United Nations A/HRC/38/35/Add.1*

General Assembly

Distr.: General
6 June 2018

English only

Human Rights Council
Thirty-eighth session
18 June–6 July 2018
Agenda item 3
Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development

Report of the Special Rapporteur on the protection and promotion of the right to freedom of opinion and expression

Overview of submission received in preparation of the Report of the Special Rapporteur (A/HRC/38/35)**

* Reissued for technical reasons on 10 July 2018.
** The present report is circulated as received.

GE.18-09119(E)
1809119

Contents

Overview of Submissions Received in Preparation of A/HRC/38/35

1. This supplemental annex accompanies the June 2018 thematic report to the Human Rights Council of the Special Rapporteur on the protection and promotion of the right to freedom of opinion and expression (A/HRC/38/35). The report examines public and private regulation of user-generated online content. The Special Rapporteur examines the role of States and social media companies in providing an enabling environment for freedom of expression and access to information online. In the face of contemporary threats such as disinformation and online extremism, the Special Rapporteur urges States to resist restrictions on expression and adopt policies targeted at fostering vibrant online space in which individuals may seek, receive and impart information and ideas of all kinds. The Special Rapporteur also examines company content moderation and argues that human rights law gives companies the tools to regulate expression in ways that respect democratic norms and counter authoritarian demands.

2. A call for submissions was issued on 15 September 2017, and requested input from States to share information concerning laws, requests and demands to regulate platforms for and creators of online content. The call also requested input from civil society, companies and all other interested parties concerning State restrictions on user-generated content and the content moderation policies and processes adopted by companies. Twenty-seven States[1], one company[2], and twenty-eight non-governmental groups and individuals[3] made submissions to the Special Rapporteur.

3. This annex identifies concerns raised by States, civil society and other stakeholders in these submissions, providing a summary of trends and concerns shared with the Special Rapporteur. Readers are encouraged to look at the submissions themselves for more detailed information. This annex should also be read in conjunction with the Special Rapporteur's report, which articulates principles that he believes should guide State and company content regulation. Finally, this annex reflects only the submissions received and should not be understood as a broader literature review related to the topics discussed in the main report.

4. The Special Rapporteur expresses sincere gratitude to those who participated in the process leading to this report. The submissions referenced in this annex and the main report may be found at the website of the mandate (http://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/OpinionIndex.aspx) or at https://freedex.org.

A. State restrictions

5. Numerous submissions addressed State restrictions of the sharing and hosting of content on social media platforms.

[1] The Special Rapporteur received submissions from the following States: Argentina, Australia, Azerbaijan, Belarus, Bosnia and Herzegovina, Burkina Faso, Croatia, Cuba, Denmark, France, Honduras, Ireland, Italy, Lebanon, Qatar, Mauritania, Mexico, Morocco, Nigeria, Peru, Poland, Portugal, Romania, Slovenia, Spain, Togo and the United States.

[2] The Special Rapporteur received a submission from Github.

[3] The Special Rapporteur received the following submissions from civil society, academia, and others: 7amleh, Access Now, Amnesty International, ARTICLE 19, Association for Progressive Communications ("APC"), Center for Communications Governance, Centro de Estudios en Libertad de Expresión y Acceso a la Informació ("CELE"), the Danish Institute for Human Rights, Digital Rights Foundation ("DRF"), Electronic Frontier Foundation ("EFF"), Emily Laidlaw, European Digital Rights Initiative ("EDRi"), FGV Direito Rio, Global Network Initiative ("GNI"), Global Partners Digital ("GPD"), International Federation of Library Associations and Institutions ("IFLA"), Laura van der Woude, Marco Mart, Natasha Tusikov, Nicolas Suzor, OBSERVACOM, Open Technology Institute ("OTI"), Ranking Digital Rights ("RDR"), Southeast Asian Press Alliance, SOCICOM, Taiwan Association for Human Rights ("TAHR"), TEDIC, and WITNESS.

Intermediary Liability for User-Generated Content

6. The report discusses various legal frameworks that protect intermediaries from liability for user-generated content.

7. In the United States, the Communications Decency Act of 1996, 47 U.S.C. § 230(c) does not impose civil liability on providers of "interactive computer service[s]" that host or publish information about others, subject to certain exceptions. United States at 3.

