# ARTICLE 19's comments

# on the proposed UN Code of Conduct for Information Integrity on Digital Platforms

2024

## Introduction

ARTICLE 19[1] welcomes the opportunity to contribute to the consultation on the proposed UN Code of Conduct for Information Integrity on Digital Platforms (the Code of Conduct), initiated by the Department of Global Communications of the United Nations.

ARTICLE 19 shares the concerns of the UN Secretary General and his vision for "an open, free and secure digital future for all", the basis on which the process of the development of the Code of Conduct was initiated. We agree that numerous forms of misinformation and disinformation, the manipulation of public opinion through propaganda and state-led disinformation campaigns, and foreign interferences, present serious challenges to the protection of human rights and democracy. We have also long advocated for the adoption of measures that would prevent concentration of power over digital space in the hands of a small group of tech companies and subsequent dangerous bottlenecks to the flow of information across society. We believe that further clarification and detailed elaboration of recommendations to address various information threats is warranted. We fully support the efforts to ensure that information ecosystems are resilient to information threats, and that ecosystems' resilience is based on principles of openness, pluralism, diversity and accountability of those who control them.

In this submission, ARTICLE 19 focuses on key conceptual issues of the proposed Code of Conduct, rather than on the proposed text of the individual principles. We believe these key issues need to be clarified and discussed before the proposal is

[1] ARTICLE 19 is an international human rights organisation which works around the world to protect and promote the right to freedom of expression and information (freedom of expression). With an international office in London and regional offices in Mexico, Brazil, Bangladesh, Tunisia, Senegal, Kenya and the Netherlands, and other national offices, ARTICLE 19 monitors threats to freedom of expression in different regions of the world, develops long-term strategies to address them and advocates for the implementation of the highest standards of freedom of expression nationally and globally. This includes promotion and protection of the right to freedom of expression in the context of digital technologies.

advanced to the first draft. Thus, although we follow the structure of the section *Towards a United Nations Code of Conduct* (the Draft Proposal) in Policy Brief 8, we consider it preliminary to suggest an actual wording of proposed principles and the text of the Code.

## The purpose of the Code of Conduct

The stated purpose of the Code of Conduct is "to provide a concerted global response to information threats", while stakeholders, in particular UN Member States, will commit to its principles and will be "invited to implement the Code of Conduct at the national level."

ARTICLE 19 makes three broader comments about this aspect of the Draft Proposal.

First, we believe that **the purpose/role of the Code of Conduct** needs to be clarified. We note that the core principles, outlined in the Draft Proposal, have been already repeatedly stated and reconfirmed by the UN Charter and treaty bodies, such as the UN Human Rights Council and the UN Human Rights Committee. The current Draft Proposal does not appear to go beyond what can already be deducted from existing standards. These include the resolutions of the UN General Assembly and the Human Rights Council, general comments of the Human Rights Committee, and other recommendations (for instance of special rapporteurs). The same can be said about the specific recommendation for Member States to ensure that responses to hate speech and disinformation are consistent with "international law" *(Respect for Human Rights* section). Again, UN Member states are already bound by existing human rights standards to protect freedom of expression. These standards and their interpretation (including in general comments and decisions of the Human Rights Committee, or in standards such as the Rabat Plan of Action), already offer the framework for restricting speech that may be considered "harmful", including hate speech and disinformation.[2]

We believe that clarifying the purpose and role of the Code of Conduct in the human rights framework is necessary to ensure coherence and consistency between existing international human rights instruments and the Code.

Second, we are not persuaded that the proposed **type of instrument** – a "Code of Conduct" – presents a suitable approach. In our experience, "codes of conduct" are typically used in various private sector fields and rely entirely on voluntary compliance based on shared ethical standards of members of the particular industry. However, the UN Member States (key stakeholders of the Code) are already bound by international human rights treaties, which they have signed and ratified. In many instances, international and regional treaties might be directly applicable and

[2] Namely, Article 20 of the ICCPR prohibits propaganda for war and advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence, and the tripartite test under article 19(3) of the International Covenant on Civil and Political Rights (ICCPR) determines the permissible extent and form of restrictions on freedom of expression.

have priority before domestic law. Equally, UN Member States are, for instance, already obliged to other "principles" of the Code of Conduct as part of their positive obligations to protect the right to freedom of expression. These include supporting independent media, or ensuring transparency and access to information for their population. We find it problematic that the proposal suggests that States would be "invited to implement them" when States are already obliged to do so.

