

Response to the UK Online Harms White Paper Consultation

Overview

The Online Harms White Paper is a wide-ranging, ambitious and systematic response to growing anxiety about online platforms' handling of harmful and illegal content. However, commentators have raised a number of concerns about the Government's proposals, highlighting potential unintended consequences and risks that may undermine the effectiveness of the proposed regulatory regime.

The complex challenges posed by online content regulation require a regulatory scalpel rather than a sledgehammer. Given the political sensitivities of regulation, and implications for online rights and freedoms, we suggest further work is needed to describe and consult on the regulatory regime before moving to legislation. The goal should be a framework that incentivises agile, balanced, accountable responses by companies to the particular problems each faces. Prescriptive, one-size-fits-all rules would make this harder rather than easier.

Our response is based on our extensive previous work in this area.¹ We focus on how such a framework may be developed to achieve proportionate, effective regulation. Specifically, we make five recommendations:

- Companies should be accountable for their systemic responses to online harms, not individual content decisions;
- The proposed regulatory regime should be supervisory, principles-based and procedural. In general, the regulator should not prescribe particular editorial or technical solutions;
- If duty of care is to be used as a legal framework, the standard of care should be procedural, with broad expectations of companies described in more detail pre-legislation. This might include, for example, standards for risk assessment, policy development, impact evaluation, complaint and appeal;
- The regulator should take into account companies' commitments to voluntary and multilateral content initiatives, and assess whether they are sufficient to comply with the UK's regulatory requirements; and
- More generally, the regulator's duties, powers, thresholds for action and accountability arrangements should be clarified.

The regulatory challenge

The issues involved in moderating online content and conduct are complex. The boundaries between legal and illegal, harmful and benign, are often impossible to draw clearly and consensually. Linking content to harm is notoriously difficult, prone to under-tested assumptions and not quick to resolve – for example in decades of debate about the effects of TV and video games on child development. While some types of harmful content are universally unacceptable, such as child sex abuse material, many – such as disinformation, manipulation and anonymous abuse – are controversial.

¹ Communications Chambers, [Keeping consumers safe online: Legislating for platform accountability for online content](#), July 2018; [Reconciling private market governance and law: a policy primer for digital platforms](#), May 2018

Consequently responses should be cautious, with speech only restricted where necessary, proportionate and prescribed by law. Competing rights and objectives need to be balanced. Hasty judgements about the need for whole classes of content to be blocked or removed are risky.

Equally, the scale of online user-generated content means it is impossible for most content decisions to be the product of careful, case-by-case consideration. The vast majority of blocked online content is identified by automated tools without any human intervention; errors are inevitable, both letting through some content that should be blocked, and wrongly blocking legal material.

There is an emerging consensus that these issues should not be resolved by tech firms acting independently, without some form of democratic scrutiny and accountability. As the Internet Association puts it,² moral and ethical judgements around content should not be simply delegated to private companies. There is a role for government to provide both industry and the public with greater clarity about the standards expected to apply to online content, conduct and speech, based on law. Users need to understand what they can expect from platforms, and platforms need to understand what is required to demonstrate sufficient protection of all their users' rights.

However, as the White Paper suggests (but could state more clearly), online harms do not lend themselves to prescriptive, rules-based regulation. Different platforms and different kinds of content pose different risks. Technical solutions to extremist content that might be needed for video sharing sites may be wholly inappropriate for messaging apps or chat services. When harms are difficult to define, heterogeneous and evolve over time, effective bright-line rules are elusive.

Platforms therefore need flexibility to respond to the particular challenges they face in an agile, iterative way. Too often, the policy debate focuses on content blocking or takedown as the preferred solution. But a wide range of other responses is possible, including deprioritisation, flagging, positive messaging, user tools, safe search settings or age verification. Again, these will be more or less appropriate in different contexts, and deliver different results on different platforms.

