DEMOS

THE ONLINE SAFETY BILL

DEMOS POSITION PAPER

ELLEN JUDSON

APRIL 2022

Published by Demos April 2022

15 Whitehall, London, SW1A 2DD
T: 020 3878 3955
hello@demos.co.uk
www.demos.co.uk

CONTENTS

ACKNOWLEDGEMENTS

Huge thanks to Kyle Taylor of Fair Vote UK and Poppy Wood, of Reset.Tech, and of course to the Demos team, especially Alex Krasodomski, Ciaran Cummins, Victoria Baines and Kosta Marco Juri, for their support in preparing this paper.

Ellen Judson

April 2022

INTRODUCTION

The Government's Online Safety Bill has finally been published (17 March 2022), almost three years since the initial Online Harms White Paper that began the debate about digital regulation in the UK. The Bill sets out a regulatory framework through which online service providers will be held accountable for how they act to reduce the risk of harms on their services.

PROTECT HUMAN RIGHTS

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to protect human rights: The Bill contains provisions to protect rights, but these are both narrow in focus, privileging rights to non-interference with freedom of expression, and too vague. It fails to engage with the numerous human rights threats that arise from how online spaces operate and are regulated. Fundamentally, without a serious commitment to protecting human rights across the Government's agenda, this Bill will also fail to do so.

PROMOTE THE DEMOCRATIC PRINCIPLES OF FREEDOM AND EQUALITY

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to promote the democratic principles of freedom and equality: The Bill seeks to protect the freedom of democratic speech. But a pervasive lack of engagement with existing inequalities in how online spaces are operated and regulated means that it is likely to fail to adequately address issues such as hate, abuse, harassment, and discrimination that threaten democratic principles and processes.

PROMOTE INFORMATION INTEGRITY

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to promote information integrity: the Bill simply overlooks the risks posed by disinformation and fails to engage with how they can be combated, delaying this discussion to years into the regulatory set-up.

THIS PAPER SETS OUT:

1. Why we need digital regulation
2. An overview of the structure of the Online Safety Bill
3. Our success criteria for a digital regulation regime; along with where the Bill risks failing to meet these criteria, with suggestions for how the Government and the forthcoming regulator (OFCOM) could address these risks

BE TECHNICALLY EFFECTIVE AND FUTUREPROOF

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to be technically effective and futureproof: the Bill is designed with specific platforms, risks and functionalities in mind, and requires actions from platforms that conflict at best, or are impossible to deliver at worst.

BE TRANSPARENT, ACCESSIBLE AND INCLUSIVE

CURRENT RATING: **AMBER** - MEDIUM CONCERN

Medium risk of failing to be transparent, accessible and inclusive: The Bill includes provisions to require the regulator to publish information and engage in research and consultation to inform its operations. However, these requirements are high-level and

leave much discretion to the regulator, giving little power to citizens and civil society to be directly involved in the regulation of platforms.

BE ACCOUNTABLE AND INDEPENDENT

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to be accountable and independent: the Bill leaves too much power in the hands of the government to direct the actions of the independent regulator.

BE HOLISTIC

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to be holistic: we are in the middle of mass Government reforms to legislation upon which the Bill depends, including data protection and human rights, which have been widely criticised for weakening rather than strengthening the protection of citizens' rights.

In summary: the outlook is not good. We are supportive of the ambition of regulation, and the broad ways in which the Bill seeks to balance reducing risks to users with ensuring that platforms are held accountable for their own systems and not for individual cases of user content violations. However, we have strong concerns about the way the Bill is currently drafted, and both the procedural elements and the likely outcomes of the regulatory regime it describes. These are not irremediable. In light of the UK claiming to be achieving 'world-leading' regulation, we would do well to learn more from the work of the European Union and civil society in developing the Digital Services Act which has already grappled with many similar issues.

This document sets out recommendations to help inform the forthcoming Parliamentary process and, should the Bill be passed, the establishment of the new framework.

WHY DO WE NEED DIGITAL REGULATION?

The power to control what happens online - what information people see, how people are able to communicate with one another, what services they are able to access - increasingly rests with the corporations who dominate the market of providing online services. Individuals have few powers, and their representative governments are always a few steps behind the tech companies, constrained by the ever-evolving context, the transnational nature of these corporations, the technical complexities of regulating, or simply lack of political will.

The ways that digital technologies are being developed and deployed threaten fundamental values, not least the free exercise of human rights, and the health and flourishing of democratic society.

Digital regulation should aim to redress both of these issues: it should provide a framework through which power can be rebalanced, and demand that where those developing and deploying digital technologies do so in a way that threatens democracy, public health, or human rights, that they should be held accountable.

Tech companies claim they address these risks themselves, but from coordinated online attacks justifying state persecution of journalists in the Philippines, exacerbating atrocities being perpetrated in Myanmar, to the President of the USA inciting violence against protestors on Twitter and later inciting an insurrection, it is evident that whatever platforms are doing, it's not nearly good enough. Meanwhile, when platforms do take action, it ranges from arbitrary, opaque and inconsistent, to downright oppressive.

As it stands, platforms are much more susceptible to political pressure to abide by arbitrary blocks or bans which garner public support in a moment of crisis. This inconsistent over-action, under-action, lack of action, unclear action, shows that we need a principled framework, centred in human rights, to set out proactive actions and procedures platforms should be expected to take: rather than whenever a crisis arises, platforms employing knee-jerk and under-resourced mechanisms to address the threat, or just letting it fall by the wayside. Measures taken should be necessary and proportionate, and a regulatory framework offers a mechanism to support making these determinations using a systems-based approach.

None of these problems are new. But the rapid acceleration of new digital technologies - which shows no signs of slowing, as we enter the age of the Metaverse and web 3.0 - has escalated these risks and threats while rights protections and regulations have struggled to keep pace.

This is not to say that digital regulation is always an unqualified good. Clearly, the other major threat to rights and freedoms comes from states taking too much power in controlling what happens online.

Internet functionalities, services and spaces are weaponised by states seeking to curtail what their population knows, can say, can access or can do. From China's use of censorship and surveillance to perpetuate atrocities against the Uighur people, to Russia throttling access to key news services and social media to attempt to quash internal dissent about the invasion of Ukraine, to journalists being persecuted in online campaigns and threatened with prison for reporting alleged 'fake news': international warfare and domestic human rights abuses

increasingly are facilitated by states seeking more power over the online world.

Regulation, therefore, must not take power from Big Tech and hand it straight to governments. The purpose of regulation should be to give users more power online not less. Checks and balances: the rule of law, judicial oversight, independent regulators, the involvement of civil society and of citizens - these are complex but absolutely essential components of a regulatory framework. We cannot be complacent about digital regulation: about the importance of getting it right, the danger of getting it wrong, and the risk of doing nothing at all.

