Response from IT for Change to the Draft 2.0

UNESCO Guidelines For Regulating Digital Platforms:

A Multistakeholder Approach to Safeguarding Freedom of Expression and Access to Information

March 2023

IT for Change

Table of Contents

Response from IT for Change[1] to the Draft 2.0

UNESCO Guidelines for Regulating Digital Platforms:

A Multistakeholder Approach to Safeguarding Freedom of Expression and Access to Information

March 2023

General Comments on the Overall Guidelines:

The second draft of the UNESCO Guidelines is a considerable improvement from the earlier draft in terms of recognizing the impact of platform content moderation and curation functions on:

- freedom of speech and access to information,
- regional and cultural differences affecting platform regulation,
- a stronger reference to the UN Guiding Principles on Business and Human Rights to hold platforms accountable, and
- emphasis on due process including judicial review echoing the three part-test on legitimate restrictions laid out in Article 19 of International Covenant on Civil and Political Rights (ICCPR).

We appreciate the Guidelines for recognizing and incorporating the following:

- Detailing the criteria for composition and scope of powers of independent regulators.
- Requiring human rights compliance and transparency in respect of both content moderation and content curation functions performed by digital platforms.
- Special attention to the employment status and conditions of work of human moderators as a factor affecting the quality of content moderation efforts.
- Significance of digital platforms conducting periodic human rights due diligence of their operations.
- Specific attention to gendered disinformation and online gender-based violence.

[1] For any clarification or queries, we can be reached at itfc@itforchange.net

However, we believe that the Guidelines are still not sufficiently attendant to:

a) Challenges of platform regulation – especially in the Global South

b) Architecture of corporate impunity that platforms benefit from and

c) Need to address communication governance through a supra-liberal framework.

The Guidelines' lack of differentiation between different types of digital platforms in terms of their functionality, user base, and size will hinder effective and proportionate responses to the distinct regulatory challenges for securing freedom of expression and information as a public good. We recommend that the Guidelines provide greater clarity on the role and mode of engagement of inter-governmental organizations and the role of civil society, media organizations, academia, and other constituencies. This is necessary to ensure that policy processes are democratic, inclusive, and accountable. We reiterate our earlier recommendation to push for interoperability and data portability in order to minimize internet fragmentation and promote diversity of information and viewpoints. Finally, we suggest that the Guidelines tighten its language in several places to ensure that terms are not left to subjective interpretation, which could defeat the purpose of developing a common framework for pinning accountability on the platforms and the government.

1. On the Objective of the Guidelines

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 10: "The Guidelines will inform regulatory processes under development or review for digital platforms, in a manner that is consistent with international human rights standards. Such regulatory processes should be led through an open, transparent, multistakeholder, and evidence-based manner. a. The scope of these Guidelines includes digital platforms that allow users to disseminate content to the wider public, including social media networks, messaging apps, search engines, app stores, and content- sharing platforms. Bodies in the regulatory system should define which digital platform services are in scope, and also identify the platforms by their size, reach, and the services they	After "a. The scope…content-sharing platforms", insert: "a….***While minimum safety requirements must be adhered to by all platforms, specific regulatory obligations in the Guidelines apply only to social media networks with the largest size and reach. For other types of digital platforms, the regulatory system should identify and define them by their size, reach, and the services they provide, as well as features such as whether they are for-profit or non-profit, and if they are centrally managed or if they are federated or distributed platforms. Regulations specific and proportionate to the nature and size of different***

provide, as well as features such as whether they are for-profit or non-profit, and if they are centrally managed or if they are federated or distributed platforms."	***platforms should be developed by the regulatory system after conducting a risk assessment.*** *"*

Justification/References

Para 10. a disregards the fact that digital platforms that allow users to disseminate content to the wider public vary in terms of their functionality, level of access or control over the information being transmitted, number and type of users (such as children), size of the platforms and scale of their operations.[2] These differences pose distinct regulatory considerations in the pursuit of safeguarding freedom of expression and securing information as a public good. For example, para 10.a mentions messaging apps as an example of digital platforms that fall under the scope of the Guidelines. Many of the messaging apps secure conversations between users using end-to-end encryption which makes it difficult to impossible to moderate or remove illegal content as compared to a public post on social media platforms, without breaking the encryption and undermining the privacy of users. Hence, there is a need for a different regulatory approach to deal with illegal and harmful content on end-to-end encrypted messaging apps as compared to social media platforms.

Further, the transparency obligations outlined in the Guidelines may be excessive or insufficient for some digital platforms, depending on their functionality, level of access or control over information, and size. For instance, the kind of information that a social media network should make transparent is different from what a search engine or an app store should in order to assess their impact on freedom of expression and information. Extensive transparency obligations will also amount to an onerous regulatory burden for micro and small digital platforms, stifling their operation and growth.

