

Background

The United Nations Secretary General, as part of his work around the strategic vision called “Our common agenda” released a series of policy briefs that should guide the implementation of the agenda.

This document is a response to the request for input that the United Nations asked about “Our Common Agenda Policy Brief 8: Information Integrity on Digital Platforms.”

The Policy brief articulates the United Nations strategy on disinformation and trustworthy information and will serve as the theoretical framework to develop a “Code of Conduct for information integrity on digital platforms” which will be developed by the United Nations’ Department of Global Communications.

The Secretary-General articulated the recommendation around nine principles:

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
9. Enhanced trust and safety

For each principle, the UN developed clear recommendations available [here](#) and highlighted in light-blue in this document, the request was to give input specifically on these rather than on the topic in general.

1. **Commitment to information integrity**

(a) 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 Wikimedia Foundation is the non-profit organization that hosts Wikipedia and other volunteer-run free knowledge projects. These projects provide free access to neutral, well-sourced information about science, culture, history, and other encyclopedic subjects.

The Wikimedia community's goal is to improve public knowledge and increase people's access to trustworthy information across the whole world through collaborative editing and governance processes. This global model for sharing neutral, well-sourced free knowledge makes Wikimedia an antidote to disinformation.

We support this recommendation, and further suggest that stakeholders should not only “refrain from using, supporting, or amplifying disinformation,” but also actively contribute to local and global information ecosystems by participating in and supporting free knowledge projects like Wikipedia.

Laws aimed at countering disinformation should be consistent with universal human rights standards and their drafting grounded in a transparent and accountable process, with meaningful participation of different stakeholders.

2. Respect for human rights

(b) Member States should:

(i) Ensure that responses to mis- and dis- information 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;

(c) All stakeholders should comply with the Guiding Principles on Business and Human Rights;

The internet is a key tool for facilitating freedom of expression and information, which empower individuals to exercise many other rights, such as the rights to education, artistic expression, economic advancement, and political participation.

The Foundation's commitment to protect the Wikimedia volunteer community-led governance model reflects the strong commitment to freedom of expression and of information built into our projects, which makes them reliable sources of knowledge for billions of people.

All stakeholders are responsible for protecting human rights principles online. These should be at the core of government policies, as well as all content moderation and privacy decisions made by digital platforms. Governments must consider freedom of expression implications of copyright or content policies which mandate automated filtering, or which incentivize top-down content moderation through harsh penalties or short removal windows.

Digital platforms have responsibilities and accountabilities under the UN Guiding Principles on Business and Human Rights. The Wikimedia Foundation refers to them in its human rights policy, and in our organizational-level Human Rights Impact Assessment. We are taking these responsibilities seriously by conducting regular due diligence, and believe stakeholders should encourage and expect a similar level of commitment from other digital platforms, proportionate to their size and resources.

3. Support for independent media

(d) 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;

(e) News media should ensure accurate and ethical independent reporting supported by quality training and adequate working conditions in line with international labour and human rights norms and standards;

Wikimedia projects and the rest of the information ecosystem influence each other. Wikipedia's content moderation principles do not allow original research: Information must be based on sources that the volunteer community has deemed reliable. The quality of content in turn suffers if the ecosystem is not composed of a plurality of trustworthy and diverse media outlets. Hence, Wikipedia both contributes directly to the information ecosystem through its open model, and also relies on this ecosystem being healthy for its content. Therefore, we actively support a healthy and diverse

independent media beyond Wikipedia, especially at the local and community level. We call for laws and regulations that protect and support independent media, research, and all forms of independent and freely accessible information, especially when shared through open licenses.

- Governments must ban the sale of private surveillance technology that has been used to surveil journalists and human rights activists.
- Governments and private media funders should support open access models such as open data repositories, openly licensed research, and cultural heritage materials, etc., and make publicly funded information accessible through free licenses.

4. Increased transparency (200 words maximum)

(f) 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;

(g) News 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;

The Wikimedia model is one of radical transparency: content is freely available, policies are collectively crafted, and the way the platform works is transparent to all. Hundreds of thousands of volunteers globally can actively participate in its functioning, including designing code to sort through information and fight against misleading content. This model offers multiple tools against disinformation, and promotes trustworthy content over profit-oriented proprietary models in which the algorithm's functioning and use of user data is often unknown.

Wikimedia projects and other community-led platforms are an example of the advantages of transparency, and should be protected and supported by legislation. Governments should protect models that increase user participation and transparency, by drafting regulations carefully to be inclusive of such models.

- To strengthen transparent and trustworthy resources such as Wikipedia, it is also essential that emerging regulation focuses on protecting open licenses. Wikimedia projects' content is created and developed under a Creative Commons open license.

Governments should balance commercial intellectual property rights interests with the public right to information, and ensure the continued existence of a robust public domain.

5. User empowerment

(h) 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;

(i) Digital platforms should ensure transparent user empowerment and protection, giving people greater choice over the content that they see and how their data is used. They should enable users to prove 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;

(j) All stakeholders should invest in robust digital literacy drives 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;

This brief highlights the “need for bottom-up solutions that empower Internet users to limit the impact of online harms on their own communities and decentralize power from the hands of the platforms,” which echoes Wikimedia's community-led model.

However, the focus is on for-profit models, ignoring other types of digital platforms. While understandable, as for-profit platforms carry higher disinformation risks, the immense benefits of non-commercial, community-led platforms like Wikipedia are often overlooked in policy debates, resulting in top-down content moderation requirements that are inconsistent with community self-governance.

Volunteers of every identity edit Wikimedia projects. They include members of at-risk categories who risk political persecution, incarceration, and even physical threats. To

ensure their safety, it is essential that their identities are protected and their privacy is respected.