OVERVIEW OF THE BILL

The Bill sets out duties that apply to online service providers [see Part 2] which:

- Facilitate user-to-user sharing of content; and/or
- Have a search engine; or
- Publish certain pornographic content.

Which specific services are in scope of regulation has not yet been determined. This will depend on:

- The nature and strength of the services' links to the UK (such as number of UK users, if the UK is a target market for the service; or the possibility of UK users and the risk of harm posed to UK individuals by content on the service)
- Whether the service qualifies as exempt (based on things like the kinds of communication they facilitate, the functionalities they offer)

What duties they will have under that regulation [see Part 3] has also not yet been determined. Services will have different duties depending on which category they fall into: category 1, category 2A (search) and category 2B (user-to-user). Which category they fall into will depend broadly on the size and functionalities of the services [see Schedule 10].

All services will have safety duties that require them to address:

- Illegal content: in particular, priority offences [see Schedule 7] which include CSEA, terrorism, threats, stalking, disclosing private sexual images, drugs and weapons sales and others

The duties will be: [see Chapter 2 (9) and Chapter 3 (24)]

- To prevent users encountering priority forms of illegal content, minimise the length of time it is online and take it down when they become aware of it
- To mitigate and manage the risks of harm from illegal content
- To specify how users will be protected from illegal content in clear terms of service, applied consistently

Services that are likely to be accessed by children will have safety duties requiring them to address: [see Part 3, Chapter 2 (11) and Chapter 3 (26)]

- Content that is harmful to children

The duties will be:

- To mitigate and manage the risks of harm to children from content that is harmful to children
- To prevent children from encountering primary priority content that is harmful to children
- To protect children in particular age groups at risk from encountering other kinds of content that is harmful to children, including priority content and other harmful content
- To specify how children will be protected from harmful content in clear terms of service, applied consistently

Category 1 services will have safety duties requiring them to address:

- Content that is harmful to adults

The duties will be: [see Part 3, Chapter 2 (13-16)]

- To specify in clear terms of service, consistently applied, how the risks of priority harmful content will be managed through content moderation and content curation
- To notify OFCOM of the incidence of any other forms of harmful content to adults on the service
- To have features which allow users to prevent users who have not verified their identity to the service from interacting with their content, and reduces the likelihood of seeing content from unverified users
- To have features which allow users to use to increase their control over harmful content
- To have systems designed to ensure protecting the free expression of democratic content is taken into account in content and user moderation decisions, to apply this the same way to a diversity of political opinion, and to set out how this is done in clear terms of service, applied consistently
- To have systems designed to ensure protecting the free expression of journalistic content is taken into account in content and user moderation decisions, to have an expedited complaints procedure relating to these decisions, and to set out how this is done in clear terms of service, applied consistently

Category 1 and 2A services will have safety duties requiring them to address:

- Fraudulent advertising online

The duties will be: [see Part 3, Chapter 5]

- To prevent users encountering fraudulent advertising, minimise the length of time it is online and take it down when they become aware of it (category 1)
- To minimise the risk of users encountering fraudulent advertising (category 2A)

Services which publish pornographic content will have to have systems in place to ensure children are not normally able to access the content. [see Part 5]

For each of the categories of harm they are required to act on, services in scope will have to: [see Part 3, 4, 5, 7]

- Conduct a risk assessment against specific harms that will be set out by Government within that category (priority harms) [see Part 3]
- Have proportional systems in place to reduce those risks, which take into account user rights to freedom of expression and privacy (including content moderation, policies and practices, functionalities of the service, design of the service) [see Part 3]
- The system and level of risk reduction will vary depending on the category of harm being addressed and whether it is a priority harm or not
- These- systems should either be those recommended in OFCOM's codes of practice or satisfy OFCOM that they are still compliant with safety duties
- Have reporting systems and complaints procedures for users [see Part 3]
- Have record-keeping duties [see Part 3]
- Provide regular transparency reports to OFCOM [see Part 4]
- Comply with OFCOM's requirements around providing information [see Part 7]
- If they fail to adequately address risks, to use certain technologies or forms of technology required by OFCOM on their services in order to do so [see Part 7]

If services fail to comply with the regulations, enforcement mechanisms range from fines to criminal liability for senior managers to restrictions on whether UK users can access the service (depending on the failure). [see Part 7] Services can appeal to OFCOM, and entities representing the interests of users can file supercomplaints with OFCOM in the case of services' failure to comply with their duties. [see Part 8]

CRITERIA FOR A SUCCESSFUL DIGITAL REGULATION REGIME AND HOW DOES THE BILL STACK UP?

Our criteria for the success of a UK digital regulation regime are as follows, and include both outcomes and process requirements:

We also set out how well the Bill meets the criteria set out above for a successful digital regulation regime.

The conclusions are not good: we have serious concerns about how the Bill is going to measure up, from serious underprotections for human rights, to a lack of technical understanding, to a lack of clarity about fundamental concepts, to inefficacy against threats to democracy, to inappropriate levels of government control.

We have rated our concerns as red or amber according to the level of risk we see that the Bill will fail to deliver on a particular objective. On none of our key success criteria do we consider that the Bill as it stands warrants a green rating.

The good news is that this next period of scrutiny of the Bill as it goes through Parliament offers an opportunity to remedy these serious flaws, and deliver a truly world-leading regime. We hope that Parliamentarians will continue to engage with civil society, academic experts, and those with lived experience of the risks both of over- and under-regulation in order to inform their work.

HUMAN RIGHTS

DIGITAL REGULATION SHOULD PROTECT HUMAN RIGHTS

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to protect human rights: The Bill contains provisions to protect rights, but these are both narrow in focus, privileging rights to non-interference with freedom of expression, and too vague. It fails to engage with the numerous human rights threats that arise from how online spaces operate and are regulated. Fundamentally, without a serious commitment to protecting human rights across the Government's agenda, this Bill will also fail to do so.

Any digital regulation regime which fails to protect human rights from threats by domestic or foreign state actors, corporations, or individuals, is a failure. Regulation should recognise the intersections between rights, and address the diversity of ways rights can be threatened. Digital regulation that exists without a bedrock of robust human rights legislation is extremely unlikely to be able to achieve this.

The two rights that are particularly mentioned in the Online Safety Bill are freedom of expression, and privacy. However, the human rights ramifications of digital regulation are much broader. Of significant concern are the ways that the risk of serious human rights abuses are affected by what happens online: whether through increased risks of violence or facilitation of state persecution. There is a risk that 'protecting rights' is interpreted solely as ensuring non-interference in a narrow set of rights: and that wider impacts of digital regulation on protecting human rights are overlooked.

