



## ORG policy responses to Online Harms White Paper May 2019

Open Rights Group (ORG) is a UK-based digital campaigning organisation working to protect fundamental rights to privacy and free speech online. With over 3,000 active supporters, we are a grassroots organisation with local groups across the UK.

ORG has actively engaged with the government's proposals for online regulation since the Internet Safety Strategy in 2017. The following policy positions have been developed through a long period of reflection and engagement with different stakeholder groups. We hope that they assist others intending to respond to the white paper consultation.

For the avoidance of doubt, this paper does not comprise our full consultation response. We will publish that in due course.

### Importance of the Internet / social media for free expression

- **The Internet in general and social media in particular play a central role in protecting free expression in society. They have particular importance for children and young people's expression and access to information.**

### Purpose and scope of regulation

- Social media companies are private entities operating for commercial profit which ultimately make decisions based not on societal good but on their own financial interests. **Their data-driven business model, powerful control over citizen speech and operation within an online environment whose unique characteristics affect and influence the reach and impact of content, activity or behaviour, together justify policy intervention.** This is a different starting point to the White Paper, and produces different potential interventions.
- **The ultimate aim of Internet regulation should be to ensure and support a digital environment that protects and respects human rights.** We fully endorse the call of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, for "smart regulation" focused on increasing and improving online companies' transparency and accountability.
- **Regulation must go beyond dealing with individual pieces of user-generated content and address the full range of issues and circumstances in play around content distribution on social media platforms. It must, in particular, be situated within a broader conversation around how to protect rights to free expression in view of the damaging effect that social media is having on democracy.**
- Regulation needs to be holistically part of creating a clearer framework covering illegal content online and unlawful offensive communications. We refer to the Law Commission's work in this area.

- **The government's proposed regulatory scope is unrealistically vast.** In order to be successful, the scheme needs to be narrower and focused on online communications platforms that handle the publication of very large volumes of user-generated material. **The government is right to rule out attempting to regulate private communications.**
- Platform experience worldwide is politically and culturally context-sensitive. The UK has strong media plurality, generally effective justice processes and well-established democratic institutions; circumstances which are not universally guaranteed. **UK regulation should only apply territorially and not seek to impose globally-applicable standards.**

#### Rights-based approach

- **Any regulatory scheme must be explicitly rooted in the international human rights framework.** This provides an established, universally-applicable standard capable of holding both companies and States to account.
- Regulation should encourage internet companies to adopt and implement the [UN Guiding Principles on Business and Human Rights](#). These establish principles of due diligence, transparency, accountability and remediation, and would commit companies to implementing human rights standards throughout their product and policy operations. **We would welcome incorporation of these principles into any regulatory framework so that they become directly enforceable.**
- **It is critical to acknowledge and understand that regulation will ultimately bite on social media users and directly impact the fundamental rights of ordinary citizens.** Regulating social media is essentially different from regulating newspapers or broadcasters because internet media platforms driven by user generated content facilitate the day-to-day freedom of expression of their users.
- **Protection of the right to free speech must infuse how legislative and regulatory schemes are developed, implemented and enforced.** If a harms-based approach is used (which we would not recommend), harms to freedom of expression must themselves be recognised as a harm, to be weighed in any balancing exercise.
- What is legal offline must remain legal online.
- **Platforms must not be obligated to generally scan or monitor content.** Proactive monitoring is inconsistent with the right to privacy and will lead to increased censorship.
- Regulation should promote non-discrimination in decision-making, both human and algorithmic.

#### Duty of care and harms-based approaches

- **The duty of care model is a poor conceptual fit for addressing the societal challenges of social media platforms and should not form the basis of regulation.** Its focus on risk rather than process does nothing to holistically improve decision-making and user/community experience at platform level. If drawn

narrowly, a duty of care risks failing to address the full range of regulatory concerns. If drawn broadly, including by extending the applicable definition of harm to include harm to individuals, vulnerable groups and society at large, the risks to free expression are particularly acute: either way, such a duty will not achieve the outcome the government desires.

- If the duty of care model is nonetheless adopted, it should focus on systemic and structural issues, addressing particularly how platforms impact on fair and open democratic processes, and considering generally whether reasonable responses to risk are being taken. The wording and scope must be very carefully formulated, and it should apply when dealing with all users of the platform, including dealing fairly with any alleged perpetrators.
- **The harms-based model is an equally poor conceptual approach. It is attractive in prioritising resources but fails to address the needs that are driving this regulation.** Regulation needs to target platform rather than user activity. It should address systemic issues including data retention and exploitation, opaque advert targeting and recommendation systems and other algorithms. A harms focus fails to acknowledge that these activities can also bring benefits. The evidence in other areas such as data protection is that the harms-based model is not able to fully tackle the negative impacts of online platforms, not least because the evidence of concrete harms can be difficult to establish.

