

# Input for “Our Common Agenda Policy Brief 8 Information Integrity on Digital Platforms”

Submission made on 18 December 2023



Workshop17  
17 Dock Road  
V&A Waterfront  
Cape Town, South Africa  
Phone: +27 21 447 6332  
[www.researchictafrica.net](http://www.researchictafrica.net)

Research ICT Africa(RIA) welcomes the opportunity to comment on the UN Secretary General's policy brief about [Information Integrity on Digital Platforms](#). The inputs below, each within the limit of 200 words, reflect RIA's work in this area, assessing the nexus of disinformation, data, and research access on the continent.

## **1. Commitment to information integrity**

RIA welcomes that the draft Code is intended for a range of stakeholders, including states and companies. Following this, however, we suggest that this section makes explicit the value of multistakeholder governance of digital developments as a fundamental condition for information integrity. This is the core proposal in "[Platform problems and regulatory solutions: findings from a comprehensive review of existing studies and investigations](#)", a policy brief for UNESCO prepared by RIA.

Since the brief notes that "Disinformation is also big business", this suggests another constituency that can be profitably referenced in the provisions of the Code – namely the entities making money out of providing disinformation services to clients. RIA thus recommends to add: "Public relations and advertising groups and their associations should adhere to codes of conduct for truth in advertising, and act against members who sell services which promote disinformation and hate speech." This helps to concretely tackle ethical and self-regulatory issues in an important area.

## **2. Respect for human rights**

Since it's hard to establish "intent" in mis-/dis-information, emphasis usually goes to the actual impact of such content in talking about harms. However:

- as it's difficult to infer causality from content, it's preferable to use the term "potential harms" – unless actual consequences have high probability. This is to avoid measures to over-curtail content without regard to its projected impact. To assess "potential" harm, the Rabat Plan of Action advises contextual consideration.
- the Code should avoid "harm" being instrumentalised to include eg. damaging image of government. Thus, harms should be explicitly specified as "harms to human rights". Human rights encompass peace, security, public health and environment, and should not be presented alongside these (as currently). Human rights also underpin concerns with content that risks humanitarian aid and climate action.

Recommendation: Instead of phrases like "harmful narratives" or "curb the harms", the Code should highlight 'potential harm to human rights'.

Operationalising the call for compliance with UN Guiding Principles on Business and Human Rights, merits explicit reference to human rights impact assessments (HRIAs) It can be encouraged that such tools help digital companies to anticipate and mitigate threats to information integrity, while civil society "shadow" HRIAs help multistakeholder governance.

## **3. Support for independent media**

Commendably, the brief calls for support for independent media, which is an alternative to, a flagger of, and investigator into, disinformation. Such a call should go further, or risk being too general as it currently stands. It is therefore recommended to add very specifically: "Governments, donors, and foundations should be urged to increase support to the viability of independent media and to the exercise of investigative journalism, including by backing the [International Fund for](#)

[Public Interest Media](#).” (for which the Secretary-General [has expressed support](#)). Further, it is recommended to add: “Digital platforms should recognise credible news outlets and adapt recommender systems to uprank their content”.

#### **4. Increased transparency**

The brief notes that at company level, gaps persist in policy, transparency and implementation, and further, that there are enormous global inequities in content moderation capacities. Of voluntary commitments such as the European Union’s Code of Practice on Disinformation, the brief says: “The real test of these new mechanisms will be in their implementation.”

Recommendation to strengthen these issues – The Code could explicitly say, “voluntary codes can be accompanied by regulatory requirements for transparency on the prevalence of coordinated disinformation on platform services, their content moderation practice, and the efficacy of their policies to counter such operations.”

[Regulatory measures like these do not interfere with platforms’ prerogative to set content policies (within the law), but instead provide the necessary independent scrutiny. Through making transparency a legal liability, platforms may be incentivised to increase their actions to anticipate and mitigate content risks to human rights. Signalling that this specific transparency may be a legitimate regulatory requirement, may stimulate companies to do so proactively to the benefit of information integrity in jurisdictions where such openness is not yet mandated]

#### **5. User empowerment**

The draft code says Member States “should ensure public access to accurate, transparent, and credibly sourced government information...”. It’s recommended to add after “information” the words: “and data, including private data, as relevant to the public interest”. [This highlights the importance of raw data as a major asset of public interest, raising issues beyond (processed) information, and points to the Covid precedent of transparency of private sector data.]

The brief urges that “all stakeholders should invest in robust digital literacy drives to empower users”. To improve impact here, it is recommended to add specifics: “Governments should promote media and information literacy, online and offline, in formal and informal educational systems; platform companies should be required to systematically amplify opportunities to build such competencies; Civil society and others should implement and monitor programmes in these areas. All should commit to the UN’s annual Global Media and Information Literacy Week”.