Additionally, platforms that have a potential for virality and amplification of content should be subjected to a higher standard of duty of care and liability as compared to other platforms. Due to these differences among digital platforms, a one-size-fits-for-all approach to regulation will fail to address the platform-specific risks to freedom of expression and information, while also potentially burdening many platforms with regulatory compliances that are in a mismatch with their functionality, user base, size and reach. Hence, a mere requirement that regulatory bodies should define and identify platforms by their size, reach, services, features, etc is not sufficient without also a requirement to tailor regulation according to these features. The Digital Services Act of the European Union differentiates between conduit, caching, and hosting services, in addition to differentiation based on size.[3] In Australia, the Online Safety Act

[2] Kumar, R., Thanugonda, K., & Guha, D. (2022). *A Framework for Intermediary Classification in India.* The Quantum Hub. https://thequantumhub.com/wp-content/uploads/2022/12/A-Framework-for-Intermediary-Classification-TQH-Dec-2022-Final.pdf
[3] European Union. (n.d.). *Digital Services Act. Article 2(f).* https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en

identifies eight types of digital services and calls for establishment of industry codes or standards for each type.[4]

In its current form, the specific obligations prescribed in the Guidelines seem to pertain mostly to the regulation of large social media companies like Facebook, Instagram, Twitter, etc. that host large-scale user-generated content, and perform moderation and curation functions. We recommend that the Guidelines follow the earlier version and apply the specific regulatory obligations only to platforms whose services have the largest size and reach while requiring all other digital platforms to adhere to minimum safety requirements and basic human rights standards and comply with platform-specific regulations developed by national regulatory bodies.

[4] Australian Government. (2021). *Online Safety Act, 2021. Section 135.* https://www.legislation.gov.au/Details/C2021A00076

2. On the States' Duties to Respect, Protect, and Fulfill Human Rights

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 27.f: "Specifically, States should: f. Refrain from imposing a general monitoring obligation or a general obligation for digital platforms to take proactive measures to relation to illegal content. Digital platforms should not be held liable when they act in good faith and with due diligence, carry out voluntary investigations, or take other measures aimed at detecting, identifying, and removing or disabling access to illegal content."	"27. Specifically, States should: ***f. Require digital platforms to adopt measures to proactively detect and remove certain categories of illegal content such as child pornography, terrorism-related content, content promoting, depicting or inciting extreme violence, drug misuse, and violent content. However, platforms should not be held liable for the failure to remove such types of illegal content, if they can demonstrate the adoption of reasonable measures and unless they failed to remove such types of content despite it being brought to notice. Digital platforms should also not be held liable when they act in good faith and with due diligence, and carry out voluntary investigations, or take other measures aimed at detecting, identifying, and removing or disabling access to other types of illegal content and content that risks significant harm to democracy and the enjoyment of human rights. States should incorporate safeguards and processes in consonance with international human rights law to ensure that such measures, when adopted by platforms, do not violate freedom of expression and access to information of users.***"
Justification/References Not asking digital platforms to take proactive measures in relation to illegal content ignores the technical capability and actual practice of social media platforms actively moderating various types of content. It also directly plays into the dumb conduit metaphor that platform companies are so eager to maintain in order to avoid social and legal obligations with respect to the content on their platform. In fact, content moderation is an enormous part of running a social media platform and is the very commodity they	

offer.[5] Further, platforms are in a better position to quickly detect and remove illegal content such as child sexual exploitation, terrorist and extreme content from their platform than State-based grievance mechanisms. Hence, it is only prudent that States recognize this moderation capability of platforms and require them to adopt measures to remove particularly grievous and harmful illegal content such as child pornography, terrorism-related content, content promoting, depicting or inciting extreme violence, drug misuse, and violent content,[6] the dissemination of which is enabled by platform affordances in the first place. At the same time, acknowledging technological limitations for 100% precision in identifying illegal content, platforms should not be held liable for each and every failure to remove content if they can demonstrate that they have acted with due diligence and taken reasonable measures within their capabilities to remove such types of content and that they have not failed to remove them despite it being brought to their notice.

It is important to note that the regulatory mandate to conduct proactive monitoring should extend only to the above-mentioned categories of illegal content, as their determination of illegality is less prone to contextual variance, unlike certain other types of illegal content such as defamation, content threatening public order, sovereignty or integrity of the State, etc. Therefore, States should not mandate platforms to proactively moderate all types of illegal content and content that risks significant harm to democracy. Today, very large social media platforms engage in active moderation of different types of illegal and harmful content in order to ensure the safety of users, regardless of regulatory mandate.[7] In many situations moderation practices result in the unfair removal of legitimate speech and suppression of certain voices, as taking a decision on these types of content requires careful consideration of cultural and contextual nuances.[8] Therefore, States should incorporate safeguards and processes in consonance with international human rights law to ensure that when platforms engage in moderation of these types of content, they do not violate the freedom of expression and access to information of users.

