CENTRE FOR MEDIA TRANSITION

Submission to UNESCO

Safeguarding freedom of expression and access to information: guidelines for a multistakeholder approach in the context of regulating digital platforms

UNESCO Conference: Internet for Trust - Towards Guidelines for Regulating Digital Platforms for Information as a Public Good, Paris, 2023

Date: 28 June 2023

About the Centre for Media Transition

The Centre (CMT) was established in 2017 as an applied research unit based at the University of Technology Sydney (UTS). It is an interdisciplinary initiative of the Faculty of Arts and Social Sciences and the Faculty of Law, sitting at the intersection of media, journalism, technology, ethics, regulation, and business.

Working with industry, academia, government and others, the CMT aims to understand media transition and digital disruption, with a view to recommending legal reform and other measures that promote the public interest. In addition, the CMT aims to assist news media to adapt for a digital environment, including by identifying potentially sustainable business models, develop suitable ethical and regulatory frameworks for a fast-changing digital ecosystem, foster quality journalism, and develop a diverse media environment that embraces local/regional, international and transnational issues and debate.

This document is an edited compilation of responses submitted via an online form on 27 June 2023 in response to UNESCO Guidelines 3.0

This submission was prepared by:

- Dr Michael Davis, Research Fellow, Centre for Media Transition
- Dr Sacha Molitorisz, Senior Lecturer, Faculty of Law
- Dr George Tian, Senior Lecturer, Faculty of Law

CONTACT

Centre for Media Transition
University of Technology, Sydney
PO Box 123, Broadway NSW 2007

cmt@uts.edu.au
+61 2 9514 9669

cmt.uts.edu.au

QUESTION 1: Should we look to a principle-based document or a document that also offers detailed regulatory guidance for digital platforms?

A high-level principle-based document is best as a foundation. This should then be supplemented by detailed regulatory guidance, in two forms: elaboration of those principles within this document; and, where necessary, further regulatory instruments, including domestic instruments. For reasons of proportionality and the nuanced nature of freedoms and rights, the elaboration of these principles must not go too far by giving excessive and undue protection to free speech and access to information, including by precluding reasonable domestic law and regulation. (See response to question 4, below).

A principle-based document can adapt to cover emerging technologies. It can also enforce fundamental standards even as it allows flexibility for different jurisdictions and countries, including with their domestic laws and regulations. Five principles are identified in the Guidelines 3.0, committing digital platforms to human rights, transparency and accountability, which are fundamental precepts of fair and just societies. which are then supplemented with considerable detail. This is a sound approach.

One key point here concerns the nature of freedoms and rights, specifically in relation to proportionality (eg, § 29(e), p. 9). As the Guidelines note, information is a public good, and UNESCO has a mandate 'to protect and promote freedom of expression, access to information, and safety of journalists'. However, a fundamental point for any proper understanding of freedoms and rights is that none is absolute. Each must be properly respected and protected by the law, and the law must explicitly recognise that each freedom and right must be respected and protected in balance. This balance must take account of other rights and freedoms (right to life and liberty, right to privacy, right to security, etc) as well as the rights and freedoms of others. Rights and freedoms thus interact in complicated ways. Often the right to free speech and the right to access information coincide. Sometimes they are at odds. If one person is very strident in expressing an opinion online, that may have a chilling effect on another. One person's speech may thus impinge on another's, which will in turn dampen access to the information contained in the second person's speech.

In Australia, the importance of limiting free speech was highlighted by the eSafety Commissioner, Julie Inman Grant, in June 2023[1]. 'We are seeing a worrying surge in hate online,' Inman Grant said. Under new laws, the eSafety Commissioner has asked Twitter to explain what it is doing to tackle hate speech [2]. The law requires a response within 28 days, or else the company faces penalties of nearly AU$700,000 a day for continuing breaches. Also in June 2023, the Australian government released a draft law to tackle misinformation [3]. These domestic laws play an important part in balancing free speech and access to information against other interests, and harmonise with the five draft principles in the Guidelines. The aim of promoting free speech and access to information is good and vital. However, there are compelling reasons for firm limits on that speech to guard against hate speech, misinformation and other corrosive speech. What's more, the non-absolute nature of all freedoms and rights means that the precise limits of the right to free speech and the right to access information are impossible to define a priori. Those limits require an in-built degree of flexibility. General high-level principles provide the best foundation, supplemented by further details in international and domestic law.

