



International Panel on the Information Environment

IPIE Submission to the United Nations

UN Code of Conduct for Information Integrity
on Digital Platforms

Technical Paper 2023.1

IPIE SUBMISSION TO THE DEVELOPING UN CODE OF CONDUCT FOR INFORMATION INTEGRITY ON DIGITAL PLATFORMS

Technical Paper TP2023.1

SYNOPSIS

This submission by the International Panel on the Information Environment (IPIE) sets out recommendations for a UN Code of Conduct for Information Integrity on Digital Platforms. The recommendations aim to extend the principles by including the following:

- Commitment to information integrity: Include an obligation for stakeholders to correct mis- and disinformation,
- Respect for human rights: Researchers should be granted access to data and systems so that they can map hate speech on digital platforms and assess countermeasures and their impact on human rights,
- Support for independent media: Independent national media that conducts fact-checking in local languages should be supported,
- Increased transparency: Researchers should have access to the data and systems of digital platforms without delay or incurring unreasonable costs,
- User empowerment: Content labelling and corrective information should be promoted as validated countermeasures to mis- and disinformation on digital platforms,
- Strengthened research and access: Digital platforms should adopt a “disclosure by default” approach to granting researchers access to data and systems, and more research in more countries and in more languages is urgently needed,
- Scaled-up responses: Research on the information environment should be encouraged in all parts of the world and in languages other than English,
- Stronger disincentives: Researchers should be given access to how digital platforms counter mis- and disinformation, and the research should be available for regulators and legislators so that they can use it to design disincentives,
- Enhanced trust and safety: Research is needed on AI in the information environment, and AI should not be used to shape the information environment unless it is safe, secure, responsible, ethical, and human rights compliant.

Introduction

This submission provides requested feedback on the online consultation for the UN Code of Conduct for Information Integrity on Digital Platforms.

The Code of Conduct is conceptualized as a “gold standard for guiding action” and is being prepared for the Summit of the Future 2024 (United Nations, 2023). The intention is to make digital platforms inclusive and safe for all, and to safeguard the rights to freedom of opinion and expression and access to information.

Policy Brief 8: Information Integrity on Digital Platforms sets out the background (United Nations, 2023). The *Policy Brief* explains that while there are many benefits to digital platforms, they also give rise to numerous problems. Elements of information integrity, such as accuracy, consistency, and reliability of information, are compromised by mis- and disinformation and hate speech. The spread of mis- and disinformation undermines scientifically established facts and threatens democratic institutions, fundamental rights, and the Sustainable Development Goals, and may ultimately pose an existential risk to humanity. Hence, as the *Policy Brief* concludes, “[t]he dangers cannot be overstated.”

A Code of Conduct is a welcomed and necessary step towards regaining and upholding information integrity on digital platforms. As an independent global organization dedicated to providing actionable scientific knowledge on threats to our information landscape, the International Panel on the Information Environment (IPIE) strives to empower policymakers, industry leaders, and civil society with invaluable insights for a better global information environment. With contributions from hundreds of leading scientists worldwide, the IPIE is uniquely placed to provide robust empirical evidence on information integrity in the global online information environment.

This IPIE submission aims to support the proposed Code of Conduct by contributing targeted recommendations for the nine principles set out in the Policy Brief: (1) commitment to information integrity, (2) respect for human rights, (3) support for independent media, (4)

increased transparency, (5) user empowerment, (6) strengthened research and data access, (7) scaled-up responses, (8) stronger disincentives, and (9) enhanced trust and safety.

The submission draws on the IPIE's reports *Platform Responses to Misinformation: A Meta-Analysis of Data* and *Trends in the Global Information Environment: 2023 Expert Survey Results* (IPIE, 2023.3). The first report provided a meta-analysis of 43 studies from 18 peer-reviewed manuscripts. The second report presented the findings of a survey of 289 scientific experts. These experts identified the absence of accuracy and a lack of diverse voices as the most prominent barriers to a healthy global information environment. Secondary issues were information literacy and education, the reliability of the digital infrastructure, and media ownership. The scientific experts raised concerns regarding gender and climate science misinformation, foreign interference in elections, and hate speech directed at minorities, women, race, immigrants, and members of the LGBTQ+ community. More than half the experts expressed misgivings concerning badly executed AI and online content moderation or poorly designed AI-powered content moderation.

Overall, the research led to the conclusion that “[t]he information environment is rapidly evolving, with algorithmic bias, manipulation and misinformation having significant impact on public life.” (IPIE (2023.3) p. 3).

1. Commitment to information integrity

All stakeholders should refrain from using, supporting, or amplifying disinformation and hate speech for any purpose, including to pursue political, military, or other strategic goals, incite violence, undermine democratic processes or target civilian populations, vulnerable groups, communities or individuals.

The IPIE considers information integrity to be vital for a healthy global information environment. Although we do not take a political or ideological position, we emphasize not only the importance of refraining from using, profiting from, supporting, or amplifying disinformation but also the obligation of platforms to actively flag and correct mis- and

disinformation to maintain information integrity. As much as possible, stakeholders should actively make use of scientific consensus, clear evidence, and accurate reporting on digital platforms.

