ANTISEMITISM
POLICY TRUST

Head Office
[redacted]

Consultation – Protecting People from Illegal Harms Online

The Antisemitism Policy Trust has taken an active interest in the Online Safety Bill process, providing evidence to parliament and to Ofcom. We welcome the consultation on illegal harms and hope that our perspective will be helpful in shaping future regulation.

Instead of using Ofcom's submission document, we have chosen to highlight specific issues and concerns that relate to online antisemitism that emerge from Ofcom's proposals. This submission was done in consultation with the Community Security Trust (CST).

Concerns Shared Across Civil Society:

The Trust was one of a number of organisations, under the umbrella of the Online Safety Act Network, to signal our concerns about the approach of this consultation and the impact this will have on user safety, setting the regulator off on the wrong footing and making future codes not fit for purpose.

Our concerns, which we understand are not necessarily part of this consultation but impact upon it, include:

- The approach to "illegal content" set out in its Illegal Contents Judgement Guidance: this does not take account of the Act's overarching "safety-by-design" approach based on obligations related to systems (not individual items of content); focuses primarily on identification of criminal conduct, rather than the content associated with a criminal offence; and uses criminal standards of proof, rather than civil standards which would be the norm in a regulatory regime.
- Burden of proof/weight of evidence: there are aspects of this issue that are problematic in relation to the illegal contents judgements guidance and there is a separate issue which is the apparently preferential weighting that Ofcom gives to evidence already collected from industry, eg "best practice" from companies, and the undefined threshold it sets for other evidence to meet for inclusion in the codes, which seems very high. The effect of this is to set the bar too low in terms of the measures with which regulated services must comply via the codes and to reinforce the status quo which

the legislation was intended to improve. Conversely, measures that Ofcom acknowledge might mitigate harm but which do not meet their (undefined) threshold for evidence are discounted. This does not align with Parliament's expectation of a systemic, risk-based regime, focused on outcomes rather than prescriptive rules.

● What proportionality means: Ofcom's approach to proportionality is primarily economic: to avoid imposing costs on companies. While the OSA requires regulated services take a "proportionate" approach to fulfilling their duties, and indeed requires Ofcom to look at resources, Ofcom is also required – among other issues – to look at the severity of harm.

● The prioritisation of users' freedom of expression above adverse impacts on fundamental rights of others: amongst other things, this has significant implications for protection of women and those from minoritised groups, for whom targeted online abuse is a means of silencing them. This is especially concerning in light of increased media attention regarding – and the government's recognition of – digital threats to democracy; increases in misogyny and other forms of online abuse limit democratic participation among those most adversely impacted by online abuse. In the past few years, high-profile racist and misogynistic online attacks on footballers and TV commentators have led to many of those targeted being hounded off platforms.

There are a number of other specific decisions on the scope of the codes of practice and the measures recommended within them which also, in our view, significantly limit the likely impact of the measures proposed in this initial consultation. Some of these include:

● Weak "safety by design" foundations: a disconnect between the evidence of harm in the risk profiles and the mitigation measures in the codes of practice.

● Lack of focus on outcomes: the regime is not outcome-orientated (eg to deliver improved safety) but focused on a prescriptive, tick-box, process-driven approach via the codes.

● Compliance expectations: the proposals take at face value evidence from the platforms that they are "doing much of this already" and Ofcom continuously emphasises the proposed measures will not incur any additional costs.

● Small vs large companies: there is a significant differentiation in Ofcom's approach to the risk assessment duties and the codes between large companies (7m+ monthly users) and small companies (everything else).

● Limited improvement in the online safety of children, women, Black women especially and other minoritised groups: while there are specific measures relating to child sexual abuse material (CSAM), overall, the impact of all the decisions taken by Ofcom above will do little to shift the dial

in terms of improving safety for children, women, especially Black women and other minoritised groups.

