CONTENT REGULATION AND HUMAN RIGHTS IN THE DIGITAL AGE

Multi-Stakeholder Roundtable on the U.K. Online Harms White Paper

9 September 2020

Roundtable Report

On 9 September 2020, GNI hosted a multistakeholder roundtable discussion on human rights and content regulation in the United Kingdom. The discussion centered on the U.K. Online Harms White Paper, the related public consultation and the government's initial response, and the implications of international human rights standards for content regulation.

GNI presented its forthcoming policy brief, "Content Regulation and Human Rights in the Digital Age," which provides a framework for considering good policy practice. Though many human rights are impacted by content regulation, controls on communication most directly impact the rights to freedom of expression and privacy. The policy brief and our discussion focused on these two rights.

During the discussion, participants, informed by the consultation response, outlined and considered various steps policymakers can take to enhance safeguards for users' rights while responding to legitimate concerns about online harms, and to propose a model regulatory framework beyond U.K. borders. Representatives from the Department for Digital, Cultural, Media and Sport (DCMS) and Home Office addressed questions regarding the plans for implementing the white paper. The moderated discussion addressed four themes: codes of conduct, duty of care, remedy, and privacy.

The conversation was held under the Chatham House Rule. Nothing in this report is attributed to any individual, institution, or affiliation, nor does it necessarily reflect GNI's position.

Table of Contents

An Introduction to the Online Harms White Paper

The Online Harms White Paper, first presented in April 2019, sets out the U.K. government's plans for a "word-leading package of measures to keep users safe online." With a focus on new "systems of accountability and oversight," the White Paper frames an outline for legislation that would allow the government to establish subject-specific expectations ("Codes of Practice") that covered companies in scope would be expected to meet to protect user safety and tackle illegal content. Those companies would be expected to meet a "duty of care" standard for protecting user safety, and oversight and enforcement would be handled by an independent regulator.

Alongside issuing the white paper, DCMS outlined a set of questions for public input, receiving 2400 written submissions from industry, media, and civil society alike. While praising the government's commitment to preserving and improving the safety of Internet users, particularly children, GNI flagged some initial concerns in our submission, noting that, among other things, without further clarification, the breadth of the approach outlined — in terms of both the scope of content and companies covered — and lack of clarity regarding the duty of care, combined with the broad enforcement powers contemplated in the White Paper, could incentivize companies toward over-removals of content and invasive monitoring of users.

Ahead of the full policy position to come later this year, DCMS and the Home Office have since issued an initial public response to this round of consultation, rounding up and addressing some of the concerns flagged. The response hinted at a narrowed scope of companies covered by the regulation, with "reasonable and proportionate" requirements targeted to business models and sizes based on the level of risk; and a general emphasis on systems-based approaches and transparency, as opposed to mandating removal of legal but problematic content. Throughout the roundtable discussion, participants explored various open questions for freedom of expression and privacy that remain following this consultation response.

A Human Rights-Based Approach to Content Regulation

Over the last few years, GNI has noted the uptick in governmental efforts that claim to address various forms of digital harm related to user-generated content — a practice we refer to broadly as "content regulation."

The analysis and recommendations in the draft "Content Regulation Policy Brief," which GNI shared confidentially with participants ahead of the event, stem from GNI's review of dozens of such efforts from jurisdictions all over the world, and draws upon the collective expertise and experience of GNI's multistakeholder membership. In this roundtable, as with previous consultations with policymakers and other experts on content regulation in the EU, India, and Pakistan, GNI continues to seek input on the brief ahead of the full launch later this year.

The brief demonstrates that the norms and principles articulated in international human rights law provide a universal, time-tested, and robust framework that can help lawmakers find creative and appropriate ways to engage stakeholders, reconcile different interests, and mitigate unintended consequences of content regulation.

The brief examines content regulation efforts for their compatibility with three key principles of international human rights law: legality, legitimacy, and necessity. It also considers proportionality as a component of necessity and extends this analysis to privacy.

Legality

The principle of legality establishes that restrictions on freedom of expression must clearly define that which is prohibited and by reference what is allowed, allowing an "individual to regulate his or her conduct accordingly."[1] Such laws must also enable those responsible for their execution to ascertain expression that is allowed and that which is not, which contributes to predictable, consistent, and non-discriminatory enforcement. This is particularly important when laws rely on private bodies, rather than democratically-accountable regulators or independent judiciaries, to adjudicate and enforce such restrictions. Here we can also see the wisdom of ensuring accessible and effective remedy as a way to mitigate the impacts of inaccurate enforcement.

Ambiguous content regulations can have a chilling effect on legitimate speech. In practice, chilling effects unfold in two ways. First, individuals who fear violating the law may shape their communications to avoid any potential implication, sometimes choosing not to speak at all. Second, companies held liable for user-generated content may be overly broad in their enforcement of the law to prevent any possible infringement.

