Guidance For Regulating Digital Platforms:

A Multistakeholder Approach

Response from IT for Change

January 2023

IT for Change

Response from IT for Change[1] to the

Guidance For Regulating Digital Platforms: A Multistakeholder Approach

January 2023

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

Index

Introduction

Securing information as a public good, that everyone is entitled to, is vital for the fulfillment of collective human aspirations.[2] Apart from its importance in dealing with emergency situations like health, disasters, and climate crisis, access to information is necessary for citizens to exercise their fundamental rights, gender equality, participation and trust in democratic governance, and sustainable development, essentially, to leave no one behind.[3] The significance of access to information is underscored by the major international human rights treaties such as the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR) that recognize the right 'to seek, receive and impart information and ideas through any media and regardless of frontiers' as an intrinsic part of freedom of expression.[4]

In today's digitally transformed information ecosystem, securing information as a public good requires us to deal with the role of digital content platforms that facilitate and mediate the sharing of knowledge and information along with other kinds of communication content such as disinformation, hate speech, entertainment and data.[5] Digital content platforms shape information flows in society through their content moderation and content curation functions, often driven by their business interests and profit motives.

Content moderation strategies of these platforms have many a time failed to remove harmful and unlawful content, resulting in egregious real world ramifications.[6] Further, the algorithms that platforms deploy for content curation tend to prioritize sensationalist content over verified information, thereby amplifying harmful, hateful, and violent content.[7] This has consequences not only for the freedom of expression of the users of these platforms but also for their right to seek and receive information. For instance, IT for Change's research has revealed that the presence of online gender-based violence on social media platforms forces women to reduce their online presence, thereby adversely affecting their freedom of expression and access to information.[8]

---

[2] Windhoek+30 Declaration. <https://en.unesco.org/sites/default/files/windhoek30declaration_wpfd_2021.pdf>

[3] Windhoek+30 Declaration. <https://en.unesco.org/sites/default/files/windhoek30declaration_wpfd_2021.pdf>

[4] Article 19 of Universal Declaration of Human Rights, 1948; Article 19 of International Covenant on Civil and Political Rights, 1966.

[5] Windhoek+30 Declaration. <https://en.unesco.org/sites/default/files/windhoek30declaration_wpfd_2021.pdf>

[6] A Genocide Incited on Facebook, With Posts From Myanmar's Military, Washington Post (Oct 15, 2018), <https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html> ; Inflammatory content on FB was up 300% before Delhi riots, says the internal report, NewsMinute. <https://www.thenewsminute.com/article/inflammatory-content-fb-was-300-delhi-riots-says-internal-report-156878>; Rina Chandran, How the Leicester Communal Clashes Were Fuelled by Online Disinformation from India, The Wire. <https://thewire.in/communalism/how-the-leicester-communal-clashes-were-fuelled-by-online-disinformation-from-india>

[7] Vosoughi, S., Roy, D. and Aral, S. The Spread of True and False News Online, 359(6380) Science 1146-1151(2018). Also see in the context of online gender-based violence, Anita Gurumurthy and Amshuman Dasarathy. 2022. Profitable Provocations. A Study of Abuse and Misogynistic Trolling on Twitter Directed at Indian Women in Public-political Life. <https://itforchange.net/sites/default/files/2132/ITfC-Twitter-Report-Profitable-Provocations.pdf>

[8] Born digital, Born free? A socio-legal study on young women's experiences of online violence in South India, IT for Change. <https://itforchange.net/sites/default/files/1662/Executive_Summary_Born%20digital-Born-free%20.pdf>

Given this worrisome context, we appreciate UNESCO's initiative to come up with a Guidance for Regulating Digital Platforms: A Multistakeholder Approach (Guidance Document) for actors to regulate, co-regulate and self-regulate digital platforms with the aim to secure freedom of expression and information as a public good and deal with content that potentially damages human rights and democracy. While one cannot overemphasize the value of this document to foster international consensus on some of the foundational principles for regulating digital platforms in line with human rights norms, we urge that the Guidance Document be informed by a) the challenges of platform regulation – especially in the Global South, b) the evidence on the potential of digital content platforms to undermine or weaken civic values and democracy, c) the architecture of corporate impunity that platforms benefit from, and d) the need to address communication governance through supra-liberal frameworks.

As discussed in the 2018 thematic report by David Kaye, the UN Special Rapporteur on freedom of opinion and expression, "ensuring company transparency and remediation to enable the public to make choices about how and whether to engage in online forums" is a crucial goal.[9] At the same time, it is also vital that institutions of law and justice are ready to address the realpolitik of digital transformation. This calls for moving beyond silos and grasping the necessary coherence and consistency of vision needed to make laws and policies adequate for development, democracy, and human rights in the 21st century. While the Guidelines may not deal with data privacy, competition, or other such legal issues, it cannot be forgotten that an inclusive, democratic and vibrant platform society is predicated on harmonizing the legal-policy regime to address the totalizing power wielded by platform companies. A whole-of-economy approach is needed to address the challenges for human rights, social justice, and equity in the platform society. This must go hand-in-hand with a commitment at the international level and from nation-states to provide individuals and communities new rights and avenues to seize the potential of platformization for democracy, solidarity, and wellbeing.

While the Guidelines comprise an important step forward, the institutional revamp necessitated by platformization requires urgent attention to financial support and assistance, especially for developing countries in order to, amongst others, (1) build regulatory capacity; (2) secure information as a public good; and (3) undertake media literacy programs. UNESCO's leadership to catalyze the institutional capabilities for a just platform society through adequate and appropriate funding is key.

Keeping these concerns in mind, we provide some comments and recommendations on the **Guidance Document,** based on our research efforts over several years on the digitally transformed information ecosystem, platform-mediated and algorithm-driven public sphere, and the legal, institutional, and design safeguards necessary to secure a just and equitable online space that can foster democratic values.

[9] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement>

Summary of Important Recommendations

The Guidance Document refers to and applies to 'digital platforms or digital platform services'. However, it is not clear what type of entities come under these two terms, except for some examples such as social media networks, search engines, and content sharing platforms. In its current form, the framework and obligations prescribed in the Guidance Document 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.

Digital platforms are of many types, including social media platforms, media sharing platforms, news aggregators, news apps, search engines etc. Different platforms perform different kinds of functions and exercise different levels of control in relation to the content that they host, and have different scales of reach. A description of the types of platforms being discussed in the Guidance Document is essential for effective regulatory governance and oversight, as each kind of platform would require its own form of differentiated regulatory tools. This can include lower thresholds in respect of some parameters for platforms that are smaller market share/ number of users or higher standards of duty-of-care to users for digital content platforms that host news content.[10]

For the purposes of this submission, our reflections and recommendations address large social media platforms that we refer to as 'digital content platforms'. Observations about platforms other than social media platforms are explicitly flagged as such.

A brief summary of our inputs and recommendations on the various themes addressed within the Guidance Document is given below:

Topic	Summary
On the Responsibilities of the Government	The Guidance Document should provide clear instructions to governments to be transparent about the requests they make to digital content platforms to remove or restrict content. Content removal requests/orders by any government must be subjected to the due processes of law, and a decision by an independent judicial authority. Also, the basis for such requests/orders must be published. In case a government deems it fit to impose any indirect restriction, such as internet shutdowns, it should submit documentation to the regulator justifying the imposition of such restriction. Further, governments should take legislative and policy measures to require digital content platforms to respect user rights to freedom of expression, the right to information, and equality and non-

[10] Philip Napoli, Social Media and the Public Interest: Media Regulation in the Disinformation Age (2019). <https://www.researchgate.net/publication/335107476_Social_Media_and_the_Public_Interest_Media_Regulation_in_the_Disinformation_Age>

	discrimination in their policies, practices, and operations. Governments should also institute an accountability framework to hold platforms liable for their actions that threaten user rights or facilitate various forms of online harms, including online gender-based violence.
On the Transparency Process of Digital Platforms	In line with the call for radical transparency by David Kaye, UN Special Rapporteur on Freedom of Expression, the Guidance Document should ask digital content platforms for more robust disclosure of information about their policies, practices, and operations. This should include transparency not only regarding the means employed for content moderation, but also the algorithmic tools used for content curation, the platform's policies and practices regarding placement of advertisements on content, number of human moderators, their expertise, and employment status, and details about the complaints received by the platforms; all with greater nuance and attention to local as well as cultural specificity of the country. Such robust declaration of information is necessary for the regulator to exercise effective oversight over the operations of such platforms, and to mandate them to make suitable changes or corrections, lest they impact the freedom of expression, access to information, and right to online participation of users adversely.
On the Content Management Policies	Digital content platforms should undertake the following for their content management policies: ● Clearly inform users about the types of content liable to be taken down under the mandate of the law, and the types that will be restricted due to their harmful nature, albeit their lawfulness. ● Remove without undue delay content that is unlawful, and take measures to restrict the reach and virality of content that is harmful albeit legal, and disclose the nature of such measures. ● Respond to government takedown requests for such content in conformity with internationally recognized standards of legitimate purpose, non-arbitrariness, necessity, and proportionality, in case of tension between national law and international human rights standards in defining illegal content. ● Take into consideration the social, political, cultural, and religious context of the country in which they are operating, particularly, the unique disadvantages and locally specific forms of online abuse faced by certain social groups falling at the intersection of marginalized identities.