[5] Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media.* p. 5.
[6] Australian Government. (2021). *Online Safety Act, 2021. Section 106*. https://www.legislation.gov.au/Details/C2021A00076
[7] Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media.* Ch. 1.
[8] Browne, M. (2017, August 22). *YouTube Removes Videos Showing Atrocities in Syria.* New York Times. https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html; Southerton, C., Marshall, D., Aggleton, P., Rasmussen, M. L., & Cover, R. (2021). Restricted modes: Social media, content classification and LGBTQ sexual citizenship. *New Media & Society*, 23(5), 920–938. https://doi.org/10.1177/1461444820904362; Kleinman, Z. (2016, September 9*). Fury over Facebook 'Napalm girl' censorship.* BBC. https://www.bbc.com/news/technology-37318031

3. The Responsibilities of Digital Platforms to Respect Human Rights

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 27.g: "Specifically, States should: g. Refrain from subjecting staff of digital platforms to criminal penalties for an alleged or potential breach of regulations in relation to their work on content moderation and curation, as this may have a chilling effect on freedom of expression."	After paragraph 27.g, add the following paragraph: ***"27.X. Develop a liability framework to hold digital content platforms and those directly in charge of and responsible for the conduct of platform businesses for enabling or facilitating harms including online gender-based violence, disinformation, hate speech, incitement to violence and for any systematic or deliberate failure to take steps to prevent or mitigate the harm, despite actual knowledge of it."***

Justification/References

Documented evidence[9] and several whistleblower revelations[10] demonstrate that digital content platforms, especially social media, have been complicit and are actively facilitating the promotion of disinformation campaigns, hate propaganda, abuse, and harassment against individuals and groups, and even subverting democratic procedures in pursuance of their business model that is predicated on the extractive logic of the attention economy. These companies often benefit from the architecture of corporate impunity and the archaic nature of laws that treat platforms like social media as dumb conduits of speech.[11] This enables them to avoid accountability for the harms and human rights violations resulting from their operation.

There is a need to break this regime of permissiveness and impunity by instituting legal frameworks and mechanisms to pin down the role of digital content platforms in perpetuating harm and holding them liable for it. The far-reaching role and functions of these platforms, especially large social media platforms, necessitate a relook at the legal assumption of 'platform neutrality'. At the same time, it is also

[9] United Nations. (2018, September 28). *Report of the Detailed Findings of the Independent International Fact-Finding Mission in Myanmar.* Agenda Item 4. 39th session of the Human Rights Council. Retrieved from https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_CRP.2.pdf; Amnesty International. (2018). *Toxic Twitter—A Toxic Place for Women.* https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1/; Manuvie, R., et al. (2022). *Preachers of Hate: Documenting Hate Speech on Facebook India.* https://pure.rug.nl/ws/portalfiles/portal/236667786/V4._Preachers_of_Hate_2022.pdf

[10] Mac, R., & Kang, C. (2021, October 3). *Whistle-Blower Says Facebook 'Chooses Profits Over Safety'*. New York Times. https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html; Forno, R. (2022, September 1). *Did Twitter ignore basic security measures? A cybersecurity expert explains a whistleblower's claims.* The Conversation. https://theconversation.com/did-twitter-ignore-basic

[11] Sinha, A. (2020). Beyond Public Squares, Dumb Conduits, and Gatekeepers: The Need for a New Legal Metaphor for Social Media. In *A Digital New Deal: Visions of Justice in a Post-Covid World*. IT for Change. https://projects.itforchange.net/digital-new-deal/2020/11/01/beyond-public-squares-dumb-conduits-and-gatekeepers-the-need-for-a-new-legal-metaphor-for-social-media/

important to ensure that the liability provisions do not cause these platforms to engage in overzealous censorship of content and to be censorship proxies for the government,[12] thereby endangering the freedom of expression of users. To avoid this, rather than holding the digital content platforms liable for each and every instance of harmful content on its platform, they should be held liable only for deliberate or intentional failure to remove the harmful content despite having knowledge of it, and for systematic or deliberate failure to take steps to prevent or mitigate harm from their operations, particularly from the operation of algorithms that amplify harmful content.

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 28. c and d: "28 c. Platforms empower users to understand and make informed decisions about the digital services they use, including helping them to assess the information on the platform. 28 d. Platforms are accountable to relevant stakeholders, to users, the public, and the regulatory system in implementing their terms of service and content policies, including giving users rights of redress against content-related decisions."	After paragraph 28. c and before 28.d, add the following paragraph: "28. c. Platforms empower users to understand and make informed decisions about the digital services they use, including helping them to assess the information on the platform. ***28. X. Platforms provide access to remedies through grievance mechanisms that are accessible directly to individuals and communities who may be adversely impacted by decisions taken by the platform thereby respecting users' right to seek redressal including in relation to content-related decisions made by the platform.*** 28.d. Platforms are accountable to relevant stakeholders, to users, the public, and the regulatory system in implementing their terms of service and content policies."
Justification/References It is essential to have a specific principle to access remedies within Paragraph 28 of the Guidelines which highlight and place onus/responsibility on platforms for their operations. The importance of an effective	

[12] Kreimer, S. F. (2006). *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link* (University of Pennsylvania, Faculty Scholarship Paper 127). http://scholarship.law.upenn.edu/faculty_scholarship/127

operational-level grievance mechanism[13] was discussed in detail as a core foundational principle in the UN Guiding Principles on Business and Human Rights. These mechanisms should be accessible directly to the individuals and communities who may be adversely impacted by harms arising out of the policies and practices of a digital content platform.[14] As discussed within the UN Report on Human Rights and Transnational Corporations and Other Business Enterprises, merely providing access to remedial mechanisms will not suffice as there should be an effective remedy. This means that every platform should incorporate both procedural and substantive aspects for an effective remedy.[15]

Operational-level grievance mechanisms perform two key functions with regard to the responsibility of business enterprises for respecting human rights. First, they support the identification of adverse human rights impacts as a part of an enterprise's ongoing human rights due diligence which is repeatedly emphasized within Paragraph 28e and Principle 5 of the Guidelines as a key principle which places responsibility on the digital platform. They do so not only by providing a channel to raise concerns but also by gathering data to understand trends/patterns in complaints within business enterprises to identify systemic problems and address them.

