



## **Disinformation and freedom of expression**

**Submission in response to the call by the UN Special Rapporteur  
on the promotion and protection of the right to freedom of  
opinion and expression for contributions to the annual thematic  
report to be presented to the Human Rights Council at its 47th  
session in June 2021**

Association for Progressive Communications (APC)

*February 2021*

# TABLE OF CONTENTS

1. Framing the problem and its impact	2
2. The purposes of disinformation campaigns	5
Gendered disinformation	7
Minorities, migrants and populations in situations of vulnerability	10
Rights defenders, activists, and their causes	11
3. For-profit disinformation: Amplification and targeted advertising	13
4. State measures taken to address disinformation and their impact	14
The dangers of criminalisation	14
Access and connectivity	18
5. Measures taken by tech companies to address disinformation and their impact	19
The challenges of content moderation: AI in content moderation	21
Micro-targeting and ad transparency	23
Transparency reports	25
6. Conclusions and recommendations	27

# Disinformation and freedom of expression

The Association for Progressive Communications (APC) is an international organisation and a network of organisations dedicated to empowering and supporting people working for peace, human rights, development and protection of the environment, through the strategic use of information and communication technologies (ICTs). APC has 57 organisational members and 35 individual members active in 73 countries, mostly in the global South. We work to build a world in which all people have easy, equal and affordable access to the creative potential of ICTs to improve their lives and create more democratic and egalitarian societies.

APC welcomes the call of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression to reflect on the impacts of disinformation on the issues pertaining to her mandate and appreciates the opportunity to provide a few comments below.

## 1. Framing the problem and its impact

The spreading of so-called “fake news” has reached new dimensions – in terms of reach, speed and volume – with the expansion in the use of ICTs and in particular with the growing access to the internet. Disinformation, however, is not a new phenomenon; propaganda and conspiracy theories have marked public debate for centuries.

To define disinformation as a modern phenomenon catapulted by social media usage is key to assessing its current impact and understanding its current framing. Filter bubbles, echo chambers, micro-targeted advertising, algorithms governance and transparency are all elements particular to the 21st century disinformation context. Another very particular characteristic of disinformation in the digital age is the diversification of actors who produce/disseminate it. In the previous era, disinformation was a prerogative mostly of the ones who owned or controlled the media. The “oligopoly” of disinformation is broken in the digital era and the implications are significant.

The discussion of the root causes and possible solutions to the disinformation problem, however, cannot be addressed without a broader look into our informational ecosystems – online and offline. This is particularly important because “[d]isinformation is a multifaceted and complex issue that can rather be understood as a symptom of a much broader information disorder.”<sup>1</sup>

As pointed out by research carried out by Golebiewski and boyd<sup>2</sup> and Shane and Noel,<sup>3</sup> disinformation is a problem that needs to be assessed from a supply/demand analysis. From the supply side, it is important to determine that many malicious actors benefit from disinformation for various and sometimes interrelated purposes, as will be addressed in the following section. From the demand side, it is crucial to recognise that much disinformation runs on the exploitation of data voids or data deficits.<sup>4</sup>

When we say that solutions to address disinformation should look at our broader informational ecosystems, we refer to the fact that the unavailability of trustworthy information on issues of public interest is one of the factors behind the growing consumption of disinformation. The correlations between disinformation and access to public information, as well as access to investigative and independent journalism and

---

<sup>1</sup> <https://www.disinfo.eu/publications/the-few-faces-of-disinformation>

<sup>2</sup> <https://datasociety.net/library/data-voids>

<sup>3</sup> <https://firstdraftnews.org/long-form-article/data-deficits>

<sup>4</sup> See footnotes 2 and 3 for definitions of data voids and data deficits.

media diversity and plurality, are particularly relevant to the issues that pertain to this Special Rapporteur's mandate.

Work carried out by civil society in different regions adds to a growing body of evidence that seems to point out that when no robust public information regimes are in place and independent media (including community and public outlets) are unable to provide diverse, quality and independent coverage, disinformation trends find especially fertile ground to spread and destabilise. Overall respect for people's freedom of assembly and association and freedom of expression, including that of activists, community leaders and human rights defenders, is also an important piece of the information ecosystem puzzle.

Understanding why people consume disinformation, and why disinformation seems to circulate more broadly and easily than true facts, helps us identify the mechanisms created to exploit it and can tell us a lot about how to prevent its harmful effects. Any measures looking forward should consider the empowerment and increased agency of users, not only as critical audiences, but as rights holders.

Disinformation is a global problem that will not be addressed properly if a fragmented approach is adopted. It is also a multistakeholder challenge that requires dialogue between different sectors – dialogue that needs to be built on transparency and participation in decision making.

APC considers it particularly concerning that such a complex and multifaceted problem has been addressed within the context of COVID-19 without the necessary procedural safeguards to ensure the needed transparency and participation. Technology companies have reacted to the growing circulation of health-related disinformation with a series of new policies, and governments across the world have proposed dedicated legislation. Often the necessary information and data needed to ground such actions is not disclosed, and limited or no multistakeholder consultation takes place, mainly under the justification of a state of emergency, sidelining human rights concerns. These concerns will be detailed in the following sections.

### Summary of key points:

- A holistic approach to understanding disinformation requires an analysis of our broader information ecosystems.
- Only such a holistic approach will allow us to identify solutions and preventive actions that build on the strengthening of the other spaces and actors that promote the flow of information, visions and ideas within our societies.
- Robust access to public information systems, a plural, accessible and diverse media context, availability of independent and qualified journalism, and safety to express ideas are all necessary elements of a healthy information system.
- Solutions to disinformation, including regulatory ones, should take into consideration the widespread character of sources of disinformation and the fact that the whole information ecosystem is permeated by it.
- Digital and media literacy programmes are key and should be part of states' efforts to counter a number of information disorders; such programmes could be carried out independently, but also embedded into the regular educational system curricula.
- Increased access to data and information held by tech companies is essential to allow us to better understand the phenomenon of disinformation.
- Any solutions sought, in particular policy and regulatory measures, should be built on truly participatory processes.