In the UK context, the government and future independent regulator will need to work hard to ensure that the codes of practice they are charged with developing avoid any ambiguity as to what specific content or behavior is prohibited, and clarify expectations for companies and the public in response.

Legitimacy

The principle of legitimacy holds that laws restricting expression can only be justified to achieve specific, enumerated purposes. These may include respect for the rights or reputations of others or the protection of national security, public order, public health or morals. While international law gives states significant latitude to determine the activities that justify restrictions, that discretion is not unlimited.

1 UN Human Rights Committee, General Comment No. 34: Article 19 (Freedoms of opinion and expression), 102nd Sess, adopted 12 September 2011, UN Doc CCPR/C/GC/34, online: < https://undocs.org/CCPR/C/GC/34>

International courts and authorities have made clear that the right to freedom of expression is broad and encompasses "even expression that may be regarded as deeply offensive."

In addition, numerous consensus United Nations resolutions establish that the same rights that are protected offline must also be protected online. Inconsistencies in the treatment of online and offline speech may be exploited by regimes and actors who do not respect democratic norms. Therefore, it is critical to protect speech equally and consistently, and to resist differentiating approaches to expression across offline and online mediums.

The White Paper's focus on categories of content that are "not unlawful but have the potential to cause harm" raises serious questions about whether such an approach will unduly restrict speech that may be deeply offensive to some but should nevertheless be protected under international law.

Necessity and Proportionality

The principle of necessity requires states seeking to restrict expression to articulate the threat imposed by a specific type or piece of speech as well as the "direct and immediate" connection between the expression and the threat.

The related principle of proportionality requires that any restrictive law, as well as the actions of administrative and judicial authorities in their application of that law, must be: (i) proportionate to the interest being protected; (ii) appropriate to achieve that protective function; and (iii) the least intrusive instrument among those which might achieve that protective function.

In the content regulation context, the principles of necessity and proportionality should guide lawmakers to think carefully about which types of services at which layers in the technology stack are most appropriately positioned to address specific concerns. Shifting liability for illegal content from creators to intermediaries rarely if ever fits this description. To the extent such regulations establish "notice and take down" mechanisms, they must ensure legal responsibility for non-compliance is predicated on clear notices about specific content that has been adjudicated to be illegal by an independent authority. Similarly, punitive sanctions, rigid timelines for content adjudication, and pre-emptive filtering requirements are also likely to run afoul of the necessity and proportionality principles, and as such are likely to prove ineffective or counterproductive.

To ensure content regulation efforts are appropriately and narrowly tailored and to guard against unintended consequences, lawmakers should look to proven approaches based on concepts like transparency, due process, and remedy. They should also consider the perspectives of and, where appropriate, provide explicit protections for specific actors such as journalists and vulnerable groups.

Privacy

Many content regulation efforts lack protections for the fundamental right to privacy at best and, at worst, actively undermine individual privacy. Requirements to proactively monitor, track, or trace content often lack consideration of associated privacy risks. In addition, compelling content hosts to proactively report user-generated content and associated data to law enforcement agencies further undermines this right. Moreover, the explicit prohibition or implicit limitation of the use of anonymity and encryption tools signals a disregard for the importance of privacy and the rights it enables.

Moderated Discussion: Key Elements of the Online Harms White Paper

Codes of Conduct:

Agenda prompt: *The codes of practice on illegal content will play a significant role in setting expectations for companies and the public, and provide the basis for enforcement. What steps can be taken to ensure definitional clarity and safeguards for fundamental rights in these codes?*

The discussion began with questions about the government's ability to define content that is clearly problematic, but remains legal under U.K. law, while providing sufficient predictability and guidance, in line with the principle of legality, for enforcement and for "individuals to regulate their conduct accordingly."

Participants noted that recent history has shown that some of the most problematic areas of content — hate speech, disinformation, violent extremism, etc. — are also the most difficult to define. Defining such terms in a truly democratic fashion and providing sufficient clarity for companies is difficult. Applying them can be even harder. As recognized in the Rabat Plan of Action, contextual factors that can be difficult for automated systems or filters to account for, are critical for determining when content is legal.

A question was asked about possible tiers of expectations and potential penalties to accommodate different company risk profiles and sizes. The initial consultation response acknowledged company size and business model distinctions, and emphasized proportionate responses. Concerns were noted about the difficulty in operationalizing these principles. The regulator may have limited capacity to cover the full scale of all covered harms, and crackdowns can prove particularly challenging to early-stage companies. As a result, flexibility and pragmatic, ongoing dialogue with covered companies will be important. The government has emphasized the principle of safety by design, targeting smaller companies for guidance to help build in protections for user safety and rights protections from the start.

The government's increased emphasis on systems-based approaches, rather than mandating certain removals, was welcomed. Participants cautioned, however, that while focusing on companies own practices and encouraging related transparency and accountability measures is commendable, the regulation must be designed to avoid incentivizing companies to monitor user content proactively or invasively in order to ensure favorable assessments from the regulator.