On Enabling Environment	Digital content platforms should conduct periodic risk assessments and submit reports of the same to the independent regulator in order to determine, identify and address any actual or potential harm or human rights impact of their operations/actions. Such assessments should also be undertaken prior to any major design changes, or major decisions or changes in operations, or new activity or relationship, or in response to a major event or any change in the operating environment. Further, digital platforms hosting news, such as news apps or news aggregators or even social media platforms sharing news content, should optimize their curation algorithms for diversity, in order to advance the goal of securing information as a public good.
On User Reporting	Users should have access to all information related to reporting and grievance redressal, including details of the representative appointed by the digital content platform in their own country, information about policies in a digestible format and in all relevant languages, and recourse/appeal mechanisms available for reasons including removal of content, suspension of an account, or any other type of action affecting users' human rights, including the right to freedom of expression and freedom from violence.
On Content That Potentially Damages Democracy and Human Rights, Including Mis- and Disinformation and Hate Speech	Algorithms, targeted advertising, and the data harvesting practices of digital content platforms, particularly, the largest social media companies, are largely credited with driving users towards extremist content and conspiracy theories. Therefore, such platforms should conduct periodic risk assessments of their curation and recommender algorithms. They should make transparent the result of such risk assessment and also the measures taken to mitigate the risks identified.
On Media and Information Literacy	Digital content platforms should implement specific media and information literacy measures for women, children, youth and indigenous groups, as well as for gender and sexual minorities.
On Data Access	In the interest of transparency leading to accountability, digital content platforms should be encouraged by the regulator to share their data with independent researchers, auditors, supervisory authorities and other external stakeholders. Given the lack of expertise of many regulators to analyze and draw insights from the data produced by platforms in their transparency reports, data sharing with researchers and auditors will help in identifying

	systematic risks from the operations of digital content platforms, paving the way for holding such platforms accountable. In the interest of safeguarding user privacy, such sharing of data should be under the strict supervision of the regulator, which will include vetting individuals/entities requesting the data. To minimize internet fragmentation and promote diversity of information and view-points, digital content platforms should allow for data portability, and make their interfaces and other software solutions interoperable, so that users can exchange information across platforms and mutually use information that has been exchanged. For the effective realization of data portability and interoperability, the regulator should, in coordination with regulators on competition, data protection and any other relevant aspect, lay down technical and operational standards and oversee their implementation by the platforms.
On Regulators' Constitution & Powers	Regulators should undertake the following to build a working structure for effective governance: ● Detail the multi-sectoral composition of members including the positions of members such as Presiding Officer, full-time, part-time members, and secretary as well as the nature of stakeholders including representation from civil society and NGO representatives. ● Prescribe qualifications of the members at different levels. ● Articulate in detail the civil measures that must be undertaken such as fines, compliance orders, and warning orders to address the identified failings of digital content platforms ● Clarify the nature and purpose of the independent third party being hired as well as outline the powers, functions and procedures to be followed through a protocol/policy. ● Require the independent third party to share details of their investigation for scrutiny. ● Publicly disclose reasons for prioritizing and selecting complaints. This is necessary to further the transparency obligations of the regulator as well as to provide for elimination of possible bias and discrimination.

Comments on the Proposed Guidance for Regulating Digital Platforms: A Multistakeholder Approach

Note to the reader:

Recommended additions in the text of the Guidance Document are in ***Bold and Italics***

Recommended deletions are denoted through ~~Strikethrough.~~

1. On the Responsibilities of the Government

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 23.4: "Be transparent about the requests they make to companies to remove or restrict content and be able to demonstrate how this is consistent with Article 19 of the ICCPR;" Paragraph 23.5: "Guarantee that any content removals are subjected to the adequate due process of law, including independent judicial review;"	"23.4 Be transparent about the requests they make to companies to remove or restrict content, and be able to demonstrate how this is consistent with Article 19 of the ICCPR. 23.5 Guarantee that any content removals are subjected to the ~~adequate~~ due process of law, including independent judicial review, ***and the publication of the basis for such removal;"***
Justification/References Recourse to the 'due process of law' implies that all established rules and procedures that eventually provide safeguards for the protection of individual rights are available. States should only restrict content pursuant to an order by an independent and impartial judicial authority.[11]	

[11] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement>

This is a mandatory requirement in several international human rights instruments. The term 'adequate' is redundant since 'due process' presupposes that acceptable human rights standards and processes for content removal are pursued.

Governments may issue takedown orders for content including those that are critical of their own policies.[12] Several international discussions at the Human Rights Council[13] have highlighted how the States restrict, control, manipulate and censor content disseminated via the Internet without any legal basis, or on the basis of broad and ambiguous laws, without justifying the purpose of such actions; and/or in a manner that is clearly unnecessary and/or disproportionate to achieving the intended aim. Such actions are clearly incompatible with the States' obligations under international human rights law, and often create a broader "chilling effect" on the right to freedom of opinion and expression. In addition to requiring that governments share and publish information about content takedown orders, it is vital that they are duty bound to share and make publicly available the basis or reasons for such removal .[14]

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 23.6 "Not impose indirect restrictions to companies (for example, internet shutdowns) for alleged or potential breaches of regulations;"	"23.6 Not impose indirect restrictions ~~to~~ ***on*** companies (for example, internet shutdowns) for alleged or potential breaches of regulations;" After paragraph 23.6, add the following paragraph: ***"23.X Submit documentation to the regulator about why a certain event is deemed a communication anomaly/irregularity requiring indirect restriction;"***

Justification/References:

Often times, in the Global South, indirect restrictions such as internet shutdowns tend to be politically motivated decisions[15] leading to disruption of mobile or internet connectivity. These restrictions are neither lawful nor are they proportionate or transparent decisions.[16] The

[12] Jack Gillum and Justin Elliott, Sheryl Sandberg and Top Facebook Execs Silenced an Enemy of Turkey to Prevent a Hit to the Company's Business, Pro Publica (Feb 24, 2021). <https://www.propublica.org/article/sheryl-sandberg-and-top-facebook-execs-silenced-an-enemy-of-turkey-to-prevent-a-hit-to-their-business>

[13] UN Human Rights Council (2011) Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue. A/HRC/17/27, para 40. <https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/a.hrc.17.27_en.pdf >

[14] OECD Best Practice Principles on the Governance of Regulators, OECD. <https://www.oecd.org/gov/regulatory-policy/governance-regulators.htm>

[15] What is an internet shutdown? A guide for South and Southeast Asia civil society, Engage Media. <https://engagemedia.org/2022/internet-shutdowns-south-southeast-asia/>

[16] Asia's Internet Shutdowns Threaten the Right to Digital Access, Chatham House. <https://www.chathamhouse.org/2020/02/asias-internet-shutdowns-threaten-right-digital-access>

right to freedom of expression, which includes the freedom to seek, receive and impart information of all kinds, regardless of frontiers, is a right guaranteed through several international human rights instruments. (Article 19 (2) of the ICCPR[17], echoing article 19 of the UDHR)

The UN Special Report[18] on Internet shutdowns highlights the critical role of transparency measures to stop shutdowns and limit their harmful consequences. It also points to the challenges in detecting and obtaining information about shutdowns, which lead to underestimations of their frequency, scope, and impact. The lack of public information on shutdowns could encourage authorities to deny any disruptions that have taken place or even deny their own role in such interventions. Hence, there must be an onus on the State to mandatorily submit documentation to the regulator about why a particular indirect restriction is taking place. Such disclosures to the regulator will enable effective regulatory governance, and would also play a key role in measuring regulatory performance in the long term.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 23.6: "Not impose indirect restrictions to companies (for example, internet shutdowns) for alleged or potential breaches of regulations;"	After paragraph 23.6, add the following paragraph: ***"23.X Take legislative and policy measures to convey a clear expectation that digital content platforms have to respect user rights to freedom of expression, the right to information, equality and non-discrimination in their policies, practices, and operation;"***

Justification/References

The UN Guiding Principles on Business and Human Rights require States to enforce laws that are aimed at, or have the effect of, requiring business enterprises to respect human rights, and periodically to assess the adequacy of such laws and address any gaps.[19] This is necessary to break the architecture of corporate impunity that allows digital content platforms to overlook human rights implications of their operations, and avoid accountability for the same.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**

[17] Article 19(2) of International Covenant on Civil and Political Rights, 1966.