It is important to highlight that access to a remedy not only lies on an individual level but also on a community level, as the centrality of rights holders is an essential component in access to effective remedies as discussed in Principle 31 of the Guiding Principles. Firstly, all remedial mechanisms and remedies must be responsive to the diverse experiences and expectations of the rights holders. Secondly, to test the effectiveness of such a remedial system, it has to be determined whether it is accessible, affordable, adequate, and timely, as the rights holders seek justice, and thirdly, all affected rights holders should have no fear of victimization in the process of seeking remedies, particularly if they are from marginalized locations or communities. [16]

A commonly expressed concern by users and civil society experts about grievance mechanisms of digital content platforms, especially social media platforms, is the limited information available to those subject to content removal or account suspension, or deactivation, or those reporting abuse such as misogynistic harassment and doxing.[17] As the UN Special Rapporteur highlights, this "lack of information creates an environment of secretive norms, inconsistent with the standards of clarity, specificity and predictability."[18] Hence, it is essential that the Guidelines give due emphasis to the process of seeking

[13] Office of the High Commissioner for Human Rights. (2011). *Guiding Principles for Business and Human Rights.* https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf
[14]Office of the High Commissioner for Human Rights. (2011). *Guiding Principles for Business and Human Rights*. Principle 31(a). https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf
[15] UN Human Rights Council. (2017). *Report of the Working Group on the Issue of Human rights and Transnational Corporations and Other Business Enterprises, A/72/162.* https://documents-dds-ny.un.org/doc/UNDOC/GEN/N17/218/65/PDF/N1721865.pdf?OpenElement
[16]Organisation for Economic Co-operation and Development. (2018). *National Contact Points for Responsible Business Conduct: Providing access to Remedy: 20 years and the road ahead.* https://mneguidelines.oecd.org/NCPs-for-RBC-providing-access-to-remedy-20-years-and-the-road-ahead.pdf
[17] UN General Assembly. (2018). *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/38/35*. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement
[18] Id.

remedies for users in relation to content or account-related decisions made by the platforms that affect users' human rights, including the right to freedom of expression and freedom from violence. Particularly, digital platforms, as discussed in the 2018 thematic report to the Human Rights Council of the Special Rapporteur on freedom of opinion and expression, must move away from generic, self-serving community guidelines and incorporate relevant principles of human rights law into content moderation standards.[19] As content decisions are taken by platforms on the basis of these standards, users must be given all information on why content related changes or interferences or takedowns occur, with a clear opportunity at the first instance, to enable them to challenge any content related decisions. This is an essential component of access to remedy on the platform and must be specifically mentioned as it is the digital platform that has a major role to play as an arbiter of content-related decisions as opposed to the regulator, who does not interfere in such decisions. Hence, access to remedy must have the highest threshold and must be recognized as a separate principle under Paragraph 28, and should not get sidelined or merged with the other principles listed.

[19] UN General Assembly. (2018). *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/38/35.* https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement

4. On the Role of Intergovernmental Organizations

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 30: "Intergovernmental organizations, in line with their respective mandates, should support relevant stakeholders in guaranteeing that the implementation of these guidelines is in full compliance with international human rights law, including by providing technical assistance, monitoring and reporting human rights violations, developing relevant standards, and facilitating multi-stakeholder dialogue."	"30. Intergovernmental organizations, in line with their respective mandates, should support relevant stakeholders in guaranteeing that the implementation of these guidelines is in full compliance with international human rights law, including by providing technical assistance, monitoring and reporting human rights violations, developing relevant standards, and facilitating multi-stakeholder dialogue. ***Intergovernmental organizations and national regulatory agencies may create modalities for engagement in order to evolve best practices.***"

Justification/References

Currently, as per the Draft Guidelines, the role of intergovernmental organizations (IGOs) is mentioned only in Paragraph 30. The role of intergovernmental organizations is unclear, particularly while working with the independent regulator. IGOs have come to play a very significant role in international political systems and global governance; the Draft Guidelines could provide more details on how IGOs can engage with national regulators with their cooperation, such as by sharing emerging insights and regulatory trends, supporting or making suggestions to the national regulator to refine institutional standards and methods, etc.

IGOs not only bring about opportunities for their member States but also exert influence and impose limits on members' policies and the way in which those policies are made. By setting international agendas, and thus influencing domestic ones, governments may be compelled to take positions on issues. Moreover, in democratic societies, norms and principles created or supported by IGOs can be used by domestic groups to push for change in national policies.[20] For instance, approaches by the European Union in creating their regulatory policies are important to understand - keeping in mind the range of nations as well as the number of regulators that operate.