---

[1] https://www.theguardian.com/australia-news/2023/jun/22/australias-esafety-umpire-issues-legal-warning-to-twitter-amid-rise-in-online-hate

[2] https://www.smh.com.au/national/australia-takes-on-elon-musk-over-spike-in-online-hate-on-twitter-20230621-p5dibk.html

[3] https://minister.infrastructure.gov.au/rowland/media-release/consultation-opens-new-laws-tackle-online-misinformation-and-disinformation

QUESTION 2: What types of digital platforms should be included in the scope of the Guidelines?

We consider that an assessment of scope should be based on risk. Some smaller platforms may be high risk. However, size and market share should be considered as an input into risk assessment.

It is important to maintain flexibility in the scope of the Guidelines so that they continue to be applicable even as technology and the communications environment undergo change. For this reason it should be clear that the types of platform listed on p.32 are only examples and are not to be taken to be an exhaustive list. The critical principle is that they are platforms that allow users to disseminate content to the wider public.

That said, this criterion is not particularly useful. For example, do app stores disseminate content to the wider public? Instead, because the guidelines are intended to set the legitimate scope for regulation of digital platforms, perhaps the scope of the Guidelines could instead focus on regulation rather than platforms, e.g. 'all regulation which regulates content on digital platforms, whether directly or indirectly'. Indirect regulation of content would, for example, address app stores' removal of apps which violate content rules in their terms of service.

§ 11 could continue to exclude particular services from the scope of the guidelines if necessary.

We do not favour limiting the scope to user-to-user platforms, as these do not exhaust the type of platforms which will be subject to regulation of content. We note here that news aggregators and other information-dissemination services are not explicitly included, though they may have rules that apply to content or take otherwise take measures to moderate content and may be subject to regulation on that basis. We also note that 'e-commerce platforms' could also be included as they often have user guidelines or policies restricting the sale of products that display or include particular types of content and may therefore be subject to content regulation, for example, misinformation, hate speech, or advertising regulation.

In addition, although this guideline focuses on human rights aspects of the digital platform, it is also noteworthy that there are some other important and relevant aspects of digital platforms, such as online sales and social media counterfeiting.[1]

For example, according to reports, sales totalling approximately $560 billion were generated through social media platforms in 2020. The significance of social media extends beyond its role in providing representation, broad reach, and accessibility. It also encompasses the power to influence potential buyers.[2] In a separate study, it was found that more than 20% of fashion-related posts on Instagram promoted counterfeit products. Additionally, research has indicated that approximately 15% of hashtags associated with luxury brands were used by accounts engaged in counterfeiting activities.[2]

---

[1] United States Trade Representative. '2022 Notorious Markets List' (Final, January 2023). Available at: https://ustr.gov/sites/default/files/2023-01/2022%20Notorious%20Markets%20List%20(final).pdf (Accessed 27 June 2023).

[2] 'Social Media Counterfeiting: The Modern-Day Nemesis of Trade Mark Proprietors' (Exc-Lon IP, 2023) at https://excelonip.com/social-media-counterfeiting-the-modern-day-nemesis-of-trade-mark-proprietors/ (Accessed 27 June 2023)

QUESTION 3: How should a multi-stakeholder approach in a regulatory process look?

A multi-stakeholder approach in a regulatory process is crucial to ensure a comprehensive and inclusive decision-making process. It goes beyond a mere nominal involvement of stakeholders and seeks to engage all relevant parties actively. However, it is important to distinguish between the development and operation of regulations, as different stakeholders may play varying roles at each stage.