[18] UN Human Rights Council (2022) Report of the Office of the United Nations High Commissioner for Human Rights on Internet shutdowns: trends, causes, legal implications and impacts on a range of human rights. A/HRC/50/55 <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/341/55/PDF/G2234155.pdf?OpenElement>

[19] UN Guiding Principles on Business and Human Rights. <https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf> Principle 3

Paragraph 23.7: "Not subject staff of companies to criminal penalties for an alleged or potential breach of regulations, as this will have a chilling effect on freedom of expression;"	After paragraph 23.7, add the following paragraph: ***"23.X Develop a liability framework to hold digital content platforms and those directly in charge of and responsible for the conduct of business 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[20] and several whistleblower revelations[21] demonstrate that digital content platforms, especially social media, have been complicit and are actively helping in promoting 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.[22] 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 important to ensure that the liability provisions do not cause these platforms to engage in overzealous censorship of content, thereby endangering the freedom of expression of users. To avoid this, rather than holding the digital content platform liable for each and every instance of harmful

[20] United Nations. 2018. "Report of the Detailed Findings of the Independent International Fact-Finding Mission in Myanmar." 39th session of the Human Rights Council, September 28, 2018. Agenda Item 4. <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/>. Ritu Manuvie 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>

[21] Ryan Mac and Cecilia Kang. 2021. Whistle-Blower Says Facebook 'Chooses Profits Over Safety'. <https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html >; Richard Forno. Sept 1, 2022. Did Twitter ignore basic security measures? A cybersecurity expert explains a whistleblower's claims. <https://theconversation.com/did-twitter-ignore-basic-security-measures-a-cybersecurity-expert-explains-a-whistleblowers-claims-189668 >; Face of Hatebook. 2021, The London Story Digital Watch. <https://thelondonstory.org/wp-content/uploads/Face-of-the-Hatebook-Report.pdf >

[22] Amber Sinha. 2020. Beyond Public Squares, Dumb Conduits, and Gatekeepers: The Need for a New Legal Metaphor for Social Media. 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/ >

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.

2. On the Transparency Process of Digital Platforms

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 25.4: "Any safeguards applied in relation to any content moderation that are put in place to safeguard freedom of expression and the right to information, including in response to government demands/requests, particularly in relation to matters of public interest, so as to ensure a plurality of views and opinions;"	After paragraph 25.4, add the following paragraph: ***"25.X Information about the number of human moderators employed and the nature of their expertise in local language and local context, as well as whether they are in-house staff or contractors;"***
Justification/References Despite a global footprint and dispersed workforce, it is seen that digital content platforms lack moderators with expertise in the local language and cultural and social context of many countries that they operate in.[23] This often results in an inaccurate understanding of content, and the application of content moderation policies in a manner that disadvantages marginalized groups with mass takedowns of content. Revelations[24] by Frances Haugen, including Facebook's lack of safety controls in non-English language markets, such as Africa and the Middle East, are only one of the many instances that reveal how AI systems were unable to detect abusive posts that promoted hate speech in a number of languages used on its platform. For example, at the height of the disinformation and hate speech campaign on Facebook against	

[23] Ángel Díaz and Laura Hecht-Felella, Double Standards in Social Media Content Moderation, Brennan Center for Justice (Aug. 4, 2021). <https://www.brennancenter.org/sites/default/files/2021-08/Double_Standards_Content_Moderation.pdf >

[24] Frances Haugen: 'I never wanted to be a whistleblower. But lives were in danger', The Guardian. <https://www.theguardian.com/technology/2021/oct/24/frances-haugen-i-never-wanted-to-be-a-whistleblower-but-lives-were-in-danger.>

Rohingya Muslims in Myanmar which led to ethnic cleansing, Facebook had employed only two Burmese-speaking content moderators[25] when almost all of the offensive posts were in Burmese.[26] While we appreciate the guidance to deploy automated language translators [Paragraph 34], the limitations of such automated tools, as the Guidance Document itself acknowledges [Paragraph 37], underscore the need for a sufficient number of human moderators with expertise in the local language. Therefore, information about the number of human moderators employed, and their expertise in local language and context is an important area of regulatory concern that needs to be made transparent.

The employment status of human content moderators is also important information to be disclosed. A report by the New York University Stern Centre for Business and Human Rights, "Who Moderates the Social Media Giants? A Call to End Outsourcing",[27] highlighted that a major part of content moderation tasks on major social media platforms such as Facebook, Twitter, Instagram, YouTube, and other Google products is done by contract workers employed by third-party vendors. These contract workers are low-paid, overworked, mentally exhausted without any psychological support, and also do not get much supervision or guidance in their work. While this is a serious labor rights issue, the poor working conditions of these workers also poses the risk of jeopardizing the critical function of content moderation. Hence, disclosure of information about the number of contract staff as content moderators will be useful in assessing the impact of outsourcing on the efficiency of content moderation.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 25.6 "Any use made of automated means for the purpose of content moderation, including a specification of the role of the automated means in the review process and any indicators of the benefits and limitations of the automated means in fulfilling those purposes."	"25.6 Any use made of automated means for the purpose of content moderation, including a specification of the role of the automated means in the review process, ***training data used for the development and improvement of such automated means, and the procedures for quality assurance or evaluation in place to improve the decisions made by them,*** any indicators of the benefits and limitations of the automated means in fulfilling those purposes***, and the measures taken to mitigate any risk to human rights of users from the incorrect decisions made by these tools."***

[25] Olivia Solon, Facebook's failure in Myanmar is the work of a blundering toddler, The Guardian, (Aug. 16, 2018). <https://www.theguardian.com/technology/2018/aug/16/facebook-myanmar-failure-blundering-toddler >

[26] Steve Stecklow, Why Facebook is losing the war on hate speech in Myanmar, Reuters (Aug 15, 2018). <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/ >

[27] Paul M. Barrett, Who Moderates the Social Media Giants? A Call to End Outsourcing, New York University Stern Centre for Business and Human Rights (June 2020). <https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report_June+8+2020.pdf >

Justification/References

Automated detection of problematic content relies on sophisticated software that matches the given content with a database of known violations, and increasingly, on machine learning algorithms that are capable of recognizing problematic content through continuous training.[28] While automated detection tools are more efficient than human review they have some limitations, as discussed below:

- Automated filtering tools fail to detect content that is not already identified as problematic content.[29] Therefore, they are not suitable to detect newer forms of violation, or emergent profanity and slang. Further, it is possible to evade detection by these tools by employing techniques such as wordplay, misspelling, insertion of punctuation marks, etc.[30]
- Automated content moderation tools are often incapable of understanding the context of a text or image, widespread variation of language cues and meaning, and linguistic and cultural particularities,[31] which may result in incorrect removal or retaining of content. For instance, texts or images relating to intimate body parts used in sex education and scientific conversations may be incorrectly removed.[32] Similarly, an offensive speech made by a public figure that should be reported in the public interest may be taken down wrongly for being 'objectionable'.[33] Thousands of videos uploaded to YouTube by civil society groups and activists to document atrocities conducted during the Syrian Civil War were removed by the platform's automated counter-terror content detection systems.[34]
- Automated content moderation tools can amplify social bias.[35] The effectiveness of machine learning algorithms to detect problematic content heavily depends on the data that is used to train them.[36] If the training data is biased, including because of language, and characterized by societal prejudices, then such bias will be reflected in the decisions made by the algorithms. Several studies have found, for example, that machine learning models reflect or amplify gender bias in the text used to train them.[37] This type of bias could

[28] Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media, p. 77 (2018).

[29] Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media, p. 99 (2018).

[30] Id.

[31] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement >

[32] Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media, p. 99 (2018).

[33] Id.

[34] Malachy Browne, YouTube Removes Videos Showing Atrocities in Syria, New York Times. <https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html >

[35] Center for Democracy and Technology, Mixed Messages? The Limits of Automated Media Content Analysis (November 2017), p. 4.

[36] Id.

[37] Id.

lead to content moderation decisions that disproportionately censor or misinterpret the speech of certain groups or those with minority views or those speaking other languages.[38] It can also lead to failure to detect many forms of abuse or harassment against minority communities. For instance, the content filtering tools of social media platforms that are not caste-sensitive have often failed to remove caste-based hate speech in India.[39] Further, it has been found that the automated content detection mechanisms of several platforms resulted in censorship of lesbian, gay, bisexual, transgender, and queer online content. LGBTQ YouTubers reported that videos on their channels were being restricted for the reason that they featured LGBTQ content. Likewise, after the introduction of the Safe Mode filtering feature on Tumblr, searches for terms like 'gay' or 'lesbian' were reported to show no results.[40]

Due to the above concerns with automated content detection and filtering tools, it is vital from an accountability point of view for platforms to be transparent about the working of such tools and the steps taken to mitigate any risk to human rights of users from the incorrect decisions made by these tools. The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression in their 2018 report, also stressed on the importance of platforms explaining the impact of automation in content moderation as part of decisional transparency. [41]

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 25.6: "Any use made of automated means for the purpose of content moderation, including a specification of the role of the automated means in the review process and any indicators of the benefits and limitations of the automated means in fulfilling those purposes."	After paragraph 25.6, add the following paragraph: ***"25.X Deployment of any proprietary recommendation algorithm to curate and recommend content to users; the logic behind determining what content appears on a user's feed and what is hidden from their view; the nature of personal data of users that is collected and used by the recommendation algorithm; and whether information collected by the digital content platform for content recommendation purposes is also***

[38] Id.

[39] Murali Shanmugavelan. March 2021. Caste-hate speech: Addressing hate speech based on work and descent. International Dalit Solidarity Network. Pg 8. <https://idsn.org/wp-content/uploads/2021/03/Caste-hate-speech-report-IDSN-2021.pdf >

[40] 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>

[41] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement >

	shared with third parties, as well as how the platform obtains information from other sources to support recommendation algorithms."

