Antisemitism Policy Trust—written evidence (FEO0007)

Lords Communications and Digital Committee
Inquiry into freedom of expression online

The Antisemitism Policy Trust is a charity that works to educate and empower parliamentarians and policy makers to address antisemitism. For more than ten years, the Trust has provided the secretariat to the All-Party Parliamentary Group (APPG) Against Antisemitism.

The Trust has conducted extensive research into online harms and produced numerous briefings and responses to inquiries on the subject. We write to contribute to the Lords Communications and Digital Committee inquiry based on our findings.

1. Is freedom of expression under threat online? If so, how does this impact individuals differently, and why? Are there differences between exercising the freedom of expression online versus offline?

1.1 The Trust contends that freedom of expression can be and is restricted for some by others choosing to abuse this freedom in order to threaten, abuse, harass and disseminate racist, radical and violent information. Those being abused are made to feel that they cannot freely and safely express themselves online for fear of being targeted with aggressive behaviour, that can cause depression and anxiety, and can spill from the online realm, offline.

1.2 Members of the Jewish community have been suffering racial abuse online and offline over a number of years. Speaking in 2015, then student Izzy Lenga spoke of the backlash to tweets she had posted exposing antisemitism. "...the backlash to my tweet has been extremely nasty and deeply upsetting. This sets a worrying precedent. I am worried about the rise of antisemitism across Europe and the world, and at points I am worried for my safety and that of my peers, but I am most concerned for the Jewish student community." Ms. Lenga is certainly not the only Jewish woman to have been attacked online. A briefing published by the Trust, 'Misogyny and Antisemitism',[1] highlighted a study compiled for our organisation which found that the platform 4chan had 630,000 antisemitic posts in 2015, rising to 1.7 million in 2017, whilst misogynistic posts had also risen significantly. The overlap of antisemitic and misogynistic posts had risen by some 180% over that three-year period. Matters are not improving. According to the NGO Glitch, almost one in two (46%) of women and non-binary people reported experiencing online abuse from the beginning of the COVID-19 pandemic[2], whilst the Community Security Trust (CST) reported that in the first six months of 2020, it had recorded its highest ever number of online incidents for a similar period. The incidents CST recorded were but a small indication of the volume of antisemitic content encountered by Jewish people online, as the organisation has strict reporting and recording parameters.[3] Unfortunately, such abuse is widespread and not limited to one particular platform. Overall, this represents an alarming increase in antisemitic rhetoric and bullying, and abused users – whose

---

1 Policy Briefing: Misogyny and Antisemitism, *Antisemitism Policy Trust*, May 2019, https://antisemitismpolicytrust.sharepoint.com/sites/AntisemitismPolicyTrust/Shared%20Documents/APT%20Policy%20Briefings/misogyny/5982%20Misogyny%20and%20Antisemitism%20Briefing%20April%202019%20v1.pdf

2 The Ripple Effect : Covid 19 and the Epidemic of Online Abuse. *Glitch*, September 2020, https://fixtheglitch.org/wp-content/uploads/2020/09/Glitch-COVID-19-Report-final-1.pdf

3 Antisemitic Incidents, January-June 2020, *CST*, https://cst.org.uk/data/file/c/5/Incidents%20Report%20Jan-Jun%202020-1.1596720071.pdf

freedom of expression is restricted for fear of retribution – must have better legal protection.

1.3 At a bare minimum, to achieve this, the same restrictions placed offline on freedom of expression, for those committing hate-speech offences, should apply online. However, we welcome the Government's proposed plans for new legislation including the introduction of a duty of care and codes of practice, its consideration of the Law Commission's review of communications (and we would hope hate crime) laws, and the protections for freedom of expression outlined in its recent response to the White Paper consultation.

3. Is online user-generated content covered adequately by existing law and, if so, is the law adequately enforced? Should 'lawful but harmful' online content also be regulated?