Freedom of expression

Threats to freedom of expression online are myriad: from platforms shutting down or moderating out content based on biased algorithms; viral and pile-on abuse silencing people; to the state determining what information or speech may or may not be accessed by users. There are freedom of expression concerns to be raised if people are systemically prevented from accessing or expressing themselves in online spaces, as these are key sites for political discourse, participation, and accessing information. Restrictions on freedom of information and association impact people's freedom of opinion and expression.

Regulation should hold platforms accountable for the risks which they are responsible for. This means the business models they pursue, the systems and processes they have in place, the design of their services and what content they incentivise, amplify, manipulate or encourage. This does not mean the content posted by users. The content is less important than the design of the systems that deliver that content.

Regulation should not be about adding content measures, but about changing the content measures which are inevitably used: moving from a framework where technical decisions are made to make shareholders more money, towards establishing a principled framework to examine those decisions, rooted in democratic values.

Privacy

Privacy is one of the most threatened rights by increased digital regulation, not least because increased digital regulation in some cases leads to increased surveillance: increased state or platform surveillance of users' identities or activity to attempt to prevent or deter crime, increasingly relying on automated technologies that are more likely to incorrectly identify members of minority groups as offenders or threats. However, privacy is already under threat from the operation of corporations, who provide free communication and information access services in exchange for mining and selling personal data, in opaque ways to facilitate targeted advertising and personalised pricing for products and services.

Successful digital regulation should address both challenges - reining in the surveillance-based business models that treat users as commodities to extract data from rather than as agents, while not

simply handing those powers over to the state to carry out instead. Proposals that replace corporate surveillance with state surveillance are inadequate.

Risks of violence and human rights abuses

Privacy and freedom of expression are not the only areas in which digital regulation should seek to protect human rights. Digital regulation should seek to prevent violence and associated human rights abuses.

It is a common claim that individual speech acts, although they can increase the risks that users face or commit violence (e.g. incitement to violence), never constitute violence in themselves. We do not agree with this position: the psychological impact of extremely harmful speech can also be significantly detrimental to someone's health and wellbeing. More relevantly for digital regulation, which should focus at a systemic rather than individual level: the prevalence, encouragement, incentivisation and normalisation of harmful speech acts (see e.g. dangerous speech), individually may not meet a criminal threshold, but collectively and at scale, can cause serious and ongoing harm, and increase the risks that users face, condone or commit violence.[1]

Inequalities, discrimination, prejudice and hate which exist offline not only exist online, but are amplified, scaled, weaponised and exacerbated online. Platforms react to emergencies in response to public pressure, but aren't preempting even clearly predictable harms against marginalised groups. They are complicit in coordinated state attacks against individuals and groups. And harms in the Global North are taken much more seriously by platforms than harms in the Global South.

An individual piece of speech is not responsible for another's actions upon reading it: but a system creating an information environment that makes abuse and harassment and violence more likely is implicatable. Similarly, the way people behave in online platforms is influenced by the design of platforms and the kind of behaviour that is accepted and normalised: meaning that harm is made more likely by poor platform design, which digital regulation should seek to address. These risks are exacerbated by for-profit business models, which create harmful incentives for platform design and practices - from promoting polarising and divisive content to mass profiling of users' personal data and inferred characteristics. Regulation is needed to rebalance these incentives in the interests of users

What happens online has concrete outcomes in the offline world, on the health and wellbeing of the public. Regulation must address these systemic failings, and not focus only on a few narrow types of illegal activity online.

HOW DOES THE BILL STACK UP ON PROTECTING HUMAN RIGHTS?

CURRENT RATING: **RED** - HIGH CONCERN

Systems-based approach

The Bill has always been claimed to promote a systemic approach, rather than a content-based approach. In the overall framework of the Bill - that is, placing duties of care on platforms, based on risk assessments of harm, to have systems and processes in place to reduce the risk of harm to users - we consider this to have been achieved. The Bill does not make platforms responsible for the presence of an individual piece of illegal or harmful content on their services.

However: this overall framing has failed to translate into the more detailed provisions of the Bill, leading to serious concerns that in practice, the regulation will focus on content over systems. Platform duties are still based in categories of content, focus on preventing access to or minimising exposure to content (rather than reducing risks of harm overall). Although risks to users from platform functionalities, design, systems and business models are included on the face of the Bill as elements to be considered within a platform's risk assessments and duties to take measures accordingly, which is welcome, these are very much presented as secondary to content duties, which are much more specific in the expectations they place on platforms [see e.g. Part 3, Chapter 2 (9)].

The Bill treats a 'systems' approach as meaning a 'systems-for-dealing-with-content' approach: and even where it has tried to introduce requirements around platform design elements, such as with the user identity verification duty [see Part 4, Chapter 1], this is in the absence of an assessment of specific risks and how they relate to specific design decisions on online services.

These are fundamental misunderstandings of how a systems-based approach should be realised in regulation.

Rights protections in the Bill

They are (the specific language varies between user-to-user and search services, but generally): when deciding on, and implementing, safety measures and policies, a duty to have regard to the importance of protecting users' right to freedom of expression and to the importance of protecting users from a breach

1 'Dangerous Speech is any form of expression (speech, text, or images) that can increase the risk that its audience will condone or participate in violence against members of another group.' https://dangerousspeech.org/about-dangerous-speech/

of any statutory provision or rule of law concerning privacy that is relevant to the use or operation of the service (including, but not limited to, any such provision or rule concerning the processing of personal data). [see Part 3, Chapter 2 (19) and Chapter 3 (29)]

There are also duties for:

- user-to-user services to have clear provisions in their terms of service about how to claim for breach of contract if terms of service are violated regarding their content [see Part 3, Chapter 2 (19)]
- all services to have user complaints procedures for if they fail to uphold their rights duties [see Part 3, Chapter 2 (18) and Chapter 3 (28)]
- And an understanding that companies will comply with their duties regarding rights if they implement the measures recommended by OFCOM in codes of practice (to be designed by OFCOM in consultation with human rights experts): and if they use other measures, they must have regards to the importance of freedom of expression and privacy. [see Part 3, Chapter 6 (45)]

Category 1 services have additional duties to carry out rights impact assessments regarding freedom of expression and privacy, to publish impact assessments, and to specify publicly what steps they have taken to protect freedom of expression and privacy [see Part 3, Chapter 2 (19)].

These duties suffer from being both overly vague and overly specific. They are not duties to protect, promote or uphold rights - only to 'have regard to the importance of' them. Moreover, the language is oddly specific: duties being tied to, for instance, 'breaches of statutory provisions or rule of law concerning privacy', rather than explicitly referencing Article 8 rights.[2] These deviations from how rights are expressed within the current Human Rights Act and European Convention leave room for concern as to how much deviation will be permitted. The first draft Online Safety Bill, though imperfect, at least explicitly identified privacy as a right - and now this language has been watered down even further to exclude this.