8. Poland's Act of 18 July 2002 on the provision of services by electronic means illustrates how members of the European Union implement intermediary liability protections established under the E-Commerce Directive. Under the Act, platforms are exempt from liability for user-generated content only if the platform "does not know that the contents placed by a user (an author of a post) are unlawful" and "without undue delay, remove[s] or make[s] it impossible to access unlawful contents after learning that the contents are unlawful." Poland at 1-2. However, Internet users are encouraged to report unlawful content directly to social media companies, and platforms that fail to remove unlawful content after notification may be found liable for storing unlawful content. Poland at 2.

9. France has a similar protection against intermediary liability but requires technical operators to put in place "easily accessible and visible" functions that enable their subscribers to report illegal content. France at 2-3.

10. Multiple states have established laws that threaten intermediary liability protections. Germany's Network Enforcement Act (commonly referred to as "NetzDG") establishes significant penalties for companies that fail to remove content that violates specified hate speech laws. Access Now at 3. In March 2017, Azerbaijan adopted a law entitled, "On Information, Informatisation and Protection of Information," which makes Internet providers and website owners responsible for a broad range of online content, including abusive speech or slander. This law also requires intermediaries to remove illegal content within eight hours of notification. Azerbaijan at 1. On 21 March 2018, the United States Congress passed The Allow States and Victims to Fight Online Sex Trafficking Act, which has been criticized as an ineffective attempt to address the sex trafficking problem that unduly restricts intermediary liability protections.

11. In Taiwan Province of China, The National Communication Commission is considering the enactment of a "Digital Communications Act" ("DCA") that would limit the liability of Internet intermediaries under a "notice and takedown" standard. TAHR at 1.

12. Various submissions raise concern about the weakening of intermediary liability protections. ARTICLE 19 argues that such protections are "under threat" and, as a result, online platforms are "removing content more than ever before". ARTICLE 19 at 1. Global Partners Digital argues that "strict liability regimes are the most likely to result in overly broad restrictions of freedom of expression" by requiring proactive content moderation, but that "conditional liability" regimes can have a similar impact. GPD at 19-20. GPD provides a series of principles for States to consider when developing liability regimes. Id. at 20.

Regulatory Mechanisms to Monitor and Restrict Online Content

13. Specialized government bodies have been established to regulate and monitor digital content shared on social media and other Internet platforms. The report considers the human rights impact of Internet Referral Units ("IRU"), which some argue circumvent legal limits on State power to restrict expression by submitting content takedown requests under companies' Terms of Service. EDRi at 1. Using the Dutch IRU as a case study, Laurens van der Woude argues that IRUs violate the State's positive obligation to protect and ensure freedom of expression. van der Woude at 31.

14. Other types of regulatory bodies also deserve human rights scrutiny. Togo, for example, has created a new department to regulate "blogs, including social networks, Facebook, WhatsApp, Twitter and information and communication services." Togo at 1.

15. In Pakistan, the Pakistani Telecommunications Authority ("PTA") has broad authority to order the blocking of content that, among other things, violates the integrity of Islam or morality, or deemed to be contempt of court. The PTA has also banned encryption

and anonymity tools, and made public announcements warning citizens to exercise self-restraint in their online activities. DRF at 6.

16. In 2015, Australia established the Office of the Children's eSafety Commissioner, which is authorized to issue notices to social media services to remove "cyberbullying material" and certain categories of offensive content. Legislation to expand the Commissioner's mandate to address the non-consensual sharing of intimate images is also being considered at the time of publication. Australia at 3.

17. In Taiwan Province of China, the Institute of Watch Internet Network ("iWIN"), a government-funded organization, handles user complaints regarding content on social media platforms. Although its recommendations appear to be non-binding, concern has been raised that iWIN's recommendations are unclear. In many cases, companies and original content providers also treat these recommendations as orders, increasing the risk of improper removals. TAHR at 1-2.

18. In France, Internet users are able to report allegedly illicit online content via a dedicated portal known as the PHAROS platform, which was developed by the French company Thales in collaboration with the Ministry of the Interior. France at 4 -5.

