

# Disinformation

**C H Powell**

UNIVERSITY OF CAPE TOWN

**Birgit Schippers**

UNIVERSITY OF STRATHCLYDE

**Irena Barkane**

UNIVERSITY OF LATVIA

**Oscar Puccinelli**

ROSARIO NATIONAL UNIVERSITY

**Jukka Viljanen**

TAMPERE UNIVERSITY

## Disinformation: the concept<sup>95</sup>

Scholars, governments and commentators are using a multiplicity of terms to describe the phenomenon of disinformation. This report uses the term ‘disinformation’ broadly: we are treating it as an online phenomenon encapsulating the elements of the ‘*deliberate* creation and sharing of false and/or manipulated information’,<sup>96</sup> ‘designed, presented and promoted to intentionally cause public harm or for profit’.<sup>97</sup> This conception excludes two related phenomena: these are, first, the spread of false or malicious information offline; and, second, the spread of unintentionally false or inaccurate information,<sup>98</sup> which is better described as misinformation.

However, our use of the term ‘disinformation’ will also include information which is technically accurate but is shared with malicious intent (also called malinformation, that is, ‘genuine information shared with the intention to cause harm’).<sup>99</sup> It can therefore include accurate information that instigates violence. A similar approach has been adopted by academics who frame disinformation as ‘viral deception’, which

---

<sup>95</sup> Report section written by C H Powell, Birgit Schippers, Irena Barkane, Oscar Puccinelli, Jukka Viljanen.

<sup>96</sup> Digital, Culture, Media and Sport Committee (2017–19), Disinformation and ‘fake news’: Interim Report (HC 363): Government Response to the Committee’s Fifth Report <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/1630/163002.htm>

<sup>97</sup> T Meyer and C Marsden, Regulating disinformation with artificial intelligence: effects of disinformation initiatives on freedom of expression and media pluralism, European Parliament, 2019 <https://data.europa.eu/doi/10.2861/003689>

<sup>98</sup> Ibid.

<sup>99</sup> C Wardle and H Derakhshan, Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking (Council of Europe report DGI(2017)09, 2017) 5; <https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77>.

contains three vectors: manipulative actors, deceptive behaviour and harmful content.<sup>100</sup> The ‘viral deception’ approach focuses on the online behaviour rather than the veracity of the content.<sup>101</sup>

While online behaviour is significant and worthy of study in itself, we consider the truth or falsity of the information shared to be a key factor in the harm that disinformation can cause. As we demonstrate below, an important aspect of countering disinformation is the development of mechanisms and processes to determine the veracity of the information available online.

## Threats posed by disinformation

Concerns over disinformation attracted significant attention in the wake of much publicised elections, such as the United Kingdom’s Brexit referendum in 2016, the US presidential election of the same year, and the Kenyan election of 2017.<sup>102</sup> The European Union (EU) describes online disinformation practices as ‘public harms’, specifically harms to the integrity of electoral processes, and ‘threats to our way of life,’<sup>103</sup> which undermine trust and confidence in democratic politics.<sup>104</sup> By eroding trust in elected governments, disinformation undermines public programmes that aim to ensure the common good. Current examples which illustrate this point are disinformation practices around Covid, as reported by the EU and monitoring bodies all over the globe.<sup>105</sup>

As we outline below, there are two important points which the Covid cases demonstrate. The first is the enormous difference made by the spreading of information *online* instead of *offline*. The second is that it is not only private individuals, but also governments, which contribute to the spreading of disinformation.

## Disinformation as an online phenomenon

Social media collates stories from multiple sources, changing the focus to the story rather than the source. This practice makes it difficult for people to judge the credibility of information, because ‘posts from publications as unlike as the New York Times and a conspiracy site look nearly identical’.<sup>106</sup> Furthermore, traditional gatekeepers are missing as readers choose their material based on endorsements and social

---

<sup>100</sup> Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, “Disinformation and freedom of opinion and expression”, A/HRC/47/25, April 13, 2021, <https://undocs.org/en/A/HRC/47/25> citing C François, “Actors, behaviors, content: a disinformation ABC” (Transatlantic Working Group, September 2019).