Moreover, OFCOM's duties towards rights are limited: there is no mention of rights in the online safety objectives that OFCOM are required to pursue [see Schedule 4], only that OFCOM must produce a statement each year about the steps they have taken to ensure their functions are compliant with Articles 8 and 10. [see Part 7, Chapter 7]. Given that OFCOM's recommendations in codes of practice will be taken as the benchmark for whether or not platforms' systems are rights-protecting, this is a worrying omission.

Requirements in the Bill that raise particular rights impacts concerns

Compounding the risk of the rights protections being vague is the fact there are many provisions in the Bill which run a strong risk of infringing on either privacy or freedom of expression.

For instance, OFCOM can also require a company to use proactive technologies on public channels, or in the case of CSEA detection, also on private channels [see Part 7, Chapter 5, 6]. These technologies can include content moderation technology, user profiling and behaviour identification technology. This is a significant power that could have huge consequences for privacy of user content, privacy-protecting functionalities such as end-to-end encryption, privacy of user characteristics and personal information, and little discussion is given to how these risks will be mitigated, or what the details of the specific requirements made of platforms will be.

Similarly, all service providers will need to take steps to protect children from encountering harmful content [see e.g. Part 3, Chapter 2 (11)] (and sites which publish pornographic content will have to have age verification measures in place) [see Part 5]. This will require significant measures taken, either to age-verify users or use age assurance technology to profile users: either way, significantly affecting the personal data platforms collect, hold, and use to make decisions about users.

Category 1 providers also will have a requirement to give the option for identity verification [see Part 4, Chapter 1]: with the result that users will be able to block content or interactions from unverified users [see Part 3, Chapter 2 (14)] and giving platforms political cover and indeed incentivisation to further restrict functionalities available to unverified users. And although the Bill leaves some discretion as to what method should be used, it does not set any minimum requirements for privacy protection.

Moreover, these requirements not only risk serious privacy violations by platforms seeking to comply

2 'Everyone has the right to respect for his private and family life, his home and his correspondence. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.' https://www.equalityhumanrights.com/en/human-rights-act/article-8-respect-your-private-and-family-life

with the regulation, but also will put marginalised groups at greater risk: those for whom it is not safe or not possible to verify their identity or their age, for instance, will face extreme restrictions of access and functionalities. Anonymity is a fundamental online protection that facilitates the realisation of other fundamental rights, including freedom of expression, privacy and safety, and should not be restricted directly or indirectly in this way.

Age-filters have also been known to over-moderate and block children from accessing vital information for them: meaning that oversight and remedy of the actual outcomes of these requirements is crucial.

RECOMMENDATIONS

Widening scope of rights protections

- The Bill should include a duty for platforms to have regard to the importance of freedom of access to information for both adults and children online, to constrain overrestriction of access to prevent harm
- Platforms' risk assessments should include not only a rights impact assessment but an assessment of how strengthening certain rights could protect users on their services
- Protections for human rights as stated in the Human Rights Act and ECHR should be included in the online safety objectives
- Platforms should be required to have duties to protect user rights in the application of all of their systems, not only those designed to comply with safety duties
- They should also include in their risk assessments how their activities, systems, business model and design (beyond just their safety procedures) might be impacting on their users' rights and require them to take steps to alleviate these risks
- The provisions for supercomplaints should specify that platforms' failures to protect privacy are also a grounds to bring a supercomplaint, rather than only harm and freedom of expression being mentioned on the face of the Bill

Reducing the risk of overreach

- The language should be standardised across the Bill relating to duties to 'prevent' or 'ensure users are not normally able to' or 'minimise the risk of' encountering certain harms, and use one framing which clearly allows for privacy protections to be in place
- Rights protection duties should set out more clearly the processes by which platforms will be expected to balance different rights where they may conflict: not a hierarchy of rights
- Privacy should be given equal consideration to freedom of expression, rather than take a secondary position as the difference in drafting implies: it should be specified that it is a right
- The identity verification duty should be scrapped
- OFCOM's requirements to use proactive technology currently stipulate that before they impose this they must 'consider matters' which include whether the use of that technology interferes' with user rights: this should be strengthened to a requirement
- OFCOM should consider in their codes of practice and audit regularly how requirements around age or identity verification online may be affecting people's freedom to access information and their right to a private life
- The Bill should reflect the wider landscape of rights that platforms should be respecting, such as the right to anonymity
- Rights impact assessments should be required to be carried out by OFCOM of codes of practice and specific measures recommended
- ISP blocking or access restrictions should only be used as an extreme last resort in the face of an overwhelming threat to public safety

FREEDOM AND EQUALITY

DIGITAL REGULATION SHOULD: PROMOTE THE DEMOCRATIC PRINCIPLES OF FREEDOM AND EQUALITY

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to promote the democratic principles of freedom and equality: The Bill seeks to protect the freedom of democratic speech. But a pervasive lack of engagement with existing inequalities in how online spaces are operated and regulated means that it is likely to fail to adequately address issues such as hate, abuse, harassment, and discrimination that threaten democratic principles and processes.

The unique power of the internet, its openness, and the facilitation of new forms of communication, is to provide greater freedom of information, association and expression than ever before. A liberal democratic, open internet ought to lower the barriers to accessing information and services to be open to all. The power of the internet to uplift marginalised voices, facilitate transnational solidarity and collective organising, and enable greater strides to be taken towards challenging injustice and rectifying inequalities.

But the inequalities that manifest offline are just as present online. Prejudice, discrimination, hate and violence put people at risk and drive them out of online spaces that they need; while biased design of technologies means that technological oppression falls heaviest on the most marginalised groups.

Abuse of women is reported with no action taken. Incitements to violence are left up on platforms. LGBTQ+ content is filtered as 'inappropriate'. Bias in how automated technologies are designed also mean that those biases are reproduced in the ways that content is curated and moderated, and what users experience and the ways users are treated. Content from marginalised communities isn't moderated correctly. Lack of contextual and language understanding, as well as lack of well paid, trained and supported moderators, has led to situations where ethnic violence has gone unaddressed, while legitimate speech (such as the use of reclaimed slurs) has been deemed toxic or harmful.

Digital regulation must recognise that positive freedoms and equality, empowering users rather than entrenching marginalisation, must be part of the goals of our online world, not simply ensuring non-interference.

HOW DOES THE BILL STACK UP ON PROMOTING THE DEMOCRATIC PRINCIPLES OF FREEDOM AND EQUALITY?

CURRENT RATING: **RED** - HIGH CONCERN

Protections against harm that threaten freedom and equality

What the platform safety duties for content that is harmful to adults will be required to tackle is difficult to know as the Bill does not state what forms of content will be priority. However, the risk assessment duties include taking into account content, functionalities of the service, and the design and operation of the service affects risks to adults. The duties also include specifying in the terms of service what content will be curated or moderated in certain ways, although no particular treatment for this content is mandated [see Part 3, Chapter 2 (12,13)].