Finally, there are two aspects of the consultation approach that are a concern:

- Speed vs comprehensiveness: the codes are a "first iteration" and will be revised. However, a "lowest common denominator" regime is very likely still to get watered down further.
- Civil society response: the size and complexity of the consultation has caused accessibility challenges for under-resourced third-sector organisations, and there is no mechanism for victims to respond.

The details offered in the rest of this submission should take into account the above context and concerns, and are additional to these.

1. Safety by Design, and Small High-Harm Platforms

Safety by Design

Safety by design is an essential factor in achieving the outcomes both parliament and Ofcom are seeking from the Online Safety Bill's implementation.

Services must plan and implement safety features in order to protect UK service users from harm. This is particularly important when considering the harm caused to individuals and groups of people with protected characteristics, who are at greater risk.[1][2] This should include, as stated in Schedule 4 (the online Safety Objectives)[3], the algorithms used by the service, the functionalities and any other features used by the service that may cause harm.

Although Ofcom has stated that it intends to 'tackle the root causes of online content that is illegal and harmful for children, by improving the systems and processes that services use to address them,'[4] the approach set out in the proposals is not likely to achieve this. One of the reasons is that

---

[1] https://www.ohchr.org/en/stories/2021/03/report-online-hate-increasing-against-minorities-says-expert

[2] https://www.surrey.ac.uk/news/research-finds-bame-and-lgbtqi-youths-most-risk-harm-online

[3] https://www.legislation.gov.uk/ukpga/2023/50/schedule/4/enacted

[4] https://www.ofcom.org.uk/__data/assets/pdf_file/0017/270215/10-23-approach-os-implementation.pdf p.5

Ofcom expects 'safety by design' features and product testing to be implemented by large or multi-risk platforms, but not by smaller platforms, even though some of these pose considerable risk and have proven to be harmful in radicalising individuals by allowing, and even encouraging the spread of illegal content.[5]

We believe that safety by design features are key to reducing harmful illegal content online, and that such features should be standard across platforms of all sizes. Though we do not wish to be prescriptive nor do we believe in a one size fits all approach, this includes, for example, the introduction of system friction (offering prompts before users proceed, for example) and proactive efforts to educate users – as detailed in the Antisemitism Policy Trust submission to Ofcom's consultation on categorisation in 2023. These measures should be in place *before* a new product is launched and after completing a risk assessment process, and product testing.

Companies operating online platforms and search engines are more likely to use these measures effectively if there is a regulatory demand for compliance. We accept that smaller services will have fewer resources to carry out this work. However, we also contend that risk to users outweighs the pursuit of profits. In addition, launching safer products could also help platforms reduce costs in the long run, including costs relating to compliance and moderation.

Small vs. Large platforms

Although the Bill states that the illegal content duties apply to all regulated services, Ofcom's proposed measures differentiate platforms according to their size and according to risk assessments, but these are carried out by the companies themselves and therefore may not reflect the levels of risk accurately, particularly where small high-risk platforms have been created in order to spread harm. This seems to contradict the approach set in the law, that category 1 services are to be assessed according to their size OR their risk/functionality.

Adding to this, the number of monthly users set by Ofcom for a company to be included in category 1 is extremely high. This will exclude large platforms with millions of users from category 1. One example is the gaming platform Robolox, which has 3.4 million average users per month, including children but will not be included in category 1. Since the war between Israel and Hamas broke on 7

[5] https://gnet-research.org/2023/06/23/how-do-terrorists-utilise-and-exploit-small-covert-online-spaces/

October 2023, there has been a proliferation of antisemitic content on Robolox targeting Jewish users.[6] The platform also faces a class action lawsuit in California for sexual content.[7]

We therefore urge Ofcom to reduce the threshold so that more services that are used by millions are captured under category 1. We also strongly recommend that the risk factors presented by platforms received more weight regardless of size. This will mean that more platforms will have duties to protect users, including implementing safety by design features.