2. Respect for human rights

Member States should: (i) Ensure that responses to mis- and disinformation and hate speech are consistent with international law, including international human rights law, and are not misused to block any legitimate expression of views or opinion, including through blanket Internet shutdowns or bans on platforms or media outlets; (ii) 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. All stakeholders should comply with the Guiding Principles on Business and Human Rights.

The IPIE's scientific evidence demonstrates that there are widespread concerns regarding the use of hate speech to target marginal, LGBTQ+ communities, minorities, and women online. Ensuring respect for human rights requires that technology companies should provide researchers with access to their data and systems to help identify human rights abuses, in particular hate speech. Member States should balance the need to protect privacy, data access, and the right to freedom of speech before imposing bans or blanket shutdowns of platforms or media outlets.

3. Support for independent media

Member States should guarantee a 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.

News media should ensure accurate and ethical independent reporting supported by quality training and adequate working conditions in line with international labor and human rights norms and standards.

The IPIE’s scientific experts are concerned that media concentration and censorship have negative consequences for the global information environment. Indeed, independent media are critical for providing corrective information that social media platforms can use in an effective, comprehensive strategy for promoting information integrity.

Our findings demonstrate how important the conditions for a healthy information environment are in this regard, because “[i]n healthy information environments, citizens have more knowledge about politics and current affairs to help them make informed decisions and hold politicians accountable. Informed citizens are more likely to participate in politics, develop well-considered attitudes towards social and political affairs, and engage in civic activities.” (IPIE (2023.2), p. 6).

4. Increased transparency

Digital platforms should: (i) Ensure meaningful transparency regarding algorithms, data, content moderation and advertising; (ii) Publish and publicize accessible policies on mis- and disinformation and hate speech, and report on the prevalence of coordinated disinformation on their services and the efficacy of policies to counter such operations.

New media should ensure meaningful transparency of funding sources and advertising policies, and clearly distinguish editorial content from paid advertising, including when publishing to digital platforms.

Transparency, especially concerning researchers’ access to data, is key to a healthy global information environment. It is not enough to frame transparency as a process of the industry self-reporting. Increased transparency will allow independent scientific study of the challenges to achieve information integrity, ways to overcome those challenges, and the methods for improving information resilience globally. Transparency should be provided without delay and without unreasonable costs being incurred by digital platforms and regulators.

5. User Empowerment

Member States should ensure public access to accurate, transparent, and credibly sourced government information, particularly information that serves the public interest, including all aspects of the Sustainable Development Goals.

Digital platforms should ensure transparent user empowerment and protection, giving people greater choice over the content that they can see and how their data is used. They should enable users to provide identity and authenticity free of monetary or privacy trade-offs and establish transparent user complaint and reporting processes supported by independent, well publicized, and accessible complaint review mechanisms.

All stakeholders should invest in robust digital literacy drivers to empower users of all ages to better understand how digital platforms work, how their personal data might be used, and to identify and respond to mis- and disinformation and hate speech. Particular attention should be given to ensuring that young people, adolescents, and children are fully aware of their rights in online spaces.

The two most effective means of user empowerment are content labelling and providing corrective information, and these are defining features of information integrity (IPIE (2023.3) p. 6). Content labelling means labelling posts, accounts, and stories with tags about fact-checking, funding, or advertising, or any other forms of tagging or flagging, including providing further context without the user having to click through to receive the additional information. Correcting information means that an organization, platform, or individual must provide accurate information without regard to whether users have preconceptions about it.

Measures that ensure the implementation of content labelling and the provision of corrective information should be comprehensive and consistent so as to strengthen user empowerment. There may also be merit in other measures, such as advertisement policies, content reporting, disclosure, redirection, security or verification, or self-fact-checking. However, the

effectiveness of these is still uncertain, and increased transparency will help affirm the role of other measures.

6. Strengthened research and data access

Member States should invest in and support independent research on the prevalence and impact of mis- and disinformation and hate speech across countries and languages, particularly in underserved contexts and in languages other than English, allowing civil society and academia to operate freely and safely.

Digital platforms should: (i) allow researchers and academics access to data, while respecting user privacy. Researchers should be enabled to gather examples and qualitative data on individuals and groups targeted by mis- and disinformation and hate speech to better understand the scope and nature of harms, while respecting data protection and human rights; (ii) ensure the full participation of civil society in efforts to address mis- and disinformation and hate speech.

Strengthened research and access to data are vital to a healthy information environment and information integrity across our digital infrastructure. The IPIE has identified a pressing need for more research from more countries in more languages, which necessitates greater access to social media data (IPIE (2023.3), p. 3). There is a particular concern that the vast majority of our sampled publications are in English and that participants in research experiments are predominantly from the USA. We support a shift from “access by request” to “disclosure by default,” and urge all digital platforms to adopt this as their approach to researchers’ access to their data and systems.