In terms of the treatment of legal but harmful content, we consider that this is a reasonable balance. Platforms already moderate and curate content, to promote engagement, or sometimes in the apparent interests of safety - where they do so, it is in the interests of users that this is stated in terms of service so they can be held accountable. Although

an individual piece of speech online may not pass the threshold for harm that would deem it illegal, at a systemic level, many such pieces of content can have a cumulative harm, including threatening people's freedoms, that is significant enough to warrant action. For example, harassment campaigns - abuse, doxxing, threats, etc. - can drive people out of online spaces due to safety or privacy fears, interfering with their own expression in that space.

We agree that if this clause is interpreted as meaning a platform must apply one of the four content treatment options to priority harmful content, (taking down the content, restricting access, limiting its recommendation or promotion, or actively recommending or promoting) there are justifiable concerns for freedom of expression. This is because, in the absence of a clear balancing test between the level of risk posed by content that meets the criteria of a specific harm and the required action for said harm, action appears to be required. . Risks arising from legal but harmful content are better tackled through a systems-based approach that drives changes to the design and functionalities of the service identified through an adults' risk assessment [see Part 2, Chapter 3, 12 (5)] beyond simply terms of service about content treatment.

Where platforms themselves are employing systems that increase the risks of harm, such as hatred or harassment, to users, these should be the subject of oversight. One of the other major threats to rights and freedoms online is when platform systems, be they content moderation, content curation, user profiling or behaviour identification, end up replicating and entrenching inequalities, with marginalised groups most affected by failures in platform design and systems (such as being disproportionately over-moderated).

The Bill has little to say about considerations of equality and nondiscrimination, beyond certain requirements about experts in equality issues being consulted by OFCOM [see Part 3, Chapter 6]. There are requirements for platforms and OFCOM to take into account, in their childrens' and adults' risk assessments and understanding of harm, how content 'particularly affects individuals with a certain characteristic or members of a certain group' [see e.g. Part 3, Chapter 2 (5); part 12 (187)].

Supercomplaints can be brought for instance, if services are 'causing significant harm to users of the services or members of the public, or a particular group of such users or members of the public' [see Part 8, Chapter 2] - suggesting that discrimination against a particular group would be grounds for a supercomplaint.

However, aside from children, there are not named groups (such as women and girls, as has been campaigned for by the VAWG sector) highlighted as disproportionately affected by online harms that services should have particular regard for the protection of.

And the risks to particular groups are characterised in terms of differing levels of risk from content, which fails to take into account how groups with different characteristics may be disproportionately negatively affected by systems and processes platforms have or put in place under the regulation, and what steps should be taken to mitigate this risk. For instance, identity verification poses particular risk to some marginalised groups, but the Bill only recognises that groups may have different abilities to 'access' identify verification, [see Part 4, Chapter 1 (58)] rather than that engaging in the very process may put people at risk.

Protections of democratic speech

Alongside duties around content that is harmful to adults, Category 1 services have duties to protect journalistic content, and to protect content of democratic importance. These provisions are allegedly designed to protect against overreach in content moderation that many are concerned will result from the safety duties. What these duties entail, is vague in the text of the Bill. Platforms must operate systems which use proportionate systems and processes designed to ensure that the importance of the free expression of content of democratic/journalistic importance 'is taken into account' when making decisions about content or about a user (e.g. banning them). The journalistic content duties also include provisions for expedited complaints procedures for platform action relating to journalistic content [see Part 3, Chapter 2, 15,16].

Journalistic content is news publisher or user generated content, that is UK linked and 'is generated for the purposes of journalism' [see Part 3, Chapter 2, 15]. Democratically important content is news publisher or user generated content, that 'is or appears to be specifically intended to contribute to democratic political debate in the United Kingdom or a part or area of the United Kingdom' [see Part 3, Chapter 2,16].

Neither of these are blanket exemptions, saying that platforms may never moderate democratically important or journalistic content. However, it is clearly at least intended that there should be a higher bar for moderating this kind of content than other sorts of content regardless of the content's potential to cause harm at scale.

The vagueness of these definitions also leads to real worries of abuse. Content relating to journalism or democratic debate can apply to almost anything: indeed, there is significant and evidenced

harm which occurs within democratic political debates online - from harassment of politicians to marginalised groups being silenced under the guise of 'debate'. The defence before the publication of the Bill that 'gender critics' would not be silenced raises alarm bells, for instance, that transphobic abuse online will be defended as 'democratic speech' even when it clearly meets the level of risk of harm that against other people would qualify for content to be moderated.

And if it does not apply to almost anything, it begs the question: why are existing freedom of expression protections not good enough to protect these types of political freedom of expression, and if not, why does adding protections in that will likely benefit the speech of politicians and journalists (who often have significantly higher online reach than other users) most, do better at protecting those rights?

There is also a general exemption for news publisher content from platform duties altogether. Platforms will have a duty to ensure that the freedom of expression of news publisher content is taken into account in moderation decisions (via the inclusion of news publisher content in the definitions of content of democratic importance and journalistic content.) Whether or not there is a strict exemption from any moderation expected, therefore, there is a clear presumption that platforms are to treat the moderation of harmful news publisher content with more hesitation than other harmful content, and this will likely inform the recommendations made through codes of practice or guidance.

And this extra protection is granted while the scope of what counts as news publisher or user generated content is extremely unclear. For instance, 'a link to a full article or written item originally published by a recognised news publisher' [see Part 3, Chapter 7] counts as news publisher content, but if a user attaches harmful commentary to the link in posting it on another site, how the platform is meant to act on the user content but not the news publisher content is unclear. The definition of 'recognised news publisher' has also been challenged on similar grounds to the above, for being too broad and allowing extremist content in under another guise: as the standards required to meet the definition are extremely low and easily achievable by websites set up to spread disinformation or hate.

As drafted, therefore, the democratic importance and journalistic content protections risk doing little extra to protect freedom of expression, but offering protections for widespread forms of disinformation and abuse, as well as weaponisation and manipulation of the 'news publisher' definition by bad actors.

The categorisation of platforms as Category 1 platforms is currently linked to the number of users and platform functionalities (though the details are undefined [see Schedule 10], with even more scope to define the conditions for 2A and 2B services. This risks leaving small but high-risk platforms out of scope for taking meaningful action on various forms of extremism, abuse and disinformation, despite the fact they pose significant risks to users.

Overall, the Bill belies a misunderstanding of how power operates in the online world. It speaks to an outdated vision of online 'marketplaces of ideas' in which the best ideas will promote high-quality democratic discourse, as long as they are unencumbered by interference. This democratic vision does not exist: structural inequalities mean that such 'free debate' only further entrenches and silences already marginalised groups, whose voices ought to be at the centre of democratic debates.