Justification/References

The transparency requirement that the Guidance Document proposes is mainly in relation to the content moderation functions of digital content platforms. While this is significant, in order to truly advance the goal of information as a public good, attention should also be paid to how these platforms curate and recommend content to the users by deploying proprietary recommendation algorithms. These algorithms that operate on the logic of the attention economy personalize the content for each user to hold their attention on the platform for a longer time. Algorithms, thus, prioritize attention-grabbing content, even if it is controversial, harmful or low-quality.[42] As a result, certain content gets traction on these platforms and is amplified by the algorithms. Rather than quality, value or integrity, content that is sensational is likely to circulate. As the UN Special Rapporteur on Promotion and Protection of Freedom of Expression, Irene Khan, highlighted in the thematic report on disinformation, the lack of transparency regarding how platforms curate content "significantly undermines people's agency and choice in relation to their information diet" and points towards "an unacceptable level of intrusion into individuals' right to form their ideas free from manipulation and right to privacy".[43]

A study by Massachusetts Institute of Technology (MIT) researchers, examining about 126,000 stories shared by some 3 million people on Twitter between 2006 and 2017, suggests that fake news was about 70% more likely to be retweeted than factual news.[44] Further, an experiment conducted with a dummy user account in India revealed that the feed of the dummy user was flooded with doctored images, fake news, and violent scenes after following Facebook's algorithm-based recommendations.[45]

This has serious implications on the flow and diversity of the information ecosystem in society, with certain voices being amplified and certain voices suppressed. In order to rectify this distortion, and to make users more conscious of their information diet, it is essential that the workings and underlying logic of platform recommendation algorithms are made transparent to the regulator, users, and general public.

Today, regulators are increasingly paying attention to the transparency of recommendation and curation algorithms of the platform. At the European level, transparency obligations for platforms' recommender systems have been introduced in the Digital Services Act.[46] Transparency

[42] Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media, p. 201 (2018).

[43] Disinformation and freedom of opinion and expression, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, April 13, 202, UNGA, A/HRC/47/25. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf?OpenElement

[44] The Spread Of True And False News Online, MIT. <https://ide.mit.edu/wp-content/uploads/2018/12/2017-IDE-Research-Brief-False-News.pdf>

[45] Facebook's algorithm led dummy user in India to fake news, hate speech within three weeks: Reports, The Scroll. <https://scroll.in/latest/1008525/facebooks-algorithm-led-dummy-user-in-india-to-fake-news-hate-speech-within-three-weeks-reports>

[46] Article 29, Digital Services Act (European Union). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en >

is also a central tenet of European Union's "Strengthened Code of Practice on Disinformation," to address the perceived impacts of platform recommender systems on political polarization, the spread of false information, and data profiling and targeting techniques.[47] The Online Safety Bill of UK is also expected to introduce new transparency requirements for platforms and their recommender systems particularly as it relates to combating illegal content and legal, but harmful content.[48]

Further from a user perspective, according to the findings of a research study,[49] users expect a transparency report on recommendation algorithm transparency to include the following details:

(i) Personal data collected and used in recommendation algorithms

(ii) Users' options for customization within a particular service and the impact of each choice

(iii) Information about how the data collected by a digital content platform for content recommendation purposes is also shared with third parties, as well as how platforms obtain information from other sources to support recommendation algorithms.[50]

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 25.6: "Any use made of automated means for the purpose of content moderation, including a specification of the role of the automated means in the review process and any indicators of the benefits and limitations of the automated means in fulfilling those purposes."	After paragraph 25.6, add the following paragraph: ***"25.X The platform's policy and practices regarding placement of advertisements on content or web-pages hosted by it."***
Justification/References	

[47] Commitment 15, 2022 Strengthened Code of Practice on Disinformation. <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation >

[48] Policy paper Online Safety Bill: factsheet, GOV.UK. <https://www.gov.uk/government/publications/online-safety-bill-supporting-documents/online-safety-bill-factsheet>

[49] Michal Luria, This is Transparency to Me" User Insights into Recommendation Algorithm Reporting, CDT Research. <https://cdt.org/wp-content/uploads/2022/10/algorithmic-transparency-ux-final-100322.pdf>

[50] Michal Luria, This is Transparency to Me" User Insights into Recommendation Algorithm Reporting, CDT Research. <https://cdt.org/wp-content/uploads/2022/10/algorithmic-transparency-ux-final-100322.pdf>

Recently, an investigation by Propublica documented how Google's sprawling automated digital ad operation placed ads from major brands on global websites that spread false claims on topics such as vaccines, Covid-19, climate change, and elections, particularly in languages other than English.[51] Considering that a significant source of revenue for many digital content platforms is advertisements,[52] such ad-placement practices that prioritize higher user engagement have made it profitable to host false, misleading and toxic content. Having transparency about the ad-placement practices of platforms will enable the regulator to call out harmful practices if it is found to fuel disinformation, and reorient the conduct of platforms. Such transparency is also important for users, as it helps them understand how content that is shown to them is monetized and by whom.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 25.5: "How users can access the complaints process;"	"25.5 How users can access the complaints process ***and data pertaining to the complaints, 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;*** ***(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) How content that is harmful, even if, lawful, is determined by the platform, and what actions will be taken with respect to such content;*** ***(v) Appeal procedure;*** ***(vi)Number of appeals and the number of cases in which the original decision was revised."***

[51] Craig Silverman, How Google Ads is funding misinformation around the world, Scroll.in. <https://scroll.in/article/1036862/how-google-ads-is-funding-misinformation-around-the-world >

[52] Megan Graham, Digital ad spend grew 12% in 2020 despite hit from pandemic, CNBC. <https://www.cnbc.com/2021/04/07/digital-ad-spend-grew-12percent-in-2020-despite-hit-from-pandemic.html >

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.[53] 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."[54] 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 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 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".[55]
- 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."[56]

---

[53] DAVID KAYE, SPEECH POLICE: THE GLOBAL STRUGGLE TO GOVERN THE INTERNET, 17 (Columbia Global Reports, New York, 2019). Also see, Paul Mozur, *A Genocide Incited on Facebook, With Posts From Myanmar's Military*, Washington Post (Oct 15, 2018). https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html; Nicolas Suzor 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 >

[54] World Press Freedom Day 2021, Information, As A Public Good, Unesco. <https://en.unesco.org/sites/default/files/wpfd_2021_concept_note_en.pdf>

[55] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E), § 4(1)(d).

[56] Article 29, Digital Services Act (European Union). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en >

- 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.[57]

3. On the Content Management Policies

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 27: "Any restriction upon content posted should be clearly set out in the platform rules, which should be implemented consistently, without arbitrary distinctions made between types of content or between users."	*" 27.* Any restriction on content posted should be clearly set out in the platform rules, ***with information about whether the restriction is in pursuance of a mandate in local law or if it is based on the platform's determination of any content as harmful, even if, legal, and implemented consistently, without arbitrary distinctions made between types of content or between users."*** ***…*** ***…*** ***27.3 Any change in the content management policies should be informed to the users from time to time.*** ***27.4 In case of unlawful content, the digital content platform should take all possible measures for removal of such content without undue delay and any other actions such as account suspension. In case of content that is determined as harmful albeit lawful, the platform should take actions short of removal such as downranking, reducing virality, and labeling the content.*** ***27.5 The rules should disclose the rationale behind placing restrictions on certain***

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

	content deemed harmful by the platform, although not considered illegal."

Justifications/References

The content management policies of digital content platforms should specify what restrictions are placed on content in pursuance of a legal mandate, and what restrictions are placed on the basis of a platform's own determination of content considered harmful. There are many types of content that may not often strictly fall under the prohibition of law but might nevertheless have harmful effects due to affordances of virality and amplification on digital content platforms. A wide range of content potentially falls into this category, including harassment, bullying, and mis- or disinformation ('fake news') that may hamper the ability of citizens to make informed decisions.[58] While the removal of such types of content may border on risks for constitutionally guaranteed freedom of expression, there are other ways to mitigate the harm from such content. In fact, the Guidance Document itself refers to these other ways i.e. restricting virality, and labeling [Paragraph 33.4]. But the regulatory framework is not clear about the categories of content that warrant removal, and those that warrant other kinds of actions.

