ANTISEMITISM POLICY TRUST

GOVERNMENT'S ONLINE HARMS WHITE PAPER RESPONSE

Contents

3 Introduction

4 Scope

5 Definining Harm

6 Duty of Care

7 Anonymity

7 Codes of Practise

7 Misinformation & Disinformation

8 The Regulator

8 Transparency

8 Researcher Access to Company Data

8 Enforcement

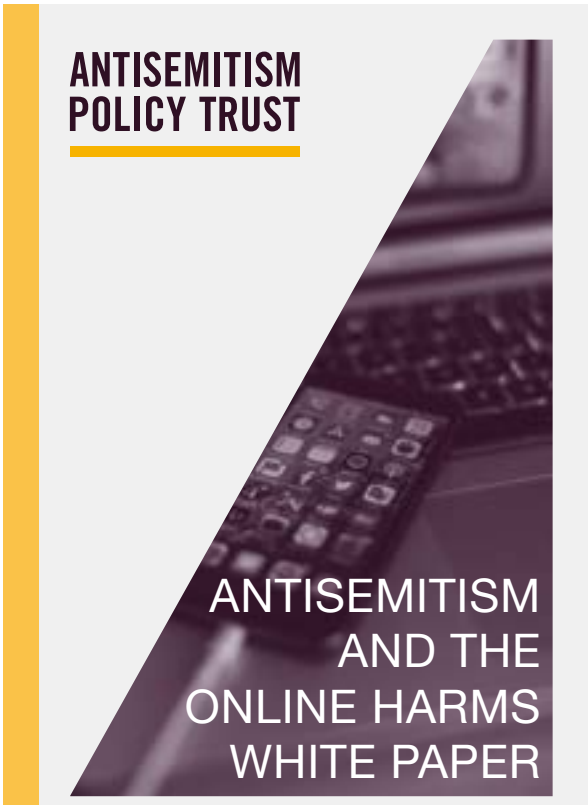
9 Technology Education

9 Conclusion

The text and illustrations may only be reproduced with prior permission of the Antisemitism Policy Trust.
Published by the Antisemitism Policy Trust, copyright © 2020
Antisemitism Policy Trust is a registered charity (1089736) [England] and company (04146486) [England and Wales]

Introduction

There is a lot to be welcomed in the Government's response to the Online Harms White Paper consultation. This follows the Initial response to the White Paper and the White and Green Papers that preceded it.[1] The Trust has followed and actively engaged throughout this process and looks forward to the introduction of the proposed Online Safety Bill.



The framing of the Government's latest response, which includes an understanding of the COVID-19 pandemic's impact on online abuse and dangers, importantly sets the UK within the global context. The Government is also right to say that "to unleash growth we need to ensure there is trust in technology". Finding the appropriate balance between freedom of expression, economic incentives and tackling harm will not be a simple task, but addressing antisemitism and other forms of hate will have numerous benefits. The promise of a duty of care, address for legal but harmful content and secondary legislation to specify categories of harm are all very welcome steps. We were pleased to see a number of the Trust's policy positions confirmed in the response. There are however areas of concern. Government says it will focus on "the biggest, highest risk online companies where most illegal and harmful activity is taking place" but it is abundantly clear that smaller and alternative platforms can be hotbeds of extremist, racist content. The prospect of the Law Commission's findings from its consultation on reform of the Communications Offences being drawn into the Online Safety Bill is also promising. However, at present, there is not enough detail on how abuse of women and intersectional abuse will be addressed. In summary, the Antisemitism Policy Trust has some reservations following the release of the Government's response, but it is a promising start.

