

Written evidence submitted by the Antisemitism Policy Trust

Digital, Culture, Media and Sport Sub-committee on Online Harms and Disinformation Antisemitism Policy Trust response to the Call for Evidence – Online Safety and Online Harms

Summary:

The Antisemitism Policy Trust has been working to inform and inspire action against online harms for the best part of two decades, often in tandem with the All-Party Parliamentary Group Against Antisemitism to which we provide the secretariat. This submission comprises reflections and recommendations based on our research and expertise.

Regarding the shift from ‘online harms’ to ‘online safety’, we analyse the changes from the Government’s Green Paper to its White Paper, the interim response and finally to the draft Bill. Whilst we believe the Bill to be a major achievement, we are concerned that there has been a shift from efforts to tackle legal but harmful content, disinformation, and measures to introduce platform liability. The use of size and functionality, rather than risk, to determine which category companies will be in, means that the Bill may well not tackle much of the radicalising and harmful content on smaller platforms which provide safe havens for extremists.

On the question of harm, we believe that the regulator should determine harm, rather than leave it to be defined by technology companies. Additionally, we think a definition of ‘significant harm’ is required, and that references to physical and psychological harm also need further clarity. We support introducing ‘priority categories’ of harmful content and expect antisemitism to be one such category.

On platform design, it is important for the government to publish a Code of Practice that includes systemic, risk-management principles, such as the one developed by the Trust in partnership with others and published by the Carnegie Trust.

One notable omission from the Bill is the issue of anonymity. Anonymity can have a negative effect on online discourse, users’ feeling of safety and on their freedom of expression. We therefore call for an approach that will limit the anonymity of abusers but preserve privacy. The exemption of search engines from Category 1 companies is another key omission. Search engine algorithms can and have directed users to an abundance of harmful, racist, and extremist content and should be regulated accordingly. We are also concerned at the absence of a senior management liability scheme, which can be an effective measure at ensuring regulatory compliance. We believe the proposed reserved powers in relation to

senior managers do not go far enough. We further argue that reference to supply chains as well as mandatory breach notices for companies failing the duty of care, should be included in the Bill.

On the Committee's question relating to inclusions, tensions, or contradictions in the draft Bill, we recommend that the Secretary of State should not have the power to direct Ofcom in relation to Government policy. We also propose that counterbalancing 'protective' duties for Category 1 companies should be reconsidered because the Bill, as drafted, risks providing protection for those who actively undermine the democratic process. The exemption of newspaper comments board from regulation is also worrying, considering the harmful information on those services.

Regarding lessons learnt from other legislation, we show that despite concerns about the German NetzDG law, there is little evidence of over blocking or negative effects on freedom of expression.

1. How has the shifting focus between 'online harms' and 'online safety' influenced the development of the new regime and draft Bill?

In order to answer this question, it is perhaps prudent to revisit some of the key elements of the Green and White Papers issued by Government which set out its original intentions for what has become the Online Safety Bill. The Government's Green Paper was clear that all users should be empowered to manage online risks and stay safe. The subsequent Green Paper consultation response highlighted the various harms that had occurred since the Green Paper itself was published – including technology-based manipulation, data misuse and misinformation. It was clear that not only illegal but legal harms including cyberbullying, online abuse, harassment, trolling, and sexting were all of concern, and the catchphrase employed was “what is unacceptable offline should be unacceptable online”.

The White paper built on the Green, proposing a modification of the online liability regime in the UK, particularly in light of Brexit. There was promise of a regulator with a “suite of powers” to take actions against those breaching what would be a statutory duty of care, including “to impose liability on individual members of senior management”. There was also reference to a code of practice relating to hate crime and a section on tackling online anonymous abuse.

These were exciting, bold, and interesting proposals. They covered a number of the concerns that the Antisemitism Policy Trust, the All-Party Parliamentary Group Against Antisemitism, the Inter-Parliamentary Coalition for Combating Antisemitism and others had been raising with Government and many others over a long period of time.

