REORIENTING RULES FOR RIGHTS

Summary of the report on online content regulation by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/38/35)

Prepared by the Association for Progressive Communications, June 2018

The internet is the greatest tool in history for global access to information and expression. Internet companies have become central platforms for discussion and debate, information access, commerce and human development. Companies running platforms are enigmatic regulators, establishing a kind of "platform law" in which clarity, consistency, accountability and remedy are elusive.

States have a significant impact on how companies deal with online content regulation. Companies face increasing pressure from states to comply with state requests (both legal and extralegal) to moderate or remove content and are also taking pre-emptive measures through, for example, adaptations to their terms of service agreements (ToS). In response to these worrying trends the UN Special Rapporteur on freedom of opinion and expression proposes a series of measures states and companies can undertake to put human rights at the very centre of online content moderation.

Do human rights principles and standards apply to online content regulation?

Freedom of expression (FoE) is protected by Article 19 of the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights. The United Nations, regional organisations and treaty bodies have affirmed that offline rights apply equally online. Any restrictions placed on the exercise of the right to FoE must be legal, necessary, proportionate and legitimate. More importantly, the restrictions placed must not undermine or jeopardise the essence of the right. Though states are primarily the duty bearers to enforce and protect these rights, non-state actors like internet companies cannot shy away from playing their part in the r ealisation of “the right to hold opinions without interference” and “the right to seek, receive and impart information and ideas of all kinds, regardless of frontiers” and through any medium. This includes the internet.

The activities of companies in the information and communications technology (ICT) sector implicate the rights to privacy, religious freedom and belief, opinion and expression, assembly and association, and participation in public life, among others. The Guiding Principles on Business and Human Rights, adopted by the Human Rights Council in 2011, place a duty on states to ensure environments that enable respect for human rights on the part of businesses, who must strive to ensure that their policies and practices adhere to the Principles in letter and spirit. By applying human rights in their work, they would not be restricted; to the contrary, it would offer a globally recognised framework for designing tools and a common vocabulary for explaining their nature, purpose and application to users and states. Human rights law also gives companies the tools to articulate and develop policies and processes that respect democratic norms and counter authoritarian demands.

What are the problems emerging from online content regulation?

STANDARDS NOT ROOTED IN HUMAN RIGHTS

Most companies do not recognise their human rights obligations and as a result do not explicitly base content standards on any particular body of law that might govern expression, such as national law or international human rights law. Few companies apply human rights principles in their operations, and most that do see them as limited to how they respond to government threats and demands.

GOVERNMENT PRESSURE AND VAGUE LAWS

While states require companies to restrict illegal content, they also often rely on censorship and criminalisation to shape the online regulatory environment. States use broadly worded restrictive laws, vague or complex legal criteria without prior judicial review, and the threat of harsh penalties to compel companies to restrict content and suppress legitimate expression. The commitment to legal compliance can be complicated when relevant state law is vague, subject to varying interpretations, or inconsistent with human rights law.

EXTRATERRITORIAL REQUESTS

Some states are demanding extraterritorial removal of links, websites and other content alleged to violate local law, which would allow censorship across borders, to the benefit of the most restrictive censors.

EXTRALEGAL REQUESTS

State authorities increasingly seek content removals outside of legal process or even through ToS requests and have established specialised government units to refer content to companies for removal. States also place pressure on companies to accelerate content removals through non-binding efforts, most of which have limited transparency, exacerbating concerns that companies perform public functions without the oversight of courts and other accountability mechanisms.

DISINFORMATION

Companies face increasing pressure to address disinformation spread through links to bogus third-party news articles or websites, fake accounts, deceptive advertisements and the manipulation of search rankings, which might not always be feasible.

USERS KEPT IN THE DARK

Company disclosure about removal discussions, in aggregate or specific cases, as a result of human evaluation is currently limited and must be reported on adequately. Users who post reported content, or persons complaining of abuse, often do not receive any notification of removal or other action or have any avenues to challenge removals. Even with appeal, the remedies available to users appear limited or untimely to the point of non-existence and, in any event, opaque to most users and even civil society experts.

AUTOMATION AND OVERBLOCKING OF CONTENT

Demands for quick automated flagging, removal and pre-publication filtering sometimes result in overblocking and disproportionate censorship. Devoid of context, this approach has led to removals of depictions of nudity with historical, cultural or educational value; historical and documentary accounts of conflict; evidence of war crimes; counter speech against hate groups; and efforts to challenge or reclaim racist, homophobic or xenophobic language.