A multi-stakeholder approach in a regulatory process involves engaging and involving various stakeholders who have an interest or expertise in the subject matter being regulated. It is a collaborative and inclusive approach incorporating diverse perspectives, knowledge, and interests into the regulatory decision-making process. It is also referred to as “multi-stakeholder process”. [1] Put simply, for the purposes of this guideline, we define Multi-Stakeholder Processes as processes that convene all stakeholder groups, which together seek solutions and develop strategies around specific objectives of digital platforms regulation. More specifically, the application of the multi-stakeholder approach in a regulatory process may include the following steps:

1. **Identifying all relevant stakeholders**: Identify all relevant stakeholders who are either knowledgeable about it or are directly or indirectly impacted by the regulation. Affected communities, experts, non-governmental organisations (NGOs), government agencies, business representatives, and industry representatives can all be included in this.
2. **Encouraging inclusive participation:** Encouraging genuine stakeholder participation in all aspects of the regulatory process. Public hearings, seminars, focus groups, roundtable talks, and written submissions are examples of this. Ensure that participants can access the knowledge, tools, and resources to contribute effectively.
3. **Ensuring balanced representation:** Aim for a balanced/proportionate representation of stakeholders to guarantee a variety of viewpoints. Consider the interests of diverse stakeholders, such as the industry, civil society, affected communities, and experts. Avoid being dominated by any group or interest.
4. **Ensuring transparency and accountability**: Maintain transparency by disseminating information regarding the regulatory process, including timelines, objectives, and criteria for making decisions. Communicate the duties and responsibilities of stakeholders and how their inputs will be considered and incorporated. Hold regulators accountable for considering stakeholder feedback and explaining decision justifications.
5. **Facilitating collaborative decision-making and dialogue:** Foster a collaborative atmosphere where stakeholders can engage in constructive dialogue, exchange ideas, and search for common ground. Encourage frank dialogue, attentiveness, and mutual regard. Facilitate opportunities for stakeholders to provide feedback, suggest alternatives, and collaborate on the development of innovative solutions.
6. **Allowing for the iterative process:** Recognise that the regulatory process may consist of multiple phases and iterations. Continuously solicit feedback from stakeholders at various process phases, allowing for adjustments based on new information or insights. Promote continuous development and learning.
7. **Coordinating implementation efforts among stakeholders:** Ensure coordination and collaboration among stakeholders during regulatory decision implementation. Encourage cooperation, the sharing of best practices, and ongoing communication in order to address any potential obstacles or unintended outcomes.

8. **Conducting periodic evaluation and review:** Establish evaluation mechanisms for the efficacy of regulatory measures and their impact on stakeholders. Conduct regular evaluations to determine whether the regulations are achieving their intended objectives and whether modifications are required based on changing conditions or stakeholder feedback.

In short, by incorporating these elements, a multi-stakeholder approach in a regulatory process can enhance inclusiveness, transparency, collaboration, legitimacy, and accountability within the regulatory process.

On the other hand, we need to pay attention to the limits of the 'multi-stakeholder approach' or 'multi-stakeholder process'. Firstly, this approach has sometimes been criticised as 'not being applicable in countries that do not have the conditions for democratic dialogue'. Secondly, the Multi-Stakeholders Process (MSPs) are often donor-driven rather than locally owned.[2][3]

- **More specifically, should the Guidelines further address the role of the media and journalism? If yes, how so?**

Yes. The role of media could be clarified in more detail (currently only addressed in section 37, which reads: 'Media and fact-checking organizations have a role in promoting the enjoyment of freedom of expression, the right to access information, and other human rights, while performing their watch-dog function'). The guideline should address the role of the media and journalism by:

- Protecting freedom of expression and press freedom online.
- Promoting media literacy and fact-checking to combat misinformation.
- Ensuring the safety and security of journalists working online.
- Encouraging ethical standards and professional conduct in reporting.
- Fostering collaboration and accountability among media organizations, governments, and stakeholders.
- Supporting independent media and fostering media pluralism.

Like the state, media also have a critical role in promoting a 'favourable environment for participation in public debate that enables freedom of expression and the right of access to information' (§ 73), including online, and in counteracting misinformation. Moreover, regulation of digital platform content will necessarily overlap in important ways with media activities and regulation.

- **More specifically, should the Guidelines provide further information about the role of Civil Society? If yes, how so?**

Yes. The role of civil society is reasonably clear (sections 34-39 of the draft guideline). The Guideline should address the role of civil society by:

- Ensuring consultation and participation of civil society organisations in decision-making processes.
- Promoting transparency and accountability in internet governance, with civil society monitoring and advocating for these principles.
- Recognizing civil society's role in safeguarding human rights and digital rights online.
- Supporting capacity building and awareness initiatives for civil society organizations.
- Ensuring the inclusion and protection of vulnerable groups through the involvement of civil society organisations.