<sup>101</sup> Khan *ibid*.

<sup>102</sup> C Cadwalladr, ‘The great British Brexit robbery: how our democracy was hijacked’, *The Guardian* 7 May 2017; Meyer and Marsden (n 2); Wardle and Derakhshan (n 4) 5.

<sup>103</sup> EU Code of Practice on Disinformation, <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>.

<sup>104</sup> EU Code of Practice on Disinformation (n 10); J Bayer, I Katsirea, O Batura, B Holznagel, S Hartmann and K Lubianiec, *The fight against disinformation and the right to freedom of expression*, European Union, 2021 [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL\\_STU\(2021\)695445\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU(2021)695445_EN.pdf).

<sup>105</sup> European Union, ‘Fighting Disinformation’, available at [https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation\\_en](https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation_en); Carlos Cortés & Luisa Fernanda Isaza, “The New Normal? Disinformation and Content Control on Social Media during COVID-19”, CELE, Palermo University, available at: [https://www.palermo.edu/Archivos\\_content/2021/cele/papers/Disinformation-and-Content-Control.pdf](https://www.palermo.edu/Archivos_content/2021/cele/papers/Disinformation-and-Content-Control.pdf).

<sup>106</sup> Wardle and Derakhshan (n 4) 12.

recommendations.<sup>107</sup> Without a ready means of ascertaining the reliability of a source of information, readers rely on friends and family members to guide them through the system.

South Africa's anti-vaccination disinformation is a case in point, where accounts with fewer than 1000 followers authored two thirds of the content containing anti-vaccination hashtags. These small accounts were responsible for 26% more volume in the anti-vaccine conversation than they were in the total vaccine conversation. With just 6% of volume coming from authors with more than 10,000 followers, anti-vaccine conversation appeared to be driven by users with small followings.<sup>108</sup>

## Disinformation spread by governments

It is worth noting that state actors can be equally guilty of spreading disinformation, for example by denying the existence or spread of the disease or by suggesting cures with no proven medical efficacy.<sup>109</sup> Among various examples of disinformation related to the Covid-19 pandemic in the Ibero-American region are the cases of two Argentine national deputies who recommended the use of chlorine dioxide to combat the virus without any evidence that it was effective against the disease and despite the contraindications published by the World Health Organization. Similarly, a Brazilian ministerial body published a video stating that the use of masks was not effective to combat the virus and that it was harmful to health, despite the fact that this was recommended by the World Health Organization. The Presidents of Brazil and Guatemala, and the mayor of Santiago, Chile, also claimed that various treatments with no medical approval were effective or that public transport does not pose a risk of contagion.<sup>110</sup>

## Disinformation and violence

Disinformation can also cause or exacerbate violence, or actively prevent its resolution. An ongoing example is found in the social media (and some state media) coverage of the war in Ukraine.<sup>111</sup> Political violence in the United States (US) and South Africa further illustrates how false information (e.g., that the 2020 US presidential election was stolen) led directly to violence, which threatened democracy directly but also caused loss of life. The attempted insurrection on 6 January 2021 in Washington had been fuelled by the inaccurate claim that Donald Trump had won the US election, and could have overthrown a democratically elected government.<sup>112</sup> The violent riots that erupted in the South African provinces of KwaZulu-Natal and Gauteng in the week of 11 July 2021 resulted in over 330 deaths.<sup>113</sup> Violence included

---

<sup>107</sup> Wardle and Derakshan (n 4) 12, citing Messing, S., & Westwood, S. J. (2014). 'Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online' *Communication Research*, 41(8), 1042-1063.

<sup>108</sup> Centre for Analytics and Behavioural Change (CABC) *Vaccine Trust Spectrum Report* <https://cabac.org.za/wp-content/uploads/2021/10/Vaccine-Trust-Spectrum-Report-Media-Release.docx-4.pdf> at 6

<sup>109</sup> Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, "Disinformation and freedom of opinion and expression", A/HRC/47/25, April 13, 2021, <https://undocs.org/en/A/HRC/47/25>.