Terrorism-related and Extremist Content

19. Several submissions raise concern about the significant pressure that social media platforms face to restrict terrorism-related and 'extremist' content, and the resulting weakening of intermediary liability protections. Access Now argues that both public and private actors' attempts to counter terrorism are typically unsuccessful due to unclear definitions, unintended consequences and a lack transparency. Access Now at 4. Poland's 10 June 2016 Act on Anti-terrorist Actions, for example, authorizes courts to order service providers to block access to specified data or services related or used to cause a terrorist event. Poland at 5-6. In France, decree n° 2015-125 of February 5th, 2015 establishes criteria for removing terrorism-related and extremist content and blocking relevant websites; complaints of unjustified removal or blocking may be reviewed by a judge. France at 3-4. Pakistan's Electronic Crimes Act, which was drafted in light of the government's National Action Plan to counter-terrorism, includes broad language that gives the Pakistan Telecommunications Authority power to restrict access to content that "glorifies an offence or a person accused of a crime, or supporting terrorism or activities of a terrorist organisation . . . " See APC at 3; DRF at 5-6.

20. States that have not directly imposed restrictions on online expression have nevertheless pressured intermediaries to adopt non-legally binding initiatives to curb extremist and other content. See Natasha Tusikov. In Australia, the government "maintains cooperative relationships with social media companies to encourage industry-led action to curtail the spread of terrorist and violent extremist content on their platforms." Australia at 4-5. In Israel, an alleged agreement between the authorities and Facebook to combat "incitement" online has reportedly led to disproportionate censorship of Palestinian users on the platform. 7amleh at 1. These initiatives delegate traditionally regulatory and police functions to private intermediaries, raising concern that they restrict freedom of expression in opaque and unaccountable ways. FGV Direito Rio at 47.

21. Growing State pressure has led companies to systematize and accelerate extremism-related content removals, including through the use of artificial intelligence. See WITNESS at 5, ARTICLE 19 at 5. However, accelerated removal processes increase the risk of unlawful or improper content removals, including the removal of evidence of war crimes. WITNESS at 7-8. For example, Facebook's algorithms have wrongfully flagged activist content as terrorist activity, such as content from Rohingya activists in Myanmar. ARTICLE 19 at 5.

Defamation, Sedition and Blasphemy

22. States increasingly seek to hold intermediaries liable for an ever-broadening range of speech-related offences, including defamation, sedition and blasphemy. In Paraguay, the Chamber of Deputies has proposed legislation that would require suppliers of applications and social networks to suspend and remove publications with offensive or defamatory character. TEDIC at 2. Civil society groups have expressed concern that the absence of

judicial review of content removal orders will lead to excessive censorship that disproportionately affects dissenting voices. Id.

23. In Southeast Asia, many government requests for content removal are based on existing sedition, blasphemy, defamation, and lèse majesté laws. SEAPA at 1. SEAPA's submission contains numerous examples of these removals, including Facebook blocking a British journalist's post about the royal family in accordance with Thailand's lèse majesté law, and Twitter suspending the accounts of four Islamic Defender Front (IDF) members' accounts after the organization called for the dismissal of an Indonesian police chief. SEAPA at 2-8.

Hate Speech

24. Several States impose criminal liability on the posting and hosting of online content that promotes or incites discrimination and hatred. Qatar criminalizes the publication of online content that creates dissension among members of society or stirs up sectarian, racial or religious strife, with penalties of up to six months imprisonment and/or a fine of up to three thousand riyals. Qatar at 2. Article 295 of Cuba's Criminal Code provides criminal penalties for individuals who promote or incite discrimination by reason of sex, race, color, or national origin. Cuba at 3. Croatia, pursuant to Article 325 of the Criminal Code, provides for a maximum three-year prison sentence for the "offense of racism and xenophobia committed by computer systems." Croatia at 1. Germany's NetzDG Law is an example of hate speech legislation that extends liability to the intermediaries that host such content.