- **More specifically, should the Guidelines provide further information about the role of the governments? If yes, how so?**

Yes. The role of government is reasonable clear in sections 27-29 of the Guideline. The Guideline should address the role of the government by emphasising the following:

- Clear legal frameworks that protect users' rights and ensure transparency and accountability.
- Proportional and targeted regulation to address issues like cybersecurity and data protection.
- International cooperation to tackle cross-border challenges and harmonise legal frameworks.
- Inclusive and multistakeholder internet governance approaches.
- Efforts to bridge the digital divide and ensure universal access to the internet.
- Respect for universal human rights online and government transparency and accountability.

---

[1] See '2.11. What is a multi-stakeholder process?' in Global Partnership for the Prevention of Armed Conflict (GPPAC), GPPAC MSP Manual: A Practical Guide for Multi-Stakeholder Partnerships (February 2018), available at: https://gppac.net/files/2018-11/GPPAC%20MSPmanual_Interactive%20version_febr2018.pdf (last visited June 14, 2023).

[2] Nicolas Faysse, 'Troubles on the Way: An Analysis of the Challenges Faced by Multi-Stakeholder Platforms', Natural Resources Forum, 30 (2006), 219–29;

[3] Julia Roloff, 'A Life Cycle Model of Multi-Stakeholder Networks', Business Ethics: A European Review, 17 (2008), 311–25, by GPPAC MSP Manual: A Practical Guide for Multi-Stakeholder Partnerships (February 2018), available at: https://gppac.net/files/2018-11/GPPAC%20MSPmanual_Interactive%20version_febr2018.pdf (last visited June 14, 2023) at 13.

QUESTION 4: During the second open consultation of the Guidelines, the content management section received a significant amount of comments. In those, there was agreement about:

- **The necessity to ensure that the Guidelines aim to safeguard freedom of expression and information in the context of any digital platform regulatory process, regardless of the regulatory goal.**
- **The importance of focusing the Guidelines on the structures and processes to moderate and curate content - and not in individual pieces of content.**
- **The importance to refer to legitimate restrictions of content as stated in internal instruments of human rights.**

Does version 3.0 suitably incorporate these points? Are we missing something else about content management?

Version 3.0 of the guidelines makes the importance of these three principles clear. However, they sometimes do so too strongly or at the expense of other principles.

Most critically, § 29(i) says that states should 'refrain from imposing a general monitoring obligation or a general obligation for digital platforms to take proactive measures in relation to content considered illegal in a specific jurisdiction or to [content that can be restricted under international human rights law]. However, platforms would be expected to take steps to restrict known child sex abuse material and live terrorist attacks.' In our view this paragraph is much too prescriptive and would seem to rule out existing legislation and other regulation designed to address digital harms such as disinformation and hate speech (e.g. the Digital Services Act and EU Disinformation Code, the Australian Code of Practice on Disinformation and Misinformation and the Online Safety Act in Australia, and the proposed Online Harms Act in the UK) as well as scams and misleading advertising on digital platforms. Given the scale and speed at which content moderation is and must be undertaken (generally requiring automation), platforms may be expected to take steps to remove or restrict the dissemination of a range of material other than known child sex abuse material and live terrorist attacks, even if a general monitoring obligation is ruled out. Of course, protection of freedom of expression is essential here, and can be preserved by, for example, (1) ensuring that states have no involvement in content-moderation decisions; (2) ensuring that platforms' content-moderation decisions follow transparent, independent and accountable rules and processes; (3) ensuring content moderation is proportional to the risk of harm; (4) ensuring user privacy (including by implementing appropriate data collection and retention rules).

Also in this context, § 21 states that 'Regulation should focus on the systems and processes used by platforms to moderate and curate content, rather than seeking to judge the appropriateness or legality of single pieces of content.' It then states that content moderation 'should follow due process and be open to review by an impartial and independent judicial body.' 'Judicial' is perhaps too strong a term here, and 'an impartial and independent body' would be adequate when combined with the complaints process set out in §§ 49(d), 90(j), 109 and elsewhere.

