



ARTICLE 19 Response to the Governing AI for Humanity Report

At the end of 2023, the UN Secretary-General's AI Advisory Body issued its interim report *Governing AI for Humanity*. ARTICLE 19 welcomes that the Report integrates international human rights law as a cornerstone Guiding Principle for AI governance. We also appreciate that it reinforces the protection of human rights in several of the risks it highlights, with explicit reference to freedom of expression. At the same time, we are disappointed that the risks and priorities articulated thus far by the Advisory Body do not adequately capture the critical and already-recognised impacts that AI has had on freedom of expression generally. The Interim Report instead focuses on the harms caused by information; only 'nudging' appears in freedom of expression risks, and 'disinformation'—which has no internationally agreed upon definition—is referenced throughout.

While ARTICLE 19 is awaiting the formal report from the AI Advisory Body in August 2024, following additional consultations, we urge the Advisory Body not to understate an entire *category* of risk, especially as the Interim Report seeks to frame AI risks in terms of the "perspective of vulnerable communities and the commons". Media and journalists are a core sector of society and more than a missing detail, and should be part of these conversations from the beginning. ARTICLE 19 continues to be ready to share its expertise and analysis of existing trends, risks, guidance, and best practices specifically as they pertain to journalists and media.

Background

In December 2023, the UN Secretary General's AI Advisory Body (the AI Advisory Body) released its [Interim Report: Governing AI for Humanity](#). The AI Advisory Body was convened to analyse and advance recommendations for international AI governance from a variety of normative and institutional perspectives.

ARTICLE 19 has for several years monitored key developments in AI, including participating in a coalition for proposals to the European Parliament and Council for the Artificial Intelligence Act, and [has addressed the intersection of emerging AI with freedom of expression](#). We previously submitted [a comment in response to the Advisory Body's call for papers](#) on global AI governance. In those comments we encouraged the adoption of a human rights-centered framework—and specifically issues related to freedom of expression and impacts on journalists and media.

We are therefore pleased to comment on the Interim Report with a view to contribute to further consultations on this important subject. We are also looking forward to sharing our expertise and analysis of existing trends, risks, guidance, and best practices with the AI Advisory Body as it continues its work.

Positive Aspects of the Interim Report

At the outset, ARTICLE 19 emphasises numerous positive elements of the Interim Report, in particular:

- **The Report explicitly mentions human rights law in the Guiding Principles.** Specifically, we applaud that Guiding Principle 5 of the Advisory Body’s work makes explicit reference to the UN charter, international human rights law, and international commitments (para 55 of the Report). We do encourage the Advisory Body to make sure that this Guiding Principle also makes mention of the [special procedures, which have already issued guidance](#) on AI and human rights.
- **The Advisory Body commits to do a “deep” dive into human rights, and consult with civil society** in next steps of its work. The Advisory Body indicates that as part of its function of developing and harmonizing standards, safety, and risk management frameworks, it will seek the “active involvement of civil society” (para 64 of the Report). It also indicates it will consult with stakeholders around the world in the coming months, naming civil society. One goal will be to “dive deep” into a shortlist of issues, one of which is human rights (para 81). We echo the importance of this multi-stakeholder approach, and the commitment to name human rights as a critical part of the work of the coming months.
- **The Report explicit mentions “human rights due diligence” as part of a “simplified schema” for interoperability of governance efforts.** The Advisory Body accurately notes that a plethora of guides, frameworks, and principles on AI governance exist across multiple layers of society. As part of an attempt to distil and consider the AI governance landscape, the Advisory Body has offered what it calls a ‘simplified schema’ with four categories within governance (data, models, benchmarks, and applications). We were encouraged to see that “human rights due diligence” appears as one of three primary components of the ‘applications’ prong. At the same time, we emphasize that human rights should also be included in some way within every prong; ARTICLE 19 echoes the [calls of the High Commissioner for Human Rights](#) to examine “AI’s entire lifecycle,” evaluating how technical standards may contribute to or undermine human rights.
- **The Report refers to addressing transparency in AI governance.** As we recommended in our submission to the AI body, ARTICLE 19 believes that the opacity of machine learning algorithms presents particular challenges for individuals, regulators, civil society, and even designers of systems, as it [is often unclear when and how systems are utilized](#) and therefore difficult to audit their human rights implications. We note positively that the Advisory Body has highlighted this opacity and named transparency as a key challenge in its work (para 34 of the Report).

Room for Work and Consultation

Impact of AI on freedom of expression and media freedom

ARTICLE 19 recalls that a coalition of UN special procedures and [other experts recently called](#) for urgent action on the “alarming” use of AI to undermine journalists and human rights defenders, as well as the mass production of synthetic content to spread disinformation or promote incitement to hatred, discrimination or violence.