In the European Union, regulatory policy has progressed under the better regulation agenda and played a crucial role in shaping the current regulatory processes. At the same time, all EU Member States have also

[20] Berg, M. (n.d.). *Conventions, Treaties And Other Responses To Global Issues – Vol. II - The Role of Inter and Non- Governmental Organizations.* https://www.eolss.net/sample-chapters/C14/E1-44-03-00.pdf

adopted their own regulatory policies. For instance, in terms of their objective of stakeholder engagement, citizens, businesses, consumers, employees, trading partners and other stakeholders can offer valuable inputs on the feasibility and practical implications of planned regulations. Meaningful stakeholder engagement in the development of regulations can lead to higher compliance with regulations, in particular when stakeholders feel that their views have been considered. For example, In 2017, the Netherlands reviewed the extent to which its internet consultation system was valued by citizens, companies, and departmental staff. The results indicated that internet consultation is systematically used by government officials, but also pointed to a number of weaknesses. For instance, the OECD Report for Better Regulation Practices Across the European Union concluded that it is not easy for citizens and businesses to understand how their consultation comments were taken into account.

The current terms mentioned in the Guidelines such as providing technical assistance, monitoring and reporting human rights violations, developing relevant standards, and facilitating multi-stakeholder dialogue are limited. There are many more ways for IGOs to engage meaningfully like ex-post performatory evaluation practices, best practices for transparency requirements from the State in case of content takedown requests etc. which can be considered by the regulator to implement within their national regulatory policies and responses, keeping in mind local and contextual requirements.[21]

[21] Organisation for Economic Co-operation and Development. (2019). *Better Regulation Practices Across the European Union.* https://www.oecd.org/gov/regulatory-policy/EU-Highlights-Brochure-2019.pdf

5. On the Constitution of the Regulator

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 43: "Officials or members of the regulatory system should: a. Be appointed through a participatory and independent merit-based process. This means that the regulator's decisions are made without the prior approval of any other government entity. b. Be accountable to an independent body (which could be the legislature, an external council, or an independent board/boards). c. Have relevant expertise in international human rights law. d. Deliver a regular public report to an independent body (ideally the legislature) and be held accountable to it, including by informing the body about their reasoned opinion. e. Make public any possible conflict of interest and declare any gifts or incentives. f. After completing the mandate, not be hired or provide paid services to those who have been subject to their regulation, and this for a reasonable period, in order to avoid the risk known as "revolving doors."""	Addition of points to Paragraph 43: Officials or members of the regulatory system should: ***"g. The regulator should be multi-sectoral, consisting of stakeholders including those at the intersection of technology and society as well as civil society and NGO representatives.*** ***h. Qualifications of the members at different levels need to be clearly prescribed."***
Justification/References Despite the expansion of how the composition of the regulator should ideally look like, there is still a lack of detail and clarity regarding the working structure of the independent regulator in the Guidelines. The regulator must provide spaces for civil society actors and NGO representatives who can provide the much-needed vigilance to safeguard the rights of users and ensure that the focus is on the processes and systems used by the platforms rather than on individual pieces of content. The regulatory body should	

also consist of members who work in the intersection of technology and society, so as to get a holistic picture, especially about the technical aspects.

One example of such a regulatory body is the Regulatory Authority for Audiovisual and Digital Communication (ARCOM), a French independent administrative agency created for the regulation and protection of access to cultural works in the digital age. ARCOM, is an independent public authority, which is composed of a board of nine members whose term of office is six years. One of the key objectives of this regulator is to ensure pluralism of views, provide for a safer internet as well as protect the rights and freedoms of users. A key aspect of the functioning of the ARCOM is the collective decision-making power as a regulator. Such a method ensures pluralistic views and also keeps freedoms online at the forefront. The National Regulatory Control Council (Nationaler Normenkontrollrat or NKR), established in Germany in 2006[22] is another example of a non-governmental body, where there are several eligibility criteria for NKR members, e.g., they may not belong either to a legislative body or a federal or state authority. Members are also mandated to have prior experience in legislative matters obtained through serving in State or society institutions, and are required to have a knowledge of economic affairs. It is important for the regulator to have a clear working structure, which is currently absent and must be expanded in the Guidelines.

[22] Organisation for Economic Co-operation and Development. (n.d.). *Case Studies of RegWatch Europe Regulatory Oversight bodies and the European Union Regulatory Scrutiny Board*. https://www.oecd.org/gov/regulatory-policy/Oversight-bodies-web.pdf