## 2. The purposes of disinformation campaigns

Information disorders take different forms, but APC is particularly concerned with disinformation. UNESCO has defined disinformation as “information that is false and deliberately created to harm a person, social group, organisation or country.”<sup>5</sup> Ireton and Posetti clarify that disinformation is particularly dangerous “because it is frequently organised, well resourced, and reinforced by automated technology.”<sup>6</sup>

Perhaps the most commonly debated disinformation campaigns are those carried out with a view to promote political and electoral gains. Much has been said about the impact of

---

<sup>5</sup> See, for example: <https://en.unesco.org/fightfakenews>

<sup>6</sup> <https://unesdoc.unesco.org/ark:/48223/pf0000265552>

disinformation on democracy, as evidenced by the 2016<sup>7</sup> and 2020<sup>8</sup> elections in the US,<sup>9</sup> the 2018 elections in Brazil,<sup>10</sup> and Brexit.<sup>11</sup>

Disinformation campaigns, however, have to be understood beyond electoral narratives. They are promoted by a variety of malicious agents with the intention to cause harm – be it to influence electoral processes, to promote authoritarian and populist agendas, to foment polarisation and promote discrimination and hatred against marginalised communities, to undermine human rights defenders and human rights processes, or to discredit science.

Disinformation causes confusion and has a chilling effect on freedom of expression and information. It directly impacts on the level of trust in the public sphere as a space for democratic deliberation. People no longer feel safe to express their ideas for fear of online harassment and of being targeted by disinformation campaigns; others feel paralysed and silenced by the puzzlement and uncertainty created by the surrounding information pollution and remove themselves from public debate concerning key issues of public interest.

The impact of disinformation is particularly pernicious against groups in situations of vulnerability or marginalisation.

Hamm has studied disinformation campaigns through three vectors: manipulative actors, deceptive behaviours and harmful content. Based on this analysis, she created the following categories related to the intentions behind disinformation campaigns:

---

<sup>7</sup> <https://knightfoundation.org/articles/seven-ways-misinformation-spread-during-the-2016-election>

<sup>8</sup> <https://bhr.stern.nyu.edu/blogs/2019/9/3/disinformation-and-the-2020-election-how-the-social-media-industry-should-prepare>

<sup>9</sup> <https://www.brookings.edu/blog/techtank/2021/01/11/the-role-of-misinformation-in-trumps-insurrection>

<sup>10</sup> <https://www.wilsoncenter.org/blog-post/brazil-fell-for-fake-news-what-to-do-about-it-now>

<sup>11</sup> <https://www.bbc.com/news/blogs-trending-48356351>

- Foreign influence – the use of disinformation by foreign actors to disrupt societies over certain issues and/or to push an agenda (foreign actor/s to a domestic audience).
- Political – the use of disinformation to undermine adversaries and/or push an agenda (domestic actor/s to a domestic audience).
- Lucrative – the use of disinformation to make a profit.
- Issue-based – the use of disinformation to serve an ideological, normative and/or financial goal.<sup>12</sup>

The author stresses that, in practice, such categories often overlap and reinforce each other; but this framework is especially interesting to highlight the importance of looking at disinformation beyond electoral contexts and, in particular, to allow a more attentive look into issue-based disinformation.

APC has observed that longstanding issue-based campaigns are particularly strong in relation to gendered disinformation, hatred against minorities and vulnerable groups, and human rights and environmental activists. These issue-based campaigns take different formats and narratives to adapt to current newsworthy stories and events.

## Gendered disinformation

EU DisinfoLab researcher Maria Sessa conducted a study in which she identified that during the pandemic, “misogynistic narratives have been retrieved and adapted to fit within the mis- and disinformation landscape around COVID-19 – an event which has had a disproportionately negative impact on women’s rights.” She systematised false information that relies on false or misleading images of women in the pandemic context and concluded that the narratives tend to produce “either a negative representation of women as enemies, in order to fuel the public debate; or a pitiful depiction of women as

---

<sup>12</sup> <https://www.disinfo.eu/publications/the-few-faces-of-disinformation>



victims in order to push an alternative agenda.”<sup>13</sup> Examples of such narratives include women framed as responsible for the spread of the virus as a result of 8 March International Women’s Day demonstrations<sup>14</sup> and women accused of taking advantage of the pandemic to push a secret gender equality agenda.<sup>15</sup>

The weaponisation of disinformation against women in public spaces has long been pointed out as a form of gender-based violence online. A recent study by the Wilson Center’s Science and Technology Innovation Program argues, however, that gendered and sexualised disinformation is a phenomenon distinct from broad-based gendered abuse and should be defined as such to allow social media platforms to develop effective responses. The authors of the study have defined it as “a subset of online gendered abuse that uses false or misleading gender and sex-based narratives against women, often with some degree of coordination, aimed at deterring women from participating in the public sphere. It combines three defining characteristics of online disinformation: falsity, malign intent, and coordination.”<sup>16</sup>

Research has demonstrated how female politicians, for example, are attacked more often than male candidates through disinformation campaigns.<sup>17</sup> Scholars and feminists point out that these attacks have the deliberate goal of preventing women from taking part in the democratic process. As early as 2017, Nina Jankowicz alerted that “[f]emale politicians and other high profile women worldwide are facing a deluge of what you could call sexualised disinformation. It mixes 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.”<sup>18</sup>

---

<sup>13</sup> <https://www.disinfo.eu/publications/misogyny-and-misinformation%3A-an-analysis-of-gendered-disinformation-tactics-during-the-covid-19-pandemic>

<sup>14</sup> <https://maldita.es/malditobulo/20200407/experto-britanico-8m-coronavirus-imperial-college-espejo-publico-antena-3>

<sup>15</sup> <https://www.nextquotidiano.it/gender-nel-piano-colao-educazione-gender>

<sup>16</sup>

[https://www.wilsoncenter.org/sites/default/files/media/uploads/documents/WWICS%20Malign%20Creativity%202021\\_0.pdf](https://www.wilsoncenter.org/sites/default/files/media/uploads/documents/WWICS%20Malign%20Creativity%202021_0.pdf)

<sup>17</sup> See, for example: <https://www.cfr.org/blog/gendered-disinformation-fake-news-and-women-politics> and <https://ssrn.com/abstract=3444200>

<sup>18</sup> <https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women>

Those who speak out on feminist issues are also particularly targeted by disinformation campaigns,<sup>19</sup> as well as the issues they convey.<sup>20</sup>

In 2019, Koki Muli Grignon, a Kenyan UN diplomat working as the facilitator during the Commission on the Status of Women in New York, a prominent annual women's rights conference at the United Nations, received thousands of emails trying to interfere with her positions and work. One of them demanded that Grignon stand against abortion and same-sex families, criticised her conduct as a facilitator, and said she was being watched.<sup>21</sup>