Copyright

25. Italy's Communication Regulatory Authority (AGCOM) was established to, *inter alia*, fight online piracy and breaches to copyright. Italy at 3. If AGCOM detects infringement of copyright law online from a server located in Italy, it can order the content removed from the website. Id. at 5. If the server is outside Italy, AGCOM can only order access providers in Italy to disable access to the website by blocking the DNS resolution or the IP address. Id. The European Union is currently considering copyright reform that has attracted criticism for its potentially adverse impact on freedom of expression. Article 13 of the proposed EU Directive on Copyright in the Digital Single Market, which is "intended to address mass unauthorized distribution of audio and visual works," will "require platforms used to share code to proactively monitor content using upload filters." Github at 8. Qatar has also established laws that prohibit the infringement of intellectual property rights on online media. Qatar at 1.

Online Disinformation, Propaganda and False News

26. Many submissions discussed the human rights impact of online disinformation, State-sponsored Internet propaganda and "fake news," as well as efforts to regulate them. Submissions addressed how disinformation interferes with the individual exercise of freedom of expression. For example, Access Now observes that disinformation campaigns via websites and social media have been employed to "discredit or silence dissenting voices, demonize opposition, and disseminate propaganda." Access Now at 5. APC notes that disinformation campaigns can disrupt democratic elections. APC at 4.

27. Growing concern about online disinformation has led to a variety of regulatory approaches. In Italy, for example, AGCOM has established the "*Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali*" to promote self-regulation to tackle online disinformation. Italy at 2-4. Taiwan Province of China has collaborated with major social media companies and NGOs to establish a third-party fact-checking organization to combat "fake news," but this arrangement has raised transparency and accountability concerns. TAHR at 1. French law prohibits the bad faith publication of falsehoods that disturbs (or is likely to disturb) the public peace. However, the government is preparing additional legislation to combat disinformation. France at 8.

28. Several submissions emphasize the importance of ensuring State and corporate transparency in their approaches to online disinformation. For example, APC raised

concern about the opacity of platform-designed tools to combat "fake news," such as Facebook's newsworthiness ranking system. APC at 4-5. Access Now has urged platforms to ensure meaningfully transparent enforcement of their terms of service to discourage fake news. Access Now at 18.

29. WITNESS notes that "fake news" and propaganda often target vulnerable populations and activists, with human rights defenders who risk their lives to document and report on human rights abuses on the ground dismissed as "fake news." WITNESS at 2. In addition to enhancing outreach to marginalized communities, platforms should hire content moderators to investigate bogus allegations of "fake news." Id. at 9.

Right to be Forgotten ("RTBF")

30. The right to be forgotten ("RTBF") is commonly associated with the right to de-list or de-link, which allows users to request the removal of a search result (from a search engine) that "appears to be inadequate, irrelevant or no longer relevant or excessive... in light of the time that had elapsed." However, it has also been expanded to include the right to erasure (for example, when users request that all their personal data be deleted when they leave a service). See Access Now at 6. For example, South Korea's implementation of the RTBF goes further than the European Union, requiring content to be deleted entirely and not merely de-listed. Id.

31. The basis of this right and its limits under international human rights law are the subject of ongoing debate. Access Now argues that the right to de-list should be limited to "circumstances where the sole objective is protection of the personal data of non-public figures," should only be wielded by the data carrier, and should never lead to the actual deletion of content. Id. at 7. ARTICLE 19 argues that implementation of the right should strike an appropriate balance between freedom of expression and data protection. ARTICLE 19 at 3. Similarly, while APC recognizes that the right to be forgotten is beneficial for data privacy, it cautions that the right should be rooted in data protection frameworks with robust "procedural safeguards and limitations that protect against the de-listing of information in the public interest." APC at 7, In contrast, IFLA opposes the right to de-list on the grounds that information should generally not be intentionally hidden, removed or destroyed. IFLA at 1.

32. Several submissions call for greater transparency concerning RTBF removals. See e.g. IFLA 2-3. ARTICLE 19 urges data controllers to take "reasonable steps" to notify content providers that their content has been de-listed under the right to be forgotten framework. ARTICLE 19 at 2-3.