Therefore, the Guidance Document should require that content management policies of platforms differentiate between unlawful content, and lawful but harmful content. While platforms should take all possible efforts to remove unlawful content, potentially harmful content could be subject to actions such as downranking, reducing virality, and labeling. Platforms should also proactively publish what factors or criteria are used in determining whether a content is potentially damaging, despite being legal. Such disclosure to the regulator will help in an external evaluation of the factor or criteria used, and pave the way for better criteria to be developed.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 27.1: "Platforms should, in policy and practice, through adequately trained and staffed personnel, ensure that, at a minimum, there is quick and decisiveaction against child sexual abuse materials, promotion of terrorism, promotion of genocide, clear threats of violence, gender-based violence and incitement to	After paragraph 27.1, add the following paragraph:: ***"27.X While framing their content management policies for determining what speech is permissible, digital content platforms should take into consideration the social, political, cultural, and religious context of the country in which they are operating, particularly, the unique disadvantages and locally specific forms of online abuse faced by certain social groups such women, children, youth, gender and sexual minorities and indigenous groups, and especially those falling at the intersection of***

[58] Reform of the EU liability regime for online Intermediaries, European Commission. <https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/649404/EPRS_IDA(2020)649404_EN.pdf>

hatred based on protected characteristics."	***marginalized identities."***

Justifications/References

The lack of cultural specificity in content moderation is exemplified by platforms' terms of service, such as Facebook's Community Guidelines and Twitter Rules, which are universal in their scope, and which usually borrow from Americanized frames of legitimacy of speech and which are not in tune with the realities of the Global South.[59] This could result in overlooking context-specific forms of abuse, and the unique disadvantages faced by a particular group or community, especially those with intersectional identities, and cultural references and textual nuances. It could also result in misrecognition and erasure of the voice of users from vulnerable social groups.

For instance, in India, caste is a protected identity marker, and caste-based discrimination and violence is widespread. Despite this, for many years the community guidelines of major social media platforms operating in India did not incorporate caste as a protected category in their hate speech and harassment policies.[60] Even when many platforms incorporated this, their forms for reporting harmful content still did not list caste.[61] This makes it difficult or even impossible to report context-specific forms of abuse.

Recognizing this, one of the foundational principles laid down by Santa Clara Principles is that the companies should ensure that their rules and policies and their enforcement, take into consideration the diversity of cultures and contexts in which their platforms and services are available and used.[62]

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 27.2 "27.2 There is often a tension between national laws and international human rights standards, which poses a challenge for any attempt to define global	"27.2 ~~There is often a tension between national laws and international human rights standards, which poses a challenge for any attempt to define global guidance on regulation. Should illegal content be defined in a jurisdiction that may violate international human rights law, the platform will be expected to report on how it~~

[59] Anita Gurumurthy and Amshuman Dasarathy. Profitable Provocations. A Study of Abuse and Misogynistic Trolling on Twitter Directed at Indian Women in Public-political Life (2022). <https://itforchange.net/sites/default/files/2132/ITfC-Twitter-Report-Profitable-Provocations.pdf >

[60] Kain, D., et al. 2021. Online Caste- Hate Speech: Pervasive Discrimination and Humiliation on Social Media. APC. <https://www.apc.org/en/pubs/online-caste-hate-speech-pervasive-discrimination-and-humiliation-social-media >

[61] Kain, D., et al. 2021. Online Caste- Hate Speech: Pervasive Discrimination and Humiliation on Social Media. APC. <https://www.apc.org/en/pubs/online-caste-hate-speech-pervasive-discrimination-and-humiliation-social-media >

[62] Santa Clara Principles on Transparency and Accountability in Content Moderation, Foundational Principle 3. < https://santaclaraprinciples.org/>

guidance on regulation. Should illegal content be defined in a jurisdiction that may violate international human rights law, the platform will be expected to report on how it responds to such requests."	~~responds to such requests.~~ ***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."***

Justifications/References

Compliance with national laws alone is inappropriate for digital content platforms that seek common norms for their geographically and culturally diverse user base.[63] Instead, human rights standards provide a useful framework for holding both States and companies accountable to users across national borders.[64]

International human rights treaties, along with setting out core rights, also set out the circumstances in which the rights may be limited or restricted.[65] 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 limitations 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);

(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

[63] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement >

[64] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement >

[65] EFF & Article 19, 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 >

be protected (paras. 14-15)."[66] 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.

4. On Enabling Environment

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 31.1: "31.1 Platforms should show what they do to provide an enabling environment that facilitates expression, that challenges false or misleading information, warns of offline consequences to speech that might be dangerous (e.g., hate speech) or simply flags different perspectives or opinions."	After paragraph 31.1, add the following paragraph: ***"31.X Platforms should conduct periodic risk assessments in order to identify and address any actual or potential harm or human rights impact of their operations. Apart from periodic review, such impact assessment should also be undertaken prior to any major design changes, or major decisions or changes in operations, or new activity or relationship, or in response to a major event or any change in the operating environment.[67] The assessment should be done internally, as well as by independent auditors. The findings of the assessment should be submitted to the regulator, along with details of the actions taken to address or mitigate the risks and human rights threats identified, and the impact of such actions."***
Justification/References:	

[66] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, UN General Assembly, A/HRC/23/40, p. Pp. 8-9. <https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf >

[67] Principle 18, Principles 17-21, Office of the High Commissioner for Human Rights. Guiding Principles for Business and Human Rights. <https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>

The regulatory framework should require platforms to conduct periodic risk assessments in order to identify and mitigate actual and potential harm to the information flow and communication ecosystem that may result from their functioning, design, policies and practices. This is essential to create an enabling environment by adopting a human-rights-by-design approach. It is to be noted that such risk assessment exercises are different from transparency and action-taken reports submitted by platforms, which are post-facto measures.

The UN Guiding Principles have specified conduct of human rights due diligence of its activities as an important obligation of businesses.[68] In addition to the periodic risk assessment, the Principles specifically call for such assessment prior to any major design changes, or major decisions or changes in operations, or new activity or relationship, or in response to a major event or any change in the operating environment.[69] The Windhoek Declaration also calls for technological companies to conduct human rights risk assessment including to identify threats to freedom of expression, access to information and privacy.[70] Similarly, The EU Digital Services Act requires that the recommender systems employed by the social media platforms be subjected to risk assessment, independent audits and risk mitigations measures to stem dissemination of illegal content, minimize negative effects for the exercise of fundamental right to freedom of expression and information, and on civic discourse, electoral processes, etc.[71]

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 31.2: "Where possible, users should be given the ability to control the content that is suggested to them - platforms should consider ways to encourage users' control over the selection of the content to be displayed because of a search and/or in news feeds. The availability to choose between content recommendation systems that display different sources and different viewpoints around trending topics should be made available to users in online platforms."	"31.2 Where possible, users should be given the ability to control the content that is suggested to them - platforms should consider ways to encourage users' control over the selection of the content to be displayed because of a search and/or in news feeds. The availability to choose between content recommendation systems that display different sources and different viewpoints around trending topics should be made available to users on online platforms. ***In the case of platforms hosting news content, the algorithms for default sorting should be diversity-optimized, prioritize credible news sources, and infuse serendipity in the interest of securing***

[68] Principles 17-21, Office of the High Commissioner for Human Rights. Guiding Principles for Business and Human Rights. <https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>

[69] Principles 18, Office of the High Commissioner for Human Rights. Guiding Principles for Business and Human Rights. <https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>

[70] Windhoek+30 Declaration. <https://en.unesco.org/sites/default/files/windhoek30declaration_wpfd_2021.pdf>

[71] Article 26, Digital Services Act (European Union). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en >

	information as a public good."

Justification/References:

The regulatory framework should require a higher standard of duty of care to the users from platforms hosting news content. This is to ensure access to plural and diverse news and information for all users so that they can effectively participate in democratic governance processes and collective debate.[72] A higher standard of care may require, as a UNESCO Report recommends, the development of curatorial responses to ensure that users can easily access journalism as verifiable information shared in the public interest, and prioritizing news organizations that practice critical, ethical independent journalism.[73] Where possible, users should be given the tools to verify the provenance and authenticity of information. Here again, transparency about the logic of news recommendation algorithms is vital. In the case of news aggregator platforms, users may be given the option to choose their own personalization settings or sorting logic.[74] But in the interest of securing information as a public good, it is recommended that the regulatory framework should require the default sorting logic of platforms to be diversity-optimized.[75]

5. On User Reporting

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 32: "[...] Where possible users should have access to a platform representative in their own country."	After Paragraph 32, add the following: "***(i) The details of the platform representative should be made available by the digital content platform to users through its policies;*** ***(ii) The role and responsibility of the platform representative should be clearly***

[72] Windhoek+30 Declaration. <https://en.unesco.org/sites/default/files/windhoek30declaration_wpfd_2021.pdf>

[73] Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression, Broadband Commission research report on 'Freedom of Expression and Addressing Disinformation on the Internet, p. 13. <https://en.unesco.org/publications/balanceact >

[74] Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse, Algorithm Watch.<https://algorithmwatch.org/en/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020-AlgorithmWatch.pdf>

[75] Lucien Hietz, Benefits of Diverse News Recommendations for Democracy: A User Study. <https://www.tandfonline.com/doi/full/10.1080/21670811.2021.2021804>

	outlined in the policies of the digital content platform; ***(iii) The regulator should play a role in monitoring the proper functioning of the platform representative and ensure that the representative keeps records and provides data to the regulator, as required, on the number of complaints and manner of resolution;*** ***(iv) If the platform representative refuses to take up a user's report, the user should have recourse to the regulator."***

Justifications/References:

The importance of an effective operational-level grievance mechanism[76] 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.[77] Details of the platform representative must be mandatorily made public by the digital content platform.[78]

Every digital content platform must clearly define the role and functions of the platform representative. The Report of the Special Representative of the UN Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises, John Ruggie, highlights the importance of how, for a complaint mechanism to be trusted and used, it should provide public information about the procedure it offers, specifically the timeframes, clarity on the types of process and outcome available and means of monitoring implementation.[79] For instance, based on a study piloting principles for effective company/stakeholder grievance mechanisms, it was concluded that a policy ensuring that users understand the progress of a complaint, while providing sufficient information about the mechanism's performance, will boost users' confidence.[80]

---

[76] Principles 29, Office of the High Commissioner for Human Rights. Guiding Principles for Business and Human Rights. <https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>

[77] Principles 31(a), Office of the High Commissioner for Human Rights. Guiding Principles for Business and Human Rights.<https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>

[78] 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>.