In 2020, openDemocracy reported on a global network of “crisis pregnancy centres”, backed by US anti-abortion groups linked to the Trump administration, that targeted vulnerable women with “disinformation, emotional manipulation and outright deceit.”<sup>22</sup> Anti-choice groups have used smartphone surveillance to target “abortion-minded women”, sending propaganda directly to a woman's phone while she is in a clinic waiting room.<sup>23</sup>

Amnesty International carried out research on the effects of such disinformation campaigns in India, concluding that “[o]nline abuse is silencing – it leads women to self-censoring, limiting what they post, anonymizing their accounts, or leaving Twitter altogether.”<sup>24</sup>

---

<sup>19</sup> See, for example: <https://www.citizenngo.org/en-af/73124-stop-ipas-killing-kenyan-babies> and <https://www.citizenngo.org/en-af/fm/171807-ban-disney-for-their-indoctrination-agenda>

<sup>20</sup> See, for example: <https://www.hrw.org/news/2018/12/10/breaking-buzzword-fighting-gender-ideology-myth> and <https://www.shethepeople.tv/news/technology-law-peddling-fake-news>

<sup>21</sup> <https://www.passblue.com/2019/05/14/the-case-of-harassing-a-un-diplomat-via-1000s-of-text-messages>

<sup>22</sup> <https://www.opendemocracy.net/en/5050/trump-linked-religious-extremists-global-disinformation-pregnant-women>

<sup>23</sup> <https://rewirenewsgroup.com/article/2016/05/25/anti-choice-groups-deploy-smartphone-surveillance-target-abortion-minded-women-clinic-visits>

<sup>24</sup> <https://decoders.amnesty.org/projects/troll-patrol-india>

The situation is even more striking when reviewed from an intersectionality lens. Female political leaders and activists from racial, ethnic, religious or other minority groups are targeted far more than their white colleagues.<sup>25</sup>

This also brings to light the close links currently seen between disinformation and hatred against particular groups. Disinformation campaigns have been promoted by populist leaders, political parties and candidates to fuel nationalism and push fringe ideas and values into mainstream conversations, sometimes leading to violence against minority communities. We have seen reports of how leaders such as Putin<sup>26</sup> and Bolsonaro<sup>27</sup> have used disinformation campaigns against the LGBTIQ+ community in their electoral campaigns.

### Minorities, migrants and populations in situations of vulnerability

In India, researcher Laxmi Murthy reports on the weaponisation of disinformation to fuel hatred against Muslims. She recounts how terms like “corona jihad” have been adopted during the pandemic, building on the lines of “love jihad”, which is “the supposed campaign of Muslim men feigning love and luring Hindu women, forcing them to abandon their faith and convert to Islam.”<sup>28</sup> According to this new disinformation campaign, Muslims “deliberately infected an innocent (Hindu) populace with the virus, with malicious forethought, indeed as a form of bio-terrorism.”<sup>29</sup>

In September 2020, APC addressed an open letter to Facebook<sup>30</sup> in which it reported “a steep rise in Islamophobic and other hate speech content against the backdrop of the

---

<sup>25</sup> <https://www.amnesty.org.uk/online-violence-women-mps>

<sup>26</sup> <https://www.dw.com/en/opinion-vladimir-putin-uses-homophobia-for-votes-in-russia/a-53778667>

<sup>27</sup> <https://www.theguardian.com/world/2020/jul/08/bolsonaro-masks-slur-brazil-coronavirus>

<sup>28</sup> [https://www.apc.org/sites/default/files/APC\\_Hate\\_Speech\\_V10\\_0.pdf](https://www.apc.org/sites/default/files/APC_Hate_Speech_V10_0.pdf)

<sup>29</sup> <https://www.thehindu.com/news/national/karnataka/tablighi-event-shobha-smells-corona-jihad/article31259288.ece>

<sup>30</sup> <https://www.apc.org/en/pubs/open-letter-apc-facebooks-complicity-enabling-hate-speech-india-must-end>

COVID-19 pandemic, where members of India's ruling party took to social media to propagate misinformation and stigma against minorities."<sup>31</sup>

In Europe, disinformation campaigns against immigrants have been reported for years, with most pieces seeking to create polarisation and fear, and set the political agenda.<sup>32</sup> Later studies point out that disinformation stories "tend to adapt and change alongside the public's main concerns."<sup>33</sup> The COVID-19 pandemic is just the most recent example. Similar stories about the spreading of the coronavirus by migrant populations have been seen in many countries in regions across the globe, such as the ones spotted in Colombian social media against Venezuelans.<sup>34</sup>

Ethnic and racial tensions have been exploited not only by national partisan forces, but also by foreign influence. Reports on Russian interference in the 2016 US elections, for example, documented Internet Research Agency (IRA) activities that specifically aimed at manipulation of Black voters.<sup>35</sup>

## Rights defenders, activists, and their causes

Another group targeted by disinformation campaigns are rights defenders, especially those representing marginalised and discriminated communities. APC member Arab Center for Social Media Development – 7amleh, for example, has reported that Israel has coordinated and cooperated with non-state actors to spread campaigns of hate and misinformation about Palestinian human rights defenders and organisations online. According to 7amleh, "The campaign was premised on racist caricatures, attempting to

---

<sup>31</sup> <https://www.aljazeera.com/news/2020/4/29/do-not-buy-from-muslims-bjp-leader-in-india-calls-for-boycott>

<sup>32</sup> <https://pathforeurope.eu/the-perpetuation-of-fear-and-disinformation-around-migration-by-the-media-in-the-eu/>

<sup>33</sup> <https://epc.eu/en/Publications/Fear-and-lying-in-the-EU-Fighting-disinformation-on-migration-with-al-39a1e8>

<sup>34</sup> <https://colombiacheck.com/index.php/chequeos/migracion-venezolana-y-salud-peligrosa-desinformacion>

<sup>35</sup> <https://www.yonder-ai.com/resources/the-disinformation-report>

paint Palestinian civil society organizations as essentially suspicious and violent, in order to discredit and defund them.”<sup>36</sup>

Delegitimation campaigns based on false or distorted information against human rights defenders in Latin America are very common against “social leaders” and other community activists and NGO representatives.<sup>37</sup>

Climate change activists have also been subject to such campaigns. Matto Mildenerger, from the University of California, has affirmed that current climate misinformation has evolved and one tactic climate sceptics use is targeting climate activists for their personal carbon footprint, in an effort “to discredit [those] who are trying to make reforms by accusing them of hypocrisy.” Researcher Jason Cook has also highlighted how new climate change disinformation campaigns have tried to promote a narrative of “the unreliability of climate experts.”<sup>38</sup>

Recent reports have also revealed the manner in which the oil sector has been using social media to promote a campaign against government measures aimed at addressing the impact of fossil fuels – the leading cause of climate change. A 2021 HEATED article points out the heavy investment in Twitter ads by the oil industry after anti-fossil fuel measures taken by the Biden administration. The article questions Twitter’s interpretation of how these ads fall outside of their policy for political ads and questions the impact on the climate change debate of paid ads in social media, something that climate groups cannot afford.<sup>39</sup> The issue of social media (political) ads will be addressed below.