6. On the Powers of the Regulator

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 46.e: "Establish a complaints process that offers users redress should a platform not deal with their complaint fairly, based on the needs of the public they serve, the enforcement powers they have in law, their resources, and their local legal context."	"46.e. Establish a complaints process, ***which focuses on systems and processes used by the platform. This complaint process should*** offer users redress should a platform not deal with their complaint fairly, based on the needs of the public they serve, the enforcement powers they have in law, their resources, and their local legal context."
Justification/References The Draft Guidelines clearly state that the regulator's focus should mainly be on the systems and processes used by platforms, rather than expecting the regulatory system to judge the appropriateness or legality of single pieces of content. Hence in the complaint process system that has to be set up by the regulator, it is essential for guidelines to address two aspects. One, the regulator in this redressal system, will only look at administrative and technical issues, which are system and process-related complaints, arising out of the platform reporting and appeal systems. This specification is required so that content-related decisions are not regulated, as this would go beyond the regulatory mandate. For instance, the Payments System Regulator in the United Kingdom also has a complaints and dispute mechanism, where the scope of the dispute is defined to include issues relating to terms of conditions of existing access to payment systems or participation in payment systems, etc. While this regulator states that, other disputes between participants, or between participants and service users may also be put forward for consideration, they reiterate that the regulator would not be a substitute or alternative to the courts in determining and enforcing private legal rights.[23] Similarly, the scope of disputes as well as the disclaimer on how the regulator's decisions can be appealed must be mentioned within this section.	

[23] Payments System Regulator. (n.d.). *The PSR Purpose.* https://www.psr.org.uk/about-us/the-psr-purpose/

7. On the Responsibilities of Digital Platforms

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 57: "It would be expected that illegal content be made unavailable solely in the geographical jurisdiction where it is illegal.20 Identification of illegal content should be interpreted consistently with international human rights law to avoid unjustified restrictions on freedom of expression."	"57. It would be expected that illegal content be made unavailable solely in the geographical jurisdiction where it is illegal.20 Identification of illegal content should be interpreted consistently with international human rights law to avoid unjustified restrictions on freedom of expression. ***In case of tension between national law and international human rights standards in defining illegal content, digital content platforms should respond to government takedown requests for such content in conformity with internationally recognized standards of legitimate purpose, non-arbitrariness, necessity, and proportionality.***"

Justification/References

Compliance with national laws alone is inappropriate for digital content platforms that seek common norms for their geographically and culturally diverse user base.[24] Instead, human rights standards provide a useful framework for holding both States and companies accountable to users across national borders.[25] This is especially important in countries with fragile democracies and authoritarian regimes that may make use of platform regulation as a mode of censorship of dissident voices.

International human rights treaties, along with setting out core rights, also set out the circumstances in which the rights may be limited or restricted.[26] Therefore, they provide useful guidance to assess the legitimacy or reasonableness of restrictions placed by States on the human rights of citizens. From a combined reading of the different permissible limitation tests set out in different articles of the ICCPR, the following have been recognized as elements of a suitable permissible limitation test for the rights to privacy and freedom of expression:

"(a) Any restrictions must be provided by the law (paras. 11-12);

[24] *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.* (2018, April 6). UN General Assembly, A/HRC/38/35. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement
[25] *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.* (2018, April 6). UN General Assembly, A/HRC/38/35. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement
[26] EFF & Article 19. (2014). *Necessary & Proportionate: International Principles on the Application of Human Rights Law to Communications Surveillance (p. 14).*
https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf

(b) The essence of a human right is not subject to restrictions (para. 13);

(c) Restrictions must be necessary in a democratic society (para. 11);

(d) Any discretion exercised when implementing the restrictions must not be unfettered (para. 13);

(e) For a restriction to be permissible, it is not enough that it serves one of the enumerated legitimate aims. It must be necessary for reaching the legitimate aim (para. 14);

(f) Restrictive measures must conform to the principle of proportionality, they must be appropriate to achieve their protective function, they must be the least intrusive instrument amongst those which might achieve the desired result, and they must be proportionate to the interest to be protected (paras. 14-15)."[27]

In order to safeguard the human rights of users, the above standards should inform the platform's response to a content takedown request in pursuance of a legal restriction inconsistent with international human rights law.

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 59: "In the case of other content that risks significant harm to democracy and the enjoyment of human rights, digital platforms should systematically assess the potential human rights impact of such content and take action to reduce vulnerabilities and increase their capacities to deal with it. For instance, companies should be able to demonstrate to the regulatory system the measures that they have in place if such risk is identified. These could be by, for example, providing alternative reliable information, indicating concerns about the origin of the content to users, limiting or eliminating the algorithmic	"59. In the case of content that risks significant harm to democracy and the enjoyment of human rights, digital platforms should systematically assess the potential human rights impact of such content and take action to reduce vulnerabilities and increase their capacities to deal with it. ***In the interest of fairness and transparency, platforms should proactively disclose, both to the users and the regulators, the criteria used by them to decide whether a content poses significant risk of harm to democracy and enjoyment of human rights. Platforms should also be able to demonstrate to the regulatory system the measures that they have in place if such risk is identified.*** These could be by, for example, providing alternative reliable information, indicating concerns about the origin

[27] La Rue, F. (2013, April 17). *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.* UN General Assembly, A/HRC/23/40, pp. 8-9. https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf

amplification of such content, or de-monetizing from advertising revenue."	of the content to users, limiting or eliminating the algorithmic amplification of such content, or de-monetizing from advertising revenue."