Other Categories of Content-based Restrictions

Gambling

33. Romania's National Gambling Office (ONJN) monitors online websites to ensure compliance with legal restrictions on gambling. Romania at 3.

Drugs

34. Romania's Ministry for the Informational Society may request the blocking of websites concerning products susceptible of having psychoactive effects (e.g. products, substances, plants etc. similar with drugs and psychotropic substances). Id. at 3-4.

Pornography:

35. Togo, Spain, Romania, Belarus and Mauritania prohibit the publication and distribution of pornographic materials online. See Belarus at 1; Mauritania at 2; Romania at 2-3; Togo at 3.

Global Removals

36. Legal requests for content removals outside the jurisdiction where the request is made raise questions about their transnational impact on freedom of expression. EFF recommends that a company facing a global removal request should weigh its "obligations

towards that jurisdiction, against its obligations to uphold the human rights of its users in other jurisdictions—and, perhaps, its conflicting legal obligations from other jurisdictions." EFF at 2-3. In the context of the right to be forgotten, APC argues that "[n]o single government should be able to decide what people in the rest of the world can see in their search results." APC at 7.

B. Company policies and processes

The Role of Intermediaries

37. Several submissions discussed the role of intermediaries in mediating and facilitating the exercise of freedom of expression and related human rights on the Internet. GPD argues that platforms are neither "creating content as such" nor "passive, neutral hosts of content generated by their users," instead operating "somewhere between these two extremes." GPD at 3. OBSERVACOM argues that large intermediaries should be "subject to public obligations" because of their significant market power and monopoly on essential services. OBSERVACOM at 3. Similarly, SOCICOM argues that private actors operating at the "content layer" are properly subject to legislation responsive to, among other things, their market effects, privacy concerns and systemic inequalities. SOCICOM at 3. APC notes that current intermediary liability regimes designed for "passive intermediaries" may dis-incentivize companies from assuming greater content moderation responsibilities for "fear of a potential loss of protection." APC at 2. Nicolas Suzor argues that the predominantly contractual relationship between users and platforms are disconnected from the latter's "central role in public communication," which demands "messy contestation" of the limits on platform power based on rule of law principles. Suzor at 6-8.

Responsibilities of Intermediaries Concerning State Restrictions

38. Several submissions address the challenges that platforms face when determining how to respond to State restrictions. EFF, for example, recognizes that while companies should generally comply with State laws when they have a physical presence in the country, they should nevertheless challenge restrictions that are inconsistent with human rights "by lawful means" and, in some cases, avoid establishing physical presence in that country. See EFF at 1-2. GNI does not require member companies to violate domestic laws when those laws are inconsistent with human rights standards, but recommends that "companies should avoid, minimize, or otherwise address the adverse impact of the government demands, laws, or regulations, and seek ways to honor the principles of internationally recognized human rights to the greatest extent possible." GNI at 2. GPD notes that company responses to government restrictions should be guided by the UN Guiding Principles on Business and Human Rights; it also argues that it is "neither realistic nor fair to expect [companies] to refuse to comply with national laws and other measures imposed by governments, even where such laws or measures are inconsistent - or may be inconsistent - with international human rights law." GPD at 6.

39. Various submissions indicated specific principles of corporate human rights responsibility that platforms should integrate into their interactions with governments. ARTICLE 19 urges platforms to challenge content restrictions that lack legal basis or are disproportionate in a court of law, appeal domestic court orders that violate human rights law, and resist informal government requests based on Terms of Service. ARTICLE 19 at 3. Access Now argues that companies should establish policy commitments to human rights, enhance staff training on content moderation, increase transparency about government requests, and undertake human rights due diligence that ensures that government requests do not circumvent freedom of expression safeguards. Access Now at 4-5.

Responsibilities of Intermediaries Concerning Content Restrictions under Terms of Service

40. As a general matter, submissions indicate that platforms have a responsibility to respect freedom of expression and related human rights when they restrict user-generated content under their Terms of Service. See e.g. GPD at 11; APC at 18. The UN Guiding Principles provide a framework for implementing the responsibility to respect,

encompassing high-level policy commitments to human rights, due diligence and remediation. Id. Multi-stakeholder fora such as GNI provide opportunities to develop industry-specific guidance based on the UN Guiding Principles. GNI at 1.