Procedural accountability

Our previous reports identified a need for new institutional arrangements to accommodate platforms' role in governing online markets. In research completed at the Oxford Internet Institute in 2017, this author labelled this 'procedural accountability', to distinguish it from the editorial obligations typically ascribed to broadcasters and publishers.³

Procedural accountability says that platforms should be held to account for the effectiveness of their systemic response to online harms; that is, the policies, code and processes they use to identify, evaluate and respond to objectives established by law. In a procedural regime, Government's role is to define the *governance* standards companies should meet, with appropriate oversight in place to ensure those standards are upheld and where necessary enforced.

This gives platforms freedom to develop systems that are appropriate for their environment, scale and risk of harm, and that can be iterated rapidly in response to changing circumstances and evidence. The regulator's task is to supervise, and to validate the effectiveness of platforms' policies,

² http://uk.internetassociation.org/wp-content/uploads/sites/2/2019/05/IA_Initial-OHWP-Response_30-09-19_UK.pdf

³ Mark Bunting, [From editorial obligation to procedural accountability: policy approaches to online content in the era of information intermediaries](#), *Journal of Cyber Policy* 3:2, pp165-186

not to prescribe particular technical solutions. Expectations of different companies should be tailored to their scale and extent of harm.

Providing regulatory certainty

The White Paper proposes a regulatory regime that is both broad – capable of addressing a range of harms, including currently unspecified issues – and also flexible, in not prescribing exactly what action is required of companies.

Breadth and flexibility allow for tailored solutions and for new and unforeseen issues to be addressed. But they cause problems for regulatory certainty, which requires that expectations of companies are clear, and regulatory action predictable. As the Secretary of State put it in his evidence to the Digital, Culture, Media and Sport Select Committee, companies need a clear structure so they can understand what they have to do.

Under a duty of care, companies may be held liable for harm if they have failed to take ‘reasonably practicable steps’ to keep users safe. Duties of care put a great deal of pressure on regulatory and/or court judgements about whether harm could have been foreseen, and whether platforms took ‘reasonable steps’ to prevent it.

It is unclear, at this stage, how these judgements will be reached. Online environments are not like the workplaces and public spaces covered by health and safety duties of care. Some forms of harm may be predictable, but others may not; and even where harm is foreseeable, there is likely to be little consensus about the steps companies should take to prevent it, which will vary from platform to platform, and issue to issue.

Regulatory uncertainty in the context of a duty of care creates a number of risks:

- Companies may choose to reduce their exposure to liability by over-blocking content, thereby undermining freedom of expression and foreclosing the benefits of open online environments;
- The duty of care may be toothless, if in practice the regulator finds it hard to decide that harm was foreseeable or that companies did not ‘do enough’; or if those decisions are easily challenged in court;
- The scope of regulation may expand over time, with the regulator adding new obligations either in an attempt to create greater certainty or in response to pressure to do more, without sufficient regard for whether additional regulation is necessary or proportionate; and
- The relationship between law and regulatory standards may be unclear.

In the next phase of work, therefore, the Government should provide more clarity in three areas, in order to provide greater certainty to industry and users, and ensure proportionality and effectiveness:

- The standard of care;
- The regulator’s duties, powers, thresholds for action and accountability arrangements;
- The relationship between the UK’s regulatory standards and voluntary or multilateral initiatives.

These areas are described in the remainder of this paper.

The standard of care

We do not take a view on whether a duty of care is the right legal framework. However, if a duty of care is adopted, resolving legal uncertainty will require greater clarity about the *standard* of care – that is, what it is reasonable to expect online companies to do to assess and mitigate the risks of harm on their platform.

We have argued that the generic expectation should be that companies adopt good governance procedures with respect to specified problems, where a risk-based threshold is met. Companies in scope could be expected to carry out a risk assessment, develop appropriate policies, evaluate their policies' impact on all their users' rights, and iterate accordingly. They should provide effective mechanisms for complaint and appeal. Both risk assessments and policy responses should be disclosed to the regulator, in response to a reasonable and evidenced request.

The White Paper contains a number of governance-related principles, such as:

- Clear terms and conditions;
- Effective and consistent enforcement of terms and conditions;
- Prompt and transparent action in response to user reporting;
- Direction of users who have suffered harm to support; and
- Regular review and iteration.