- We agree that data protection and other identity safeguards are a good start toward protecting user privacy online. This principle should be extended further, so that users' access to online spaces does not require proof of their identity.
- Governments and international organizations should work to expand stakeholder support for public goods and bottom-up, community-based models.

Policymakers should assess legislation for human rights risks, and risks of harming digital public goods models.

6. Strengthen research and data access

***(k)** 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;*

***(l)** 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;*

All Wikimedia data is free and available to all. Anyone can search and reuse content on Wikipedia and other Wikimedia projects. This transparency fosters academic research. The Foundation also has an internal research team and also partners with academic researchers through exchanges of analysis and data, particularly with academics working on trustworthy information.

We also work with civil society, including formal volunteer communities (i.e., affiliates) and individuals to gain better visibility into disinformation threats to our platforms. Wikipedia is active in more than 300 languages; English Wikipedia alone receives more

than ninety billion website views each month. Working with civil society and academia is essential to better understand potential disinformation threats, particularly in geographies that require complex cultural and language skills.

- Digital platforms should ensure that academic research is promoted and supported, especially in regions in which they have less visibility and in minor languages. Researchers should be given access to data while respecting user privacy.
- Digital platforms should be encouraged to collaborate with independent researchers and exchange information about disinformation appearing on their platforms.

7. Scaled up responses

(m) 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 to 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;

The Foundation works actively to establish and foster broad coalitions on information integrity, including as co-chair (with the Government of Denmark) of a newly established multi-stakeholder Freedom Online Coalition (FOC) Task Force on Trustworthy Information Online.

We regularly conduct human rights due diligence with external partners to understand the impacts our projects and activities can have on underrepresented and vulnerable groups, including a previous organization-wide Human Rights Impact Assessment and a recent Child Rights Impact Assessment. These reports focus also on the potential spread and impact of false and misleading information.

Due to the participatory nature of our projects, we are naturally inclined to work with external partners to increase understanding of threats and possible responses.

We support the proposals in this section and pledge to continue supporting multi stakeholder work to counter false and misleading information online.

8. Stronger disincentives

(n) Digital platforms should move away from business models that prioritize engagement above human rights, privacy and safety;

(o) 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;

(p) News media should ensure that all paid advertising and advertorial content is clearly marked as such and is free of mis- and dis-formation and hate speech;

Unlike many of the platforms which may be targeted by laws regarding advertising, the Wikimedia projects are free of commercial advertising and available under a free license to reuse and republish with attribution. Our non-profit, volunteer-run content governance model means that the Foundation does not have the same incentives to encourage the creation of or drive traffic to particular content. The Foundation also minimizes the data collected on editors and readers of Wikimedia projects, does not target users with content based on detailed profiles, and does not sell data to third parties.

Many content regulations focus on addressing concerns about harms caused by large, commercial social media platforms with targeted advertising business models. When drafting any laws directed at digital platforms, including privacy legislation, it is important that legislators account for unique models, including volunteer-run, public interest models like the Wikimedia projects. Government and industry alike must work to encourage the adoption and sustainability of a variety of platform models, including non-profit models which do not rely on commercial advertising.

9. Enhanced trust and safety

(q) 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

(r) 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 Wikimedia model for content moderation is different from that of other large online platforms. Communities co-create and co-moderate content, set and enforce rules, while the Foundation supports them in addressing needs, obstacles, and threats. The Foundation cannot directly intervene in most content moderation processes or decisions without destroying the community's content governance model.

Instead, the Foundation provides support, from sharing research material to investigative support, technical tools for detecting vandalism, as well as specialists focused on volunteers' safety and protection, and intervenes only in a very narrow range of cases, such as child sexual abuse material or threats of imminent physical harm where volunteer involvement would be unethical or illegal.

Wikimedia's unique Trust & Safety model empowers our communities to self-govern and address challenges in contexts appropriate for specific geographies, languages, and subject matter. However, community content moderation takes time, and inflexible regulation that focuses on quick takedowns can be fatal for Wikimedia's decentralized governance model.

- Governments and international organizations should take alternative models of content moderation in consideration when drafting all technology policy, including for the deployment of AI, and ensure that regulation safeguards them as public goods.

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

Trustworthy information online requires collective work, as well as the use of diverse sources that take into consideration different perspectives, especially those from

people and communities who are currently at risk or have been marginalized in the past.

Projects like Wikipedia that support community engagement are essential for this, and should be included in policy conversations. Digital public goods such as Wikipedia and other Wikimedia projects should also be supported because they clearly advance the Sustainable Development Goals (SDGs).

- Governments and international organizations should protect and support the right of communities to create and maintain online platforms dedicated to public information and other digital public goods.
- Governments and international organizations should work through multi stakeholder processes to develop standards and definitions for what constitutes a digital public good, ensure the right to freedom of expression online is protected, and that people are not punished for accessing or sharing factual information.
- Governments and international organizations should invest in and promote the creation and distribution of content that is relevant to local and communities in a variety of languages.

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

We support a multi-stakeholder policy making model, and stress the importance of local, bottom up contribution and engagement. It would be important for the success of the Code of Conduct to create spaces for national dialogues—perhaps by sectors—so that all stakeholders feel that they are part of the solution, and able to make concrete commitments. In our experience, in the 21st Century top-down, rigid policy-making processes are unlikely to succeed. It is essential for policy-making processes to encourage and empower grassroots ideas and energy through bottom-up organization of dialogues tailored to a range of contexts, priorities, concerns and challenges. Such processes take longer to play out and can be unpredictable, but in the end are more enduring and hold greater legitimacy for more people who depend on their success.