Summary of key points:

- Special attention should be given by platforms and governments to long-term issue-based disinformation campaigns, especially those targeted against

<sup>36</sup> [https://7amleh.org/storage/Hashtag\\_Palestine\\_2019\\_English\\_20.4.pdf](https://7amleh.org/storage/Hashtag_Palestine_2019_English_20.4.pdf)

<sup>37</sup> <https://www.amnesty.org/download/Documents/ACT3060112017SPANISH.PDF>

<sup>38</sup> <https://www.igi-global.com/chapter/understanding-and-counteracting-misinformation-about-climate-change/230759>

<sup>39</sup> <https://heated.world/p/twitters-big-oil-ad-loop-hole>

specific groups and themes, including human rights, women's rights and environmental issues.

- Gendered disinformation should be considered as a different phenomenon, separate from gender-based online violence, which requires specific monitoring and solutions.

### 3. For-profit disinformation: Amplification and targeted advertising

One growing and worrisome trend observed in the past years is the growing “professionalisation” of the disinformation sector. Disinformation campaigns are today planned and executed by specialised agencies and brokers. A recent BuzzFeed investigation uncovered how this has become a highly lucrative business: “Around the globe, politicians, parties, governments, and other clients hire what are known in the industry as ‘black PR’ firms to spread lies and manipulate online discourse.”<sup>40</sup> The article points out the strength of the “marketing and politics” sector in places such as the Philippines, where freedom of expression has been under constant threat; there, many legitimate-seeming agencies offer black PR services that include fake social media accounts, websites, and coordinated harassment campaigns. Much of this fake and misleading content targets journalists, human rights defenders and opposition voices, such as Maria Ressa<sup>41</sup> and Senator Leila de Lima.<sup>42</sup>

In 2019, there were already reports of the easy access to Russian trolls, who could be hired for “as little as a few thousand dollars.”<sup>43</sup> Today, however, the situation has reached new levels. The Oxford Internet Institute’s *2020 Global Inventory of Organized Social Media Manipulation*<sup>44</sup> mapped recent trends in computational propaganda across 81 countries and found evidence of the deployment of computational propaganda on behalf

---

<sup>40</sup> <https://www.buzzfeednews.com/article/craigsilverman/disinformation-for-hire-black-pr-firms>

<sup>41</sup> <https://www.wired.com/story/maria-ressa-disinformation-more-insidious-than-we-think>

<sup>42</sup> <https://www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war>

<sup>43</sup> <https://www.nbcnews.com/tech/security/trolls-hire-russia-s-freelance-disinformation-firms-offer-propaganda-professional-n1060781>

<sup>44</sup> <https://comprop.oii.ox.ac.uk/research/posts/industrialized-disinformation>

of political actors in 48 of them. According to the study, almost USD 10 million has been spent on political advertisements by “cyber troops” operating around the world.

The Global Disinformation Index has been mapping ad-funded disinformation and estimates that advertisers have provided around USD 25 million to nearly 500 English-language coronavirus disinformation sites in 2020.<sup>45</sup>

Summary of key points:

- Measures to curtail disinformation should specifically address the different “markets” established to profit from it.

## 4. State measures taken to address disinformation and their impact

### The dangers of criminalisation

Countries worldwide are using legitimate concerns about online disinformation to deepen their control over the internet and people. These policy and legislative initiatives share some similarities: they give discretionary powers to executive bodies to decide whether a piece of content is false or misleading, and give these bodies the power to issue fines, corrections or even hand out prison sentences for creating, publishing or disseminating pieces of content. In these cases, creators, disseminators and publishers of disinformation are the main targets of these regulatory initiatives. These criminalisation efforts often do not distinguish between lawful and unlawful expression, limiting the exercise of freedom of expression and allowing governments to exercise greater control and discretion.<sup>46</sup>

For example, in 2018, Malaysia adopted the Anti-Fake News Law, which provides prison

---

<sup>45</sup> [https://disinformationindex.org/wp-content/uploads/2020/07/GDI\\_Ad-funded-COVID-19-Disinformation-1.pdf](https://disinformationindex.org/wp-content/uploads/2020/07/GDI_Ad-funded-COVID-19-Disinformation-1.pdf)

<sup>46</sup> <https://www.apc.org/sites/default/files/OnlineContentToRegulateOrNotToRegulate.pdf>

sentences for up to 10 years for knowingly creating, distributing or publishing “fake news”, defined to include “news, information, data and reports” that are “wholly or partly false.” In Egypt, Article 19 of the 2018 Media and Press Law grants the Supreme Media Council the authority to “suspend any personal website, blog, or social media account that has 5,000 followers or more if it posts fake news, promotes violence, or spreads hateful views.” Under this framework, bloggers can be subjected to prosecution for publishing false news or incitement to break the law. In 2019, Singapore adopted its Protection from Online Falsehoods and Manipulation Act, which empowers any Singaporean government minister to issue a range of corrective directions, including fines and prison sentences, against online “falsehoods” deemed to be against the public interest. The act states further that false statements cannot be transmitted to users in Singapore through the internet, or through systems “that enable the transmission through a mobile network” of text and multimedia messages. This potentially gives the government power to police encrypted chat apps like Signal and WhatsApp. Sri Lanka contemplated the adoption of amendments to the penal and criminal procedure codes to criminalise the dissemination of “false news” where it is deemed to affect “communal harmony” or “state security”.<sup>47</sup>

In Africa, disinformation laws exist in Kenya, Uganda and Tanzania, among other countries. Several of these disinformation laws fail to meet the threshold of necessity and proportionality, as required by the three-part test for a justifiable restriction of rights under international law.