The interim response to the Government's White Paper signaled a change in approach. There was a

noticeable shift to discussion of freedom of speech, a focus on business and a reduction in the scope for taking action against legal but harmful content. The full White Paper response went further still, limiting the number of companies captured by the duty to address legal but harmful content but maintaining a responsibility under the proposed duty of care to address content that gives rise to a reasonably foreseeable risk of harm. The Government's focus on the wider concept of safety, was perhaps at the expense of some of the ambition in relation to addressing harms. That is not to say that freedom of speech or a focus on business are bad, quite the contrary.

The draft Bill itself includes some of the specific details, for example, Category 1 companies which have a duty to address legal but harmful content must simply ensure such material is 'dealt with' through Terms and Conditions which do not, on the face of it, need to meet a particularly high standard. The more general duty of care to address reasonably foreseeable harms is replaced with more specific duties across user-to-user and search services, with 'balancing' duties for services in Category 1.

We have therefore moved from a position in which there were proposals for individual management liability, a broad and strong duty of care, action on anonymity, and a review of platform liability to a narrower set of actions with weaker penalties. The Code of Practice in relation to Hate Crime which was floated in the White Paper is now expected to be issued by Ofcom, and it was deeply disappointing that a draft Code was not published alongside the draft Terrorism and Child Sexual Exploitation and Abuse Codes. The Bill is still a major advancement on the current situation, and we have enjoyed constructive engagement with Government about its content, but it is nonetheless disappointing not to see bolder proposals being brought forward in some areas. That is the consequence of the change in approach.

A case study might help explain where the differences are in practice. On a call-in to the TalkSport radio station, one caller referenced the Jewish owner of Tottenham Hotspur Football Club in relation to the prospects of footballer Harry Kane and said: "He's a Jew, he's not going to let him go for nothing, is he?". This racist comment was not broadcast on the radio, but had it been, Ofcom would almost certainly have found it to be in breach of the Broadcasting Code rules on Harm and Offence which cover "generally accepted standards". Material causing offence, according to the Code, is justified by the context – there was none in this case. However, whilst the content was edited from the radio broadcast, it was livestreamed without interruption on YouTube, left on the platform for a long period of time and then shared widely on other social media.

At best, under the new legislation, Ofcom would look at YouTube's Terms and Conditions and check whether these are being applied consistently and whether they 'deal with' antisemitism. YouTube argues that they do but there are countless examples of anti-Jewish racism on the platform.¹

If the draft Bill included a more general duty of care, to address reasonably foreseeable harms, the focus might more clearly be on systems to prevent harm. The system in this case was livestreaming, and the ability of a platform to regulate livestreamed content. It might be that YouTube makes errors in the future, or that there are disputes over decisions in regulating livestreamed content, but a systems failure would be identified quicker, and solutions would address not individual content but rather more widespread concerns.

To take a second example, the racist twitter messages posted by the musician Richard Kylea ‘Wiley’ Cowie Jr pointed to the requirement for a platform, in this case Twitter, to have in place measures to quickly act on accounts with large followings which breach community standards. The upstream focus on systems will be more effective than downstream content moderation at tackling problems. The Trust’s concern is that without the overarching duty to address reasonably foreseeable harms, such as those exemplified by the cases above, and appropriate penalties, the Bill may miss the opportunity to incentivise good practice by technology companies when it comes to safety, especially Search engines which have no duty whatsoever to address legal but harmful content, despite evidence that they cause harm.²

The exemptions from addressing legal but harmful content and some of the counter-duties, are curious. We are concerned that adults will continue to be exposed to a substantial amount of information that is legal, but can be extremely harmful by causing psychological distress, promoting self-harm and radicalisation or indoctrination into extremist and racist ideologies. There are vast numbers of antisemitic, misogynist, homophobic, far-right and Islamist extremist materials, as well as disinformation other harmful content online and it is not clear from the Bill how these would be addressed, when present on Category 2 platforms. Platforms like Bitchute, Gab, 4Chan and many others have long hosted antisemitic content that might not fall foul of Communications or Harassment laws but would be addressed in other regulatory environments (TV, radio) or private arenas (football grounds, for example). Our view is that the platform categorisation should be determined according to risk, not size and functionality alone.