Legislation could clarify these broad principles, but specific expectations will require further iterative work between the regulator, companies and relevant stakeholders, including other public bodies. There will be an element of 'learning by doing', with the regulator and companies working together to assess whether the required standard has been achieved on a case-by-case basis. Penalties should be avoided in this learning period, with companies only becoming liable for sanctions once regulatory 'case law' has helped to clarify the standards.

The regulator's mandate and Codes of Practice

Greater visibility of regulatory powers, duties and thresholds for action would help clarify the regulatory framework and prevent scope creep. This could include, for example:

- Defining the evidential threshold of harm that must be reached for the regulator to intervene;
- Requiring the regulator to define precisely the type of content or harm to be addressed, after consultation, consistent with relevant legal definitions, and with accountability to Parliament for content and conduct that is undefined or unclearly defined in law;
- Setting different tiers for regulatory expectations, with smaller firms or start-ups having fewer or no obligations unless there is evidence of significant risk of harm;
- Obliging the regulator to put in place independent evaluation and advice mechanisms to assess the impact of regulation on competition, innovation and fundamental rights.

A central role for the regulator will be developing Codes, or a Code, of Practice. Section 7 suggests that these may be detailed and vary by type of content. For example, it is suggested that the regulator should provide guidance on proactive steps to prevent new and known child sex abuse or terrorist material, or links to it, being made available; action on illegal material in live streams; systems to prevent further circulation of images of children that may fall below the illegal threshold but leave them vulnerable to abuse; measures to identify which users are children; identification of violent or violence-related content to prevent its availability or further sharing; and so on.

A regulator that is directed to provide detailed Codes of Practice on these matters is likely to stray into a level of prescription that sits uncomfortably with the principles-based, supervisory regime that we described above. Any such Codes may (a) set an impossible standard, (b) become rapidly outdated, (c) only be relevant to some companies, and/or (d) favour larger companies who already have, or can afford to invest in, complex technical solutions. Consistency may be seen as a virtue, but detailed rules tend to favour some over others: the truly ‘level playing field’ is a mirage.

A less prescriptive approach will be more future-proof, and capable of applying across multiple harms. Where evidence exists of a significant risk of harm, the regulator may require a company to disclose its policy for managing the risk, assess its effectiveness and iterate accordingly – but should not, in general, prescribe particular editorial or technical solutions.

Given the breadth of its remit and the sensitivity of its role, the regulator’s own accountability arrangements will need careful thought. We suggests these should include independent external scrutiny, due process obligations enshrined in legislation, and accountability to Parliament, particularly with regard to types of content that are not clearly defined and/or are legal but harmful. We agree with the suggestion in the White Paper that there should be a statutory mechanism for companies to appeal against a decision of the regulator. Any proposals to extend the scope of regulation to new harms or legal content should be subject to scrutiny by an appropriate body, which could also be Parliament.

Voluntary and multilateral initiatives

The White Paper recognises the extensive existing commitments and voluntary initiatives by companies, including the Global Internet Forum to Counter Terrorism, the UK Council for Internet Safety and the WePROTECT Global Alliance. There are also well-established bodies such as the Internet Watch Foundation and newer initiatives such as the Internet Commission.

There is a risk, if regulatory standards differ from those applied by such initiatives, that the latter will be undermined, especially given that the UK is only one of many jurisdictions considering regulation. Fragmented national standards create disincentives to participate in voluntary initiatives, but these can – if they meet suitable standards of accountability and transparency – deliver a more effective, harmonised solution. The UK can set a lead by stating that participation in industry or multilateral initiatives that meet the UK’s procedural standards will be sufficient to comply with the UK’s rules.

Conclusion

The approach proposed here and in our previous reports is supervisory, principles-based and procedural. We believe that this is better suited to the broad, flexible regulatory regime envisaged by the White Paper than a rules-based approach. We also believe it is consistent with the vision described by the White Paper. Regulatory certainty can be provided by clarifying the expectations of companies, and the regulator’s mandate, powers and accountability arrangements.

We hope this submission is useful and look forward to continuing to contribute where we can to help shape an effective, proportionate regulatory response to this important policy priority.

Mark Bunting

July 2019