With the pandemic, some African countries included disinformation provisions in their COVID-19 prevention, containment and treatment policies.<sup>48</sup> As the African Declaration on Internet Rights and Freedoms Coalition expressed, governments in the region have failed to demonstrate the necessity of limiting the right to freedom of expression through the criminalisation of false news, or that this is linked to addressing specific public interest needs.<sup>49</sup>

---

<sup>47</sup> Ibid.

<sup>48</sup> [https://www.apc.org/sites/default/files/AfDecPositionPaperCOVID19\\_EN.pdf](https://www.apc.org/sites/default/files/AfDecPositionPaperCOVID19_EN.pdf)

<sup>49</sup> Ibid.



In South Africa, for example, the government has criminalised the dissemination of false information through regulations published in terms of the Disaster Management Act 57 of 2002, which seeks to address the publication of any statement published through any medium, including social media, with the intention to deceive any other person in respect of information relating to COVID-19. The offence is punishable with an unspecified fine, imprisonment of up to six months, or both. The regulations were followed by directives from the Minister of Communications and Digital Technologies compelling communications service providers to “remove COVID-19 related fake news from their platforms immediately after it is identified as such.” Within days of its passing, several individuals were arrested for spreading false information about COVID-19. In one case relating to a COVID-19 interview, the Broadcasting Complaints Commission of South Africa fined two broadcasters ZAR 10,000 (USD 660). While various activists initially raised their voices in support of governments’ efforts to halt the spread of the disease, they also cautioned against overly restrictive conditions that limit human rights, including freedom of expression, access to information and public accountability.<sup>50</sup> In Zimbabwe, new provisions in section 31 of the Criminal Law Codification and Reform Act criminalise the publishing or communication of false statements that are prejudicial to the state. In the context of rapidly accelerating state responses to COVID-19 disinformation, the Collaboration on International ICT Policy for East and Southern Africa (CIPESA), PROTEGE QV, and other groups launched an interactive map to track and analyse disinformation laws and policies in Sub-Saharan Africa. This tool assesses whether laws, policies and other state responses are human rights-respecting.<sup>51</sup>

In March 2020, the Palestinian president declared a state of emergency in all of the Palestinian territories to confront the coronavirus and, on 22 March 2020, the State of Emergency Law was approved by presidential decree. This law stipulates, among other provisions, penalties for any person who violates the decisions, instructions and

---

<sup>50</sup> <https://www.apc.org/en/blog/regulating-freedom-association-amidst-covid-19-response-south-africa>

<sup>51</sup> <https://cipesa.org/2020/06/coalition-of-civil-society-groups-launches-tool-to-track-responses-to-disinformation-in-sub-saharan-africa>

measures applied by the official authorities to achieve the goals of the state of emergency. These penalties include one year of imprisonment and fines, without prejudice to any other penalties stipulated in other Palestinian laws. New presidential decrees for the state of emergency were announced in April and June. As APC member 7amleh pointed out, the emergency legislation provided more space for restrictions on freedom of expression and privacy, since it includes broad terms with no legal provisions, safeguards or measurable standards. Article 3 of the State of Emergency Law, for example, prohibits posts, statements or news on social media related to the state of emergency that are not based on an official source. The law establishes that anyone who violates this shall be punished with detention and fines. 7amleh expressed concerns regarding the fact that, under these measures, people were being deprived of their right to provide oversight and criticise the performance of official authorities in dealing with the pandemic.<sup>52</sup> 7amleh also raised concerns about the ability of Palestinian citizens of Israel to access safe, accurate information as well as the Israeli Ministry of Health's response to the coronavirus. According to 7amleh research, material in Hebrew was posted three times as often as in Arabic, and whereas posts in Hebrew rely on doctors and epidemiologists, Arabic posts to a large degree rely on religious leaders.<sup>53</sup>

Also in the midst of the COVID-19 health crisis, some troublesome initiatives proliferated in Latin America. In Paraguay, a bill that establishes fines for anyone who knowingly or wilfully disseminates, by any means, false information that generates social panic linked to the epidemiological alert or declaration of a health emergency was presented in the Congress.<sup>54</sup> In Brazil, a draft bill on disinformation was also introduced during the pandemic crisis. Civil society groups voiced alarm over the lack of transparency in the discussion process, scarce multistakeholder participation and poorly drafted provisions that may amount to serious harms to freedom of expression and privacy.<sup>55</sup> As APC

---

<sup>52</sup> <https://7amleh.org/2020/05/07/digital-rights-in-palestine-between-emergency-and-pandemic>

<sup>53</sup> <https://7amleh.org/2020/06/28/7amleh-publishes-monitoring-of-discrepancies-between-the-israeli-ministry-of-health-s-coronavirus-campaigns-in-arabic-and-hebrew>

<sup>54</sup> <https://www.tedic.org/en/worrisome-regulation-on-disinformation-in-times-of-covid19>

<sup>55</sup> <https://www.apc.org/en/pubs/brazilian-disinformation-bill-threatens-freedom-expression-and-privacy-online>

member Intervozes – Brazil Social Communication Collective expressed, draft bills in Brazil that deal with disinformation are aimed at criminalising users for sharing false content and this, according to the organisation, is “a simplistic approach, which ignores the complexity of the problem, the debates and consensus obtained during the processing and approval of the Civil Rights Framework for the Internet, and the recommendations of international bodies.”<sup>56</sup>

In Bolivia, the government issued a decree “to deal with COVID-19” which, among other measures, stated that those who “misinform or generate uncertainty for the population will be prosecuted.” Peru also decided to establish penalties of up to six years in prison for those who spread false news. Civil society warned that in Venezuela, the accusation of “causing anxiety” was used to put people in jail in the context of the pandemic.<sup>57</sup>

Civil society groups worldwide claimed that many of these initiatives are deliberately defined vaguely to allow maximum discretion for the governments to target critics and that many regulatory proposals to counter online harms run the risk of unintentionally creating more harm than the initial harm they are meant to tackle.<sup>58</sup> APC has argued that, given the human rights issues at stake, state-driven interventions to regulate content online should be subject to particular precautions. Any legitimate intervention must aim for a minimum level of intervention in accordance with the principles of necessity and proportionality, be based on an inclusive consultation process with all relevant stakeholders, and not strengthen the dominant position of the large incumbents.