§ 21 also states that review should follow the three-part test on legitimate restrictions to freedom of expression as laid out in article 19(3) of the ICCPR and the prohibition of advocacy to hatred that constitutes incitement against discrimination, hostility or violence as laid out in article 20(2) of the ICCPR; including the six-point threshold test for defining such content outlined in the Rabat Plan of Action.

In our view this is too prescriptive and would be better expressed with the phrasing that review should be 'guided by' the relevant articles of the ICCPR, and in any case 'should take no action that contravenes them'. This is because the three-part test is very broad and high-level, and review bodies should not be restricted from following other decision frameworks as long as they do not contravene the articles of the ICCPR.

QUESTION 5: Future Proofing. How can we ensure that the guidelines are flexible enough to adapt to new and emerging technologies?

The guidelines can be effectively future-proofed with the adoption of high-level general principles, as proposed in our answer to question 1. Instead of seeking to identify all the minutiae in an attempt to cover every future innovation, regulation ought to aim to promote fairness, egalitarianism and equity. Indeed, the high-level principles contained in the Guidelines 3.0 are well-suited to adapt for new technologies and innovations. These principles commit platforms to human rights, transparency and accountability, which underpin fair and just societies. At the very least, the implementation of these high-level principles would constitute an excellent start. At best, they could achieve more, when supplemented with further detail in the Guidelines, as well as further laws and regulatory instruments.

The implementation of high-level general principles recognises that legislating and regulating at the micro level is challenging, if not impossible, in a digital age. Rather, sweeping provisions are required. In some cases, the law already does this, with some success. In 2022, prohibitions in the Australian Consumer Law against 'misleading and deceptive' conduct led to one of Australia's biggest privacy wins, after the Federal Court ordered Google to pay $60m in penalties for making misleading representations about the collection of location data [1]. As one of us wrote in 2020 in relation to privacy, both domestic and international law in a range of areas ought to outlaw misleading and deceptive conduct, and also ought to buttress consent, mandate fairness, outlaw coercion and mandate transparency, as well as working to support and enforce human rights [2]. The adoption of the principles proposed in the Guidelines would be a welcome step in this direction.

Another point here is that any protections of free speech and access to information must be developed with an holistic outlook. This means that free speech and access to information cannot be considered in isolation, but rather must be considered in the context of our digitally inflected and tech-saturated lives. The challenges of privacy, misinformation, hate speech, jurisdiction, trade and more all overlap, to some extent, with the issues raised by free speech and access to information. Regulating for one area needs to take account of other areas, which is yet another reason for supporting high-level general principles (including human rights) that apply for all these challenges, and then supplementing these high-level principles with further detail, including in the form of domestic laws and regulation that align with the high-level principles.

---

[1] https://www.accc.gov.au/media-release/google-llc-to-pay-60-million-for-misleading-representations

[2] https://www.crikey.com.au/2020/04/24/weekend-read-net-privacy-surveillance/

QUESTION 6: Groups have been consulted to introduce a gender and intersectional approach in the Guidelines. Are there specific elements that should be considered to ensure the guidelines are sensitive to gender and intersectionality?

Above, we noted that all freedoms and rights are necessarily limited. No freedom or right is absolute, and nor should it be. The notion of absolute freedom is not only an impossible goal, it is anathema to the notion of a society founded on respect and civility. This becomes particularly apparent when we turn our focus onto race, age, sexual orientation, gender identity, Indigeneity, other cultural and linguistic diversity, disability and socio-economic status. These aspects of identity have been named by international human rights instruments as the most common markers of privilege or marginalisation. Accordingly, these aspects of identity have a significant impact on a person's experience of being in the world, both physically and online. While nearly one in five Australians has experienced online hate, First Nations and LGBTQI+ people are particularly vulnerable, experiencing online hate at more than twice the rate of the national average [1].