[79] UN Human Rights Council (2011), Report of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises, John Ruggie : Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, A/HRC/17/31. <https://digitallibrary.un.org/record/705860?ln=en>.

[80] Principle of Transparency and Predictability, UN Human Rights Council (2011), Piloting Principles for Effective Company/Stakeholder Grievance Mechanisms: A Report of Lessons Learned- Report of the Special Representative of the Secretary-General on the Issue of Human Rights and Transnational Corporations and Other Business Enterprises, John Ruggie: Guiding Principles on

The regulator must also play a role in monitoring the role of platform representatives by ensuring that they are effectively available for users. Further, ensuring that the platform representatives follow platform policies which require them to provide data on complaints and the manner of resolution, will also help in a regular analysis of the frequency, patterns and causes of grievances which can in turn enable the regulator to identify and influence policies, procedures or practices that should be altered to prevent future harm by digital content platforms. This is also an effective manner to understand and get feedback on the operational-level grievance mechanisms from those directly affected.[81]

Evidence of non-functioning[82] platform representatives in India has resulted in the creation of a legal entity called the grievance appellate committee or committees that will have the power to override the decisions of social media companies. Likewise, in the current framework of the Guidance Document, it is essential to hold platform representatives liable for their non-functioning as well as provide alternatives for users if they cannot approach the platform-appointed representative. It is also necessary to make it mandatory for platforms to provide details of the representative publicly, including the details of how to contact the officer.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 32: "[...] Companies in particular, in addition to the platform providing information about its policies accessible in a digestible format and In all relevant languages, it should show how it allows users to report potential abuses of the Policies, whether that be the unnecessary removal of content, the presence of violent or threatening content, or of any other content which is in breach of the policies."	"32.…Companies in particular, in addition to providing information about its policies accessible in a digestible format and in all relevant languages, should show how it allows users to report potential abuses of the policies, whether that be the unnecessary removal of content, the presence of violent or threatening content, or of any other content which is in breach of the policies. ***In case of content takedown in pursuance of a government order, the platform should inform the user of the reason for such takedown and what recourse they have against such government request under the law of the country. The platform should also preserve records of such government requests."***

Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework, A/HRC/17/31/Add.1. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G11/133/55/PDF/G1113355.pdf?OpenElement>

[81] Principles 29, Office of the High Commissioner for Human Rights. Guiding Principles for Business and Human Rights. <https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf>

[82] Press Trust of India, Social Media Platforms Not Adequately Redressing Grievances: Government, The Outlook.<https://www.outlookindia.com/business/social-media-platforms-social-media-platforms-not-adequately-redressing-grievances-government-news-204328>

Justification/References:

In the digital paradigm, as Frank La Rue, former Special Rapporteur on the freedom of opinion and expression, pointed out in his 2011 Report to the UN General Assembly (A/HRC/17/27), "the Internet has become a key means by which individuals can exercise their right to freedom of opinion and expression".[83] Keeping in line with the rationale of Paragraph 23.4 and 23.5, the onus also rests on the digital content platform to inform the user about why there has been a removal of content regardless of whether it is necessary or unnecessary. In situations where the takedown orders come directly from the government, users must be informed of the reasons for which the content has been taken down as part of their right to hold opinions, receive and seek information and exercise the freedom of expression as per Article 19 of the ICCPR.[84]

In 2018, David Kaye, the UN Special Rapporteur on freedom of opinion and expression called for "radical transparency"[85] for both online platforms and States; the kind of transparency which includes knowing what rules the States and companies use to moderate content, the rules regarding content, how those rules are applied, what kind of appeals process exists, and what kind of accountability exists for wrongful take down of content. The Special Rapporteur in his report suggested how transparency reporting should extend to government demands under the platform's terms of service and how companies should preserve records of requests made under these initiatives and the communications between the digital platform and the requester, and explore arrangements to submit copies of such requests to a third-party repository.[86]

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 32.2: "There should also be an effective user complaints mechanism to allow users meaningful opportunities to raise issues of concern. This should include a clear and	After paragraph 32.2, add the following paragraph: ***"32.X There should be processes in place that permit users to appeal against the decisions to remove content, keep content up that was flagged by users, suspend an account, or take any other type of action affecting users' human rights, including***

[83] Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue, UN General Assembly, A/HRC/23/40, p. Pp. 8-9.<https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf >

[84] Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue, UN General Assembly, A/HRC/23/40, p. Pp. 8-9.<https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/RegularSession/Session23/A.HRC.23.40_EN.pdf >

[85] UN Expert: Content moderation should not trample free speech, UN OHCHR.
< https://www.ohchr.org/en/stories/2018/07/un-expert-content-moderation-should-not-trample-free-speech>

[86] Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement >

understandable reporting channel for complaints and users should be notified about the result of their appeal."	***the right to freedom of expression.*** ***(i) The Appeal process should be easy and accessible to users, and information on how to access such processes and the timelines for the same should be published for users' benefit;*** ***(ii) Users should be informed about the nature of the body that hears the appeal - whether it is a body created within the digital content platform or an industry self-regulatory body;*** ***(iii) Users should be notified of the results of the review, and sufficiently informed of the reasoning to allow them to understand the decision;*** ***(iv) If content is taken down to comply with government requests, users should be informed about the recourse that they have against it, such as approaching a regulatory or judicial body."***

Justification/References

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.[87] 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.'[88] Hence, it is essential that the Guidance Document gives due emphasis on the appeal process and provides detailed guidance regarding the processes and safeguards available to the 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. It is instructive to refer to the Santa Clara Principles that recognize the appeal process as one of the core operational principles for platform companies and provides best practices to achieve a robust appeal mechanism.[89] An equally important information about the appeal process is the nature of the body that hears the appeal. Such a body may be one that is created by the digital

[87] Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018, p. 18. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement >

[88] Id.

[89] Santa Clara Principles on Transparency and Accountability in Content Moderation, Operational Principle 3. < https://santaclaraprinciples.org/>

content platform itself, such as the Oversight Board of Meta, or it may be an industry self-regulatory body.[90]

6. On Content That Potentially Damages Democracy and Human Rights, Including Mis- and Disinformation and Hate Speech

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 33.2: "Platforms should say how they define and respond to a wider set of damaging content through a systematic risk assessment. The regulatory system should assess if platforms are consistently applying their systems and processes to effectively enforce their own standards (including the protection of legitimate speech) which should be aligned to international human rights standards."	" 33.2 Platforms should ***~~say~~ disclose*** how they define and respond to a wider set of damaging content through a systematic risk assessment, ***especially of the curation and recommender algorithms that they employ, and their data collection practices***. The regulatory system should assess if platforms are consistently applying their systems and processes to effectively enforce their own standards (including the protection of legitimate speech) which should be aligned to international human rights standards."
Justification/References: Algorithms, targeted advertising, and data harvesting practices of the largest social media companies are largely credited with driving users towards extremist content and conspiracy theories, prioritizing unverified news, and news that provokes an emotional response.[91] This	

[90] Oversight Board. <https://oversightboard.com/>

[91] Disinformation and freedom of opinion and expression, Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan, April 13, 202, UNGA, A/HRC/47/25, p. 14. <https://digitallibrary.un.org/record/3925306?ln=en >

undermines the right to form an opinion and freedom of expression.[92] The use of algorithms can be attributed to the business model of platforms that is predicated on the quest for virality and maximum attention, which is exploited by bad actors to spread mis- and disinformation.[93]

Hence, in addition to being transparent about the deployment of curation and recommender algorithms and their logic, digital content platforms should also conduct a risk assessment of such algorithms and publish the results of such assessment.[94] This is vital to enable researchers and activists to understand the way in which algorithms promote certain kinds of content and to come up with measures to reorient the algorithmic logic so that it does not facilitate the spread of content that potentially damages democracy and human rights, including mis- and disinformation and hate speech.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 33.4: "Platforms should demonstrate how they would respond to potentially damaging speech – either by providing alternative reliable information, flagging concerns about the quality of this information, curbing its virality or any other means. Content removal or de-platforming of users should be considered only when the intensity, and severity of content that has the intention to harm a group or individual occurs. The platform should also be explicit about whether it partners with outside organizations or experts to help it make these kinds of decisions particularly in countries or regions where the platform itself has little local knowledge."	"33.4 Platforms should demonstrate how they would respond to potentially damaging speech – either by providing alternative reliable information, flagging concerns about the quality of this information, curbing its virality or any other means. Content removal or de-platforming of users should be considered only when the intensity, and severity of content that has the intention to harm a group or individual occurs. ***Platforms should also proactively publish what according to their policy is construed as potentially damaging content that may otherwise be legal.*** The platform should also be explicit about whether it partners with outside organizations or experts to help it make these kinds of decisions particularly in countries or regions where the platform itself has little local knowledge."