3.1 The Trust maintains, based on research and evidence, that user-generated content is not covered adequately by existing law. We have contributed to the consultations on proposed changes by the Law Commission to section 127(1) of the Communications Act 2003 and section 1 of the Malicious Communications Art 1998, and to hate crime law. We have long argued that legislation needs consolidating and hope that the Online Safety Bill will help in this regard. It will be interesting to see the way in which the Bill harmonises existing legal protections, for example those afforded to service users in the Equality Act, with the new protections it will afford.

3.2 We also welcome the consideration given to legal but harmful content in the proposed Bill. Facebook, to take one example, has already started applying what we deem to be appropriate and necessary restrictions to legal but harmful content by banning content such as Holocaust denial and materials relating to the QAnon movement. We have explained the impact of Holocaust denial, which remains legal in the UK, on both the propagators and the victims of this hateful lie.[4]

3.3 We have found a strong correlation between online content that radicalises and spreads disinformation and racist materials, and offline acts of terrorism, radicalisation and violence. We have outlined in our materials that the Christchurch mosque attack in 2019, the Pittsburgh synagogue attack in 2018 and the Finsbury Park mosque attack in 2017 were committed by far-right extremists who were radicalised, at least in part, by online extremism.[5] There are many examples of online radicalisation in which the content consumed is not explicitly illegal but might well contravene a duty of care to a platform's users. If this content is not regulated, there is a significant and ongoing threat that it will lead to further acts of violence and curtailment of freedoms for others both online and offline. It is crucial that the UK provides a legal basis for addressing such content with both sensible limitations on freedom of speech and expression and safeguards for it. Either way, it should not be left to private companies to determine for Britain what harmful material is, that should be for parliament to decide.

3.4 Free speech is already not unlimited. There are limits in the UN's International Covenant on Civil and Political Rights, in the European Convention on Human

---

4 Policy briefing; The effects of Holocaust denial, *Antisemitism Policy Trust*, https://antisemitism.org.uk/wp-content/uploads/2020/10/Holocaust-Denial-October-2020.pdf

5 'Policy Briefing: Online and Offline Harms: The Connection.' *Antisemitism Policy Trust*, August 2020. https://antisemitism.org.uk/wp-content/uploads/2020/08/Online-Harms-Offline-Harms-August-2020-V4.pdf

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. The BBFC, our film certifier, uses "discrimination" as a category that it considers when classifying potentially harmful content. This can result in a higher age classification where the viewers are judged too young to be able to critically understand the racist or discriminatory commentary. The BBFC also refuses to classify content which is likely to cause "harm risks to potential viewers and, through their behaviour, to society". For instance, the BBFC refused to classify the online film 'Hate Crime' in 2015 because it consisted of nothing but an extended scene in which a Jewish family is subjected to racist abuse, sexual and other violence in their own home. The BBFC concluded there was a risk that some viewers may be entertained by the abuse, and associate with the attackers. This is a sensible limit on the freedom of speech and expression.

3.5 As regards enforcement, we are concerned that, given the global nature of the internet, enforcement is difficult. The Home Affairs Committee has, on numerous occasions[6], highlighted to YouTube and others the existence of materials from proscribed groups on their platforms. We are aware of examples, in particular, of smaller and alternative platforms, which can be a safe place for extremists and radicals (and from which content can carry across to larger platforms and cause wider harms) on which illegal materials have been available. The Community Security Trust has produced reports about Bitchute and other platforms which fit this description.[7]

3.6 The case of far-right activist Joshua Bonehill-Paine is an example of enforcement difficulties. Bonehill-Paine operated in a number of different areas, predominantly online but was known to a number of police forces and had cases assessed under the jurisdiction of multiple regional Crown Prosecution Service areas. Though eventually apprehended, and sentenced for the racially aggravated harassment of former Labour MP Luciana Berger, the online elements of the case led to enforcement problems.[8]

5. What model of legal liability for content is most appropriate for online platforms?

5.1 Social media companies commission, edit and curate content for broadcast or publishing and as such, are benefiting from an absence of liability. The companies at present pick and choose their status depending on how it suits them to be defined at a given time. The legal case between app Six4Three and Facebook saw the latter company argue in court that it was protected as a publisher under the first amendment for making editorial decisions not to publish content, whilst claiming protection under the Communications Decency Act because it was not a publisher, something it repeatedly claims in public.[9] Social media platforms are not simply hosts, nor neutral or mere conduits as they do apply community guidelines, albeit inconsistently.