1 https://www.gov.uk/government/consultations/online-harms-white-paper

Scope

The Trust supports the Government's proposed scope. It is right that Ofcom will take a risk-based and proportionate approach to regulation. We do however wish to better understand what the proposed focus on "companies whose services pose the biggest risk of harm" will resemble. Whilst Facebook and Twitter might have broader reach into public discourse, smaller platforms like Telegram and Bitchute have significant potential for radicalisation.[2] For example, the Institute for Stategic Dialogue found that in 60.1% of the channels propagating white supremacist content it monitored on Telegram, there was support for Terrorists or Terrorist organisations. [3]

We are also pleased that search engines will be in scope. We have previously detailed how small adjustments to search algorithms can lead to significant reductions in harmful prompts, for example Google's adjustment of the "…are Jews…evil" prompt. We are also pleased that instant messaging services and closed groups will be in scope.

As we set out in our White Paper response, there is significant potential for closed groups to be used for extremist activity. We hope that any Codes of Practice will reflect this reality. Finally, we commend the proposed review of out-of-scope services - which might become new havens for hate[4], and the extension of the regime to some user-generated adverts which might be used to promote hate.

2 https://jigsaw.google.com/the-current/white-supremacy/the-problem/

3 https://www.isdglobal.org/wp-content/uploads/2020/06/A-Safe-Space-to-Hate2.pdf

4 https://www.voxpol.eu/download/report/Beyond-the-Big-Three-Alternative-platforms-for-online-hate-speech.pdf

Defining Harm

The legislative approach to defining harm, that is to have a general definition of harmful content and activity covered by the duty of care, is reasonable. We are pleased with the specific inclusion of content or activity that "gives rise to a reasonably foreseeable risk of harm" as informing the duty of care. Furthermore, we are strongly supportive of the proposed secondary legislation to identify 'priority categories' of harmful content, posing the greatest risk to individuals. We would fully expect antisemitism to be in that list given the demonstrable impact it has as a motivator for, and indicator of, extremism. At the very least, we would expect any definition of harms to include reference to those with protected characteristics under the law. Antisemitism was not specified or explicitly referenced in the detail of the White Paper response, in relation to which harms will be listed, although 'hate content' was included, and we would like reassurance on this point.

The requirement Government plans to introduce on companies to understand the risks of harm arising from their services and to address these is critically important, and with the exception of management culpability in penalties, outlined below, we are comfortable with the position Government has taken on publisher liability, in not increasing existing liability. At present, the UK position follows Europe's e-Commerce Directive through which companies become liable for failure to remove illegal content 'expeditiously'.

The Government states that in order to safeguard freedom of expression, it will establish "differentiated obligations on companies in scope with regard to different categories of content and activity". It continues that "only a small number of high-risk, high-reach Category 1 services will have to address legal but harmful content and activity accessed by adults on their services". As outlined earlier, we are concerned that this categorisation not exclude smaller or alternative sites or applications including; Telegram, Bitchute, 8Chan, Gab and many other sites which host extremist and racist content with implications for real world harm.

It is right, however, that the Government highlighted this legal but harmful of content as requiring address. The online bullying and abuse, disinformation and advocacy of self-harm that the Government referenced in its response is appalling. This content is also relevant to antisemitism and other forms of racism. Antisemitic content, including Holocaust Denial or anti-Jewish coronavirus conspiracies, has significantly damaging effects, stifles users' freedoms and feeds toxic online environments.

We remain supportive of the Government's proposals for the regulator to issue codes of practice for systems and processes which will contribute to fulfilling the duty of care. However, we would also like to see a code of practice on hate crime and wider harm given further prominence, and to this end are working with the Carnegie Trust on a model code.

Duty of Care

The Trust is strongly supportive of the introduction of a duty of care, and the focus on systems and processes. The details provided by Government in its Annex to the White Paper response reflect a great deal of what the Trust has asked for, including a duty of care, Codes of Practice with Governmental oversight, a regulator with duties to consult, address for legal but harmful content and specificity of particular harms, and we are supportive of the Government's general approach with risk-assessments and safety by design at its heart. The use of risk-assessment and safety by design, in particular, will address the current situation where organisations like ours are repeatedly asked to assist companies to our detriment and their benefit. As above, we remain concerned about the approach to categorisation of companies, and hope smaller companies will be required to address legal but harmful content, in certain circumstances.