41. The corporate and State roles in respecting and protecting freedom of expression should be mutually reinforcing. CELE argues that, in the Inter-American system, permitting intermediaries to interpret and enforce Terms of Service in an "arbitrary, obscure or ambiguous way" could amount to a violation of the State's duty to prevent violations of freedom of expression through reasonable means. CELE at 6.

42. Mainstream corporate human rights responsibility discourse has attracted criticism for providing weak or inadequate protection of human rights. Emily Laidlaw cautions that human rights standards rooted in corporate social responsibility frameworks typically lack "the standards and compliance mechanisms needed to be a credible and sustainable framework for speech regulation in the communications technology sector." Laidlaw at 12. Research conducted by The Danish Institute for Human Rights shows that company commitments to human rights only extend to "protecting against external threats from governments." Danish Institute for Human Rights, Rikke Jørgensen, *Framing human rights: exploring storytelling within internet companies* at 4.

Transparency

43. Transparency is a critical element of the corporate responsibility to respect human rights, particularly in the context of due diligence and providing remedies for unlawful or improper content removals. Many submissions indicate that online platforms lack meaningful transparency about their standards and processes for restricting content and challenging improper restrictions. Users therefore lack the information necessary to regulate their conduct on these platforms. See e.g. WITNESS at 3, OBSERVACOM at 11, ARTICLE 19 at 19.

44. In particular, company explanations of why content has been removed or otherwise restricted are frequently opaque or non-existent. ARTICLE 19 observes that users are only notified of restrictions after the fact, on a discretionary basis instead of as a matter of consistent policy. ARTICLE 19 at 9. In the context of the right to be forgotten, a search for de-listed names on Google's search engine typically returns a generic statement indicating that content may have been removed. EFF at 6-7. There is also a general lack of clarity regarding how users should access complaints mechanisms to challenge wrongful removals. ARTICLE 19 at 9.

45. Submissions proposed a variety of recommendations for establishing meaningful standards of transparency. ARTICLE 19 urges companies to publish their internal guidelines for content removals, and data about Terms of Service enforcement in a disaggregated format. ARTICLE 19 at 9. EFF observes that it is possible for platforms "to identify the removal of information by inserting a note at the location from which the information was removed," citing takedowns under the US Digital Millennium Copyright Act as an example of this approach. EFF at 6. WITNESS recommends platforms to inform users of what content was removed, the reason for removal, and any relationship between State requests and removals in direct messages to users. WITNESS at 10. According to Github, it publishes, in real time, all notices that lead to takedowns. GitHub at 2-3. RDR argues that governments should encourage or require a high level of transparency from Internet companies operating in their jurisdiction and must themselves also be transparent about the demand and requests they submit to platforms. RDR at 2.

Bias and Non-Discrimination

46. Without clear and predictable standards for content restrictions, Terms of Service enforcement, many submissions suggested, may replicate and amplify offline biases and patterns of discrimination against users based on their protected characteristics, such as gender, sexual orientation, race, ethnicity, religion and national origin. Such discrimination may be amplified in "local or hyper-local" contexts that companies do not adequately understand. Center for Communication Governance at 4. Submissions provided numerous examples of such incidents: see e.g. Access Now at 10; Amnesty International at 2; APC at

11; WITNESS at 3; TEDIC at 5. Government officials may also exploit ambiguities in a platform's content restriction standards to target critics, political opposition and activists. WITNESS at 5; SEAPA at 1. CELE observes that the challenges of combating abuse on platforms stem in part from variances in definitions of abusive behavior and approaches towards regulating such behavior. CELE at 2.

47. Various submissions encouraged platforms to provide users with the tools they need to protect themselves against abusive behavior online. Such tools include "filters, blocklists, and reporting mechanisms," EFF at 3, and additional layers of password security for users disproportionately at risk of human rights violations, Access Now at 12. Such tools are typically more effective than the "blanket laws or policies that attempt to regulate speech." EFF at 3. In the context of real name requirements, platforms should allow the use of pseudonyms in "appropriate circumstances." Access Now at 12.