RECOMMENDATIONS

- Platforms and OFCOM should have duties to consider in their risk assessments and codes of practice how platform systems as well as harmful content can pose risks to members of groups with different characteristics
- There should be additional duties for platforms and OFCOM to carry out equalities and impact rights assessments on the systems and measures being recommended and put in place
- The exemptions for content of democratic importance should be removed, and the freedom of expression protections strengthened. Platforms should state in their terms of service how they intend to balance harms with public interest, and enforce this consistently.
- The definition of a recognised news publisher should include higher thresholds
- At a minimum, it should be clarified that platforms are not being prevented from taking action on news publisher content shared on their sites in pursuit of their safety duties e.g. if content violates their terms of service
- Which category platforms are in should include an assessment on the basis of risk, not only of number of users and functionalities.

INFORMATION INTEGRITY

DIGITAL REGULATION SHOULD: PROMOTE INFORMATION INTEGRITY

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to promote information integrity: the Bill simply overlooks the risks posed by disinformation and fails to engage with how they can be combated, delaying this discussion to years into the regulatory set-up.

From the Cambridge Analytica scandal to Russian disinformation campaigns to voter suppression in the US, the January 6th insurrection incited on social media to China banning memes making fun of the state leader: what happens online, in our key sites of political discourse and political participation, news and information, has direct impact on democratic processes and citizen empowerment across the world.

With this greater freedom of and democratising of information also comes the correlating threats: disinformation and censorship both allow bad actors to weaponise the functionalities of internet services to disrupt the integrity of the information environment that people access. And platforms choose not to take meaningful action on the fundamental structures of their services that help foster misinformation in the interests of preserving profit: and conceal information about the level of threat of disinformation.

Digital regulation must address both these threats, and not tackle one at the expense of the other.

HOW DOES THE BILL STACK UP ON PROMOTING INFORMATION INTEGRITY?

CURRENT RATING: **RED** - HIGH CONCERN

Harmful disinformation and misinformation

Despite disinformation posing a demonstrable and ongoing threat to public safety, public health, the rights of marginalised groups, social cohesion and democracy, this Bill has remarkably little to say.

OFCOM must establish a committee, to report within 18 months, to provide advice on how service providers should deal with disinformation, and how OFCOM should approach disinformation in requiring information from platforms and how they should carry out their media literacy duties [see Chapter 7, 130].

There is a new false communications offence, which is a new individual criminal offence [see Part 10, 151] rather than forming a meaningful part of the new regulation, and in any case, as an individual offence, would not be an appropriate basis or method for tackling amplified disinformation or the diversity of forms of disinformation that include manipulating true information.

There are other clauses through which action on disinformation might be required, but this is extremely unclear (such as the ability to direct OFCOM to give priority to other objectives in the case of a threat to public safety [see Part 9, 146].

It is plausible that the duties around content that is harmful to adults [Part 3, Chapter 2 (13)] could include duties around disinformation, although as currently drafted disinformation would be difficult to designate a priority harm since only some forms of disinformation are especially harmful (as opposed to, say, abuse or harassment). Moreover, the action required would only be to act on disinformation insofar as it is prohibited in platforms' terms of service.

In short, the Bill as it currently stands requires effectively no action on harmful disinformation, and does not provide a clear roadmap to achieving action.

RECOMMENDATIONS

- OFCOM should prepare a code of practice which sets out steps platforms are expected to take to combat the risks associated with harmful disinformation and misinformation
- Coordinated inauthentic activity online should be designated a priority harm

TECHNICAL EFFECTIVENESS AND FUTUREPROOFING

DIGITAL REGULATION SHOULD BE TECHNICALLY EFFECTIVE AND FUTUREPROOF

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to be technically effective and futureproof: the Bill is designed with specific platforms, risks and functionalities in mind, and requires actions from platforms that conflict at best, or are impossible to deliver at worst.

For these outcomes, it is not enough that they are the intention of the regime. They must be evidenced outcomes. Lip service paid in regulation to the importance of safeguarding rights or promoting health is meaningless if concrete improvements are not in fact made.

This means that requirements of a digital regulation regime must be reasonably technically achievable given current technologies, but also be able to be applied to a range of current and future technologies. As such, they should be principle-based, with specifics determined by the regulator in an iterative and evaluative process which can easily be adapted and updated as understanding, risks and technologies evolve. Requirements which are designed with the risks and processes of a particular platform in mind (e.g. Facebook) or are designed from an idealistic standpoint without appreciation for the technical ramifications of what is being required are going to be both ineffective and encourage overreach.

HOW DOES THE BILL STACK UP ON BEING TECHNICALLY EFFECTIVE AND FUTUREPROOF?

CURRENT RATING: **RED** - HIGH CONCERN

There are elements in the Bill which have understandable intentions, but which are unlikely to be able to be implemented to any meaningful level of success.

Specific measures set out across platforms

In particular, the 'user empowerment' duties [see Part 3, Chapter 2 (14)] require that Category 1 services 'include features which adult users may use or apply if they wish to increase their control over harmful content', to reduce the likelihood they will encounter harmful content or alert the user to harmful content. Similarly, requirements that apply to all platforms regardless of individual risk and the functionalities of each platforms should not be designed with one set of risks and functionalities in mind: such as the requirement that 'a provider of a Category 1 service must offer all adult users of the service the option to verify their identity', [see Part 4, Chapter 1] which facilitates users preventing interactions with non-verified users [see Part 3, Chapter 2 (14)].

Both of these duties are to have specific systems in place to reduce the risk of harm to users, but which have significant risks of overreach and infringement upon user rights, without acknowledgement of this or the duties being conditional on the specific risks on a platform or a rights impact assessment of having such systems in place. This flies in the face of the purpose of the Bill, which is to ensure

platforms have systems in place tailored to specific risk profiles and in adherence with evidence-based recommended measures.

Priority illegal offences

The extension of priority illegal offences [see Schedule 7] which platforms will have duties which are likely to require proactive identification [see Part 3, Chapter 2 (9) and Chapter 3 (24)] also raises similar concerns, given its extension beyond the original offences of terrorism and CSEA, where there is a clear technical mechanism for identifying illegal content at scale. This leaves a significant degree of uncertainty about what the standards for identifying and removing illegal content will be, and how these will be balanced against the risk of overreach at scale.

Conflicts with rights protections

There is also little acknowledgement of where conflicts might arise from the technical requirements of the Bill. For instance, notices can be issued to require accredited technology to be used to identify and take down CSEA content on public or private channels [see Part 7, Chapter 5 (103)]. There is no mention of whether if channels are end-to-end encrypted, the technology that will be required will be compatible with preserving the integrity of end-to-end encryption or not; and how platforms are expected to meet their duties to user privacy in light of these requirements and other potential requirements to use proactive technologies and age verification to mitigate various harms.