Justification/References:

Such disclosure of methods and criteria to the regulators and the public will help in reassessing the factors that platforms take into consideration and to improve their determination of potentially damaging speech, with minimal restriction of freedom of expression and access to information.

[92] Id.

[93] Kate Jones, Online Disinformation and Political Discourse Applying a Human Rights Framework, Chatham House, November, 2019, p. 34. <https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf>

[94] Please refer to our third recommendation under the Transparency Section.

Examples of potentially damaging albeit legal content could include misogyny, glorification of eating disorders, hatred against certain communities which may not meet the threshold of legal liability, and even information in relation to sale of products as has been included under the definition of "illegal content" under Article 3(h), Digital Services Act in order to address concerns regarding sale of counterfeit goods.

As discussed in the 2018 thematic report to the Human Rights Council of the Special Rapporteur on freedom of opinion and expression, social media companies should move away from generic, self-serving community guidelines and incorporate relevant principles of human rights law into content moderation standards.[95] It may not be possible to prescribe all relevant considerations that platforms take into account in making content decisions. At the same time, decisions by digital content platforms should not stem from opaque or 'secret rules', but from transparent criteria and processes that adhere to human rights norms.[96]

7. On Media and Information Literacy

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 34.1 "Platforms should implement specific media and information literacy Measures for women, children, youth and indigenous groups."	"34.1 Platforms should implement specific media and information literacy measures for women, children, youth, ***gender and sexual minorities,*** and indigenous groups."
Justification/References As mentioned in the UN Reports by the Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity, Victor Madrigal-Borloz, there has been a steep rise in the use of platforms by extremist political leaders and religious groups to	

[95] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement >

[96] Catherine Buni and Soraya Chemali, *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.>

promote bigotry, dehumanize persons based on their sexual orientation or gender identity (SOGI), and foster stigma and intolerance among their constituencies.[97] The freedom of trans and gender-diverse persons from policing of their gender identity and expression and the right of self-determination must be recognized and guaranteed through the workings of the independent regulator. Raising awareness of violence and discrimination against persons on the basis of their sexual orientation or gender identity, and identifying and addressing the root causes of violence and discrimination is part of the 2019 UN mandate[98] of the Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity.

8. On Data Access

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 38: "[...] There need to be safeguards with providing data access that ensures the protection of privacy and respect of commercial confidentiality. For platforms to build reliable interfaces for data access, there will need to be alignment among regulators that can determine what is useful, proportionate and reasonable for research and regulatory purposes."	"38[…] ***The regulator should be empowered to mandate platforms to share data with independent researchers, auditors, other regulators and oversight bodies, subject to safeguards to ensure protection of privacy.*** For platforms to build reliable interfaces for data access, there will need to be alignment among regulators that can determine what is useful, proportionate, and reasonable data sharing for research and regulatory purposes, ***including vetting of independent researchers, auditors or research projects for which data access can be granted."***

[97] UN Human Rights Council (2021), The law of inclusion, Report of the Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity, Victor Madrigal-Borloz, A/HRC/47/27. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/123/16/PDF/G2112316.pdf?OpenElement>; UN Human Rights Council (2021), Practices of Exclusions, Report of the Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity, Victor Madrigal-Borloz, A/76/152. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/192/14/PDF/N2119214.pdf?OpenElement>

[98] UN Human Rights Council (2016) Resolution on Protection against violence and discrimination based on sexual orientation and gender identity, A/HRC/RES/32/2<https://documents-dds-ny.un.org/doc/UNDOC/GEN/G16/154/15/PDF/G1615415.pdf?OpenElement>; Resolutions on sexual orientation, gender identity and sex characteristics, United Nations OHCHR.< https://www.ohchr.org/en/sexual-orientation-and-gender-identity/resolutions-sexual-orientation-gender-identity-and-sex-characteristics>

Justification/References

In keeping with the call for radical transparency by David Kaye, former UN Special Rapporteur on freedom of opinion and expression, platforms should be encouraged to share their data with independent researchers, auditors, supervisory authorities and other external stakeholders as these actors can help in identifying systematic risks from platform operations. This is especially helpful when regulators may lack the technical expertise to analyze and draw insights from the data produced by platforms in their transparency reports. Such sharing of data with different stakeholders will also help in understanding various trends relating to the use of digital content platforms by different sections of users and thereby help in relevant policy formulation. EU's Digital Services Act mandates platforms to comply with reasoned requests from the Digital Service Coordinator for the sole purpose of conducting research that contributes to the identification and understanding of systemic risks.[99]

Sharing of platform data, which may also include non-personal and anonymized data of users, raises concerns of privacy. Hence, in addition to requiring platforms to share data with external researchers, regulators should closely oversee and streamline the process of data sharing by platforms.

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 38: **"[...] There need to be safeguards with providing data access that ensures the protection of privacy and respect of commercial confidentiality. For platforms to build reliable interfaces for data access, there will need to be alignment among regulators that can determine what is useful, proportionate and reasonable for research and regulatory purposes."**	**After paragraph 38, add the following as a new paragraph:** ***"XX. 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."***

[99] Article 31, Digital Services Act (European Union). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en>

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 than the dominant platforms, and consequently, diverse information and viewpoints that users can have access to. This ties into para 15.3 in the Guidance Document that underscores the need for digital platforms to align to a common framework and develop coherent systems across regions to minimize internet fragmentation and protect freedom of expression and enhance the availability of accurate and reliable information in the public sphere. 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,[100] the Digital Markets Act requires digital platforms to provide effective portability of data generated through the activity of a user and provide them tools to facilitate the exercise of data portability.[101] Further, the Digital Markets Act also requires digital platforms to interoperate with each other.[102]

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.

9. On Regulator's Constitution & Powers

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 41: "Any regulatory system, whether a single body or multiple overlapping bodies, charged with managing online content (overseeing systems and processes) needs to be independent and free from economic or political pressures or any external influences.	"41. Any regulatory system, whether a single body or multiple overlapping bodies, charged with managing online content (overseeing systems and processes) needs to be independent and free from economic or political pressures or any external influences. Its members should be appointed through an independent merit-based process of appointment, overseen by

[100] Article 20, General Data Protection Regulation (European Union). <https://gdpr-info.eu/>

[101] Article 6, Digital Markets Act (European Union). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0842&from=en>

[102] Article 6, Digital Markets Act (European Union). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0842&from=en>

Its members should be appointed through an independent merit-based process of appointment, overseen by an oversight body (which could be the legislature or an independent board /boards). Its members should not seek or take instructions from any external body, whether public authority or private actor."	an oversight body (which could be the legislature or an independent board /boards). ***The regulator should be multi-sectoral, consisting of positions including Presiding Officer, full-time, part-time members, and secretary. The composition should also include stakeholders from related industries in the technology sector as well as civil society and NGO representatives. Qualifications of the members at different levels need to be clearly prescribed."***

Justification/References

There is a lack of detail and clarity regarding the working structure of the independent regulator in the Guidance Document. The nature of the regulator and some key characteristics – such as the composition, nature of positions like Presiding Officer, etc. – must be detailed. 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 [Paragraph 46]. The regulatory body should also consist of members from related industries in the technological sector so as to get a holistic picture, especially about the technical aspects.

One example of such a regulatory body is the National Regulatory Control Council (Nationaler Normenkontrollrat or NKR), established in Germany in 2006.[103] This is 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. Members are also mandated to have prior experience in legislative matters obtained through serving in State or society institutions, and required to have a knowledge of economic affairs. The Finnish Council of Regulatory Impact Analysis requires a formal membership criteria which is that the Council as a whole must have expertise in legislative drafting and the necessary expertise to assess different impact areas. In India, the Competition Commission of India, a regulatory body for competition related issues, mandates the requirement of a chairperson and not less than two and not more than six other members appointed by the Central Government.[104] As given in all these examples, every regulator must have a clear working structure, which is currently absent and must be expanded in the Guidance Document.

[103] Case Studies of RegWatchEurope Regulatory Oversight bodies and the European Union Regulatory Scrutiny Board, OECD. <https://www.oecd.org/gov/regulatory-policy/Oversight-bodies-web.pdf>

[104] Competition Act, 2000. <https://www.cci.gov.in/legal-framwork/act>

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 49: "The regulatory system will have the power to call in any digital platform service deemed not to be complying with its own policies or failing to protect users and after discussion , may recommend a specific set of measures to address the identified failings. Any such judgment should be evidence-based; the platform should have an opportunity to make representations and/or appeal against a decision of non-compliance; and the regulatory system should be required to publish and consult on enforcement guidelines and follow due process before directing a platform to implement specific measures. Failing to comply with this stage could lead to penalties which are proportionate, dissuasive, and effective (but excluding personal criminal liability)."	" 49. The regulatory system will have the power to call in any digital platform service deemed not to be complying with its own policies or failing to protect users and after discussion, may recommend a specific set of measures to address the identified failings. ***These measures will include imposition of fines, compliance orders, and warning orders.*** Any such judgment should be evidence-based; the platform should have an opportunity to make representations and/or appeal against a decision of non-compliance; and the regulatory system should be required to publish and consult on enforcement guidelines and follow due process before directing a platform to implement specific measures. Failing to comply with this stage could lead to penalties which are proportionate, dissuasive, and effective (but excluding personal criminal liability)."