We were particularly pleased to see the attention to detail on implementation, enforcement and transparency of Terms of Service and that companies in scope "will be expected to consult with civil society and expert groups when developing their terms and conditions". The Trust already engages with numerous social media companies and would welcome the opportunity for further such engagement. We were also pleased to see recognition of the benefits to "those disproportionately affected by online harms, including groups with protected characteristics as they are currently more likely to experience harm associated with such content or activity online".

Anonymity

Whilst it is pleasing that the regulation will address anonymous abuse that is illegal through the duty of care, we believe the Government's stated approach to be insufficient. As we detailed in our briefing on anonymous abuse, a large percentage of antisemitic incidents from online sources are anonymous and we believe that legal but harmful materials being spread by anonymous accounts should be in scope for all platforms in respect of the duty of care. Anonymity, as our briefing makes clear, is important. It can ensure protection for whistle-blowers or victims of domestic or other abuse, for example. However, it is a privilege and should be subject to appropriate safeguards that prevent hate actors from abusing it.

In relation to illegal anonymous abuse, our view is that if a crime or a libel has been committed in the UK on regulated platforms, and they cannot or will not provide proof of identity, where a magistrate's court order demands it (subject to an appropriate burden of proof), then a range of options should be considered. The Trust believes that the civil or criminal liability should pass to the platform itself (this would be in line with existing measures in the e-Commerce Directive), and fines or other corrective measures could be put in place. Ultimately, companies should apply the 'Know Your Client/Customer' principle, familiar to those in the financial sector. In summary, this is the principle that organisations or individuals make efforts to verify the identity and risks in relation to the relationships or transactions they are entering into or undertaking.

The Trust submitted evidence to the Committee on Standards in Public Life in respect of electoral abuse, including on the role of social media in facilitating abuse of candidates, and on the simplification and update of electoral offences, and are pleased to see the Committee's recommendations being actioned.

Codes of Practise

We were pleased to see additional clarity on Codes of Practice in the White Paper response, including that these will be statutory, that Ofcom will consult relevant parties during the drafting process, that the absence of a Code of Practice does not absolve companies of responsibilities to act, and that there will be oversight from Government Ministers. It is somewhat disappointing that an interim code on hate crime and wider harms was not produced, given the levels of existing harm. We still maintain that the special powers afforded the Home Secretary in relation to Terrorism and Child Sexual Exploitation and abuse should also stand in respect of any hate crime code of practice.

Misinformation & Disinformation

The Trust is supportive of the measures outlined in the White Paper response but would like the expert group proposed by Government to include those with a firm understanding of antisemitic and anti-Muslim conspiracy theories, given the threat to Jewish and Muslim people caused by this specific type of hateful disinformation. These experts might also contribute to the safety by design framework the Government will introduce.

We are also strongly supportive of the Cabinet Office Defending Democracy programme, which seeks to bring together work to safeguard our British democratic processes, and have submitted evidence on the introduction of a digital imprints regime. The Trust highlighted the findings of the All-Party Parliamentary Inquiry into Electoral conduct including that electoral imprints be extended to online and other election communications, including for non-party groups or campaigners.

The Regulator

The Trust has long been supportive of Ofcom being established as the independent regulator, funded by industry, and supported parliamentary oversight; we are pleased to see confirmation of both. Specifically, parliamentary oversight of the codes of practice and priority categories of harms is helpful. We are very pleased that the regulator will be required "to take a consultative approach, including on the production of codes of practice". We also believe the balance between independence and appropriate oversight has been struck, and support review of the regime within the two - five year timeframe.