5.2 The Government previously discussed and undertook a review of liability[10][11], ultimately considering the matter too complex to resolve. In response to the

---

6 Home Affairs Select Committee, 13 March 2018. https://committees.parliament.uk/committee/83/home-affairs-committee/news/100755/google-questioned-on-failure-to-remove-national-action-content/

7 Community Security Trust, Hate Fuel report, June 2020. https://cst.org.uk/news/blog/2020/06/11/hate-fuel-the-hidden-online-world-fuelling-far-right-terror

8 Ibid. pp.10-11.

9 Levin, S. 'Is Facebook a Publisher? In Public it says No, but in Court it Says Yes.' *The Guardian*, 3 July 2018, https://www.theguardian.com/technology/2018/jul/02/facebook-mark-zuckerberg-platform-publisher-lawsuit

Committee on Standards in Public Life it said "The Government has been clear that social media platforms are no longer just passive hosts, and we need a new approach. We need to think carefully about what level of legal liability social media companies should have for content on their sites, and we need to fully understand the consequences of any changes."[12]

5.3 The current definitions of publishers and the language used in legal framing do not match the development of the technology we now use. Platforms like Facebook and Twitter have so much content streaming through their service that to introduce liability for individual posts might not be possible. Platforms systems and moderation cannot keep pace with the rate of upload, and despite their claims, they often only act after an event.

5.4 The platforms have, however, been proven to have aspects of publisher liability in certain circumstances, where they have had prior knowledge and failed to act expeditiously in respect of illegal materials, and so full immunity is not a given. However, designating them as publishers, at least on a national basis, no longer seems credible and might in fact disadvantage the UK if this wasn't an internationally agreed approach. This is aside from jurisdictional concerns, whereby cases might need to be filed, or action taken, in Ireland (the EU/UK base) or America, which operate under a different ruleset. To this end, maintaining the status quo whereby courts have deemed liability under certain circumstances may be a preferable option to redefinition in narrow regional parameters. This would also go some way to preventing social media preference for their companies to be entirely envisaged as 'good faith operators/good Samaritans' in acknowledging that sometimes they do act in bad faith.

5.5 Codifying the existing case law emanating from Europe will be difficult. Platforms will already contest whether or not they 'have knowledge' of wrongdoing and will argue that it could be against natural justice to hold them accountable for something they had knowledge of and later addressed. However, as above, there is a gap at present whereby platforms shaping information are not widely responsible for harms occurring on their service. Social media companies must be accountable to a regulator for failing to apply minimum standards, acting irresponsibly, or breaching a future statutory Duty of Care in relation to takedown according to terms of use. The regulator would; potentially licence the firms or otherwise have them notify to a regulator they had users in a particular jurisdiction; have the ability to issue strong fines; and individual senior management liability would be introduced. The latter point is perhaps the most important if wider liability is unachievable and the Trust does not believe this matter should be reserved as is the current proposal.

5.6 We welcome the new proposed British duty of care, European and indeed potential American efforts which might bring additional liability for providers for failing to act within defined parameters. As Baroness Kidron has said, by introducing liability and having a better regulatory approach, we can revolutionise the way our online world works and better safeguard of future and prepare our children for the digital experiment we are all a part of.

