudoscience" group was reported to include 78 million people, but nonetheless represents a form of micro-targeting, as opposed to regular targeting, as it enables the group to receive a tailored message based on one or several specific characteristics.

Transparency is needed to address this issue (discussed further below), but because micro-targeting or recommender systems entail the use of personal data points or surveillance, privacy and data protection rules are particularly relevant. Strict rules on the use of certain categories of personal data, explicit consent for such usage, and effective enforcement for violations may disrupt the virality of disinformation.

Policymakers should refer to the General Data Protection Regulation (GDPR) when developing such legislation. The GDPR requires that individuals' personal data be treated lawfully and transparently by specific entities. It also protects, with some limited exceptions, the processing of various types of special categories of personal data, including individuals political opinions, racial or ethnic origin, religious or philosophical beliefs, etc.[5] The independent Data Protection Authorities (DPAs), which oversee compliance with GDPR, have struggled to enforce or level sanctions against companies for lack of compliance with these provisions. However, properly enforced legislation would serve as a strong deterrent to companies and could significantly limit micro-targeting.

Recommendation 6: States should consider legislation that requires platforms to allow users to understand what types of controls or algorithms are in place, including the data points used to make recommendations. This should be provided in an accessible format to inform users.

Recommendation 7: States should consider legislation that requires platforms to let users have a say in whether algorithmic recommender systems are applied, and have the option to turn them off, or instead only have them base recommendations on specific data points agreed to by the user.

[4] A. Sankin, "Want to Find a Misinformed Public? Facebook's Already Done It." *The Markup* (May 7, 2020), available at: https://www.newamerica.org/oti/reports/getting-to-the-source-of-infodemics-its-the-business-model/targeted-advertising-and-covid-19-misinformation-a-toxic-combination

[5] General Data Protection Regulation (GDPR), Art. 9.

Some states have already begun to consider this type of legislation, including Germany's proposed law (Medienstaatsvertrag).[6] If approved, this law would impose binding obligations on platforms to disclose the selection criteria used to determine the sorting and presentation of content. Social media platforms would also need to disclose the way that criteria are weighed, the functioning of the algorithm, and provide users with an opportunity to both understand and modify based on their individual preferences. We do not suggest that this particular proposal be used as a model for regulation, but make reference to it here as a means of highlighting the ability for states to act on our recommendations.

Recommendation 8: Platforms should adopt measures that limit the virality of false or misleading content shared on messaging apps, and research further options to quell the spread of such content without undermining privacy or freedom of expression.

Recommendation 9: States should consider measures that encourage companies to allow their users to report disinformation, even on private or encrypted channels. They should also encourage companies to conduct further research on limiting the virality of disinformation on their services in a rights respecting manner.

Platforms should adopt measures that add friction to the ease in which disinformation and misinformation may spread on their services. WhatsApp's limitation on the number of members of groups, or flagging of forwarded messages, while not silver bullets, are helpful in stemming the flow of disinformation.[7] However, there are many options available that may slow the spread of fast-growing and unchecked content, including by providing users with warnings about the unverified nature of specific pieces of content, labeling accounts that have a history of sharing false or misleading information, or elevating authoritative information. Platforms should adopt and implement such measures and research further options that respect individuals rights to privacy and freedom of expression. States should also encourage companies to allow their users to report disinformation, and encourage companies to conduct research on limiting the virality of disinformation on their services in a rights respecting manner.

[6] Medienstaatsvertrag, Rundfunkkomission Der Lander, available at: https://www.rlp.de/fileadmin/rlp-stk/pdf-Dateien/Medienpolitik/04_MStV_Online_2018_Fristverlaengerung.pdf

[7] M. Singh, "Whatsapp's New Limit Cuts Virality of Highly Forward messages by 70%" TechCrunch, (April 27, 2020), available at: https://social.techcrunch.com/2020/04/27whatsapps-new-limit-cuts-virality-of-highly-forwarded-messages-by-70/.