## Access and connectivity

In many parts of the world, access to the internet is limited to access to social media and messaging platforms and applications, provided at low or no cost to users due to agreements between these platforms and mobile telecommunication companies (zero

---

<sup>56</sup> <https://app.rios.org.br/index.php/apps/onlyoffice/s/8cwkrStRPbtTfNg>

<sup>57</sup> <https://www.derechosdigitales.org/14405/desinformacion-y-salud-publica-en-tiempos-de-pandemia>

<sup>58</sup> <https://www.apc.org/sites/default/files/OnlineContentToRegulateOrNotToRegulate.pdf>

rating and similar schemes). These types of connectivity do not provide meaningful access to the internet – that is to say, such users do not have access to a truly free, open, interoperable, reliable and secure internet. They will not have access to diverse and plural sources of information and will be particularly exposed and vulnerable to filter bubbles and echo chambers. Addressing the disproportional influence of disinformation today also passes through efforts to provide people affordable and appropriate internet access.

Summary of key points:

- General prohibitions on the dissemination of information based on vague and ambiguous concepts and definitions should be avoided.
- Disinformation should be clearly defined and differentiated from other information disorders, especially when regulations are adopted to address it. In this regard, states should apply the three-part test of legality, necessity and proportionality to any measures taken, considering that attempts to curtail information disorders may significantly impact on the exercise of the rights to expression and opinion. An assessment of “intentionality” is particularly relevant in this regard.
- As highlighted by the previous UN Special Rapporteur on freedom of expression, “the penalization of disinformation is disproportionate, failing to achieve its goal of tamping down information while instead deterring individuals from sharing what could be valuable information.”<sup>59</sup>
- Governments should promote digital inclusion, including universal and affordable access to the internet for all.

## 5. Measures taken by tech companies to address disinformation and their impact

With the continuous spread of disinformation – including with new peaks during the COVID-19 pandemic of health-related disinformation and conspiracy theories – internet platforms have been facing increasing pressure from governments and other actors to take down content that is harmful. Platforms have reacted with different initiatives aimed at countering so-called “fake news”; as highlighted by Meyer and Hanot, “Recent

---

<sup>59</sup> <https://undocs.org/A/HRC/44/49>

measures have included promotion of authoritative sources, alongside an increase in automation of content moderation.”<sup>60</sup>

PEN America tracked policy, operational and functional updates made by major social media companies in response and relation to the pandemic during the first six months of 2020 and provided a searchable database of such measures. The investigation points out that “Facebook and Twitter have each focused heavily and publicly on stopping the spread of conspiracy theories, coordinated disinformation campaigns, and hate speech, while at the same time maintaining a steady clip of new feature releases and public good campaigns. Others, including LinkedIn, Snapchat, and Reddit, have generally focused less on responding directly to COVID-19 and instead provided updates related to the operation of their own platforms and the thematic interests of their users. Pinterest has stood out by attempting to comprehensively redirect users searching for COVID-related terms to authoritative content – keeping the platform focused on well-being and positivity, including through what they’ve termed ‘compassionate search’ features.”<sup>61</sup>

Although these are signs of a more reactive and responsive industry, it is important to remember that many of these changes are also “public relations” reactions aimed at reputation control and management, and their real effect is yet to be observed and studied in the long run. More meaningful and impactful changes targeting the business model of these companies, in particular their exploitation of personal data and the obscure use of algorithms, remain to be seen.

Recent events have also called our attention to the fact that the major social media platforms have particular responsibilities given their enormous numbers of users and the fact that in many parts of the world they constitute virtual oligopolies. However, stricter community rules in these platforms may lead to a massive migration to smaller ones, which have no proper rules in place, or even those that particularly appeal to more radicalised groups.

---

<sup>60</sup> <https://www.disinfo.eu/publications/how-platforms-are-responding-to-the-disinfodemic>

<sup>61</sup> <https://pen.org/the-first-wave-social-media-platforms-responding-to-covid-19>

Another important note is that, in many countries, the initial coordination and circulation of disinformation campaigns take place first through messaging services – especially WhatsApp<sup>62</sup> – to only later become viral in social media. Given that these services make use of end-to-end encryption and are not subject to content moderation scrutiny,<sup>63</sup> particular measures should be designed to address their role in the amplification of disinformation.<sup>64</sup>

### The challenges of content moderation: AI in content moderation

Research from APC member Intervezes shows that digital platforms lack policy and structured processes on the issue of disinformation and that they have been developing specific and reactive actions to combat the phenomenon.<sup>65</sup>

Intervezes research found that companies do not present a unified concept of disinformation and this opens room for unilateral and unclear decisions. Moreover, says Intervezes, “None of the companies has reported having a specific structure to address the issue of disinformation, which may hinder the coordination of initiatives within each organization.” Intervezes’ work also highlights the platforms’ business models that rely on the spread of this type of content, and the lack of transparency regarding how companies deal with “disinformative content”. It points out the obstacles in effective communication with users on this subject: since information is scarce and scattered, there is also no evaluation of the effectiveness of what has been implemented or a proper process – a mechanism that could enable users to be notified and be able to defend themselves in processes of content moderation or in possible sanctions applied by the platforms on their posts or accounts.

---

<sup>62</sup> <https://www.reuters.com/article/us-brazil-election-whatsapp-interpretor/facebooks-whatsapp-flooded-with-fake-news-in-brazil-election-idUSKCN1MU0UP>

<sup>63</sup> <https://www.buzzfeednews.com/article/pranavdixit/whatsapp-misinfo-privacy-policy>

<sup>64</sup> See, for example: <https://www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news>

<sup>65</sup> <https://intervezes.org.br/publicacoes/10-ways-to-combat-disinformation>

As APC has stated, processes developed by intermediaries should be transparent and include provisions for appeals, users should be informed of repeated posts carrying disinformation that they share, and if there is a systemic pattern, companies should take action.<sup>66</sup>

In the 2019 edition of the annual Global Information Society Watch (GISWatch) report, Luis Fernando García Muñoz affirms that increasing pressure for stricter content moderation with the aim of curbing disinformation and incitement to violence has produced surging investment in the development of automated content moderation technologies. The pandemic also provoked an increased reliance on artificial intelligence (AI) tools by platforms, removing misinformation and apparently inaccurate information about COVID-19 at an unprecedented rate. The increased use of AI has been fostered by the growth in remote work modalities adopted for content moderation workers in view of the sanitary restrictions imposed during 2020.<sup>67</sup> EU DisinfoLab provides an illustration of this volume: “Over the period from April to June 2020, a total of 11,401,696 videos were removed from YouTube. Of these, only 552,062 (or 4.84%) were reported by humans.”<sup>68</sup>