Third, it is not clear what will be the **system of enforcement or oversight** over the implementation of the Code of Conduct.

ARTICLE 19 therefore recommends that the scope, purpose and objectives of the Code of Conduct be clarified from the outset. Namely:

- We believe that respect for human rights should not be merely one of the core principles of this new instrument but the purpose of the instrument. This means that the instrument should specify that the aim of addressing information threats is to secure the protection of human rights, in particular the right to freedom of expression. This is not merely an issue of phrasing, but the overall framing of the instrument. The ultimate goal of the new instrument should be to provide stakeholders with detailed recommendations of measures needed to create a healthy information ecosystem (that is, "a balanced and resilient system of information creation, exchange, flow, and utilization").[3]

- In terms of proposed type of the instrument, we suggest considering similar instruments as "guiding principles" or an "action plan".[4] Here, the objective would be to produce a detailed compendium of measures that the UN Member States and digital companies should adopt to fulfil their obligations under the existing human rights framework. It should also specify that nothing in the new instrument should be read as limiting or undermining any legal obligations States may have undertaken or be subject to under international law with regard to human rights.

## Key concepts used in the Code of Conduct

ARTICLE 19 appreciates that the Draft Proposal of the Code of Conduct expects all stakeholders to commit to "information integrity" and states that all stakeholders "should refrain from "from using, supporting or amplifying disinformation and hate speech for any purpose…" (*Commitment to information integrity* section).

[3] See the Center for Democracy and Technology, Envisioning a Healthy Information Ecosystem, 2023.

[4] Similar to the Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework ('Guiding Principles'), which is also based on key pillars and provides more detailed directives for stakeholders; or the Rabat Plan of Action which specifies measures States must adopt under their obligations under Article 20 para 2 of the International Covenant on Civil and Political Rights.

This principle might be well-intended, but we find the suggested wording problematic. Namely:

- The term "**all stakeholders**" might encompass a very wide array of actors – including states, digital companies of different sizes, media outlets or even individuals. Each of these entities have different types of obligations under international law and international human rights law and the Draft Proposal does not seem to distinguish between them in several areas. For instance, we do agree that digital companies have responsibilities to respect human rights, independent of State obligations or the implementation of those obligations (*c.f.* e.g. The Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework). At the same time, we note that the ultimate responsibility to protect human rights lies with States, which might justify some legislative and other regulatory measures towards companies. The instrument should therefore clearly specify who are stakeholders of the Code, and recognise that their obligations would differ based on the nature of their entity.

- Similarly, we observe that **key terms** – such as disinformation or hate speech – are not defined in the Draft Proposal. We recall that these terms do not have an agreed definition in international or regional human rights law. Under international law the right to freedom of expression is not limited to information that is "truthful", making the regulation of disinformation and hate speech[5] highly complex and requiring nuanced approaches. In many instances, "disinformation" and hate speech can be forms of expression protected under international human rights standards.[6] In our experience, there is a lack of understanding about the applicable international freedom of expression standards, and national standards often either do not comply with the international ones and/or are prone to abuse. We worry that this broadly worded commitment could also be abused by States to restrict protected expression with the excuse that they are doing it in the name of their commitment to information integrity. The new instrument should at least clarify that international human rights framework, under which the respective expression can be restricted, is complex and requires careful assessment.[7]

- Further, we find that **some terminology** used in the text of the Draft Proposal suggests a uniform veracity standard. We understand the importance of "accurate", "consistent", and "reliable" information and the intention to promote such content. However, we reiterate that under international human rights standards, information cannot be censored or restricted merely on the basis of

[5] For 'hate speech,' ARTICLE 19 specifically cautions against using the definition set out in the United Nations Strategy and Plan of Action on Hate Speech. This broad definition also captures lawful expressions, and is therefore too vague for use in identifying expressions that may legitimately be restricted under international human rights law. See e.g. ARTICLE 19, Hate Speech Explained, 2015.