The emphasis on protecting ‘democratically important’ content, prohibiting platforms in Category 1 from discriminating against political ideologies and viewpoints, also marks a shift from the Online Harms White Paper towards a wider conceptual approach to safety. The Bill exempts content that is ‘specifically intended to contribute to democratic political debate’³ and journalistic content. We

¹ <https://www.thejc.com/news/uk/shame-of-youtube-jew-hate-1.517411>

² <https://antisemitism.org.uk/wp-content/uploads/2020/06/APT-Google-Report-2019.1547210385.pdf>

³

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safet

wholeheartedly agree that preserving democratic principles is critical. However, the way this section is framed could give extremists opportunities to exploit such exemptions and disseminate harmful content online. For example, a far-right activist might demand their material be re-platformed after being barred for posting extremist or racist content, or disinformation, claiming bias or discrimination by the platform. In this context, the draft Bill does not provide information as to how to deal with material that promotes disinformation and extremist materials that undermine the democratic process, rather than preserving it, as the Bill intends to do. What would happen in the case of someone spreading misogyny in an electoral race against a candidate from the Women's Equality Party? The vagueness of what constitutes journalistic and political content also raises concerns about the ability of service providers to effectively moderate harmful content that falls under this category.

In summary, the Bill still represents a welcome and necessary move forward to protecting users in a presently largely unregulated space. There are important considerations required in relation to free speech and economic prosperity, but the shift of explicit focus from harms has potentially limited some of the bolder options for action discussed in the Government's earlier proposals for legislation in this area.

2. Is it necessary to have an explicit definition and process for determining harm to children and adults in the Online Safety Bill, and what should it be?

The Trust stands by the principle that it should be the British Parliament, or a regulator acting with its authority, which should determine harms occurring online and that this should not be the arbitrary whim of technology companies. As outlined above, we were supportive of a broad duty of care, encompassing an obligation to take steps to address reasonably foreseeable harms, similar to that contained within the Health and Safety at Work Act or the proposals made by the Carnegie Trust. The Carnegie Trust's subsequent suggestions for defining harms are also detailed and merit serious consideration.⁴

The Bill uses the term 'significant harm' repeatedly; in respect of jurisdiction, the proposed powers for Ofcom and the role of Super-complainants, for example, but without providing a definition. There is no explanation as to how 'significant harm' differs from 'harm' and what legal implications, if any, this has. This requires clarification. Furthermore, the detail on harm in Clause 11 in relation to legal but harmful content and Category 1 platforms requires further explanation.

The Bill also references physical and psychological harm in very broad terms. The Threshold for

[y_Bill_Bookmarked.pdf](#), 13

⁴ https://d1ssu070pg2v9i.cloudfront.net/pex/carnegie_uk_trust/2019/04/08091652/Online-harm-reduction-a-statutory-duty-of-care-and-regulator.pdf

determining harm is described in the explanatory notes, but this needs further clarification and emphasis. Psychological harm is a fairly subjective term and using vague language will cause future problems for moderators, the regulator, and members of the judiciary, who will be required to distinguish between common and tort law interpretations.

There is a specific omission in the Bill regarding the harm caused to electoral candidates and public figures. The Antisemitism Policy Trust provided the secretariat to the All-Party Parliamentary Inquiry into Electoral Conduct.⁵ That All-Party report discusses the threats electoral candidates might face when running for office and recommended amongst other measures, the use of digital imprints on electoral material. We provided evidence to the Committee on Standards in Public Life (CSPL), including details of the all-party report. The CSPL's own report on intimidation in public life referenced the all-party report and made specific recommendations about social media.⁶ However, there is no address for the CSPL proposals in the draft Bill. If the Government's 'Defending Democracy' programme, run through the Cabinet Office, is set to cater to these issues, then that should be made explicitly clear, including whether separate legislation will be brought forward to address attacks on electoral candidates and others. It seems odd that this subject is absent from the draft legislation.

We were and remain strongly supportive of 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. We were pleased to see the Secretary of State give an indication this would be the case when appearing before the full DCMS Select Committee.⁷ At the very least, we would expect any definition of harms to include reference to those with protected characteristics under the law, and together with partner organisations, we are calling for greater and specific recognition of intersectional harms, including antisemitic misogyny.