<sup>110</sup> Center for Studies on Freedom of Expression and Access to Information 'Are public official's lies unsustainable or do they have far reaching effects? A study on the obligations of the State and its officials to prevent the proliferation of disinformation' August 2021, [https://www.palermo.edu/Archivos\\_content/2021/cele/papers/Disinformation-and-public-officials.pdf](https://www.palermo.edu/Archivos_content/2021/cele/papers/Disinformation-and-public-officials.pdf).

<sup>111</sup> Mart Susi, Wolfgang Benedek, Gregor Fischer-Lessiak, Matthias C. Kettemann, Birgit Schippers, Jukka Viljanen (eds.), *Governing Information Flows During War: A Comparative Study of Content Governance and Media Policy Responses After Russia's Attack against Ukraine* (Hamburg: Verlag Hans-Bredow-Institut, 2022), GDHRNet Working Paper #4, DOI: <https://doi.org/10.21241/ssoar.78580>.

<sup>112</sup> See, generally, <https://www.aljazeera.com/program/the-listening-post/2022/6/18/the-spectacle-and-scrutiny-of-the-jan-6-hearings>.

<sup>113</sup> <https://www.aljazeera.com/news/2021/7/22/south-africa-unrest-death-toll-jumps-to-more-than-300>.

damage to vital infrastructure and led to a failure of the rule of law as citizens took up arms and policed their own neighbourhoods while social media fuelled political and racial divides.<sup>114</sup>

## Current countermeasures to disinformation

Since 2017, the United Nations (UN) Special Rapporteur for Freedom of Opinion and Expression, the Representative for Freedom of the Media of the Organization for Security and Cooperation in Europe (OSCE), the Organization of American States (OAS) Special Rapporteur for Freedom of Expression and the Special Rapporteur on Freedom of Expression and Access to Information of the African Commission on Human and Peoples' Rights (CADHP) have collaborated to release joint declarations on disinformation and freedom of expression.<sup>115</sup> In separate endeavours, the European Union produced soft and hard law instruments as well as legislative proposals to address the challenges faced by disinformation.<sup>116</sup>

Current approaches tend to impose different burdens on state officials and private parties. Where private parties are concerned, the focus is often protecting freedom of expression, which needs to be maintained particularly during elections. Government, by contrast, does not have a right to freedom of expression, but a duty first to ensure it does not disseminate false or misleading information itself, second, to respond to and remove disinformation posted by non-state actors when this disinformation can cause a certain level of harm, and, third, to ensure that a single voice does not dominate the marketplace of ideas.

This approach is supported by the jurisprudence of the Inter-American Court, which holds that officials are charged with a more onerous duty to verify the facts, due to their function and the position they occupy in a democratic society. The Inter-American Court ruled in two resounding cases against Venezuela that the exercise of freedom of expression is not the same when it comes to a merely private subject as opposed to public officials, since in a democratic society it is not only legitimate, rather, it is sometimes the duty of state authorities to rule on matters of public interest. This duty of special care is particularly accentuated in situations of greater social conflict, disturbances of public order or social or political polarization, precisely because of the set of risks that they may imply for certain people or groups at any given time.<sup>117</sup>

<sup>114</sup> <https://www.bloomberg.com/news/articles/2021-07-20/south-african-economy-set-to-take-3-4-billion-hit-from-riots>

<sup>115</sup> <https://www.osce.org/fom/302796>; <https://www.osce.org/files/f/documents/1/e/379351.pdf>; [https://www.oas.org/basic\\_documents/declarations](https://www.oas.org/basic_documents/declarations); [https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclarationDigitalAge\\_30April2020\\_EN.pdf](https://www.ohchr.org/sites/default/files/Documents/Issues/Opinion/JointDeclarationDigitalAge_30April2020_EN.pdf); <https://www.ohchr.org/en/special-procedures/sr-freedom-of-opinion-and-expression/resources>; <https://www.ohchr.org/sites/default/files/2022-05/Gender-Joint-Declaration-Freedex.pdf>.