Evidence and the importance of safety by design features in large and high risk platforms

There is a close reciprocal link between large and small platforms, with users moving from large platforms to smaller ones, where they become increasingly radicalised into extremist ideologies. They then 'migrate' back onto large platforms where they can radicalise others more widely and also target users with abuse. As Ofcom rightly points out in its proposals: 'We have found that a lot of terrorism content is first posted on smaller U2U services and then linked to from larger, higher-reach services.'[8]

A report by the Community Security Trust (CST) on the radicalisation of young people towards far-right ideologies found that 'Instagram is a useful tool for young racial nationalists, providing them with a powerful opportunity to recruit.'9 The report also found that far-right extremists use mainstream platforms to obtain wide audience reach, to spread their ideology in a way that circumvents those platforms' moderation, and then 'funnel' young people into more extreme smaller platforms that have been known to turn a blind eye to illegal terrorist and violent content with little or no moderation, such as Telegram.

---

[6] https://www.jewishnews.co.uk/roblox-doing-all-it-can-to-tackle-proliferation-of-antisemitic-content-on-platform/

[7] https://www.pcgamer.com/roblox-faces-class-action-suit-from-parents-about-sexual-content-and-grooming-it-is-illegal-to-expose-minors-to-these-kinds-of-things-and-its-not-slowing-down/

[8] https://www.ofcom.org.uk/__data/assets/pdf_file/0019/271243/volume-2-illegal-harms-consultation-1.pdf p.22

[9] https://cst.org.uk/data/file/2/4/We%20are%20Generation%20Terror.1639414296.pdf p.3

The report found that far-right extremists have migrated from the large, mainstream social media platforms towards smaller platforms, such as Gab, Bitchute and Telegram, where antisemitic content flourishes.10 This includes illegal material that calls for violence and terrorism against Jews.

On Telegram for example, CST found posts glorifying far-right terrorists including Thomas Mair and David Copeland, and posts calling for the killing of Jews. The platform 4chan was also found to host threads containing explicit calls to kill Jews. Similar posts, containing violent, antisemitic comments, were found on Bitchute, including images with phrases such as, 'all Jews must die,' and images of people aiming weapons accompanied by threatening language. The information is easily accessible, and CST concluded that: 'the quantity and spread of this incitement poses an urgent and ongoing terror threat to Jewish communities.'

Small platforms that host illegal content have been linked to terror attacks. A briefing by the Antisemitism Policy Trust, published in August 2020, provided examples of the connection between online and offline harms, citing examples of attacks against Jewish targets (for example, the Pittsburgh Synagogue attack) and against Muslim targets (for example, the Finsbury Park and Christchurch mosque attacks). In all of these, attackers participated in extreme online forums where they were either radicalised to the point of attacking Jews and Muslims, or inspired others to commit acts of hateful violence. The terrorist who killed eleven congregants and injured six others in a synagogue in Pittsburgh in 2018, promoted his hateful, antisemitic agenda on the social media platform Gab – where he also posted minutes before attacking the synagogue. Based on testimonies after the attack, he had consumed large amounts of racist and other material online which had incited him to violence. Another briefing by the Antisemitism Policy Trust, on anti-Jewish misogyny, referenced a study by the American NGO Media Matters, which found a staggering 180% increase in posts containing both antisemitism and misogyny on the far-right anonymous message board 4chan between 2015 and 2017.

As mentioned in Ofcom's proposals, certain functionalities pose high risk.[11] One of these is live streaming of terror attacks. More recently, we have witnessed in relation to the recent pro-Palestinian demonstrations that were live-streamed on services, including on TikTok for example, chants and

---

[10] https://cst.org.uk/news/blog/2020/06/11/hate-fuel-the-hidden-online-world-fuelling-far-right-terror

[11] https://www.ofcom.org.uk/__data/assets/pdf_file/0019/271243/volume-2-illegal-harms-consultation-1.pdf p.23

placards that express antisemitic hate speech, and priority offences such as glorifying terrorism and support for proscribed terror organisations. Having sufficient systems capability to review all livestreams is certainly not the case for all platforms at present. Analysing content, for example, posters or speeches, in realtime within livestreams seems at best an idealist prospect at the present time. During Hamas's attack on Israel on 7 October 2023, terrorists live-streamed their torture and murder to people on social media platforms using mobile phones and GoPro cameras. The extremely violent content – illegal in itself – was also later used to glorify terrorism against Jews.