RIA’s discussions on the Global Digital Compact around the continent have brought out issues about lack of understanding among state agencies to tackle gender-related harms. A recommendation should therefore be made for platform companies to help capacitate stakeholders such as judiciary, law enforcement, and researchers to investigate misinformation-related hate speech, especially narratives targeted towards communities as well as [gender-based hate and inequalities](#) linked to the digital economy.

#### **6. Strengthen research and data access**

The brief notes that “data access for researchers is also an urgent priority on a global scale”. Interpreting this as about the “supply side” of data, attention is also needed to the “demand side” – as noted by key RIA [consultations](#) and [research](#) results. The draft code addresses this issue by saying: “Member States should invest in and support independent research ...”. This can be

strengthened by adding the following text: “Such support should be through independent mechanisms with appropriate safeguards and vetting procedures, and it should also be extended to researchers acquiring appropriate tools and building their digital research capacity.”

In “Next Steps”, the brief signals that the UN will develop “tailored communication strategies to anticipate and/or rapidly address threats before they spiral into online and offline harm”.

Recommendation: To this phrase can be added: “To optimise this foresight, digital platforms should partner with UN agencies, funds and programmes to provide data access to enable tracking and pre-emptive actions”. Alternatively, the point about platforms sharing data with UN actors can be made in this subsection of the Code (on Strengthen research and data access).

## **7. Scaled up responses**

The draft Code calls for stakeholders to “Form broad coalitions on information integrity, bringing together different expertise and approaches to help to bridge the gap between local organisations and technology companies operating at a global scale”.

Recommendation: this could be supplemented more concretely with a sentence such as: “Such coalitions are particularly important for information integrity in terms of elections, conflict situations, health crises and natural disasters”.

In “Next Steps”, the policy brief notes the possible “establishment of an independent observatory made up of recognized experts”. Since there are several such initiatives, the Code could – under the section on Scaled Up responses - encourage their co-operation amongst each other, and with UN agencies, Funds and Programmes.

## **8. Stronger disincentives**

The brief recognises that a problem is the current business models of most digital platforms which reward and amplify mis- and disinformation and hate speech. The draft Code calls for companies to “Ensure meaningful transparency regarding algorithms, data, content moderation and advertising”. It further states that “Digital platforms should move away from business models that prioritise engagement above human rights, privacy and safety”.

Recommendation: To make this more operationalisable, it can be added: “Steps can especially be taken to ‘dial down’ algorithmic amplification and recommender systems ahead of, or in times of, crisis, and to give prominence to authoritative content including credible news. In addition, data-driven profiling can be moderated and third-party cookies deprecated in order to reduce manipulative potential in advertising and content feeds.”

Also recognised in the brief is that selling a seal of approval as if it were authenticity, is a serious concern. To this can be added concerns with providing undisclosed promotion of content from those who purchase this seal (as Elon Musk has publicly stated). A recommendation here is for stakeholders to support verification of identity, plus watermarking and credible labelling that shows content provenance.

## **9. Enhanced trust and safety**

The draft Code calls for “urgent and immediate measures to ensure the safe, secure, responsible, ethical and human rights-compliant use of artificial intelligence and address the implications of recent advances in this field for the spread of mis- and disinformation and hate speech”.

A recommendation is to add: “This call applies to foundation models and applications in the field of generative AI, as well as in classificatory AI deployments in recommender systems and adtech.”

**10. Other (proposal for a new principle not already addressed)**

While the code is focused on content, it would be relevant and legitimate to propose support for platform pluralism, including support for decentralised and non-profit alternatives to the major companies. This can provide more choice to users, particularly if there are platforms in local languages and at a scale where safety (as the respect of human rights) is manageable. Whether this entails the need for effective competition / subsidy policies or regulation for interoperability between platforms is a further consideration. However, the Code currently risks the analogy of focussing on curbing fossil fuel and ignoring the importance of developing renewable energy. For future proofing, this is very important.

**11. In no more than 200 words, please provide any additional suggestions for methodologies of implementation.**

A code such as this is normative, with potential influence rather than enforcement. This means that its significance extends only in as much as it is promoted and publicised to reach the status of a benchmark. Without this, its value will easily dissipate, and the key duty bearers left free to ignore its calls. The UN and its agencies can support multistakeholder panels to conduct assessments in given countries, and produce reports on shortfalls and successes in relation to the Code. These can be tabled at appropriate fora, and linked to appropriate days on the UN calendar.

==