<sup>116</sup> See e.g., European Parliament Resolution of 3 May 2018 on media pluralism and media freedom in the European Union (2017/2209(INI)); Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European democracy action plan COM(2020) 790; the European Declaration on Digital Rights and Principles for the Digital Decade COM(2022) 28; the Charter of Fundamental Rights of the European Union (2012/C 326/02); Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016; the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Shaping Europe's digital future' COM(2020) 67; the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions '2030 Digital Compass: the European way for the Digital Decade' COM(2021) 118; Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC; Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act); Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts; Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising.

<sup>117</sup> I/A Court HR. Case of Ríos et al. V. Venezuela. Preliminary Objections, Merits, Reparations and Costs. Judgment of January 28, 2009. Series C n° 194, <[https://www.corteidh.or.cr/docs/canesc/articulos/seriec\\_194\\_esp.pdf](https://www.corteidh.or.cr/docs/canesc/articulos/seriec_194_esp.pdf)> and Case of Perozo et al. V. Venezuela. Preliminary Objections,

In our survey of legislative and policy proposals, we deal first with the normative framework and then the practical measures suggested by these bodies and initiatives.

## The Normative System

### Non-state actors: Rights

Various documents seek to protect the freedom of expression of non-state actors, in particular by discouraging states from using vague or indeterminate criteria when they restrict freedom of expression. The 'Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda' (the Joint Declaration) suggests that states impose restrictions on the right to freedom of expression only if they be provided for by law, serve one of the legitimate interests recognized under international law, and be necessary and proportionate to protect that interest.<sup>118</sup> The Joint Declaration allows restrictions to prevent advocacy of hatred on protected grounds that constitutes incitement to violence, discrimination or hostility (with reference to Article 20(2), International Covenant on Civil and Political Rights). It also protects intermediaries distributing third party content and requires an independent oversight mechanism to rule that third party content should be removed before the intermediaries can be held liable for it. Under the Joint Declaration, state-mandated blocking of entire websites, IP addresses, ports or network protocols are recognised as an extreme measure which can only be justified where provided for by law and where necessary to protect a human right or other legitimate public interest, including in the sense of that it is proportionate, there are no less intrusive alternative measures which would protect the interest and it respects minimum due process guarantees. Finally, governments may not impose content filtering systems which are not controlled by the end users. The 'Joint Declaration on Freedom of Expression and Gender Justice' supports the prohibition of hate speech, applying it specifically to discrimination and sexual and gender-based violence.

The importance of freedom of expression is further supported by the Santa Clara Principles on Transparency and Accountability in Content Moderation, and, in the EU, by the European Parliament's Resolution on Media Pluralism and Media Freedom and the European Charter of Fundamental Rights of the European Union (the Charter).<sup>119</sup> The latter document, the EU's key human rights instrument, provides for the right to privacy (Article 7), protection of personal data (Article 8), and the protection of freedom of expression and information (Article 11). The EU Digital Services Act (DSA)<sup>120</sup> updates the liability rules for intermediaries and introduces added due diligence obligations for very large online platforms. Its balanced approach to the liability of intermediaries seeks to establish effective measures for tackling illegal content and societal risks online. The DSA also aims to set a benchmark for a regulatory approach to online intermediaries worldwide if they offer their services in the EU's single market. In return, online

---

Merits, Reparations and Costs. Judgment of January 28, 2009. Series C No. 195, <[https://www.corteidh.or.cr/docs/cases/articulos/seriec\\_195\\_esp.pdf](https://www.corteidh.or.cr/docs/cases/articulos/seriec_195_esp.pdf)>.

<sup>118</sup> A position reinforced by The "Joint Declaration on Media Independence and diversity in the digital age", the "Joint Declaration on Politicians and Public Officials and Freedom of Expression".

<sup>119</sup> Charter of Fundamental Rights of the European Union (2012/C 326/02).

<sup>120</sup> Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC.

intermediaries will benefit from the legal clarity of the liability exemptions and from a single set of rules within the EU.