---

10 'PM Speech at Davos.' *gov.uk*, 25 January 2018, https://www.gov.uk/government/speeches/pms-speech-at-davos-2018-25-january

11 'PM Speech on Standards in Public Life.' *gov.uk*, 6 February 2015, https://www.gov.uk/government/speeches/pm-speech-on-standards-in-public-life-6-february-2018

12 'Government response to the Committee on Standards in Public Life Review of Intimidation in Public Life.' *Cabinet Office*, 8 March, 2018. https://www.gov.uk/government/publications/government-response-to-the-committee-on-standards-in-public-life-review-of-intimidation-in-public-life

6. To what extent should users be allowed anonymity online?

6.1 The Trust recommends that online anonymity be restricted in order to protect the safety and freedom of expression of users from anonymous bullies, racists and extremists. The Trust has conducted research into online anonymity and found that a growing body of evidence established a positive correlation between online anonymity and the expression of extremist, racially biased and prejudiced hate-speech.[13] Anonymity can lead to group polarisation and increase aggressive user behaviour by producing an environment unconstrained by social norms.

6.2 The Trust is, as we have set out, concerned by the high level of online antisemitic abuse and the spread of antisemitic conspiracy theories; much of it spread under a veil of anonymity. This has been found to radicalise people and affect violence, hate crimes and terrorism offline. The *Trust's* aforementioned briefing on the connection of online and offline harms provides numerous examples of this type of content.[14] Apart from its radicalising potential, hate speech and bullying can cause emotional strain, depression, anxiety and fear to victims of online abuse.

6.3 The internet currently offers anonymous abusers and spreaders of radical ideologies some degree of protection by allowing them to hide their identities. Limiting anonymity, or incentivising against harm from anonymous accounts, can help victims regain a sense of control and confidence. It is also likely to reduce the overall volume of online abuse.

6.4 The Trust recommends that platforms should be able to determine the degree of anonymity they give users, within the legal framework that the government will set. The platform should also determine how it chooses to incentivise users, primarily those who choose to have a permitted level of anonymity, against producing hateful content. This allows for whistle-blowers, victims of domestic abuse, and others to remain anonymous online on the platforms in scope of the proposed regulator. Forcing users to register with their own names is unpopular.[15] However, action to guard against hate emanating from anonymous accounts would, in the view of the Trust, fall within the reasonably foreseeable harms captured by a statutory Duty of Care on platforms. This approach is also sits squarely within existing expectations for financial institutions which must 'know your customer'.

6.5 In addition, online companies should stipulate in their terms and conditions that users engaging in hate speech and other abusive behaviour might be banned from using the platform and their identity may be revealed to law enforcement. This would act as a deterrent for offenders and better guarantee the right of users to a safe environment, free from hate speech, bullying and trolling. Users will be less inclined to use hate speech and other abusive language, images or videos, if their identity is known to a host and if they are in danger of wavering their right to anonymity when their behaviour violates the host's terms and conditions, or the law.

6.6 Maintaining some degree of anonymity should still be possible, as long as they keep to the platform's terms and conditions. This will give all users confidence that another user is a real and known individual.

---

13 'Regulating Online Harms: Tackling Anonymity.' *Antisemitism Policy Trust*, 2020, https://antisemitism.org.uk/wp-content/uploads/2020/12/Online-Anonymity-Briefing-2020-V10.pdf

14 'Policy Briefing: Online and Offline Harms: The Connection.' *Antisemitism Policy Trust*, August 2020. https://antisemitism.org.uk/wp-content/uploads/2020/08/Online-Harms-Offline-Harms-August-2020-V4.pdf

15 Google Plus Ends Real Name Policy After Three Years.' *NBCNews*, 16 July 2014, https://www.nbcnews.com/tech/social-media/google-plus-ends-real-name-policy-after-three-years-n156841

6.7 If a crime or a libel has been committed in the UK on the 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 transposed e-Commerce Directive), and fines or other corrective measures could be put in place. We would suggest giving the platforms a year to become compliant.