Currently, it is easier for platforms to detect and remove content that is uploaded to the service rather than live-streamed, and it is crucial that large and high-risk small services are required to prioritise assessing the risks caused by their functions and employ effective mitigation measures to potentially high risk functions. It is also important that such requirement are not limited by other considerations such as cost to services.

Considering the evidence, we believe that it is imperative for Ofcom to place the responsibility for mitigating the risks of functionalities and design choices, which have been shown to cause harm, on the regulated services, and that this should also include small high risk platforms. Taking measures to address harms found by risk assessment and that emerge from functionality and design should be a requirement. Testing for product safety should also be a requirement, and the effectiveness of complying with both requirements should be monitored by Ofcom. Safety by design should be a basic requirement on all regulated companies and its foundations have to be robust and effective.

2. Approach to illegal content judgements guidance

The guidance focuses on items of content that require assessment as to whether they should be taken down, instead of focusing on a safety by design approach that we believe should be at the heart of the new regulation. The approach also sets the standard of proof that a criminal offence has taken place by uploading illegal content at a very high threshold in order to avoid over-removing content. Whilst we understand and appreciate efforts to protect freedom of expression, this will slow or prevent the effective removal of illegal content before it is seen by a large number of users.

The Trust strongly believes in a systems-based approach, and designing systems in a way that reduces harms and the prevalence of illegal content from being published, rather than placing an emphasis on take-down, which risks criticism from free-speech advocates and also depends on effective moderation – something that has been inconsistent across platforms and has usually not been

effective enough. We believe that designing systems with safety in mind will produce platforms that promote more positive engagement between users. This can help not only avoid the upload of illegal content, but of content that does not cross the legal threshold but is nonetheless harmful to users.

Much of the antisemitic content online is legal, but it promotes racism and prejudice against Jews, and helps radicalise people into more extreme anti-Jewish attitudes. This contributes to rising levels of illegal content including threats against Jewish users, glorification of terrorism and violence against Jews – as has been detected in recent months,[12] and other abusive behaviour and illegal hate speech. This also created an unsafe environment that has an adverse effect on users' mental and emotional health and violates their freedom of speech.[13]

We are disappointed that Ofcom requires that a criminal offence has taken place each time content is posted (rather than anchoring content with an offence which has already been acknowledged as such), There is a limited view of relevant content twinned with proof being set at the criminal level– at odds with what is a civil regulatory regime. It is also unfortunate that Ofcom has not considered any of the existing non-priority offences. Further to this we believe some reference should have been considered to the Equality Act, which would help Ofcom to take the Act into consideration. It is welcome that protected characteristics get some mention but it feels like this is an attempt to have ones cake and eat it.

3. Proportionality

Ofcom has placed an emphasis on avoiding imposing costs on companies. However, costs need to be balanced with the need to create services that are safer for users and the harm that those services cause to individuals and to our society. Since companies tend to place profits at the top of their priorities,[14] and creating a safe product is more costly than creating an unsafe one (which in itself can

---

[12] https://www.adl.org/resources/blog/platforms-struggling-curb-online-hate-amidst-war-israel-and-gaza and https://www.isdglobal.org/digital_dispatches/rise-in-antisemitism-on-both-mainstream-and-fringe-social-media-platforms-following-hamas-terrorist-attack/

[13] https://committees.parliament.uk/writtenevidence/9377/html/

[14] https://abcnews.go.com/Politics/social-media-companies-prioritizing-profit-harmful-content-senate/story?id=93360057

boost profits, because it encourages a kind of content that promotes engagement), given the option, services may do the minimum required from them legally.[15][16]

The focus on cost and the suggestion that small services have fewer resources and are therefore not expected to have the same duties as large services with relation to mitigating and managing risks, allows small high-risk services to continue cause extensive harms.