### **Non-state actors: Obligations and Responsibilities**

The 'Joint Declaration on Media Independence and diversity in the digital age' states that media outlets and online platforms should enhance their professionalism and social responsibility, e.g., by adopting codes of conduct and fact-checking systems and putting in place self-regulatory systems or participating in existing systems, to enforce them. The 'Joint Declaration on Freedom of Expression and Elections in the Digital Age' states that digital media and online intermediaries should make a reasonable effort to address dis-, mis- and mal-information and election related spam, including through independent fact-checking and other measures, such as advertisement archives, appropriate content moderation and public alerts.

In Latin America, electoral laws and the regulation of political parties, has been established and extend to candidates for public office.<sup>121</sup> These are intended to prevent 'dirty campaigns', including the use of libel and slander, intrusion into a candidate's private life, or inventing 'information'.<sup>122</sup>

The EU's Digital Services Act (DSA) imposes obligations on digital service providers, such as social media or marketplaces, to tackle the spread of illegal content, online disinformation and other societal risks. These requirements are meant to be proportionate to the size and risks that platforms pose to society. Some of these obligations include measures to counter illegal content online and to react quickly, while respecting fundamental rights, including the freedom of expression and data protection. The DSA also compels online platforms to establish a transparency and accountability framework, for example by providing clear information on content moderation or the use of algorithms for recommending content (so-called recommender systems). Further, it bans targeted advertising on online platforms through profiling children or the use of special categories of personal data such as ethnicity, political views or sexual orientation. It also prohibits misleading practices aimed at manipulating users' choices and gives users with the choice to not receive recommendations based on profiling. The DSA has the potential to significantly improve the mechanisms for removing illegal content and disinformation and effectively protecting users' fundamental rights, but there are many implementation challenges ahead.

### **State actors: general obligations**

The 'Joint Declaration on Freedom of Expression and Fake News, Disinformation and Propaganda' holds that State actors should not make, sponsor, encourage or further disseminate statements which they know or reasonably should know to be false (disinformation) or which demonstrate a reckless disregard for verifiable information (propaganda), and should take care to ensure that they disseminate reliable and

---

<sup>121</sup> Mexico: article 247.2 of the General Law of Electoral Institutions and Procedures; Argentina: article 140 of the National Electoral Code; Honduras: arts. 146 and 148 of the Electoral and Political Organizations Law, in addition to an extraordinary agreement of 2018 of the National Electoral Chamber by which a "Register of social media accounts and official websites of candidates, political groups and highest authorities is created"; Brazil: Article 9 of Resolution 23.610 / 2019 of the Superior Electoral Court and article 323 of the Electoral Code; Peru: Article 42 of Law No. 28094 - Law of Political Organizations.

<sup>122</sup> There were complaints about manoeuvres of this type in the presidential elections in Mexico (2000 and 2006), Colombia (2014) and the Dominican Republic (2015) and in the referendum to modify the Bolivian Constitution regarding presidential re-election (2016), where extensive and unsubstantiated references were made to illegal campaign finance, corruption in the concession of public works or the private life of the candidates.

trustworthy information, including about matters of public interest, such as the economy, public health, security and the environment.

The ‘Joint Declaration on Politicians and Public Officials and Freedom of Expression’ focuses on the duties of public officials not to disseminate disinformation. Amongst other measures, it encourages states to provide for disciplinary measures to be imposed on public officials who, when acting or perceived to be acting in an official capacity, make, sponsor, encourage or further disseminate statements which they know or should reasonably know to be false and to ensure that public authorities make every effort to disseminate accurate and reliable information, including about their activities and matters of public interest. It further encourages the prohibition of hate speech, that is ‘any advocacy of hatred that constitutes incitement to discrimination, hostility or violence’.

The policy documents also emphasise that governments are under an obligation to counteract the spread of disinformation in ways that fall short of direct prohibitions of speech. Thus the ‘Twentieth Anniversary Joint Declaration: Challenges to Freedom of Expression in the Next Decade’ considers private control as itself a threat to freedom of expression, calling for measures that address the ways in which the advertising-dependent business models of some digital technology companies create an environment which can also be used for viral dissemination of, inter alia, deception, disinformation and hateful expression. It also urges human rights-sensitive solutions to the challenges caused by disinformation, including the growing possibility of ‘deep fakes’, in publicly accountable and targeted ways, using approaches that meet the international law standards of legality, legitimacy of objective, and necessity and proportionality.