An increased reliance on AI-driven content moderation with the risk of false positives,<sup>69</sup> limitations in capturing nuances and contextual specificities, and without transparency, accountability and due process pose serious risks for freedom of expression online. APC together with other organisations asked companies, in the context of COVID-19, to commit to preserve all data on content removal during the pandemic, including but not limited to information about which takedowns did not receive human review, whether users tried to appeal the takedown (when that information is available), and reports that

---

<sup>66</sup> <https://www.apc.org/en/pubs/apc-input-public-consultation-santa-clara-principles-transparency-and-accountability-content>

<sup>67</sup> <https://www.theverge.com/2020/3/16/21182011/youtube-ai-moderation-coronavirus-video-removal-increase-warning>

<sup>68</sup> <https://www.disinfo.eu/publications/how-platforms-are-responding-to-the-disinfodemic>

<sup>69</sup> <https://www.theverge.com/2020/9/21/21448916/youtube-automated-moderation-ai-machine-learning-increased-errors-takedowns>

were not acted upon, and to produce transparency reports that include information about content blocking and removal related to COVID-19, among other things.<sup>70</sup>

APC has acknowledged that it is unrealistic to expect companies to perfectly differentiate between speech that should be protected and speech that should be erased. We are not opposed to the use of artificial intelligence to identify problematic content for human review. However, when such automated processes are used, it is inevitable they will “make mistakes” and therefore their use should be more transparent, all content removal should be subject to human review, and users should have easy recourse to challenging removals which they believe to be arbitrary or unfair.<sup>71</sup>

## Micro-targeting and ad transparency

Micro-targeting is a marketing strategy that uses people’s data to segment them into small groups for content targeting. While it allegedly helps deliver content that “is interesting and helpful to you”, it is particularly problematic when used in the distribution of inaccurate and biased information. How people and organisations collect and use such data is largely an unregulated activity in most markets.<sup>72</sup> And across the world, micro-targeting has been increasingly used for political objectives, in particular during electoral campaigns.<sup>73</sup>

In addition to being a threat to privacy, “micro-targeting could lead to a fragmentation of the marketplace of ideas. Fragmentation happens when the public loses track of overarching themes, and instead focuses on the single issues that are relevant to them personally, which are the topics delivered through micro-targeting techniques.”<sup>74</sup> With

---

<sup>70</sup> <https://www.apc.org/en/pubs/open-letter-covid-19-content-moderation-research>

<sup>71</sup> <https://www.ohchr.org/Documents/Issues/Opinion/ContentRegulation/APC.pdf>

<sup>72</sup> Dobber, Ó Fathaigh and Zuiderveen argue that in the EU, “[w]hile there are no specific rules on such micro-targeting, there are general rules that apply. We focus on three fields of law: data protection law, freedom of expression, and sector-specific rules for political advertising.” See:

<https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe>

<sup>73</sup> <https://www.wired.com/story/how-trump-conquered-facebook-without-russian-ads>

<sup>74</sup> <https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe>



micro-targeting, advertisers can curate the content predicted to be the most relevant to specific groups of people (also known as “filter bubbles”).

In early 2020, Privacy International stressed the risks of this practice, explaining that “[p]articularly in countries where there is history of political violence, campaigning based on intensive collection of personal data is untested ground fraught with great risk. Collecting data on ethnicity or political affiliation with no limitations or safeguards is open to abuse. Plus, many countries still lack sufficient laws and regulatory mechanisms to safeguard data protection and privacy affected by this level of data generation and processing, particularly sensitive personal data such as political views or ethnicity.”<sup>75</sup>

In 2019, APC expressed concern about Facebook’s promises to protect European users from targeted disinformation campaigns during the European Parliamentary elections<sup>76</sup> while at the same time blocking access to ad transparency tools.<sup>77</sup>

Although recent progress has been observed in relation to transparency in political ads, such as advertiser verification processes and ad repositories, social media platforms regulate it in a fragmented manner that leads to what Privacy International has referred to as a “transparency divide”. That means that the benefits of such increased transparency and safeguards have not been equally distributed among the global user base of these platforms. In February 2021, APC joined more than 60 organisations in an open letter to Facebook and Google addressing this issue.<sup>78</sup>

In May 2020, Ranking Digital Rights launched an extensive study relating micro-targeting to the spread of disinformation during the early months of the COVID-19 pandemic. According to the organisation, “Content moderation is a downstream effort by platforms to clean up the mess caused upstream by their own systems designed for automated

---

<sup>75</sup> <https://privacyinternational.org/news-analysis/3735/why-were-concerned-about-profiling-and-micro-targeting-elections>

<sup>76</sup> <https://www.apc.org/en/pubs/open-letter-facebook>

<sup>77</sup> <https://www.propublica.org/article/facebook-blocks-ad-transparency-tools>

<sup>78</sup> <https://www.apc.org/en/pubs/open-letter-facebook-and-google-equal-and-better-transparency-regarding-political-advertising>

amplification and audience targeting.”<sup>79</sup> As early as 2017, media reports had already called attention to the risks of targeted advertising, which allowed, for example, advertisers to direct their content to “Jew haters”.<sup>80</sup> In February 2021, The Markup reported that Google’s advertising system allowed employers or landlords to discriminate against non-binary and some transgender people.<sup>81</sup>

## Transparency reports

In 2010, Google released its first tech industry transparency report. Since then, 70 companies have adopted the practice.<sup>82</sup> These reports have been an important source of information, in particular about government requests for user data. However, after 10 years of practice, despite its expansion throughout the sector, we have seen little innovation in the content of the reports, and the last couple of years may have seen a stagnation of this expansion.<sup>83</sup>

APC believes these documents should include more detailed data on content moderation practices. Social media platforms are private spaces that *de facto* have, today, a public function. Also, companies still need to do more to comply with their responsibility to ensure that their policies for restricting content are being applied in a non-discriminatory and equitable manner. Platforms report very little about when they remove content or restrict users’ accounts for violating their terms of service, and often fail to provide adequate notice: users are not adequately informed about the rules they have violated.<sup>84</sup> Moreover, companies enter into agreements with states to operate locally, and the terms