#### Regulatory models

- Laws protecting human rights apply equally online as offline; consequently, regulation must comply with the legality, legitimacy and necessity provisions established in Article 19 ICCPR and other international laws and treaties.
- State regulation of social media takes decisions about the limits of free expression out of the hands of independent judicial authorities. Extensive state regulation will lead to mass government enforcement of private censorship. This model of regulation is not tolerated for the press; it is hard to see how and why it could be justified for online communications of millions of individuals.
- Independent self-regulation is problematic as it permits privatised censorship to continue without adequate human rights protection.
- **We advance co-regulation as a means to advance meaningful accountability and procedural improvements at company level, which would better protect human rights both where content is wrongfully removed and when it remains in place.** Co-regulation creates public accountability for the kind of regulation that takes place whilst maintaining distance from state interference and the setting of inappropriately restrictive norms.
- Genuine co-regulation requires that the regulatory body be robust, independently managed, financially independent from both government and industry and with the power to make decisions that are final and respected. A statutory footing is required for all to have confidence that the scheme is effective and accountable.
- **Regulation should focus on reviewing internal company processes and auditing decisions.** Given the volumes of online content, we would expect powers of audit

and review of process to be extremely important to produce robust results. This includes ensuring that platforms conduct sound content moderation with an effective appeals process to rectify mistakes.

- **Regulation should encourage companies to align their terms of service more closely with human rights law.**
- Regulation should not entrench monopoly positions of Internet companies but support diversity in the online ecosystem.
- In its policy development process, the UK government should take account of overseas regulatory initiatives, notably the French government proposals.
- We welcome the call in the white paper to “promote a culture of continuous improvement among companies and encourage them to develop and share new technological solutions rather than complying with minimum requirements.”

#### Liability and enforcement

- **We strongly caution against attaching liability to platforms for third-party content.** While well-meaning, proposals such as these contain serious risks, such as requiring or incentivising wide-sweeping removal of lawful and/or innocuous content. Imposing time limits for content removal, heavy sanctions including personal liability for non-compliance or incentives to use automated content moderation processes also only heightens this risk.
- We understand the need to remove some content speedily, e.g. live-streaming of criminal acts, but excessive requirements in this respect pose risks to fairness and due process.

#### Evidence and definitions

- **Any policy intervention must be underpinned with a clear, objective evidence base which demonstrates that actions are necessary and proportionate.** Regulation impacting on citizen’s free speech needs to be based on evidence of harm traceable to specific pieces or types of content, activity or behaviour, rather than expectations or social judgements that these may be related to possible harms. It will be challenging to develop a regulatory scheme that fulfils this criteria.
- **The limitations of research in this area must be taken into consideration when assessing the weight to be given to evidence.** Risk encounters cannot easily be measured except by asking children directly, which raises ethical (children might be unaware of harm until asked specifically) and measurement (risk of under- or over-reporting) questions. Research is also not able to predict which children will experience harm as a result of encountering risk. Risk refers to the probability of harm, and e.g. encountering hostile messages or pornographic images is not necessarily harmful. Some risks may also be rare but severe in their consequences, and this, too, is difficult to assess. Since children are no more homogeneous than the adult population, a host of factors affect the distribution of risk and harm, vulnerability and resilience.

- **Any policy intervention must be defined and limited by precise terminology. Imprecise language risks dangerous overreach.** If the harms-based model of regulation is used, tighter identification/definition of types of harms and their natures is vitally needed.
- **All relevant stakeholders, including civil society and smaller/niche platforms, should be fully engaged throughout the Online Harm White Paper consultation period, and able to participate in the design and implementation of any legislative/regulatory measures which are finally adopted.**

### Transparency and accountability

- **Regulation should be primarily and predominantly aimed at radically improving transparency and accountability on the part of social media platforms and others involved in the moderation and removal of online content.**
- Transparency goes beyond reporting on raw numbers of content removed. Operational-level transparency should cover how and on what basis rules and policies are made, what factors inform content-related decisions and provision of hypothetical case examples showing how rules are interpreted and applied across a range of scenarios.
- **Transparency needs to include information around political advertising, with sufficient public information provided so that relevant third-parties can be held accountable.**
- **Independent audits are essential for effective regulation.** Audits go beyond ensuring moderation decisions are accurate and that inaccurate decisions and trends of decision-making are detected and resolved. They are needed throughout all parts of the systems, as the questions are about volume and impact of systems as much as particular aspects and decisions. This implies a familiarity with commercially sensitive information, so would be potentially more effective in a co-regulatory framework.
- Regulatory standard-setting for content moderation should be guided by the [Santa Clara Principles](#).
- Platforms should also be required to provide user-accessible information about the policies they have in place to respond to unlawful and harmful content, how those policies are implemented, reviewed and updated to respond to evolving situations and norms, and what company or industry-wide steps they have or are planning to improve these processes.
- **Content removal must be subject to precise, accessible and consistently-applied rules. Users must have effective ability to contest decisions made to remove or not remove content with appeals heard by an independent human decision-maker. A right to an effective appeal is essential for companies to fulfil human rights obligations.**
- Accountability includes developing quality standards for training content moderators.

- If external actors are able to complain and remove material in bulk, there should be penalties for unjustified threats.
- Algorithms and automated decision-making should not be developed or used in a way which would risk adverse impacts upon users' human rights (such as the right to non-discrimination). There should be greater transparency over the use of algorithms so that users have a basic understanding of when they are used and what their effects are.

#### Regulation as a means to build public trust

- Platforms are private companies and operate differently according to internal company identity and policy. The diversity in the platform ecosystem is positive and support innovation. Nonetheless, consistency in compliance with fundamental rights and transparency across platforms would increase public trust.
- It is important for police action and prosecution to follow where criminal activity is suspected/indicated. Trust in regulation is built by there being real-world consequences for unlawful activity.
- Regulation should focus on systemic issues. Separately, an independent dispute resolution mechanism should be established to facilitate mediated conflict resolution between platform users. This would improve individual access to effective remedy in appropriate cases without overburdening the courts and support improved civil discourse on platforms.