Duty of Care

Agenda prompt: *The duty of care underpins steps companies will be expected to take to ensure safety of their users. How will it account for the distinct role companies with differing business models and sizes play in protecting user safety? How can it set predictable and consistent expectations for companies, particularly for legal but harmful content?*

Participants continued discussing the regulation of content that is legal but harmful, as the duty of care will presumably offer the regulator a framework for assessing companies' practices on a broad set of protected content and conduct. The regulator should strive to be transparent and consistent in how it assesses companies practices, participants noted, particularly in areas where codes of practice don't exist, and facilitate ongoing dialogue with the public and legislators about implementation. An over-reliance on quantitative metrics could pose particular risks, as they may not accurately reflect practice and can offer perverse incentives for companies. The global precedent this law might set, as emphasized by GNI and recognized in the government's initial consultation response, is important to consider, as more restrictive governments might put greater pressure on companies to alter their policies and procedures under similar regulatory models.

Remedy

Agenda Prompt: *What incentives can be built into the law, not only for removal of harmful content, but also in incentivizing proportionate responses? Where over-removals do occur, what remedy will be available for users — vis-a-vis both companies and governments?*

Participants shared concerns that certain forms of enforcement mechanisms can contribute to overblocking, particularly by incentivizing automation. The government has noted in the interim consultation response that redress mechanisms will be required for both content takedown and for reporting harmful content. In establishing systems and safeguards to prevent overblocking and prioritize freedom of expression, the potential for mandatory due diligence models was underscored.

Participants emphasized how robust processes for remedy for affected users, without robust systems for transparency and accountability, are insufficient. It is difficult to expect individual users to fully utilize appeal systems, particularly for government-ordered restrictions. Meanwhile, companies should not be tasked with filling roles as judge, jury, and executioner. Given the interrelationship between elements of remedy, due process, and transparency and accountability, it is all the more significant that scope questions are clarified as early as possible to allow companies to implement the requirements

thoughtfully and build sufficient protections for users' rights. Finally, it was noted that narrower application of penalties also makes them more proportionate.

At the same time, participants acknowledged that systems for remedy can sometimes be abused, and the discussion moved to safeguards that should be built into remedial processes. A participant noted that the same structured mechanisms that might facilitate remedy for groups representing marginalized populations, should also be enabled for free expression groups, and that this should apply vis-a-vis both companies and governments.

Privacy

Agenda prompt: *To what extent will the law apply to private messaging services? How might this impact encryption and anonymity, which human rights experts have pointed to as critical for protecting both privacy and freedom of expression?*

While much of the day's discussion focused on principles of international human rights law relating to restrictions on freedom of expression, it is important not to overlook privacy risks. Concerns were raised that the White Paper could set expectations for monitoring private communications, including with respect to private, encrypted messaging services. Given public scrutiny about the role some communications companies play in monitoring users' content and conduct, participants encouraged the government not to legally mandate or otherwise incentivize such activity.

Participants also cited the need to further codify some of the privacy protections, and pointed to open questions about the interoperability between the White Paper and the Data Protection Act 2018. Some participants encouraged deliberation by policymakers and the public about the proper uses of data for public purposes, rather than leaving these determinations to companies or the regulator. Even for private messaging services, there are distinctions in harms that can be mapped to different risk environments to clarify the purposes and requirements for data collection. And tying back to the duty of care, a question was raised about whether the White Paper's approach would effectively mandate proactive, general monitoring by covered companies.

Conclusion

GNI is extremely grateful for the candid discussion from the diverse group of representatives in attendance, including representatives from Her Majesty's Government. We also thank Richard Wingfield and Molly Land in particular for their support with facilitating the event. We look forward to continuing discussions with relevant parts of the government and hope for additional opportunities for public consultation.

Participants

Alissa Starzak
Cloudflare

Bernard Shen
Microsoft

Charles Bradley
Global Partners Digital (GPD)

Collin Kurre
BT Group

Dan Mount
Ofcom

Dom Hallas
The Coalition for a Digital Economy (Coadec)

Elisona Shala
Department for Digital Culture, Media and Sport (DCMS)

Emma Ascroft
Verizon Media

Emma Llansó
Center for Democracy and Technology

Gabrielle Guillemin
Article 19

Graham Smith
Bird and Bird LLP

Ian Brown
FGV University

Jim Killock
Open Rights Group

Jon Higham
Ofcom

Jonny Shipp
Internet Commission

Global Network Initiative Staff

Katy Minshall
Twitter

Matt Stokes
DCMS

Michael Tunks
Internet Watch Foundation

Moira Oliver
BT Group

Molly Land
UConn Human Rights Institute

Naomi Standing
DCMS

Richard Wingfield
GPD

Ruth Smeeth
Index on Censorship

Silkie Carlo
Big Brother Watch

Steve Crown
Microsoft

Victoria Nash
Oxford Internet Institute