Justification/References

The Guidelines talk about two types of content in respect of which digital platforms should take action - illegal content and content that risks significant harm to democracy and enjoyment of human rights. Through references in footnotes and appendix, the Guidelines give an indication of what it means by 'content that risks significant harm to democracy and enjoyment of human rights'. The Guideline also mentions hate speech, disinformation and misinformation, and content that portrays online online-based violence as examples. While this is an improvement over the earlier draft of the Guidelines which provided no guidance on what potentially harmful content means, the phrase is still open to interpretation, which is inevitable as whether a content poses significant risk to democracy and enjoyment of human rights would depend on factors such as the socio-political and cultural context of a particular country or region. However, this subjectivity should not give unaccountable power to platforms to take decisions in relation to content hosted by them. In order to ensure that the decisions by digital content platforms do not stem from opaque or 'secret rules', but from transparent criteria and processes that adhere to human rights norms,[28] platforms should be required to proactively disclose to the users (through the Community Guidelines and Terms of Services) and to the regulators about the criteria used by them to decide whether a content poses significant risk of harm to democracy and enjoyment of rights. Such disclosure will also enable external evaluation of the factors or criteria by regulators, researchers, and other stakeholders and pave the way for better criteria to be developed with minimal restriction of freedom of expression and access to information.

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 70.j: "Digital platforms should publish information outlining how they ensure that human rights and due process considerations are integrated into all stages of the content moderation and curation policies and practices. This publicly available information should include: (j) Information relevant to complaints about the removal,	*After paragraph 70.j, add the following paragraph:* ***"70.X. Data pertaining to the complaints and their resolution, including:*** ***(i) Types of complaints received in relation to content hosted, the category of rule that is violated by such content, and the data for each type;***

[28]Buni, C., & Chemali, S. (2016, April 13). *The Secret Rules of the Internet*. The Verge. https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech

blocking, or refusal to block content and how users can access the complaints process."	***(ii) Action taken by the platform on the complaints received and the number of links and/or extent of information removed or made inaccessible;*** ***(iii) Time taken to resolve the complaint;*** ***(iv)Number of appeals and the number of cases in which the original decision was revised."***

Justification/References

Many concerns are raised about how dominant digital content platforms, particularly social media platforms, deal with user complaints about the content that they host – inconsistent and biased enforcement of community guidelines, delayed or no response to user complaints, inadequate action to mitigate the harm from the content, etc.[29] The grievance redressal mechanism of platforms is often the first port of call for users to mitigate the harm from unlawful and harmful content online. Therefore, it is vital to have transparency not only about how users can access the complaints process, but also about the types of complaints received and the number of complaints under each type, the criteria that platforms employ in dealing with user complaints, time taken for resolution, and the specific action taken by platforms in relation to a complaint. This level of transparency will enable the regulators to require specific actions to be taken by platforms to improve their complaint redressal mechanism in general, and specifically with respect to certain types of content. It will also help the regulators to ascertain the platform's accountability for any undesirable consequences resulting from content that has been reported.

As UNESCO rightly points out: "The transparency of these online platforms would in itself constitute a sharing of information as a public good, by making available data that is not yet in the public realm – both proactively and on demand."[30] Many jurisdictions have included a legislative mandate for platforms to publish compliance/transparency reports regarding the details of complaints received and action taken thereon. Some examples are below:

- India's Intermediary Liability Rules, 2021 require platforms to "publish periodic compliance report every month mentioning the details of complaints received and action taken thereon, and the number of specific communication links or parts of information that the intermediary has

[29] Kaye, D. (2019). *Speech Police: The Global Struggle to Govern the Internet.* Columbia Global Reports.
Also see,Mozur, P. (2018, October 15). *A Genocide Incited on Facebook, With Posts From Myanmar's Military.* The New York Times. https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html; Suzor, N., et al. (2018*). Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online.* Gender-Based Violence Online. https://www.researchgate.net/publication/327962592_Human_Rights_by_Design_The_Responsibilities_of_Social_Media_Platforms_to_Address_Gender-Based_Violence_Online_Gender-Based_Violence_Online/link/6030763f92851c4ed5837306/download
[30] Unesco. (2021). *World Press Freedom Day 2021, Information, As A Public Good.* https://en.unesco.org/sites/default/files/wpfd_2021_concept_note_en.pdf

removed or disabled access to in pursuance of any proactive monitoring conducted by using automated tools or any other relevant information as may be specified".[31]

- Article 13 of the Digital Services Act of the European Union requires platforms to publish reports, at least once a year, indicating "the number of complaints received through the internal complaint-handling system referred to in Article 17, the basis for those complaints, decisions taken in respect of those complaints, the average time needed for taking those decisions and the number of instances where those decisions were reversed."[32]
- Germany's NetzDG law mandates operators of social networks to submit biannual reports on their handling of complaints about criminally punishable content. These reports must contain information, for example, on the volume of complaints and the decision-making practices of the network, as well as about the teams responsible for processing reported content. They must be made available to everybody on the Internet.[33]

[31] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), § 4(1)(d). European Union. (2020). *Digital Services Act. Article 29*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en

[32] Article 29. (n.d.). *Digital Services Act (European Union).* EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en

[33] § 2, Netzwerkdurchsetzungsgesetz (NetzDG Act) (Germany). https://www.bmj.de/DE/Themen/FokusThemen/NetzDG/NetzDG_EN_node.html