Our scientific experts expressed concerns regarding the lack of access to platforms’ data that makes it difficult to study, and therefore to draw conclusions about, the information environment. The lack of access also makes it difficult to assess the impact of countermeasures, such as online content moderation. We therefore support the principle that

by default, large-scale public-facing information systems should provide researchers with access to the data and metadata they need to evaluate information integrity.

7. Scaled-up responses

All stakeholders should: (i) allocate resources to address and report on the origins, spread and impact of mis- and disinformation and hate speech, while respecting human rights norms and standards and further invest in fact-checking capabilities across countries and contexts; (ii) form broad coalitions on information integrity, bringing together different expertise and approaches to help bridge the gap between local organizations and technology companies operating at a global scale; (iii) promote training and capacity-building to develop understanding of how mis- and disinformation and hate speech manifest, and to strengthen prevention and mitigation strategies.

As part of the principle on scaled-up responses, the IPIE would like to add that there should be encouragement of research on the global information environment in all parts of the world, across many cultural contexts, and in languages beyond English.

8. Stronger disincentives

Digital platforms should move away from business models that prioritize engagement above human rights, privacy and safety. Advertisers and digital platforms should ensure that advertisements are not placed next to online mis- or disinformation or hate speech, and that advertising containing disinformation is not promoted.

News media should ensure that all paid advertising and advertorial content is clearly marked as such and is free of mis- and disinformation and hate speech.

While the IPIE's analyses have validated the use of content labelling and corrective information as counter measurements against mis- and disinformation and hate speech, there are numerous other countermeasures which still need to be evaluated. More research is urgently needed to test the effectiveness of an array of possible countermeasures. As part of

the principle for stronger disincentives, we would like to add an obligation for digital platforms to provide researchers with access to their data and systems, including any data on how behaviorally targeted advertising and psychometrics are used. Researchers should also be provided with access to data on how mis- and disinformation are identified and countered by the digital platforms. This research should be made available to inform regulators and legislators that are designing disincentives to suppress mis- and disinformation and hate speech in the information environment. Unfortunately, until transparency and researcher access to data is achieved, it is difficult to evaluate a large number of the proposals for disincentivizing interference in information integrity.

9. Enhanced trust and safety

Digital platforms should: (i) ensure safety and privacy by design in all products, including through adequate resourcing of in-house trust and safety expertise, alongside consistent application of policies across countries and languages; (ii) invest in human and artificial intelligence content moderation systems in all languages used in countries of operation, and ensure content reporting mechanisms are transparent, with an accelerated response rate, especially in conflict settings.

All stakeholders should take 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.

The IPIE is concerned that at the time of writing there is limited robust empirical research on unfolding AI advancements, particularly regarding how these may be used to spread mis- and disinformation and hate speech on digital platforms. The IPIE is also concerned that AI technologies are being adopted to counter mis- and disinformation and hate speech, for example by automated AI-driven content moderation, without robust transparency and assessment of intended and unintended consequences. There is a risk that these systems will

infringe fundamental rights or suppress information that should be readily available. We would therefore like to add to this principle that measures designed to make AI safe, secure, responsible, ethical, and human rights-compliant must be based on empirical evidence-based research and globally meaningful auditing standards. Stakeholders should also be obliged to adopt effective measures to counter algorithmic manipulation, bias, and misinformation in the global information environment.

References

International Panel on the Information Environment, 2023. *Platform Responses to Misinformation: A Meta-Analysis of Data*. SR2023.2. Zurich, Switzerland: IPIE.

International Panel on the Information Environment, 2023. *Trends in the Global Information Environment: 2023 Expert Survey Results*. SR2023.3. Zurich, Switzerland: IPIE.

United Nations (2023), *Policy Brief 8: Information Integrity on Digital Platforms* (Our Common Agenda).

Contributors

Drafting Author: Ann Kristin Glenster (Consulting Scientist, Norway/UK). Reviews: Philip Howard (IPIE Chair, Canada/UK), James Ward (Researcher, USA/UK). Copyediting: Beverley Sykes.

Copyright Information



This work is licensed under an Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Technical Papers

Technical Papers are prepared on specialized topics for which an independent scientific review or technical perspective is deemed essential. Most Technical Papers are used to inform scientific consensus building. Some Technical Papers are prepared as submissions to, or at the request of, international organizations or governments. Technical Papers are produced by

individual authors at the request of the IPIE, but any opinions, findings, conclusions, or recommendations expressed in Technical Papers are those of individual authors and may not reflect the views of the international scientific community.

About the IPIE

The International Panel on the Information Environment (IPIE) is an independent and global science organization committed to providing the most actionable scientific knowledge about threats to the world's information environment. Based in Switzerland, the mission of the IPIE is to provide policymakers, industry, and civil society with independent scientific assessments on the global information environment by organizing, evaluating, and elevating research, with the broad aim of improving the global information environment. Hundreds of researchers from around the world contribute to the IPIE's reports. www.IPIE.info.

For more information, please contact the International Panel on the Information Environment (IPIE), secretariat@IPIE.info. Seefeldstrasse 123, P.O. Box, 8034 Zurich, Switzerland.



International Panel on
the Information
Environment

Seefeldstrasse 123
P.O. Box 8034 Zurich
Switzerland