8. How do the design and norms of platforms influence the freedom of expression? How can platforms create environments that reduce the propensity for online harms?

8.1 The Trust supports the 'safety by design' approach. Applying safeguarding measures such as privacy and easy reporting procedure for users can help protect users from harm. This must include transparency about the safeguarding mechanisms used in order to increase consumer confidence as well as company accountability and scrutiny. Companies should regularly communicate with key stakeholders, including NGOs working to address hate.

8.2 Websites or platforms should have a concise and clear codes of conduct, approved by users when they sign up. The codes of conduct can help shape positive community starts and behaviours by making clear that bullying, harassment and any form of illegal abuse will lead to immediate sanctions. These sanctions should be listed and the process must be transparent. For the rules to be effective, the company must enforce its code vigorously and consistently and make it easy for users to complain and block, as well as to appeal a decision. A visible presence of a human moderator, where possible, can also contribute to positive interactions between users.

8.3 The use of alorithms is particularly important. Work carried out by the Trust together with the CST showed that when Google amended it's search prompt from 'Are Jews….Evil' there was a 10% reduction in searches for that phrase.[16] Engaging NGOs to benefit from this and other similar learning, and which are already engaged in partner work with social media platforms can yield multiple benefits for technology companies.

8.4 A number of platforms have embarked on campaigns to educate their users about online harms. The Antisemitism Policy Trust has worked on excellent campaigns with Facebook, Twitter and TikTok, for example, these campaigns can have a very positive effect on the user audience for a platform.

10. How can content moderation systems be improved? Are users of online platforms sufficiently able to appeal moderation decisions with which they disagree? What role should regulators play?

10.1 In order to improve content moderation, companies must allocate sufficient numbers of moderators, proportionate to the company size. Moderators should be appropriately supported and safeguarded due to their exposure to harmful and illegal content. They should also receive appropriate training to be able to expertly judge content. Trusted organisations like the Antisemitism Policy Trust or

[16] Report: Stephens-Davidowitz, S, Antisemitism Policy Trust and Community Security Trust, 2019 https://antisemitism.org.uk/wp-content/uploads/2020/06/APT-Google-Report-2019.1547210385.pdf

the Community Security Trust should be included in the design or quality assurance of relevant aspects of such training – as should other key organisations in respect of other forms of hate.

10.2 Machine learning and Artificial Intelligence tools cannot sufficiently replace human review and oversight. The use of technological means for content moderation can be effective in quickly sifting through large amounts of data and flagging harmful content, thereby supporting the work of human moderators rather than replacing it. Companies should have processes in place to ensure that where machine learning and artificial intelligence tools are used, they operate in a non-discriminatory manner and that they are designed in such a way that their decisions are explainable and auditable.

10.3 We further recommend that companies employ a separate notification system for each type of harmful and illegal content to ensure the correct moderators, trained in their specialist subjects and on related language and cultural context considerations (where proportionately reasonable), are able to review the appropriate content. It is crucial that moderators also take into account national laws and the Terms of Service when judging content. This would also assist with reporting mechanisms.

10.4 Any content that has been deemed to be illegal must be removed within 24 hours of becoming aware of such content. However, a company must also take action on content which is not deemed to be illegal but is considered to break their Terms of Service or Community Guidelines. Actions can include removal of content; termination or suspension of an account – depending on the severity of the content and on whether the account holder is a repetitive offender; or a strike, if a striking system is in place.

10.5 To improve moderation, systems of assessment and feedback to the initial reporter and the owner of content that has been flagged and actioned should be in place to ensure transparency of decision making. Users should be kept up to date with the progress of their reports and receive clear explanations of decisions taken. They should also be able to appeal a decision if they think is was wrong.

10.6 Companies that outsource their content moderation, must ensure that the vendor adheres to the Terms of Service and Community Guidelines, as well as to the law. The vendor should use similar transparency and standards expected from companies.

7 January 2021