8. On the Data Access for Research Purposes

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 72: Digital platforms should provide access to non-personal data and anonymized data for vetted researchers that is necessary for them to undertake research on content to understand the impact of digital platforms. This data should be made available through automated means, such as application programming interfaces (APIs), or other open and accessible technical solutions allowing the analysis of said data.	After paragraph 72, add the following paragraph: ***"X. Digital content platforms should provide users with access to their data and to insights drawn from their data through the operations of the platform, including through continuous and real-time access, and the ability to port the data to other platforms. Platforms should make their interfaces and other software solutions interoperable so that users can exchange information across platforms and mutually use information that has been exchanged. The regulator should, in coordination with regulators on competition, data protection, and any other relevant aspect, lay down standards for data portability and interoperability for digital content platforms to comply with and oversee their implementation by the platforms. The regulator will have the power to ask social media platforms to comply with this requirement of enabling open data access."***
Justification/References Data portability and interoperability are crucial to ensure that users are not tied down to one digital content platform and have the real choice and ability to migrate to other platforms that serve their interests better. This will facilitate the emergence of newer digital content platforms that may have alternative value propositions to the dominant platforms, and consequently, diverse information and viewpoints that users can have access to. Many jurisdictions are giving increasing attention to data portability and interoperability, the EU being the case in point. In addition to the data portability obligation under the General Data Protection Regulation,[34] the Digital Markets Act requires digital platforms to provide effective portability of data generated through the activity of a user and provide	

[34] European Union. *Article 20, General Data Protection Regulation*. https://gdpr-info.eu/

them tools to facilitate the exercise of data portability.[35] Further, the Digital Markets Act also requires digital platforms to interoperate with each other.[36]

For the effective realization of data portability and interoperability, regulators should lay down technical and operational standards and oversee their implementation. In the development of these standards, the regulator should coordinate with the relevant regulators who are overseeing other aspects that will be affected by interoperability and data portability such as market competition and data protection.

Till recently, Twitter made available its platform data to independent researchers so that they can analyze everything from online trolling to the spread of misinformation. These studies have been critical to understanding the tactics used by scammers, foreign influence campaigns and other malicious actors trying to manipulate social media. However, in February 2023, Twitter announced that they would no longer support free access to the Twitter API, both Version 1 and Version 2, and that a paid basic tier will be available instead. [37] Such a change is bound to adversely impact research efforts on Twitter manipulation and have devastating consequences for transparency, accountability, and the public good. To tackle such situations, the Guidelines should give regulators the power to order digital platforms within their jurisdictions that have closed data systems to revoke such policies and provide open platform data to citizens. Otherwise, the result of such practices will be profit over rights and profit over communities that goes directly against the transparency obligations of the digital platforms proposed in the Guidelines.

[35] European Union. *Article 6, Digital Markets Act*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0842&from=en
[36] European Union. *Article 6, Digital Markets Act*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0842&from=en
[37] TwitterDev. [@TwitterDev]. (2023, February 2). Twitter. https://twitter.com/TwitterDev/status/1621026986784337922

9. On Platforms Empowering User Reporting

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 76: "The user reporting system should give high priority to concerns regarding content that threatens users, ensuring a rapid response, and, if necessary, by providing a specific escalation channel or means of filing the report. This is particularly important when it to comes to gender-based violence and harassment."	"76. The user reporting system should give high priority to concerns regarding ***particularly grievous and harmful illegal content such as child pornography, pro-terror content, content promoting, depicting or inciting extreme violence, drug misuse, and violent content***, ensuring a rapid response, and, if necessary, by providing a specific escalation channel or means of filing the report. ***Such high priority should also be given to content that endangers the safety and human rights of users, such as gender-based violence and harassment.***"
Justification/References Platforms receive a huge volume of user reporting in relation to the content that they host. Hence, it is essential to have a mechanism to prioritize complaints concerning the content that require urgent and timely response to mitigate the harm that it produces. However, 'content that threatens users' is a vague and overbroad criterion which may be used by bad actors to claim priority in responding to content that merely annoys or inconveniences them. In order to ensure that the user reporting system works efficiently and responds rapidly to the most illegal and harmful content that endangers the safety and human rights of users, the prioritization criteria should be clearly defined. We suggest that the user reporting system should give priority to particularly grievous and harmful illegal content such as child sexual exploitation material, pro-terror content, content promoting, depicting, or inciting violence, etc,[38] and content that endangers the safety and human rights of users, such as gender-based violence and harassment that affect the freedom of women and girls. For example, non-consensual distribution of intimate images can cause significant harm and rights violation, especially for women and girls if it is not immediately acted upon and the image removed from the platform.[39] These types of content should receive immediate and preferential attention from the platforms.	

[38] Government of Australia. *Section 135, Online Safety Act, 2021.* https://www.legislation.gov.au/Details/C2021A00076
[39] Dunn, S., & Petricone-Westwood, A. (2018). *More than 'Revenge Porn' Civil Remedies for the Non-consensual Distribution of Intimate Images.* 38th Annual Civil Litigation Conference. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3772050