[6] *Ibid.*; see also ARTICLE 19, Response to the consultations of the UN Special Rapporteur on Freedom of Expression on her report on disinformation, 2021.

[7] *Ibid.* See also Hate Speech Explained, *op.cit.*

the degree of its "falsehood" or "accuracy". It is key to strictly link any restrictions of information in the digital space to a clear legitimate aim recognised in international human rights law: protection of the rights or reputations of others, national security, public order, public health or morals. All of these legitimate aims must be interpreted narrowly, and none of them allows for restriction of expression simply because it includes untrue information.

- ARTICLE 19 is also mindful that **disinformation and hate speech are often state-led**. They originate directly from representatives of state and public institutions, actors tied to authorities, and their various proxies. We believe it might be more appropriate to directly require *States* to refrain from resorting to "disinformation" or "hate speech" as part of their positive obligation to promote and protect freedom of expression and other human rights. As the same time, we observe that there are no restrictions on **extraterritoriality of restrictive measures**. When information threats are instrumentalised as tools of malicious foreign interference, UN Member States retain the right, and in certain cases carry the responsibility, to counteract these threats. However, restrictive responses to information threats should be geographically limited to the State mandating such measures, consistent with international principles of comity under international law.

## Regulatory measures to protect fundamental human rights

ARTICLE 19 appreciates that the second part of the section *Respect for human rights*[8] mandates Member States to "undertake regulatory measures to protect the fundamental rights of users of digital platforms, including enforcement mechanisms, with full transparency as to the requirements placed on technology companies." However, we observe that this section does not provide any clarity about the type of framework envisaged to achieve these objectives.

Subsequently, in the section *Stronger disincentives*, the Draft Proposal stipulates that "digital platforms should move away from business models that prioritize engagement above human rights, privacy and safety". It does not seem that this principle is linked to the role of member states to introduce the regulation that would mandate companies to end their exploitative practices. We find references to platforms' business models important in the context of information threats as research shows they promote lower diversity and quality of content. At the same time, the nature of "disincentives" that should be adopted by platforms – and whether states should have a role in them – is unclear. We note that ARTICLE 19 has elaborated a series of proposals for addressing these problems, including through improving competition and diversity of digital markets, and we encourage the drafters to consider some of our recommendations in this area.[9]

[8] We also note that the title of the section appears to contracting the protection element spelled out in the section. .

[9] See e.g. ARTICLE 19, Taming Big Tech: A pro-competitive solution to protect free expression, 2021; and proposals for the EU Digital Services Act and the Digital Markets Act.

ARTICLE 19 also believes that it is important to recognise that from the perspectives of human rights, freedom of expression, and democracy itself, another major concern is the power of very few large companies over information flow. We have also previously recommended that a tiered approach in this area might be necessary. In other words, large social media platforms could be made subject to more stringent obligations than smaller players.[10] This approach is also envisaged in the UN Guiding Principles. The Draft Proposal does not indicate any principles in this area.

Further, the Draft Proposal does not clarify enforcement mechanisms that Member States should adopt. ARTICLE 19 has long argued that the broader context in which regulatory proposals are made is particularly important. We note that in many Member States, regulatory bodies are not independent, whether in law or practice, and remain powerful avenues for governments to exercise control over information flows. This is not a theoretical concern, but a real threat both in established democracies and in countries where the rule of law is weak or under threat. UN Member States should be therefore obligated to ensure independence of all regulatory bodies enforcing adopted legislation.