We were also particularly pleased to see the emphasis on Ofcom's proposed co-operation and interaction with other bodies, including the prospect of co-designation. It might be that, for example, Ofcom could design another body with relevant expertise for a particular function, as it did with the Advertising Standards Agency for the regulation of video-on-demand advertising content. We have long argued for such status for the Extremism Commission and British Board of Film Classification in respect of online harms.

We are pleased to see confirmation of the super-complaint function whereby substantial evidence of systematic issues affecting large numbers or specific groups of people can be heard. We are reassured that, in exceptional circumstances, specific platform functionality might be considered.

Transparency

We are supportive of efforts to widen transparency from companies in scope. We believe that, for Category 1 companies, details of the types of hate materials reported and removed should be compiled, and detailed in transparency reports, with specific reference to those with protected characteristics under law where they are not already doing so.

Researcher Access to Company Data

Proposals for research into online harms and best practice guidance on research activity are welcome. Effective research allows for a better understanding of specific issues, for example work the Trust is carrying out with the Community Security Trust and the Woolf Institute. We are aware that some companies, like Twitter [5]and Facebook [6], already facilitate some researcher access to their platforms but ensuring this is common practice will be beneficial to those seeking to address online harms.

Enforcement

The focus on encouraging compliance and industry engagement is one we support. We also understand the need to balance the attractiveness of the UK as a technology sector and effective enforcement.

The powers to issue directions for improvement and non-compliance notices are welcome. Civil fines up to £18 million or 10% annual turnover, whichever is higher, is also welcome. The last-resort powers, of disrupting UK business activities, including against a parent company, wherever that company is based, is a good backstop and the different levels of activity set out are sensible.

It is, however, extremely disappointing that Government has reserved the right to introduce criminal sanctions for senior managers who fail to respond fully to regulator demands, and will not do so for at least two years, if at all. Though a last resort, it is crucially important as an incentive and the Government should reconsider its position.

5 https://developer.twitter.com/en/solutions/academic-research

6 https://research.fb.com/

Conclusion

We look forward to the Online Safety Bill in 2021 and hope that with further measures on anonymity, oversight of a hate crime and harms code by the Home Secretary, the commencement of senior executive liability and work to ensure category two groups are not excused from addressing legal but harmful content, that it will be as strong as it can possibly be.
This is a once in a generation opportunity, it is important we get it right.

Technology Education

The proposed focus on media literacy is also welcome, including equipping users with skills to manage risks online and critically appraise information all the while liaising with, supporting and promoting the safety tech industry. We await details of the Media Literacy Strategy before taking a view as to its effectiveness.

In Review: Our key concerns

- Reassurance that a focus on "biggest" "highest risk" online companies will still capture small and alternative platforms that represent significant risk and contain harmful materials.
- A duty on category 2 companies that fall into the above category to be responsible for reviewing legal but harmful content.
- Further detail on abuse of women and intersectional harms.
- That codes of practice will contain details of the potential for action to be taken on closed groups in which extremist/terrorist materials are being shared or acts planned.
- That antisemitism or protected characteristics be included in the proposed identified categories of harm.
- Further prominence, including Home Secretary oversight for a code of practice on hate crime and wider harm.
- Action on anonymity, to include potential liability for failure to provide proof of identity in certain circumstances.
- Government's expert group on mis-and dis-information to include those with expert understanding of antisemitic and anti-Muslim conspiracy theories.

The Antisemitism Policy Trust's mission is to educate and empower parliamentarians, policy makers and opinion formers to address antisemitism. It provides the secretariat to the British All-Party Parliamentary Group Against Antisemitism and works internationally with parliamentarians and others to address antisemitism. The Antisemitism Policy Trust is focussed on educating and empowering decision makers in the UK and across the world to effectively address antisemitism.

Contact APT


www.antisemitism.org.uk


@antisempolicy


Antisemitism Policy Trust


mail@antisemitism.org.uk

The Antisemitism Policy Trust is a registered charity (1089736) [England] and company (04146486) [England and Wales]