As we noted previously in our submission to the Advisory Body, AI impacts multiple groups—who are often subject to intimidation, harassment, and threats of violence in a [transforming media environment](#)—via means such as [bot network harassment](#), doxing or the use of generative AI to [create materials for blackmail](#) as well as AI-based surveillance (see above). AI can also be utilized to ‘de-anonymize’ individuals, undermining journalist-source relationships.

At the same time, AI impacts the work of newsrooms in novel ways. These include automated news creation, promoting broader dissemination (such as quickly translating stories for new audiences), or curating [access to stories based on reader patterns](#). These should not be used as a pretext for media regulation, and as such ARTICLE 19 suggests the Advisory Body monitor any attempts of governments to regulate the media. ARTICLE 19 recommends media self-regulation on how it deploys AI, to promote a pluralistic media environment.

ARTICLE 19 is concerned that the freedom of expression risks articulated so far by the Interim Report ignore these, instead solely focusing on the risks posed *by* expression rather than *to* expression. For example:

- The only risk listed following “freedom of expression” is so-called “nudging”. [Nudging refers to methods](#) in which artificial intelligence may be used to influence and manipulate decision-making.
- The phrases “disinformation”, “hostile information”, and “information integrity” are also referenced on numerous occasions throughout the Interim Report (Box 3, paras. 2, 24). These are raised to reference the potential harms on elections or other aspects of society. “Deep fakes” are a particular challenge in this regard.
- Finally, “filter bubbles, declining trust in news, information” are listed as risks to “social cohesion” (Box 3).

ARTICLE 19 acknowledges that there are serious debates to be had with all relevant stakeholders with respect to how artificial intelligence impacts so-called harmful information. However, these debates must not lose sight of the other ways that artificial intelligence is a freedom of expression issue, or encourage regulation of media and a framing that focuses on *limiting* information. Stakeholders must also not lose sight of other potentially harmful forms of information, such as

hate speech, which may [be exacerbated by AI systems](#) that struggle with contextual nuance, and require particular attention to human rights standards to appropriately address.

We are mindful of the Advisory Body’s observation that at this stage, “[p]utting together a comprehensive list of AI risks for all time is a fool’s errand” (para 30). However, media and journalists are more than a minor detail or specific example of a risk; they are an entire category that should be examined as a core part of society. We also note that merely adding the word ‘journalists’ after ‘nudging’ under freedom of expression would not properly address this either. This is because freedom of expression is only listed under the category of ‘individuals’, whereas media rights impact and concern all of society. Ultimately, journalists and media should not just be included on a list of topics or examples, but be a key part of the framing and conversation.

Challenges of Regulating Disinformation

ARTICLE 19 also notes that responses to ‘disinformation’—a term which has no internationally agreed-upon definition—have already led to significant infringements on freedom of expression. It is important that approaches to AI accordingly protect and promote freedom of expression, and this view is consistent with prevailing standards within the UN. In 2022 the Secretary General issued [a report on disinformation](#) from a framework of human rights and fundamental freedoms, in which he emphasized that “State responses to disinformation must themselves avoid infringing on rights, including the right to freedom of opinion and expression.”

The Secretary General’s report followed a strong call by the General Assembly that countering disinformation “requires” State responses to be in “compliance with international human rights law” and accordingly did not include criminal measures as an appropriate response. The General Assembly, in [Resolution adopted on 24 December 2021](#), explicitly reiterated the need:

to ensure that efforts to counter disinformation promote and protect and do not violate individuals’ freedom of expression and freedom to seek, receive and impart information, and noting that media and information-related technology literacy can help to achieve this through independent and free media, awareness-raising and a focus on the empowerment of people.

Elaborating on effective solutions, the call emphasized the need to address disinformation in a multistakeholder fashion that includes civil society, media, and business, through “education, capacity-building for prevention and resilience to disinformation, advocacy and awareness-raising.”

The Special Rapporteur on freedom of expression also issued a [comprehensive report on international standards surrounding disinformation](#) in 2021. In that report, she found that so-called “false news” laws typically failed to meet the three-pronged test of legality, necessity and legitimate aims set forth in Article 19(3) of the ICCPR.

Next steps

ARTICLE 19 applauds the inclusion of human rights as a central part of the guiding principles of the Interim Report, and the references that are made to freedom of expression. We hope that the Advisory Body will include media and journalists, particularly if themes such as ‘disinformation’, ‘nudging’, or other instances of the ‘harms’ of expression itself are included. We also urge the Advisory Body to identify the media and journalists as important stakeholders in the life cycle of the Advisory Body’s work, and a lens through which prevailing issues should be considered. Further, the Advisory Body should recognize that any attempts to address disinformation must respect and promote freedom of expression and be done in consultation with media, civil society, and relevant stakeholders.

ARTICLE 19 remains available throughout the consultation process to help continue the “dive deep” commitment to human rights and civil society engagement that the Advisory Body has expressed.