Justification/References

Regulatory accountability mechanisms for corporations are an effective tool for oversight of digital content platforms and to end corporate impunity. However, current accountability initiatives in the international arena are largely hypothetical,[105] and there is hence a strong case to be made for institutional competencies at the national level. One successful legislation is the French law on 'Duty of Vigilance', which places the onus on large companies in France to identify and prevent risks to human rights and the environment that could occur as a result of their business activities.[106] Although France is the first country to adopt binding legislation of this sort, this development is part of a larger movement across Europe.[107] The law provides for a formal notice mechanism to order the company to comply with its vigilance obligations, including orders such as

[105] Notes: Ongoing TNC binding principles. Binding Treaty, Business and Human Rights. <https://www.business-humanrights.org/en/big-issues/binding-treaty/> See also, Note: Limitations of ratione personae of the ICC as well as difficulty in implementing UNSC resolutions domestically. Regis Bismuth. 2010. Mapping a Responsibility of Corporations for Violations of International Humanitarian Law Sailing between International and Domestic Legal Orders. Denver Journal of International Law and Policy, pp 120. <https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1210&context=djilp>

[106] France's Duty of Vigilance Law, Business and Human Rights Centre. <https://www.business-humanrights.org/en/big-issues/corporate-legal-accountability/frances-duty-of-vigilance-law/>

[107] Note: Similar legislation is currently being considered in Switzerland, where the necessary signatures have been collected for a referendum on mandatory human rights due diligence. Parallelly, igniting the same principles even the Dutch Parliament in 2017 adopted the Child Labour Due Diligence Bill, which if approved by the Dutch Senate, the law would require companies to

creation of a company vigilance plan. Penalties for non-compliance imposed by a court and civil liability measures that can be taken by an individual against the company for non-compliance of vigilance orders are useful to hold companies accountable for their actions.[108] Similarly, the regulator should detail the nature of the measures to take to address the failings and must detail them including imposition of fines, compliance orders, and warnings.

Paragraph Number & Proposed Text	**Revised/Recommended by ITfC**
Paragraph 50: "It will have the power to commission a special Investigation or review by an independent third party if there are serious concerns about the operation or approach of any platform or an emerging technology."	"50. It will have the power to commission a special Investigation or review by an independent third party if there are serious concerns about the operation or approach of any ***digital content*** platform or an emerging technology." ***50.1 The regulator should clarify the nature of the independent third party being hired and the purpose for which the particular entity is being hired.*** ***50.2 The regulator should outline the extent of powers, functions and procedures to be followed by the independent third party through a protocol/policy.*** ***50.3 The independent third party carrying out the investigation or review should be free from external or political influence and should mandatorily share details of their investigation with the regulator for scrutiny."***
Justification/References	

identify whether child labour is present in their supply chains and – if this is the case – develop a plan to combat it. Similarly, even the United Kingdom, in 2016, adopted the Transparency in Supply Chain Clause of the Modern Slavery Act. This provision requires companies domiciled or doing business in the UK to report on the measures they take to prevent slavery or human rights trafficking in their supply chains. See also, Switzerland to hold referendum on proposed human rights due diligence law, Business and Human Rights Centre. <https://www.business-humanrights.org/en/latest-news/switzerland-to-hold-referendum-on-proposed-human-rights-due-diligence-law/> Dutch Senate votes to adopt child labour due diligence law, Business and Human Rights. <https://www.business-humanrights.org/en/latest-news/dutch-senate-votes-to-adopt-child-labour-due-diligence-law/> Transparency in Supply Chains etc. A practical guide, Guidance issued under section 54(9) of the Modern Slavery Act 2015, UK Government. <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1040283/Transparency_in_Supply_Chains_A_Practical_Guide_2017_final.pdf.>

[108] France's Duty of Vigilance Law, Business and Human Rights, <https://www.business-humanrights.org/en/big-issues/corporate-legal-accountability/frances-duty-of-vigilance-law/>

Clarity about the independent third party that a regulator can hire is currently not addressed in the Guidance Document. For instance, there are multiple independent private entities that can be hired including an auditor, legal teams or even government investigation entities. In India, an example of a government investigation agency is the Serious Fraud Investigation Office ("SFIO"), an entity constituted under the Ministry of Corporate Affairs, which is the pioneer in conducting investigations in cases of fraud in companies.[109] The SFIO has the power to commence an investigation on the report of the Registrar within the Ministry in matters involving public interest on request by the State/Central Government or any department. Another example is under the Companies Act 2013, where the Registrar of Companies is empowered to conduct inspection, inquiry and investigation into the affairs of the company, on a scrutiny of any document filed by a company or on receiving any information in this respect.[110] Commonly, for such investigations, auditors, human resources or compliance personnel, in-house counsel, external lawyers and forensic/accounting firms are often seen as suitable options for conducting investigations.[111]

It is also important to clarify the manner in which the independent third party will be used if the regulator wants to commission a special investigation or review. The regulator must outline within a policy the reasons/purpose of hiring an independent third party. Some reasons to start an investigation or review could include an inquiry based on internal complaints made by any member of the organization, external complaints raised by stakeholders, complaints from whistleblowers who may or may not be a part of the said organization, disclosures made in the course of audit or on site inspections or suo motu cognizance of such offences taken by the regulator.[112] However, the scope of the same has to be detailed through a policy framework by the regulator.

If the objective of the regulator is to be independent and free from external or political influences while deciding to conduct investigations or reviews, it is essential to draw up clear and concise policies for the third party, including the manner of requesting information and protocols (including on-site inspections[113]) for the same, maintaining records of investigation details, privilege and confidentiality in investigations, and coordination with state investigation agencies. For example, a warning[114] through proper communication channels should be given to employees of digital content platforms before they are interviewed as part of the investigation, while ensuring a similar standard such as the attorney-client privilege regarding the information.

---

[109] Section 211, Companies Act 2013.

[110] Section 206-229, Companies Act 2013.

[111] The Contours of Conducting Internal Investigations in India, Nishit Desai. <https://www.nishithdesai.com/fileadmin/user_upload/pdfs/Research%20Papers/The_Contours_Of_Conducting_Internal_Investigations_In_India-PRINT-1.pdf>

[112] The Contours of Conducting Internal Investigations in India, Nishit Desai. <https://www.nishithdesai.com/fileadmin/user_upload/pdfs/Research%20Papers/The_Contours_Of_Conducting_Internal_Investigations_In_India-PRINT-1.pdf>

[113] Note: EU Digital Services Act's transparency requirements are part of a pervasive regulatory structure that even includes (at Article 54) on-site inspections very much like bank supervisory visits designed to provide some public oversight.

[114] UpJohn Warnings, American Bar Association. <https://www.americanbar.org/content/dam/aba/publications/litigation_journal/spring2016/upjohn-warnings.pdf>

Paragraph Number & Proposed Text	Revised/Recommended by ITfC
Paragraph 53: "Given the likely volume of complaints , the regulatory system will be expected to prioritize those complaints that demonstrate importance and relevance, systemic failings and/or substantial user harm. In this event the relevant regulator will have the power to Intervene and require action, including on an interim/urgent basis if necessary."	"53. Given the likely volume of complaints , the regulatory system will be expected to prioritize those complaints that demonstrate importance and relevance, systemic failings and/or substantial user harm. In this event the relevant regulator will have the power to Intervene and require action, including on an interim/urgent basis if necessary. ***The regulator should also demonstrate publicly why a certain complaint is being taken up and provide reasons for the same for elimination of bias and discrimination."***
Justification/References Digital content platform companies are increasingly taking actions on content issues without significant disclosure about those actions. As discussed in the 2018 thematic report by David Kaye, the UN Special Rapporteur on freedom of opinion and expression, "companies remain enigmatic regulators, establishing a kind of "platform law" in which clarity, consistency, accountability and remedy are elusive."[115] Ideally, companies should develop a compendium of its history of cases that would enable users, civil society and States to understand how the companies interpret and implement their standards. [116] An example of a form of decisional transparency can be seen through the workings of the Oversight Board of Meta as they publish the cases that they select publicly.[117] The Guidance Document can recommend that the independent regulator go one step ahead and mention the reason for the decision too and not keep the case selection process internal like how the Meta's Oversight Board does.[118]	

[115] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN General Assembly, A/HRC/38/35, Apr. 6, 2018. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement>

[116] UN Human Rights Council (2018) Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression A/HRC/38/35, para 63. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement>

[117] Oversight Board announces new cases related to Nigeria and India, Oversight Board.<https://www.oversightboard.com/news/389976813317756-oversight-board-announces-new-cases-related-to-nigeria-and-india/>

[118] Rulebook for Case Review and Policy Guidance, Oversight Board. <https://oversightboard.com/sr/rulebook-for-case-review-and-policy-guidance>