### **State actors: elections**

The ‘Joint Declaration on Freedom of Expression and Elections in the Digital Age’ encourages states to ensure that any restrictions on freedom of expression that apply during election periods comply with the international law three-part test requirements of legality, legitimacy of aim and necessity. This entails that there be no prior censorship of the media, administrative blocking of media websites or internet shutdowns. Limits on the right to disseminate electoral statements should conform to international standards, including that public figures should be required to tolerate a higher degree of criticism and scrutiny than ordinary citizens. The media should also be exempted from liability during election periods for disseminating statements made directly by parties or candidates unless the statements have specifically been held to be unlawful by an independent and impartial court or regulatory body, or the statements constitute incitement to violence and the media outlet had a genuine opportunity to prevent their dissemination.

In the EU, the Draft Regulation on the transparency and targeting of political advertising aims to address obstacles to the cross-border provision of online political advertising services in the internal market as well as problems for democratic processes in the context of the internal market.<sup>123</sup>

## **Practical measures**

It is evident from the general principles set out above that disinformation cannot simply be suppressed or criminalised. This is both to protect human rights and because mere censorship would not be an effective

---

<sup>123</sup> Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising

counter to the problem. As we have seen, states can also be guilty of disinformation. When states seek to promote the truth, they may not be trusted because the disinformation already disseminated has turned part of the population against them.

In practical terms, disinformation cannot be countered without an ongoing process of persuasion and ongoing interaction between states and non-state actors. Furthermore, non-state actors themselves need to identify and challenge misinformation. In this regard, information pluralism and the prevention of monopolies enables non-state actors to contest claims made online.<sup>124</sup> Further, because disinformation spreads extremely quickly, it is imperative to address disinformation online before it becomes viral. Without reliable reference points for the validity of online information, there can be no informed citizenry able to make decisions required in a healthy democracy.<sup>125</sup>

Practical steps against disinformation include:

1. Monitoring and fact-checking: carried out by internet communications companies, academia, media, civil society, and independent fact-checking organizations.
2. Investigative responses, which establish the accuracy of online content and provide insights into disinformation campaigns, including its origins, key actors, degree of spread, and affected communities.
3. Curatorial responses, primarily editorial and content policy and ‘community standards’.
4. Technical and algorithmic responses, implemented by the social media platforms, video-sharing and search engines themselves, but also through third party tools (e.g. browser plug-ins) or experimental methods from academic research, using algorithms and/or Artificial Intelligence (AI) to detect and limit the spread of disinformation, or provide context or additional information on individual items and posts.
5. De-monetization responses, designed to stop profit and disincentivise the creation of clickbait, counterfeit news sites, and other kinds of for-profit disinformation.

Additional measures, which focus specifically on the targets of the disinformation, include:

6. Ethical and normative responses carried out at international, regional and local levels involving public condemnation of acts of disinformation or recommendations and resolutions aimed at thwarting these acts and sensitizing the public to the issues.
7. Educational responses which promote citizens’ media and information literacy, critical thinking and verification in the context of online information consumption, as well as journalist training.

---

<sup>124</sup> For a thorough explanation of the importance of information pluralism, see *NIT S.R.L. v. The Republic of Moldova*, *NIT S.R.L. v. The Republic of Moldova*, 5.4.2022.

<sup>125</sup> Maria D. Molina, S. Shyam Sundar, Thai Le and Dongwon Lee. “Fake News” Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content, *American Behavioral Scientist*, 2021, Vol. 65(2) 180–212, October 14, 2019.