Automation

48. In order to handle the overwhelming volume of content created and shared on platforms, most companies rely on the use of automation to moderate content. However, opaque and unaccountable use of algorithmic decision-making and other forms of automation have led to over-blocking, discrimination and bias. Amnesty International at 2; Access Now at 14-15; ARTICLE 19 at 7-9; OTI at 9. Excessive reliance on algorithms has unduly restricted depictions of violence in conflict zones, curtailing efforts to document evidence of war crimes, Access Now at 14-15, and illegitimately targeting content provided by human rights defenders, WITNESS at 4.

49. Multiple submissions recommend that automated content moderation processes should integrate appropriate levels of human review to address complex issues of context. Access Now at 14-15; ARTICLE 19 at 8; GPD at 17; IFLA at 6; and WITNESS at 4. In particular, Access Now indicates that "companies should not rely exclusively on automated systems" and that companies "should implement a procedure that combines use of algorithms and human evaluation, and, crucially, is situated within a framework that is grounded in international human rights law and standards." Access Now at 15.

Appeals and Remedies

50. Under the UN Guiding Principles, non-judicial grievance mechanisms should be legitimate, accessible, predictable, equitable, transparent, rights-compatible, a source of continuous learning, and based on engagement and dialogue. UN Guiding Principles, Principle 31. Various submissions discuss how platforms should develop mechanisms for appealing and remedying improper content restrictions consistently with these principles. At a minimum, users must be notified that their content was restricted and the basic reasons for such restriction. ARTICLE 19 at 7. Terms of Service should also clearly identify how users can "appeal mistaken or inappropriate restrictions, takedowns or account suspensions." EFF at 5. Remedial mechanisms should also be "broad enough to cover a range of complaints that users may submit." RDR at 16. Platforms should use data gleaned from their appeals process to continuously improve policies and processes in order to minimize adverse impact of freedom of expression, while engaging in dialogue with all relevant stakeholders to ensure confidence and legitimacy in an appeals process. If necessary, companies should consider adjudication of appeals through a third-party system to avoid conflict or bias. GPD at 15.

51. State regulation may have a positive role in ensuring companies provide meaningful remedies for improper content restrictions. South Korea, for example, requires companies to provide an appeals mechanism when content is removed in response to defamation claims. RDR at 17.

C. Privacy

52. Although the report focuses on platform content regulation and its impact on freedom of expression, it is noteworthy that a few submissions discussed the privacy concerns associated with large-scale data collection and analysis on these platforms and other digital spaces mediated by private parties. For example, IFLA raises concern about

the human rights impact of the digitization of library records by third-party vendors and the collection of e-book histories. IFLA at 4. Several States have established specialized government bodies that address privacy concerns associated with commercial data collection and analysis. Like all other EU countries, Italy has established a Data Protection Authority (DPA), an independent administrative authority that handles individual complaints concerning unfair or unlawful data processing and supervises compliance with privacy and data protection legislation. Italy at 4-5.

D. Summary of Special Rapporteur's consultations

53. The Special Rapporteur's report also benefited from a range of consultations with civil society, companies, academics and other stakeholders. The following material summarizes some of these consultations.

March 2018 Consultation (Middle East and North Africa)

54. In a video conference with civil society representatives based in the Middle East and North Africa, participants raised concern about the lack of transparency and accountability concerning content moderation on social media platforms, and the lack of consultation with civil society in the region on issues associated with moderating local content. For example, Twitter appears to have an algorithm that triggers the suspension of an account with a significant number of followers that are "bot" accounts. Bad actors trying to game this algorithm have mobilized masses of "bot" accounts to follow accounts belonging to prominent activists and human rights defenders, triggering their suspension. Even though civil society has attempted to contact Twitter about these incidents, the company has not been sufficiently responsive. Participants also urged companies to provide clear guidance on the processes for flagging, removal and appeal.

55. Participants expressed concern about the revolving door between government agencies and major social media companies, which enhances the risk of government or political bias in how company policies and processes are developed and implemented in the region.