Meaningful powers

The information that OFCOM is able to require [see Part 7, Chapter 4] also do not appear to have been designed with the facilitation of meaningful algorithm inspection in mind, that will be necessary for the regulator to be adequately assessing the systems that platforms have in place (including content moderation, user profiling, or behaviour identification), and are focused much more on transparency about policies and incidence of harm than on details of decision-making processes.

There are also reviews [see e.g. Part 9, 149] built into the framework, but these are infrequent and lack involvement of independent third parties to support the evaluation, testing and iteration of how measures are recommended.

RECOMMENDATIONS

- The user empowerment duties and user identity verification requirements should be removed from the Bill, and instead be considered during the drafting of the Codes of Practice.
- Language around users being 'prevented' from accessing certain forms of content, illegal or legal, should be amended (as it already is in some areas of the Bill - [see e.g. Part 3, Chapter 2, (9) to 'minimising' or 'mitigating' risk to avoid mandating the use of technical systems that will overmoderate.
- Requirements to use proactive or accredited technologies or processes like age verification should include clear privacy standards. Regarding Age Assurance, the Joint Committee report sets out useful recommendations as to incorporating privacy-protecting requirements.
- The information requirement provisions should be strengthened to facilitate and require algorithmic audit.
- There should be a process through which independent accredited third parties, such as researchers, should be able to bring to OFCOM's attention the need to amend a code of practice or guidance based on evidence of systemic failures in prescribed measures.

Procedure

There are key procedural requirements which the regime should meet, regardless of the outcomes or focus of the regime.

TRANSPARENCY, ACCESSIBILITY, INCLUSIVENESS

DIGITAL REGULATION SHOULD BE TRANSPARENT, ACCESSIBLE AND INCLUSIVE

CURRENT RATING: **AMBER** - MEDIUM CONCERN

Medium risk of failing to be transparent, accessible and inclusive: The Bill includes provisions to require the regulator to publish information and engage in research and consultation to inform its operations. However, these requirements are high-level and leave much discretion to the regulator, giving little power to citizens and civil society to be directly involved in the regulation of platforms.

The principles, processes and decisions of a regulator, how they are made, what information they are gathering, what information they have received, what instructions they have been given by government, who they have consulted and what they are recommending and why - all of this should be transparent and public by default, with exceptions made only in extraordinary circumstances, rather than the other way around.

The body or bodies in charge of digital regulation should ensure that their decision-making processes are not only transparent but are actively accessible to a wide range of stakeholders - at a minimum, that the information they provide is accessible and in plain English, but also that they are actively seeking to engage with and include different groups in their work: looking to models of co-regulation and power-sharing to ensure that the regulatory regime does not simply replace one form of top-down control of the internet for another, but works to actively democratise it.

HOW DOES THE BILL STACK UP ON BEING TRANSPARENT, ACCESSIBLE AND INCLUSIVE?

CURRENT RATING: **AMBER** - MEDIUM CONCERN

Public access to information relating to the regime

OFCOM must [see Part 7] publish transparency reports based on platform transparency reports; publish their codes of practice and guidance [see Part 3, Chapter 6], reports on reviews of the overall incidence and severity of online harms; risk profiles and the register of categories of services; an annual report; enforcement decisions; and they may publish other reports about online safety matters. There are exceptions, however, when material can and must be kept confidential. Moreover Ofcom have only to publish their required information 'in such manner as OFCOM consider appropriate for bringing it to the attention of the persons who, in their opinion, are likely to be affected by it' [see Part 11, 168]: leaving a significant amount of room for discretion in how publications are prepared and disseminated.

Platforms have transparency duties about information they must publish, terms of service, complaints procedures, and publicly available statements on positive steps which a platform has taken to uphold its safety and rights duties. These must be clear and accessible. There are additional transparency duties about risks of and actions taken by Category 1 platforms on content that is harmful to adults [see Part 3, Chapter 2, 3]

It is appropriate that the regulator should have powers to access and produce information that might not be appropriate for publication. However, the vague specifications in the Bill currently mean that the level of detail about, for instance, how and

why OFCOM has reached certain decisions, is likely to be very discretionary. Similarly, platforms must publish certain things, but there are not minimum standards for how detailed that must be, leaving a lot of room for information to be concealed.

Public involvement in OFCOM's decision-making processes

OFCOM has duties to arrange for research into public opinion and users' experiences of online services [see Part 7, Chapter 7]. OFCOM also must consult a wide range of people in the course of their preparation of guidance, codes of practice &c [see Part 3, Chapter 6]. Supercomplaints can also be brought to OFCOM where services are significantly adversely affecting users [see Part 8, Chapter 2].

Again, this leaves a significant amount of discretion to OFCOM as to not only who and how many people are consulted, but the manner of the consultation, which can be more or less accessible and inclusive. Moreover, supercomplaints can only be brought by 'eligible entities', the criteria for which are yet to be determined by the Secretary of State. It is unclear how far this will empower users and enable groups of users to raise complaints of systemic harm. It is vital that the criteria are broad so that interest groups are not able to be excluded on political grounds.

Independent scrutiny of platform compliance and risk profiles

An essential part of transparency, to hold both platforms and the regulator to account, is independent researchers being able to access platform data in order to understand what risks are present on different services, how they are being managed, and whether the changes being made are actually having the impact they claim to be. Currently, access to this kind of data is patchy and inconsistent, dependent on the goodwill of the platforms and easily revoked.

But all the Bill has to say on this is that OFCOM must produce a report within two years assessing the extent of independent researcher access to platform data and how this could be improved [see Part 7, Chapter 7 (136)].

RECOMMENDATIONS

- Securing independent access to platform data, by, for instance, vetted civil society researchers, is essential and must be included in the Bill. Instead of their report, OFCOM should be required to produce a code of practice on platforms' facilitation of independent access to data, the minimum levels of data access they must facilitate and the appropriate safeguards that should be included.
- Platforms and the regulator should be required to publish greater detail about the justification and processes involved in their decision-making, not only summaries of the outcomes
- There should be more minimum standards established for the process of OFCOM's consultations, not only who should be consulted
- Clarity and accessibility requirements should apply to OFCOM and the Secretary of State's publications as well as platforms
- The criteria for entities eligible to bring supercomplaints must be broad and established in consultation with users and groups who represent them

ACCOUNTABILITY AND INDEPENDENCE

DIGITAL REGULATION SHOULD BE ACCOUNTABLE AND INDEPENDENT

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to be accountable and independent: the Bill leaves too much power in the hands of the government to direct the actions of the independent regulator.