There is a globally agreed, international standard for defining antisemitism – the International Holocaust Remembrance Alliance (IHRA) definition of antisemitism. The Trust does not recommend its inclusion in legislation, as it was not designed for such purposes. We would however strongly recommend Ofcom has reference to it in the future, as it does now in other areas of its work.

⁵ https://antisemitism.org.uk/wp-content/uploads/2020/06/3767_APPG_Electoral_-Parliamentary_Report_emailable.pdf

⁶

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/666927/6.3637_CO_v6_061217_Web3.1_2_.pdf

⁷ <https://committees.parliament.uk/oralevidence/2185/pdf/>

3. Does the draft Bill focus enough on the ways tech companies could be encouraged to consider safety and/or the risk of harm in platform design and the systems and processes that they put in place?

The Trust believes, as set out in the answer to the first question, that there should be a broader duty on platforms to consider and address reasonably foreseeable harms. We are not convinced the current duties do this to the extent that they should. To this end, the Trust worked to develop what became a draft Code of Practice published by the Carnegie Trust⁸ which includes some of the more detailed systemic, risk-managed principles we thought should be in place in guiding development of technologies. As outlined, we were disappointed a version of this Code or any other on the same topic was not published by Government alongside the Terrorism and CSEA Codes. We think it would be helpful for Government to adopt and promote the code on Hate Crime in advance of Ofcom receiving its powers.

4. What are the key omissions to the draft Bill, such as a general safety duty or powers to deal with urgent security threats, and (how) could they be practically included without compromising rights such as freedom of expression?

Anonymity

The issue of online anonymity is entirely absent from the draft Bill, to its detriment. Anonymity has many benefits. For example, by protecting users from the LGBTQ+ community, victims of domestic violence, and making it safer for whistle-blowers and political dissidents of oppressive regimes to speak out. However, it also has a significant negative impact on online discourse. The ability to be anonymous contributes to trolling, bullying and hate speech, which has a harmful psychological impact on users. This has created a threatening and toxic environment in which many users, including and specifically women and ethnic and religious minorities, suffer constraints to their freedom of speech as a result. Research by the Antisemitism Policy Trust has found that anonymity has a widespread impact on online abuse towards Jewish individuals and on the wider Jewish community.⁹

Based on our findings, we recommended using a “Know Your Client/Customer” approach, similar to that used in the financial sector. In our view this would guard the privacy and identity of users who wish to remain anonymous but restrict the right to anonymity for users who choose to target others through illegal activity. Our premise is that users will be less inclined to use hate speech and other abusive

⁸ <https://www.carnegieuktrust.org.uk/blog/draft-code-of-practice-in-respect-of-hate-crime-and-wider-legal-harms/>

⁹ <https://antisemitism.org.uk/wp-content/uploads/2020/12/Online-Anonymity-Briefing-2020-V10.pdf>

language, images, or videos, if their identity is known to an online platform and if they are in danger of waving their right to anonymity if their behaviour violates the platform's terms and conditions, or the law. To that end, we would recommend platforms include clear rules on anonymous accounts and enforce these with vigor. We also recommend setting the bar for removing anonymity high, requiring a burden of proof for a magistrate to order disclosure of identity. We understand there are concerns about data security. Technological solutions, including middleware, might be considered as might efforts in other jurisdictions to address these issues. One example is the Pan Canadian Trust Framework (PCTF) developed by the Digital ID & Authentication Council of Canada (DIACC) and the Pan-Canadian Identity Management Sub-Committee (IMSC) of the Joint Councils of Canada.

We are aware and supportive of other approaches, including those proposing verification systems which would allow users the opportunity to only engage with verified accounts on platforms hosting user-to-user services, but would like to see action that goes further than this.