However, the Act does not direct Ofcom to place its focus on costs when measuring proportionality, and we believe that doing so does not fit the legislative intent, which is predicated on risk. Whilst we appreciate that costs are a factor ,, helping services minimise cost at the expense of making safer products should not be the guiding principle. Considering the large amount of illegal content online, including content that promotes terrorism and other forms of violent extremism that cause considerable harm to our communities, our democracy, our values and our safety, we believe it is Ofcom's role to act on the best interests of the British public rather than on the best interests of companies based predominantly outside of the UK. Prioritising cost considerations can provide services, big and small, with excuses to avoid complying with regulation. It creates a major loophole and we therefore urge Ofcom to consider risk as the major factor when judging proportionality.

As set out in section 1 of our submission, small high risk platforms host an abundance of priority and non-priority illegal content, creating a toxic and unsafe environment that has a wider effect on our society. Ofcom mentions in the consultation that terrorists move from large services to smaller ones that have fewer resources for content moderation.[17] The lack of moderation provides terrorists with more freedom to post content that radicalises users and promotes terrorism. These findings seem to be at odds with Ofcom's other claim, that:

> *On most services, and especially small services, we consider it unlikely that there are accounts operated by proscribed organisations, nor do we expect there to be any illegal terrorist content. In these cases, we do not envisage such services needing to incur any costs in advance of receiving a complaint or otherwise becoming aware that a proscribed group*

---

[15] https://www.wsj.com/articles/the-facebook-files-11631713039

[16] https://www.ft.com/content/62805ce1-ac7d-4ef9-bf4b-99876960af08

[17] https://www.ofcom.org.uk/__data/assets/pdf_file/0019/271243/volume-2-illegal-harms-consultation-1.pdf p.37

may have an account on their service. Such services would not need to train staff in advance or develop processes, and would only incur the costs of assessing an account in the unlikely event they were made aware of one.

Certain smaller platforms have been hosting extensive volumes of terrorist content. For example ISIS had exploited the small information-sharing platform justpaste.it[18] As terror organisations like ISIS and Hamas[19] are banned from large services (although content that supports and glorifies them remains on these platforms[20]), they will likely find ways to disseminate their messages on smaller platforms if those escape obligations due to proportionality measures that prioritise cost over risk. Cost seems to be over-emphasised to a degree where, according to Ofcom, small platforms would not even need to train staff to recognise terrorism content, which should be a basic requirement for training to moderate priority illegal content on all services.

The consideration of cost should be measured against the evidence regarding the prevalence of content in a service and the effect it has, with the evidentiary threshold reduced. As we set out earlier in our submission, creating products with safety features can reduce risks and help companies reduce costs later on, which is why a systems based approach can be cost effective.

As new harms emerge, or there is an influx of illegal harms, as we have seen in the recent significant spike in antisemitic hate speech, threats and disinformation in response to the war between Hamas in Gaza and Israel, it is important to consider the time it will take Ofcom to decide on proportional responses. The speed at which new harms emerge and spread online once again highlights the importance of safety by design measures and underlines the need for proportionality to consider risk over costs.

4. Human rights

The Online Safety Act directs Ofcom to consider a variety of rights protected by the European Convention of Human Rights (EHRC).

---

[18] https://www.theguardian.com/world/2014/aug/15/-sp-polish-man-website-isis-propaganda-tool

[19] https://time.com/6330005/the-oct-7-massacre-revealed-a-new-hamas-social-media-strategy/

[20] https://www.isdglobal.org/digital_dispatches/rise-in-antisemitism-on-both-mainstream-and-fringe-social-media-platforms-following-hamas-terrorist-attack/

Of all the relevant rights, freedom of expression is one that has the most relevance to online anti-semitism. We are concerned that Ofcom's proposals are indicative of an approach that places the rights of users and their freedom of expression over the rights of victims and their own freedom of expression. Ofcom's approach to blocking users also considers limitations to these users' freedom of expression and freedom of association, but not the freedom of expression of their intended victims which could be violated by users' illegal speech. Article 17 of the European Convention on Human Rights is as important as Article 10, speech has floors and ceilings. So too, when taken into account properly, freedom of speech shouldn't mean freedom of expression for the loudest/largest/most egregiously offensive within the law, but rather freedom of expression for everyone, including those that are bullied out of online spaces by the loud, aggressive and mendacious users.