ARTICLE 19 believes that in this section, the new instrument should include recommendations for clear interventions that would decentralises the power of the few dominant companies and include examples of regulatory interventions in the field of competition law and policy that can shape digital markets. ARTICLE 19 has developed a number of recommendations in this area.[11]

At minimum, if the Code of Conduct contains recommendations about regulatory measures for digital companies, it should at minimum specify that:

- The overarching principles of any regulatory framework of digital companies must be transparency, accountability, and the protection of human rights. The latter means that the legality, legitimacy, necessity and proportionality principles must be upheld throughout. In addition, any such framework must regulate companies rather than expression of the users of platforms, be based on robust evidence in order to adopt the most appropriate solutions; and

- UN Member States should adopt proactive action to prevent monopolisation or undue media concentration that may negatively impact the diversity and pluralism of the information environment.

[10] See ARTICLE 19, Watching the Watchmen: Content moderation, governance, and freedom of expression, 2021.

[11] See e.g. Taming Big Tech, *op.cit.*; ARTICLE 19, Submission to DG COMP call for contributions about Shaping Competition Policy in Age of Digitalisation, 2018; ARTICLE 19, Submission before the USA Federal Trade Commission, Consumer protection and Competition in the 21st century; and submissions in the course of drafting and negotiations of the EU Digital Markets Act. .

## Other issues

ARTICLE 19 also finds that the current text of the principles in the Draft Proposal consists in part of high-level principles and in some parts more detailed recommendations (e.g. regarding blanket shutdowns and bans on media outlets). Again, in some parts, the proposals repeat the recommendations from numerous recommendations of human rights bodies (e.g. the Human Rights Council resolutions) while providing no further specifications of measures that should be adopted. For instance:

- The section on *Support for independent media* stipulates that Member States should guarantee "free, viable, independent and plural media landscape with strong protections for journalists and independent media, and support the establishment, funding and training of independent fact-checking organizations in local languages." It does not provide more guidance on principles or framework under which the system of "establishment, funding and training" should be provided; e.g. that allocation of resources should follow certain transparency, independence and oversight safeguards. In ARTICLE 19's experience, allocation of state resources to media outlets is often a measure of state control and must always be non-discriminatory, and must be based on fair and neutral criteria, that it will never be used to promote official figures, political content or viewpoints expressed by media actors.

- ARTICLE 19 also observes that the importance of free, independent, and diverse media, one of the key pillars of responses to information threats, is understated in the Proposed Draft. The Proposed Draft contains some recommendations for "news media" in the *Support for independent media* section, but does not elaborate on the different types of media and role of genuine public service media in healthy media ecosystems. We urge the drafters to address this gap.

- The section *User empowerment* mandates that digital platforms give "people greater choice over the content that they see and how their data is used" but does not illuminate the types of measures platforms must undertake to achieve this. Here, ARTICLE 19 notes that this is exactly the section where recommendations should be more detailed and elucidate specific obligations, e.g. the requirement for large platforms to unbundle their hosting and content moderation functions and ensure they are interoperable with other services.[12]

- Similarly, the section *Increased transparency* would benefit from more detail when discussing meaningful transparency of digital platforms. The transparency obligations of platforms should not be limited to policies and reporting on mis- and disinformation and hate speech (especially in the light of the lack of definition of these terms). Transparency obligatoins should apply also to, *inter alia*, distribution of content, companies' terms of service and

[12] See Taming Big Tech, *op.cit.*

community standards, human and technological resources used to ensure compliance, online advertisement or decision-making.[13]

## Next steps

ARTICLE 19 hopes that the development of this new instrument continues to be subject to global **multi-stakeholder consultations**, guided by principles of genuine transparency, openness, inclusion, equality, participation, and accountability. A wide variety of voices need to be represented at these discussions, ranging from Global Majority actors to human rights organisations, development actors, digital platforms, and the media and advertisement industry.

We stand ready to continue our engagement in this process and look forward to further consultations.

[13] See e.g. ARTICLE 19, ARTICLE 19's Recommendations for the EU Digital Services Act, April 2020; ARTICLE 19, At a glance: Does the EU Digital Services Act protect freedom of expression?, 11 February 2021.