It is important to recognise that privacy and complete anonymity rarely exists online. Technology companies gather an enormous volume of personal information about the users of their services and sell those for profit. As such, placing formal restrictions on anonymity will make a marginal difference in users' actual privacy, but can have a major impact on creating a safer and more positive online environment. Using a digital identity may be another safe way for individuals to maintain anonymity unless they commit an offence. The technology for creating digital identities already exists, and the Cabinet Office in collaboration with DCMS is already looking into creating a framework for digital identity governance that will safeguard privacy and prevent identity theft. This framework might also apply to the prevention of, and action to tackle, online anonymous abuse.

Search engines

As outlined above, search engines are included in the draft Bill, but will not be designated as Category 1 services. Search engines have a duty of care in relation to illegal material and content that children are exposed to through their services. However, they do not have a duty to remove legal but harmful content that adults are exposed to. Considering our research¹⁰ into Google's 'safe search' option, which has found to be inadequate when it comes to producing a large amount of antisemitic content, and related concerns we have raised with Microsoft Bing about its platform, we believe that it is a mistake not to include some search engines, in particular the ones most commonly used by children and adults, in Category 1.

There is no specific reference to voice technology in the Bill, and certainly, if Alexa or Siri were

¹⁰ <https://antisemitism.org.uk/wp-content/uploads/2021/05/Unsafe-Search-Report.pdf>

directing people to harmful materials and answers – as the former was¹¹- one would hope this would be addressed. At present, the Bill would not, for example, deal with the Alexa service telling people that the Holocaust did not happen in response to a question.

Senior Management Liability

Senior management liability is included in the draft Bill, but only as a reserved and fairly limited power relating to specific information retrieval. Fines are important but will be written off as the cost of doing business by major companies. Senior management liability, as we have in financial services, can be a powerful and effective tool that will encourage companies to comply with the law and face consequences. It is therefore our position that Ofcom should be granted full enforcement powers with associated criminal sanctions under Chapter 6 of the draft Bill in relation to senior management liability, for breaches of the duties of care, in extremis.

Supply Chains

Platforms, particularly with user-to-user generated content, employ services from third parties. In the past, this has included the provision of Gifs (moving pictures), for example and might extend to the hiring of external content moderators. There are examples in UK legislation, for example the Bribery Act, in which a company is liable if anyone performing services for, or on the company's behalf, is found culpable of specific actions. Reference to supply chains, and similar culpability would be appropriate.

Mandatory Breach Notices

Clause 96 details provisions for publishing decisions by Ofcom but there is no provision to mandate publication of a breach notice by a service. The Trust believes publications, like newspapers directed by IPSO or others, should have details of their breaches of a duty of care available to view on the platform.

5. Are there any contested inclusions, tensions or contradictions in the draft Bill that need to be more carefully considered before the final Bill is put to Parliament?

Powers of the Secretary of State

The draft Bill empowers Ofcom to consult on, and to produce, numerous codes of practice. The Secretary of State's power to direct Ofcom to modify its codes of practice to bring them in line with

¹¹ <https://news.sky.com/story/mps-demand-amazon-explain-why-alexa-offers-messages-from-antisemitic-websites-and-conspiracy-theories-12142340>

government policy is unnecessary and a bad precedent and should be removed. Ofcom should be held to account by parliament for its decisions, as it is in existing scenarios.

Counterbalancing ‘Protective’ Duties for Category 1 Platforms

As discussed above, there are counterbalancing ‘protective’ duties which apply to Category 1 platforms which have the duty to address legal but harmful content. These protective duties cover rights, such as free speech and privacy, content of democratic importance and journalistic content. As the Trust has set out, whilst perhaps well intentioned, the drafting of these duties appears problematic. There are those that actively undermine the democratic process, or cause harm using it as cover. For example, might standing in an election allow a racist activist to claim their content should effectively be re-platformed once barred from a service? What about those standing on a neo-Nazi platform more broadly? And we have set out concerns about misogyny in answer to the first question. The protective duty in relation to journalistic content presents similar and additional concerns. Journalistic content is poorly defined and might be read, as the anti-racism NGO Hope Not Hate has suggested, as content “generated for the purposes of journalism”.¹² To this end, citizen journalists’ content achieves the same protections (through Ofcom) and enhanced routes to appeal (through platforms) as content from any major national publication. There are examples of far-right activists self-identifying as journalists, and news companies like InfoWars which spread hateful and dangerous conspiracy theories. There is little assurance on the face of the Bill that content produced by such individuals and companies might not be offered special protections, or otherwise benefit from these duties as currently drafted. We are also not convinced by the explanations offered to the Trust and others by officials, that platforms will have to perform a balancing act in respect of harms and content.