---

<sup>79</sup> <https://www.newamerica.org/oti/reports/getting-to-the-source-of-infodemics-its-the-business-model/>

<sup>80</sup> <https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>

<sup>81</sup> <https://themarkup.org/google-the-giant/2021/02/11/google-has-been-allowing-advertisers-to-exclude-nonbinary-people-from-seeing-job-ads>

<sup>82</sup> <https://www.accessnow.org/transparency-reporting-index/>

<sup>83</sup> <https://www.theatlantic.com/technology/archive/2019/09/what-happened-transparency-reports/599035/>

<sup>84</sup> This initiative by the Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE) contributes to illustrating how confusing the terms of use and terms of agreements can be: <https://letrachica.digital>

of these agreements, which have implications for content regulation, are often completely unknown.<sup>85</sup>

In general, platforms should provide greater transparency and accountability regarding the following:

- The implementation of content moderation guidelines.
- The rejection of reports of online abuse and disaggregated data on reports received.
- Data on accuracy of human and automated detection.
- The departments and staff responsible for responding to content and privacy complaints.

Summary of key points:

- Content moderation practices adopted by tech companies to address disinformation should comply with due process and transparency standards.
- Although safe harbour provisions concerning the liability of intermediaries for third party content are important to ensure freedom of expression online, tech companies should be held accountable for any information they manipulate and curate.
- Terms and conditions need to be easily accessible and understandable in different languages.
- Remedies should be available to users to appeal tech companies' content moderation and other decisions.
- Measures that seek to increase the self-determination, agency and autonomy of users are also important.
- Companies should abide by the UN Guiding Principles on Business and Human Rights – companies have the responsibility to respect human rights – and human rights standards should guide content moderation efforts.

---

<sup>85</sup> [https://www.apc.org/sites/default/files/APCInputSantaClaraPrinciples\\_June2020.pdf](https://www.apc.org/sites/default/files/APCInputSantaClaraPrinciples_June2020.pdf)

## 6. Conclusions and recommendations

APC considers that the main responsibility for respecting, protecting and fulfilling human rights lies with states. They should refrain, therefore, from outsourcing to companies the moderation of rights. That said, companies have the responsibility to respect human rights, as clarified in the UN Guiding Principles on Business and Human Rights.<sup>86</sup>

APC details below some further recommendations directed to states and tech companies.

To states:

- States should implement any measures, including legislation introduced to address disinformation, in a manner that complies with international human rights law. Any restrictions on freedom of opinion and expression must be consistent with Article 19 of the International Covenant on Civil and Political Rights.
- General prohibitions on the dissemination of information based on vague and ambiguous concepts and definitions should be avoided. Criminalisation should be avoided.
- Disinformation should be clearly defined and differentiated from other information disorders, especially when regulations are adopted to address it. In this regard, states should apply the three-part test of legality, necessity and proportionality to any measures taken, considering that attempts to curtail information disorders may significantly impact on the exercise of the rights to expression and opinion. An assessment of “intentionality” is particularly relevant in this regard.
- States should also take active steps to address disinformation targeted at vulnerable groups. Particular attention should be given to the specific targeting of women and its impact, from an intersectionality perspective.
- Any intervention carried out by states must be based on an inclusive consultation process with all relevant stakeholders.

---

<sup>86</sup> [https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR\\_EN.pdf](https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf)

- States should not enter into undisclosed contracts with companies – addressing issues from copyright to blasphemy to violent extremism, among others – to voluntarily remove content from their platforms.
- States should consider setting regulatory standards on transparency requirements.
- Robust data protection legislation should be adopted and enforced to protect internet users' right to privacy and other human rights, including freedom of expression and information.
- States should also establish robust access to public information systems, as detailed by international standards, including the UN Human Rights Committee's General Comment 34 and the 2020 UN Human Rights Council resolution on freedom of opinion and expression.<sup>87</sup> Pro-access to public information measures could include frequent public service announcements, increased public-interest reporting and support for fact-checking platforms.
- States should enable and ensure a context that is conducive to media pluralism, diversity and independence.
- Concerted efforts to bring the benefits of meaningful internet access to all people are necessary. In the immediate and short term, barriers related to costs and data capacity should be addressed. In the longer term, it is critical that attention be urgently given to affordable and appropriate internet access for communities. A radical change of policy, logic and strategy is necessary to address the needs of the billions of people in developing countries who still suffer from digital exclusion.

To tech companies:

- A human rights-based approach should guide companies' content moderation processes (not just in how they respond to requests for takedowns, but throughout the entirety of their operations). This approach should be guided by, among others,

---

<sup>87</sup> <https://undocs.org/en/A/HRC/44/L.18/Rev.1>

the principles of accountability, equality and non-discrimination, participation and inclusion, transparency, empowerment, sustainability, and non-arbitrariness.

- Increased transparency should include, among others, the criteria used by algorithms to curate content feeds and search query results, as well as those applied in targeted advertising.
- Improved policies should be adopted concerning political advertising, including to allow further monitoring, screening and research.
- In relation to the use of AI in content moderation, its use should be more transparent, all content removal should be subject to human review, and users should have easy recourse to challenging removals which they believe to be arbitrary or unfair.
- Enhanced due process should be ensured to every user, including to be expeditiously notified of content takedowns, the reason for the takedown, and the option to appeal a company's takedown decision. The Manila Principles provide a framework for this.<sup>88</sup>
- Further attention should be given to building and applying alternatives to the taking down of content that is not illegal. Options that empower users to make informed choices about the content that they see should be explored.
- Respect for data protection and privacy legislation and international standards should be at the centre of the companies' business models.
- Equitable and unbiased application of community standards across communities and countries should be urgently improved.
- Human rights impact assessments should be carried out on an ongoing basis to review new products, updating of companies' policies, and expansion into new jurisdictions. Human rights impact assessments should include all human rights that companies' policies may impact, beyond freedom of expression and privacy.
- Multistakeholder processes, building on existing multistakeholder initiatives, with input from different parts of the world, could be adopted to develop global guidelines or norms to address the challenge of harmful content within a rights-

---

<sup>88</sup> <https://www.manilaprinciples.org>

respecting framework. This multistakeholder process could explore whether establishing a more traditional self-regulatory framework would have positive consequences.

---