The Antisemitism Policy Trust believes that freedom of expression is crucial for an open, stable democracy that values liberal ideals. However, we also maintain that freedom of expression should not be absolute and consideration must be given to competing freedoms (as the UN and EU Human Rights conventions/declarations make clear), such as the freedom of religion, association and freedom from harm. We support the current legal restrictions on freedom of expression in the UK. We see freedom of expression as all-encompassing, and therefore it must include those whose freedom to express themselves may be otherwise restricted or curtailed for fear of abuse by hateful online trolls or spaces. Freedom of speech is not simply about preventing censorship but rather about ensuring everyone has their right to speech.

Jews who are targeted with antisemitic abuse on and off-line, that is currently protected by the right to freedom of expression, such as Holocaust denial and antisemitic conspiracy theories, experience harmful effects. Antisemitism often does not only harm an individual victim, but affects the wider Jewish community in the UK. Abuse has a silencing effect on the individual who has been abused, and often on the group of the protected characteristic or minority community this individual belongs to. If Jews are trolled and abused on a platform (as they disproportionally tend to be), other Jewish users may feel compelled to limit their activity on the platform for fear of suffering similar abuse. This is a major violation of their freedom of expression.

In a response to an inquiry by the Human Rights (Joint Committee) on freedom of expression, the Trust maintained that: 'hate speech is not covered adequately by existing law. Current legislation on

hate speech and hate crimes is overly complex and fragmented which can cause problems with e forcement.'21

Hate speech should be broadened to include some currently-protected speech and the legal gaps that have been exploited by extremists need to be addressed. Action against it should also be better enforced to ensure victims feel safer and in order to protect their rights. This will make more of the antisemitic expressions that are currently allowed online and contribute to abuse of Jews, radicalising people against Jews and to violating the freedom of speech of users who are Jewish, illegal, and place them within the proposed regulation.

Although we do not support widespread blocking of accounts, especially for minor or one-off infringements, we believe that blocking repeat offenders and those who disseminate harmful illegal materials has value to the online community of a specific service by making it a safer space. Blocking a user should not be arbitrary but in direct response to a users' behaviour, in spite of a platform's terms and conditions and knowing that their behaviour could result in de-platforming.

5. Risk analysis and mitigation measurements

The Trust believes that risk assessment and mitigation requirements are key to improving online safety, and it is imperative these are carried out effectively. Ofcom should require companies to measure the effectiveness of their mitigation practices and report these with complete transparency about how they measured effectiveness. Ofcom should also be able to require services to improve their practices if these are found to be ineffective.

There is an abundance of evidence to demonstrate the prevalence of illegal content online, and evidence as to the harm caused by illegal or legal but harmful to individuals, communities and

[21] https://committees.parliament.uk/writtenevidence/21580/pdf/

minority groups such as Jews.[22][23][24][25][26] There is however a lack of evidence as to how effective mitigation measures are, which is key to understanding what helps reduce risks and what does not, and to adopting good practices. There could perhaps be a proactive effort by Ofcom to share good practices and measures that have been proven to work with services to help them improve safety (this could also help them cut costs, although there is no guarantee that measures used successfully in one service will be suitable for another, so assessments will still need to be carried out).

As a result of Ofcom's approach to proportionality, it is likely that small platforms will be exempt from employing mitigation measures, unless Ofcom adopts a risk-based approach. The dangers of high risk small platforms have been discussed in this submission.

6. Risk assessment

An effective risk assessment process is fundamental to improving safety. This requires governance, scrutiny and transparency. Ofcom's proposals indicate that Ofcom is content with the collaboration from larger platforms. However, there is not much evidence to prove that their governance and oversight structures are effective.