For marginalised groups, the theory and reality of free speech diverge. In theory, free speech enables each member of a society to voice their opinions and concerns. This line of thinking was particularly prominent in the early years of the world wide web, when digital media was regarded as a powerful democratising force. The reality has proved somewhat different, however. While in some cases the democratising potential of the internet has led to profound social benefits, in other cases it has spawned misinformation, hate speech, vitriolic abuse and real-world violence. In reality, members of marginalised groups are served a much smaller slice of the free speech pie. For one thing, the vulnerability of marginalised groups to hate speech creates a chilling effect on their speech. It is worth noting, however, that privilege and marginalisation are not binary. A person can be privileged in some aspects of their identity, and yet marginalised in others. This makes the interplay of free speech with gender and intersectionality all the more nuanced.

With this in mind, we support the 'specific measures to counter online gender-based violence' as outlined in § 60 (pp. 18-19). However, as well as gender identity, the Guidelines should also encompass race, age, sexual orientation, Indigeneity, other cultural and linguistic diversity, disability and socio-economic status. As noted, these aspects of identity are markers of privilege or marginalisation, and they are irrevocably bound up in notions of free speech. The Guidelines ought to make explicit that free speech does not extend to speech that is violent towards or that vilifies anyone, including specifically any member of these potentially vulnerable groups. More generally, the Guidelines ought to spell out in clear terms areas of legitimate and justifiable restriction to free speech under international law regarding aspects of identity.

---

[1] eg, https://www.esafety.gov.au/sites/default/files/2021-08/LGBTQI%2B%20cyber%20abuse%20resource%20development%20-%20Report.pdf

QUESTION 7: Beyond the previous questions, please add any general or specific comments to the different sections or paragraphs of the guidelines.

General Comments:

Further to our comments above, the guidelines note that it is up to states to define legitimate restrictions on content (within international law). However, it gives platforms no guidance on what to do in cases of state overreach, except in § 90(g), which is just a transparency requirement.

Comments on specific paragraphs:

§ 25: The guidelines could be more explicit in encouraging states not merely to safeguard access to information, but also to promote access to high-quality news and information that enables citizens to exercise their freedom of expression and other rights to their full potential. This would help reinforce § 29(a) on strengthening civic space and § 73 on creating a favourable online environment for participation in public debate that enables freedom of expression and the right to access information.

§ 60: Further to our comments in answer to question 6, the measures to counter online gender-based violence should be generalised to cover other kinds of discrimination, vilification, hatred or violence against particular groups. E.g. community feedback mechanisms designed to address bias in generative AI would also be useful in addressing these other kinds of harm.

§ 73: This paragraph notes that there is a responsibility to create a 'favourable online environment for participation in public debate that enables freedom of expression and the right to access information'. But there is no detail on what is meant by a favourable online environment. Perhaps this could read 'favourable online environment for equitable, free and informed participation in public debate for all, including freedom of expression and the right to access quality news and information, without being subject to abuse, hatred or vilification'.

§ 75: The following could be added to this paragraph: 'Platforms operating in multi-language environments should ensure that content moderation is available across all languages and follows these guidelines equally across all languages.' Should cross-reference § 101.

§ 79: This paragraph should provide more detail on measures to counter mis- and disinformation, and set these out explicitly as responsibilities for platforms to take action, rather than measures that 'could be' taken. It should also specify the need for risk assessment and proportionality in application of measures rather than the identification of content (as in § 78).

§ 83: The specification of user engagement seems unnecessary. The paragraph could read, 'when any aspect of the design of the platform's services could result in the amplification of [content that can be restricted...]'.

§ 85: This paragraph seems overly restrictive. Making available more than one content-curation mechanism is a possible solution to some problems of content moderation but that does not mean it should be enforced. This paragraph could instead read, e.g. 'provide users with options to adjust content recommendation and moderation systems'. Platforms could also be required to give users options to manage collection of personal data and the extent to which content recommenders respond to explicit or inferred preferences.

§ 86: Mandatory notification should apply to removal of content but it would seem unnecessary to notify users when content is algorithmically demoted.

§ 95: This paragraph seems unnecessarily focused on reporting and transparency given it is under Principle 4. Instead it should be phrased as a responsibility for platforms to help users navigate the online environment and to promote a favourable online environment for civil public discourse and debate.

§ 104: This paragraph should be explicit about the specific consent required for children's access and use, including data collection, account setup, etc.