No regulatory regime is going to be perfect. There will be times that a government seeks to use regulation in its own interest, or to advance its own agenda: tech companies will do their best to meet the letter of the requirements while avoiding the spirit where it is inconvenient or costly to them. Different interest groups will see different things as success or failure. Transparency is crucial to be able to address these challenges, but is not sufficient: there must be sufficient power-sharing and oversight from different bodies, as well as the involvement of multiple stakeholders, that allow for course corrections, and for any overreach or inefficacy to be addressed. The independence of a regulator from the government is crucial to this.

HOW DOES THE BILL STACK UP ON BEING ACCOUNTABLE AND INDEPENDENT?

CURRENT RATING: **RED** - HIGH CONCERN

Secretary of State Powers

Since the original draft Bill was published, it's been a rare point of consensus across civil society that the Secretary of State powers give the Government excessive power to direct the operations and recommendations of the independent regulator, in a way that risks inappropriate Government overreach.

The revised Bill has made few improvements. The Secretary of State has powers including (but not limited to):

- Setting strategic priorities that OFCOM must bear in mind in their decisions about the online safety regime [see Part 9]
- Directing OFCOM to modify a draft of a code of practice for reasons of public policy, or (b) in the case of a terrorism or CSEA code of practice, for reasons of national security or public safety [see Part 3, Chapter 6 (40)].
- Reasons must be given for these changes except where the SoS considers that doing so would be against the interests of national security, public safety or relations with another government.
- The modified codes of practice must be laid before Parliament, which can be passed by the affirmative procedure only for changes made for reasons of public policy, or by the negative procedure otherwise [see Part 3, Chapter 6 (41)].
- Designating the harms which are considered 'priority' harms to children and adults, which platforms will have greater duties to address [see Part 3, Chapter 7]
- Determining the criteria for which entities are able to bring supercomplaints [see Part 8, Chapter 2]
- Determining the conditions under which a platform will meet the threshold to be within a particular category [see Schedule 10]
- Directing OFCOM in exercising their media literacy functions, to give priority for a specified period to specified objectives designed to

address a threat either to public health and safety or to national security [see Part 9, 146]

- Issuing guidance to OFCOM [see Part 9, 147]
- Powers to amend the Bill, including powers to repeal some existing exemptions for services out of scope [see Part 11]

RECOMMENDATIONS

We support Carnegie's recommendations that:

- 'We suggest that the Secretary of State's powers to direct OFCOM on the detail of its work (such as codes) are removed for all reasons except National Security.'

BEING HOLISTIC

DIGITAL REGULATION SHOULD BE HOLISTIC

CURRENT RATING: **RED** - HIGH CONCERN

High risk of failing to be holistic: we are in the middle of mass Government reforms to legislation upon which the Bill depends, including data protection and human rights, which have been widely criticised for weakening rather than strengthening the protection of citizens' rights.

A digital regulatory regime cannot live or die by one Bill or Act alone. It relies on an interlocking network of legislation, from criminal law to human rights law to data protection to product safety and competition legislation, as well as wider policy decisions around education, digital access, media literacy, not to mention international technological standards. Weakening any of these will weaken the whole regime.

The UK regulation also has significant implications for other models of digital regulation worldwide, and if it is not compatible with defending democracy, may well set a dangerous international precedent.

HOW DOES THE BILL STACK UP ON BEING HOLISTIC?

CURRENT RATING: **RED** - HIGH CONCERN

Framework of criminal law in the UK

Platforms' most stringent duties around content relate to content which is illegal under UK law: some of this is defined by existing offences, and new communications offences are being added through the Online Safety Bill. UK laws, however, are not perfect: from gaps in the law that mean some groups are seen to be underprotected, to issues of structural inequality, and in particular racial disparities, evident in how laws are enforced. These problems will only be translated across to how platforms implement their safety duties unless these problems are acknowledged, and proactively addressed and rectified.

Framework of data protection legislation in the UK

The privacy expectations on platforms make explicit reference to statutory provisions and the rule of law concerning privacy, including personal data. However, the Government is currently reforming the UK's data protection regime: meaning that if protections for data generally are weakened, the safeguards in the Online Safety Bill against platform overreach through their privacy duties are accordingly weakened, as would be efforts to improve online safety through competition policy.

Framework of human rights legislation in the UK

The backstop to what this means in practice is likely to come down to the wider framework of human rights legislation operational in the UK. The Government is currently proposing to reform the Human Rights Act in a move near-universally condemned by civil society (you can read our submission to the consultation here). In particular, the proposed reforms seek to prioritise freedom of expression above privacy and other human rights concerns, which seems reflected in the Government's approach to online safety, and recently the Government has stated their intention that freedom of expression should be a legal 'trump card'. This prioritisation of non-interference in speech risks interfering with crucial action from platforms to combat extremism, disinformation and abuse, without meaningfully strengthening positive promotion of freedom of expression for those groups whose speech is most marginalised. We do not believe there should be a hierarchy of rights, and that the Human Rights Act remains the strongest framework to provide a backstop to the rights protected in the Online Safety Bill.

RECOMMENDATIONS

- The Government's proposed reform of the Human Rights Act should be abandoned
- The Government should not seek to reform GDPR in a 'pro-growth, innovation-friendly' manner which in practice undermines the strength of user data protection
- After the regulatory regime has been set up, the Law Commission should carry out a review of where existing laws are leading to discriminatory outcomes in platform safety duty compliance (as opposed to where platforms are failing to comply with their safety duties)

We welcome engagement on any aspect of this paper: please contact ellen.judson@demos.co.uk for further information or discussion

Licence to publish

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.

b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.

c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.

d 'Original Author' means the individual or entity who created the Work.

e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.

f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b to distribute copies or phono-records of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phono-record of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicence the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.

b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital file sharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

c If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, you must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

a By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

b Except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licenced on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

a This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

b Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

a Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

b If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

c No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

d This Licence constitutes the entire agreement between the parties with respect to the Work licenced here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

DEMOS

Demos is a champion of people, ideas and democracy. We bring people together. We bridge divides. We listen and we understand. We are practical about the problems we face, but endlessly optimistic and ambitious about our capacity, together, to overcome them.

At a crossroads in Britain's history, we need ideas for renewal, reconnection and the restoration of hope. Challenges from populism to climate change remain unsolved, and a technological revolution dawns, but the centre of politics has been intellectually paralysed. Demos will change that. We can counter the impossible promises of the political extremes, and challenge despair – by bringing to life an aspirational narrative about the future of Britain that is rooted in the hopes and ambitions of people from across our country.

Demos is an independent, educational charity, registered in England and Wales. (Charity Registration no. 1042046)

Find out more at **www.demos.co.uk**

DEMOS

PUBLISHED BY DEMOS APRIL 2022

15 WHITEHALL, LONDON, SW1A 2DD
T: 020 3878 3955
HELLO@DEMOS.CO.UK
WWW.DEMOS.CO.UK