Although companies collaborate with Ofcom, their performance in mitigating risks is still not effective enough. A study by the Institute for Strategic Dialogue (ISD) found a 'major and sustained rise in antisemitic posts' on 'X' following the takeover by Elon Musk.[27] During the recent war between Israel and Hamas, YouTube[28] and Facebook[29] for example, have been found to host large amounts of antisemitic content, some of it illegal including glorification of terrorism, support for a

---

[22] https://antisemitism.org.uk/wp-content/uploads/2020/08/Online-Harms-Offline-Harms-August-2020-V4.pdf

[23] https://committees.parliament.uk/writtenevidence/9377/html/

[24] https://edmo.eu/publications/how-misinformation-and-far-right-groups-sparked-a-riot-in-dublin-after-the-stabbing-of-three-children-at-a-school/

[25] https://actearly.uk/radicalisation/online-safety/

[26] https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence

[27] https://www.isdglobal.org/isd-publications/antisemitism-on-twitter-before-and-after-elon-musks-acquisition/

[28] https://www.isdglobal.org/isd-in-the-news/online-antisemitism-soars-with-a-51-fold-surge-in-antisemitic-youtube-comments-following-hamas-attack/

[29] https://decoding-antisemitism.eu/publications/sixth-discourse-report/

proscribed terror organisation, hate speech and threats. This is an indication that there are fundamental flaws in the platforms' risk assessment or the implementation of measurements to reduce risk.

Ofcom's proposals also make the assumption that larger platforms mean bigger risk because of the wider reach that they have. As referenced in section 1 of our response, some smaller platforms are extremely high risk. They can radicalise people who later go on to larger platforms and disseminate illegal content or result in real-world harms such as terror attacks. We therefore believe that it is a wrong assumption and that the seriousness of the harm should be considered just as importnat as the quantity of the people exposed to harmful content.

We recommend further that Ofcom not use the rigid 'tick-box' nature of measuring harm as set out in volume 4, but allow a more flexible approach to allow platforms to adjust to their structure, user-base and experience that draw on best practices.

7. Identity Verification

As stated in the illegal harms consultation, anonymity can encourage engagement in harmful behaviours.[30] We agree with Ofcom that anonymity is important for whistleblowers and allowing those with protected characteristics to express themselves online more freely. We therefore recommend that services should be required to offer users an option to verify their identity if they wish.

In addition to identity verification, we agree with Ofcom that services should include functionalities that include options to block and mute accounts. However, this option will often be used *after* a user has already been exposed to harms, including abusive and threatening behaviour and harassment. This option should therefore be in addition to identity verification, not instead of it.

8. Access to Information and Transparency

Additionally, we urge a policy that encourage greater transparency from services by allowing researchers greater access to content. Content that is published in open forums or groups can help find evidence of illegality, ensuring that users' privacy is retained

[30] https://www.ofcom.org.uk/__data/assets/pdf_file/0019/271243/volume-2-illegal-harms-consultation-1.pdf p.3

Currently there is an effort by some companies to undermine the work of researchers. Since Twitter was taken over by Elon Musk and renamed X, it has become far more difficult for researchers to find information on the platform. Those that managed to gather information, have found that following Hamas's attack on Israel on 7 October, X hosted the most antisemitic content of all the 'big-5' platforms (X, TikTok, Instagram, Facebook, and YouTube).[31] In trying to mitigate this PR nightmare, especially after warnings from advertisers that they would boycott X, Musk brought new measures to fight some of the antisemitic content on the platform. He also claimed that an audit found that X has the least amount of antisemitism. This is hard to verify since access to primary information is severely restricted. The recent case brought by Must against the Centre for Countering Digital Hate (CCDH) is an example of the risk posed to research that can benefit society by highlighting risks present on online services, from large companies seeking to silence such studies.

[31] https://cyberwell.org/wp-content/uploads/2024/01/Denial-of-October-7-Social-Media-Trend-Alert-CyberWell.pdf