56. Participants also expressed concern about the reliance on automation to flag or remove extremist or terrorism-related content in the absence of clear definitions of extremism and terrorism. According to one participant, thousands of videos documenting the Syrian war have been removed from YouTube since it announced the use of automated tools to enhance the speed and efficiency of content removals. Participants urged companies to provide meaningful information about how and under what circumstances they rely on automation to moderate content. Even though human review should be a critical aspect of content moderation, participants emphasized that merely adding "warm bodies" to the content moderation workforce was not enough; meaningful training, at the least, must also be provided.

March 2018 Consultation (Africa)

57. In a video conference with civil society representatives from Eastern, Central, and Southern African regions, participants raised concern about the lack of transparency concerning content removal standards and processes. Many voiced a sense that large global platforms did not demonstrate sufficient respect or understanding for the particularities of African cultures. For example, images of topless tribal women are often removed on grounds of nudity. Language barriers were another concern, as takedown notices and appeals processes are not always available in the user's native language, and often unclear and confusing. Participants expressed that platforms should work with local communities to review flagged posts that raise issues of context, particularly when they contain a mix of English and non-English content. Participants also suggested providing Terms of Service in as many local languages as possible.

58. While the lack of company engagement with local communities was frequently cited as a cause for concern, participants acknowledged Facebook and Google's engagement efforts. However, participants noted that while local representatives of both companies were willing to listen to their concerns, they also indicated that critical decisions were

usually escalated and made at global headquarters. Concerns were expressed that companies did not sufficiently incorporate local input into high-level executive decision making.

59. Participants also discussed a range of blunt State restrictions on online expression. In South Africa, for example, Parliament is considering legislation that would empower the Film and Publication Board to rate content prior to its publication, raising concerns of pre-censorship and upload filtering. The persistence of Internet shutdowns in the region also raises serious human rights concerns. Participants urged companies to adopt transparent guidelines on how they respond to government requests and other forms of government pressure.

February 2018 Consultation (Latin America)

60. In a video conference with civil society representatives based in Latin America, participants expressed concern about the systematic failure of large social media platforms to take into account local and regional particularities in the enforcement of global standards for content removal. For example, photos of South American indigenous/native tribal members have been repeatedly removed on grounds of nudity. Participants suggested that this lack of sensitivity to context could be partly attributable to the lack of local presence and meaningful and sustained engagement with civil society, government officials, and other local and regional stakeholders.

61. Another overarching concern was the spread of "fake news" and proposals to regulate it, both of which could undermine civic engagement and participation during upcoming elections in the region. Participants urged caution in the development of regulatory frameworks to prevent the spread of "fake news," given the high risk of censorship. One participant indicated that government initiatives to combat "fake news" were often opaque and inaccessible to the public.

62. Participants also discussed approaches to guaranteeing meaningful access to remedy when content is improperly removed on social media. One suggestion was the development of an independent administrative authority (such as an ombudsman) to handle user complaints and convene consultations between social media companies and civil society. Others discussed the need to provide publishers with legal avenues to challenge discrimination or censorship arising from content curation on platforms. While several participants expressed reservations about the proposals discussed, most agreed with the need for meaningful discourse on access to remedy.

October 2017 Consultation (South and Southeast Asia)

63. In October 2017, the Special Rapporteur attended a conference in Bangkok, Thailand, that focused on digital freedom of expression issues in South and Southeast Asia. During this conference, the Special Rapporteur received input from several civil society groups that helped inform the main report. Participants provided keen insight into trends in social media use and adoption in the region.

64. Participants alerted the Special Rapporteur to an increasingly repressive and punitive environment for online expression in the region. In particular, participants expressed concern that defamation, blasphemy, sedition, and *lèse majesté* laws were being used to impermissibly restrict legitimate expression online. In a growing number of countries, legislative proposals to curb "fake news" and extremist content also increase the risk of government overreach and censorship.

65. Participants also discussed the role of private companies in online content moderation. The lack of meaningful company engagement with local civil society groups was an overarching concern. Users that raised questions to large social media platforms about the specific reasons for a suspension, ban, or removed content usually did not receive satisfactory responses. Participants also expressed significant concern about platforms' acquiescence to government demands, and the general lack of transparency for appealing platform decisions to remove content.