HATE SPEECH AND MARGINALISATION OF VULNERABLE GROUPS

Company policies on hate speech, harassment and abuse do not clearly indicate what constitutes an offence. The vagueness of hate speech and harassment policies has triggered complaints of inconsistent policy enforcement that penalises minorities while reinforcing the status of dominant or powerful groups. Steps taken by platforms have resulted in the suppression of lesbian, gay, bisexual, transgender and queer expression, as well as advocacy against repressive governments, reporting on ethnic cleansing, and critiques of racist phenomena and power structures. Misogynist or homophobic harassment designed to silence women and sexual minorities and incitement to violence of all kinds continue to thrive in online spaces, which has a significant impact on the offline realities of the people targetted.

REAL NAME POLICY

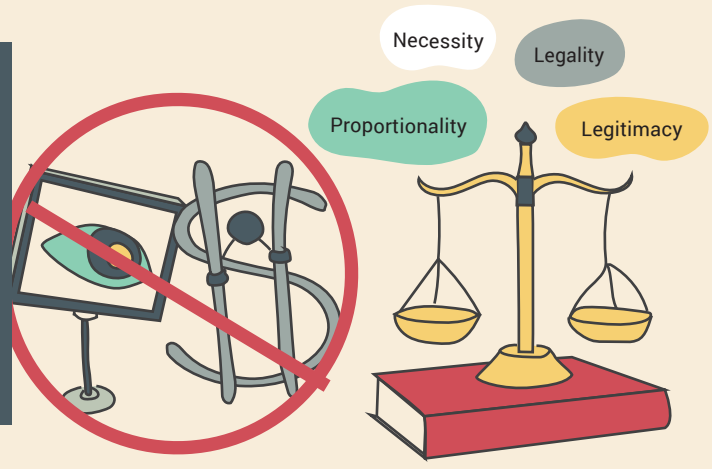
Strict insistence on real names not only exposes bloggers and activists using pseudonyms to grave physical danger, but has also led to blocking of the accounts of vulnerable users and activists, drag performers and users with non-English or unconventional names.

What should states and companies be doing to address these problems?

Recommendations for states

Recommendations for ICT companies

Recommendations for states



- Ensure that limitations on FoE meet established conditions of legality, necessity, proportionality and legitimacy, including on content deemed to advocate hatred and incite discrimination, hostility or violence.

- Repeal any law that criminalises or unduly restricts expression, online or offline.

- Refrain from establishing laws or arrangements that would require the “proactive” monitoring or filtering of content by companies. This would violate the right to privacy and amount to pre-publication censorship.

- Make regulation “smart”, not heavy-handed or viewpoint-based but focused on ensuring that companies are transparent, provide remedy, and that users can make choices about whether and how to use online forums.

- Ensure commercial policies and licensing comply with internationally accepted human rights standards.

- Refrain from imposing disproportionate sanctions such as heavy fines or threats of imprisonment on publishers of content that states consider to be illegal. This results in companies introducing broad and pre-emptive rules which have a chilling effect on FoE.

- Refrain from adopting models of regulation in which government agencies, rather than judicial authorities, become the arbiters of lawful expression.

- Issue content restrictions through an order by an independent and impartial judicial authority in accordance with due process and standards of legality, necessity and legitimacy.

- Create an environment in which companies are incentivised to uphold human rights principles.

- Avoid delegation of regulatory functions to private actors that lack basic tools of accountability.

- While seeking removals that may be applicable in multiple jurisdictions, make such requests in every relevant jurisdiction, through regular legal and judicial processes.

- Support scalable appeal mechanisms that are consistent with human rights standards.

- Publish detailed transparency reports on all content-related requests issued to intermediaries and involve genuine public input in all regulatory decisions/measures.

Recommendations for ICT companies

HUMAN RIGHTS PRINCIPLES, POLICIES AND ASSESSMENT

- Recognise international human rights law as the authoritative global standard for ensuring FoE on their platforms, not their own private interests or the varying laws of states. Revise ToS and community standards accordingly.

- Direct all business units, including local subsidiaries, to resolve any legal ambiguity in favour of respect for FoE, privacy and other human rights.

- Make content standards clear and specific. Provide examples to help users interpret and apply specific rules.

- Commit to maintain platforms as spaces where users, consistent with human rights law, develop opinions, express themselves freely and access information.

- Conduct rigorous human rights impact assessments on all products and policies. Include meaningful consultation with users and civil society and seek comments from interested users and experts, especially from the global South. Enable confidentiality of inputs.

- Adopt the Guiding Principles on Business and Human Rights, along with industry-specific guidelines, e.g. those developed by civil society, intergovernmental bodies and the Global Network Initiative.

DEVELOPMENT OF POLICY ON MISINFORMATION AND MEDIA

- Refrain from placing restrictions on news content that may threaten or limit independent and alternative news sources or satirical content.