8. Empowerment and credibility labelling efforts around building content verification tools and web content indicators, in order to empower citizens and journalists to avoid falling prey to online disinformation.<sup>126</sup>

## Concluding suggestions

Any attempt to counter disinformation has to comply with the principles which we have set out in the section regarding the normative system. These include an ongoing and public discussion on the proportionality analysis by which appropriate levels of intervention in the right to freedom of expression and also the right to privacy are determined. While it is important to have a normative framework, such a framework requires effective implementation, and we have set out possible mechanisms in the above section regarding practical measures. We will conclude with five additional observations.

First, while states are an important cog in the anti-disinformation machinery, they find themselves in an uneven power relationship with transnational online media and, in particular, very large online platforms when it comes to demand and control of information. States have a duty under human rights law to protect their citizens from the harm caused by disinformation. But, both due to the power imbalance between themselves and other stakeholders, and because their own credentials are not above suspicion, they cannot act alone.

Second, further actions need to be taken at global level to limit the huge power large technology companies have over people and democracies. The EU DSA, which introduces new mechanisms for removing illegal content and disinformation while seeking to protect users' fundamental rights, including freedom of expression, could serve as a global benchmark for regulatory approaches to online intermediaries at the global level.

Third, private content providers need well trained, well supported and well paid content monitors who scan and delete posts that violate clearly articulated platform guidelines. Such platforms should be required by law to ensure the validity of the information shared on them as quickly as possible. In this regard, we warn against a lenient approach to the intermediaries through whom information is shared.

Fourth, if content providers do not address disinformation, states should work with other stakeholders to identify and counter the disinformation before it can go viral or do significant harm. We recommend the creation and operation of experienced and well-resourced units that respond to disinformation as required. They should identify disinformation and take action to remove and mitigate it. This takes place on the platforms by taking posts down, reporting accounts, requesting bans, as well as sharing and generating accurate information to dispel the disinformation.

Given the extremely harmful potential of disinformation, we propose that criminal prosecution can and should follow where disinformation caused particular harm or posed a particular threat.

Fifth, any meaningful anti-disinformation programme needs to constrain the state as well as private actors. This is both because the state may not always have, or deserve, the trust of the general populace, and also because, even in those cases where state bodies may (currently) be trusted to identify and suppress

---

<sup>126</sup> These responses are proposed in 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression', Broadband Commission research report on 'Freedom of Expression and Addressing Disinformation on the Internet' <<https://broadbandcommission.org/publication/balancing-act-countering-digital-disinformation/>>.

disinformation in good faith, unaccountable and concentrated power can lead to abuse. The institutional design of the anti-disinformation therefore needs to include checks and balances to counter the power of both the state and big tech. We suggest that expert units be staffed by people who are not connected to any social media distributors; that is, people with training in the area of disinformation but no financial or political incentive either to suppress or promote particular narratives.<sup>127</sup> These independent units need to monitor and analyse the distribution of information on social media. Such teams should be allocated to specific areas of interest and specialization, including disinformation, incitement to violence, hate speech and xenophobia (broad types of information).

These teams should be located outside of the state, but will need to work in conjunction with organs of state to handle crises and identify emerging threats. Co-operation is necessary because the civil units may not have the capacity to remedy disinformation (which generally requires a real-time response) or to address forms of disinformation which threaten to develop into a crisis such as an insurrection. Government may be needed to provide the person-power and forms of state-sanctioned coercion when strictly necessary. It should thus have access to the knowledge produced by such bodies, and should have dedicated units gaining it and analysing it to be able to respond in a multisectoral manner (there is, for example, a different set of skills required to combat vaccine hesitancy than there is to prevent posts which are organizing an insurrection).

Co-operation between state and non-state actors should not, in itself, promote abuse of power by the state provided the institutional structure keeps the state in dialogue with other actors and accounting to civil society. A healthy institutional structure will facilitate the interaction of NGOs, the private sector, international organisations, such as the United Nations, and governments. Disinformation is a global phenomenon, and these forms of co-operation allow for a global response.

---

<sup>127</sup> The South African NGOs 'Centre for Analytics and Behavioural Change' and 'Real411' are good examples of such independent fact-checkers. See <https://cabac.org.za/> and <https://www.real411.org/>.