Newspaper Comments Boards

There is an explicit exemption for content present on the website of a recognised news publisher. This is deeply problematic. The Antisemitism Policy Trust has worked with Government and others for many years, to highlight the abuse on newspaper website comment forums. For example, as secretariat to the APPG, the Trust worked with the Department of Communities and Local Government (now MHCLG) towards a guide delivered by the Society of Editors in 2014¹³ which was inspired by discussion of this form of harm. The exemption in Clause 18 relating to newspaper comments boards should be removed or, at worst, amended to ensure publications have measures in place to address harm on relevant boards.

¹² <https://www.hopenothate.org.uk/2021/06/07/hope-not-hates-response-to-the-draft-online-safety-bill/>

¹³ <https://www.societyofeditors.org/wp-content/uploads/2018/10/SOE-Moderation-Guide.pdf>

6. What are the lessons that the Government should learn when directly comparing the draft Bill to existing and proposed legislation around the world?

A big part of the debate on the forthcoming Online Harms Bill will be around freedom of expression. The same debate took place in Germany when it advanced its NetzDG legislation. There were concerns that the threat of high fines, up to €50 million, will act as an incentive for companies to delete first – and ask questions later, leading to over blocking and stifling free speech. Despite flaws in that text, information based on transparency reports and court rulings appears to indicate that ‘over blocking’ of content has not materialised.¹⁴ Censorship en masse by over-zealous platforms has not been the result of the German law. In fact, studies show that online hate speech and harassment had a more profound effect on restricting other users’ participation and freedom of expression, than the law has.¹⁵ There is a specific concern in Germany, about the negative effects that hate speech – which has been disproportionately directed at politicians and women – had on the two groups’ freedom of expression.¹⁶ Therefore, the argument that legislation will curtail free speech largely misses the point. Free speech is not unlimited. There are limits in the UN’s International Covenant on Civil and Political Rights, in the European Convention on Human Rights and here in Britain. Legal but harmful content is regulated in other areas already and the harms are defined by parliament or regulators which divine their authority from it.¹⁷ Failure to regulate might in fact have the inverse effect and benefit those seeking to impinge on the free speech of minorities or particular groups.

The German Government’s concern with the rise in far-right extremism and with an increase in online hate speech in general, and its effect on real life violence, including the targeting of asylum seekers, Jews, and Muslims, is one of a number of factors that has led to discussion of amending the NetzDG legislation. The plan is to amend not only the NetzDG but other laws too, such as the Criminal Code and the Telecommunications Act.¹⁸ One set of amendments to the NetzDG law, passed in June 2021, introduced greater transparency by companies and improved, user-friendly procedures for flagging unlawful content.¹⁹ This is something the Government should consider – both the efforts to amend NetzDG and the broader principles, which might lead to further engagement with the Law Commission. The Commission’s work on Communications Offences has been published but its review of Hate Crime Law has not. A Bill which learns the lesson of efforts overseas would, of course, be better informed and hopefully more effective.

¹⁴ <https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1782&context=iplj>, p.1129-30

¹⁵ Ibid.

¹⁶ <https://inform.org/2021/04/24/germany-new-law-against-right-wing-extremism-and-hate-crime-judit-bayer/>

¹⁷ https://www.counterextremism.com/sites/default/files/CEP-CEPS_Germany%27s%20NetzDG_020119.pdf

¹⁸ <https://policyreview.info/articles/news/germany-amending-its-online-speech-act-netzdg-not-only/1464>

¹⁹ <https://www.loc.gov/item/global-legal-monitor/2021-07-06/germany-network-enforcement-act-amended-to-better-fight-online-hate-speech/>

Prepared by Danny Stone MBE, Chief Executive and Dr Limor Simhony Philpott, External Affairs and Policy Manager.