RESPONDING TO STATE REQUESTS

- Ensure that requests cite specific and valid legal bases for restrictions and are issued by a valid government authority, in writing.

- Always seek clarification on requests that could potentially restrict fundamental rights, such as FoE. Solicit assistance from civil society, peer companies, relevant governmental and international bodies, and explore all legal options for challenging such requests.

- Route requests from states, including those made under ToS, through legal compliance processes. Assess their validity based on national and international human rights standards.

- Include granular data in reporting state requests (e.g. related to defamation, hate speech, or terrorism) and actions taken (e.g. partial, full, country-specific or global removal; user account suspension or removal). Provide specific examples as often as possible.

TRANSPARENCY

- Ensure and document transparency at all stages, from rulemaking to implementation. Make transparency reports easy to understand and available in English and local languages.

- Disclose responses to government requests and to requests based on ToS.

- Preserve records of all requests and subsequent communications between the company and requesters. Consider submitting copies of requests to a third-party public repository such as the social media council mentioned below.

- Develop “case law” to frame the interpretation of rules so that users, civil society and states understand how companies interpret and apply their standards.

- Avoid secretive arrangements with states on content standards and implementation.

- Provide meaningful and consistent transparency about enforcement of policies governing contentious issues, such as hate speech.

- Make public information on the results of automated content moderation, human moderation including flagging by users and trusted or specialised “flaggers”.

- Explain the selection of people performing human evaluation, e.g. specialised flaggers (some which work in units established by governments). Disclose how they interpret legal and community standards.

- Explain how the public interest is defined, and what other factors are used in decisions to take action against content.

- Explain how newsworthiness is determined.

- Develop best practices on how to provide transparency on interactions between states and companies.

NOTICE AND APPEAL

- Provide “counter-notice” procedures that permit users to challenge content or account removals.

- Institute robust remedy programmes which may range from reinstatement of content to settlements related to reputational or other harms.

CONTENT CURATION

- Disclose details concerning approaches to curation. If companies rank content on social media feeds based on interactions between users, they should reveal what data is collected and how this informs ranking.

- Provide all users with accessible and meaningful opportunities to opt out of platform-driven curation.

USER EDUCATION AND AWARENESS

- Consistently provide sufficient information to users on the development and evolution of rules.

CONTEXT, CONSULTATION, DIVERSITY AND GROUPS AT RISK

- Increase engagement and consultation with users, civil society and digital rights organisations.

- Avoid real-name requirements. Protect users' anonymity by default. Online anonymity (for example, through use of pseudonyms) is often necessary for the physical safety of users at risk.

- Actively consult local community groups to help with taking cultural and artistic contexts into account (for example when assessing content featuring nudity).

- Consider the concerns of communities historically at risk of censorship and discrimination, e.g. linguistic minorities and LGBTQ individuals and groups.

- Describe in greater detail contentious and context-specific rules. Disclose data and examples that provide insight into how violations are assessed and responded to.

- Strengthen and ensure professionalisation of human evaluation of flagged content, including by seriously committing to involve cultural, linguistic and other forms of expertise in every market where they operate.

- Provide protections for human moderators consistent with human rights norms applicable to labour rights.

- Diversify company leadership and policy teams to bring local or subject-matter expertise to how content is approached.

AUTOMATED CONTENT MODERATION

- Take into account the challenges of automated content moderation, such as assessing context, meaning, variation in language cues, linguistic and cultural particularities.

- Technology developed to deal with considerations of scale should be rigorously audited and developed with broad user and civil society input, keeping in mind differences across user bases.

INDUSTRY-WIDE ACCOUNTABILITY MECHANISMS

- Given their impact on the public sphere, all companies that moderate content or act as gatekeepers must open themselves up to public accountability and should make the development of industry-wide accountability mechanisms a top priority.

- Work with one another and civil society to explore establishing an independent social media council to enable complaints and remedy for rights violations across platforms and borders at an industry-wide level.

This illustrated summary was adapted by APC from the report by the United Nations Special Rapporteur on freedom of expression and opinion (UNSR) on online content regulation. The content of this publication should not be attributed to the UNSR and we strongly encourage readers to refer to and read his formal report at *https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ContentRegulation.aspx*

The UNSR's report provides a comprehensive overview of the major human rights concerns raised by commercial content moderation. It examines requests for moderation by states, as well as company standards and processes, and includes recommendations to states on how to ensure they uphold rights and follow due process. It also outlines measures that companies should incorporate to emphasise transparency, due diligence, regular public input and engagement, and access to remedy.

Reorienting rules for rights: Summary of the report on online content regulation by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/HRC/38/35)

APC
ASSOCIATION FOR PROGRESSIVE COMMUNICATIONS