Tech Against Terrorism
Written evidence to the Draft Online Safety Bill (Joint Committee)'s inquiry, Draft Online Safety Bill.

Submitted: 15 September 2021

1. BACKGROUND

On 12 May 2021, the UK Department for Digital, Culture, Media & Sport published the draft Online Safety Bill (OSB), aiming to counter harmful content online and announced in a White Paper in April 2019. In July 2021, a Joint Committee on the draft was established by the House of Lords and House of Commons to consider the OSB and publish a report on its findings in December 2021. In late August-September, a call for written evidence to the draft OSB was published by the Joint Committee. The Joint Committee accepted written evidence on the draft OSB from July to September 2021.

2. TECH AGAINST TERRORISM'S CORE ARGUMENTS AND RECOMMENDATIONS

Chief among the harms addressed by the Online Safety Bill is the harm caused by terrorist use of the internet. As an organisation concerned with balancing practical efficacy and conformity to international standards of human rights in the development of counterterrorism policy, we believe our insights will be of value to HMG in presenting an effective and human rights-compliant bill to Parliament.

The arguments in our response to the consultation on the Online Safety Bill ("OSB") may be summarised as follows:

1. **The OSB lacks consideration for smaller platforms.** Despite the OSB differentiating between different categories of service, the bill currently lacks precision on the threshold conditions for each category. To provide clarity for platforms affected by the requirements in the Bill, the OSB should include a taxonomy of platform size in its categorisation of services.

2. **Imposing stringent legal requirements with no regard for platform size will harm the diversity and innovation that drives the tech sector.** Stringent legal requirements will disproportionately impact smaller tech companies with fewer resources to support compliance, whereas larger tech platforms will be able to allocate the resources necessary to comply with the Bill.

3. **The draft OSB undermines the rule of law and due process.** It does so by outsourcing the responsibility of adjudicating what constitutes illegal content (including terrorist content) to tech companies without appropriate judicial oversight. Online counterterrorism efforts should be led by democratically accountable institutions and the OSB should reflect this.

4. **The OSB lacks proper and operable definitions of what is considered "harmful online content" or of what might constitute terrorist content**. The same goes for "democratic" and "journalistic" content, which are currently not definitions capable of operational effect, and risk being weaponised by terrorists and violent extremists. This may render the Bill incapable of achieving its purpose to make internet use in the UK the safest worldwide.

5. **HMG should ensure that appropriate support mechanisms are created for platforms of all sizes and resources to ease compliance with the Bill.** In particular, provisions in the Bill relating to risk assessments and the use of proportionate systems to counter illegal content will not be practicable for smaller platforms without support. HMG should continue supporting initiatives such as Tech Against Terrorism which provide policy and practical support to tech platforms in countering terrorist use of the internet.

3. CONSULTATION RESPONSE

Will the proposed legislation effectively deliver the policy aim of making the UK the safest place to be online?

Tech Against Terrorism welcomes the UK Government's willingness to act against the spread of illegal content online, including terrorist content. However, **we have concerns that the Bill, as it stands, will not be effective in fulfilling its stated purpose and risks undermining the rule of law, freedom of expression online and tech sector diversity.**

- **We are concerned that the OSB will create a fragmented regulatory landscape.** The Bill creates new legal obligations for tech platforms and new enforcement requirements for Ofcom, and also places the onus of adjudicating legality and harmfulness on tech platforms. In our view, additional resources should be made available to enforce existing legislation capable of tackling illegal and harmful content, and to prosecute those creating and sharing that content [1] before enacting a new law which will only further fragment the regulatory landscape.

- **We are also concerned that the Bill undermines freedom of expression online**. The broad scope of the Bill and the lack of precise or operable definitions of illegal and harmful content in the current draft means that platforms will have to decide for themselves what content is 'justiciable' and by what standard that content should be adjudicated. These standards will vary hugely across the tech sector ecosystem, depending largely on the resources available for moderation, and may come to appear arbitrary; this inconsistency may prove highly corrosive of public trust in their right to freedom of expression if their speech is unevenly curbed by private entities at the behest of Parliament. . Deciding what is legal or illegal online is the duty of democratic institutions and independent judicial authorities, not private tech companies

Is the "duty of care" approach in the draft Bill effective?

- The "duty of care" approach makes commendable provisions for platforms to assess the risks to users and to detail how they counter illegal content and protect freedom of expression online. The provisions on risk assessments and clear Terms of Service (ToS) are particularly commendable in encouraging platforms to comprehend the ways in which their services can be abused to disseminate illegal content and in being more transparent about their response. **However, we are concerned that some duty of care provisions outlined in Chapter 2 and 3, such as the provisions relating to illegal content duties (Section 9) and risk assessments (Section 7(1)) will be difficult to implement for smaller tech platforms.**

[1] For instance, laws on countering terrorism, hate crimes, and incitement to hatred or violence

- **We are also concerned that the duty to protect journalistic and democratic content outlined in the "duty of care" approach risks being abused by terrorist and violent extremist actors, as well as being challenging for platforms to implement.** Without clear definitions and practical guidance, tech companies will struggle to adjudicate on what is illegal or terrorist content as opposed to permissible in a democratic society. This in turn risks an inconsistent application of the law. Tech companies will also likely struggle to adjudicate on what is terrorist content that is shared for terrorist purposes versus terrorist content that is shared for journalistic purposes, particularly given that terrorist actors often disguise propaganda as journalist content. HMG should provide more context and guidance to tech platforms on how to determine what is journalistic or democratic content. In addition, this should be done with safeguards for human rights and freedom of expression in place.

- **We are also concerned that some of the "Safety duties about illegal content" (Chapter 3, Section 21) are effectively encouraging tech platforms to increase their use of automated tools to constantly monitor their services.**[2] Whilst automated content moderation has its benefits, current solutions are not nuanced enough to correctly assess whether certain pieces of content are in fact terrorist material or otherwise harmful. An increased reliance on automated moderation solutions raises the risk of false positives in taking down content that is legal, and raises questions about accountability in removal decisions. The use of automated solutions to detect and remove terrorist content is also not straightforward. These solutions cannot replace consensus on what constitutes a terrorist organisation, and need to be informed by HMG's official proscriptions. Moreover, automated solutions are often resource-intensive and most smaller platforms will not have the capacity to deploy them or will have to use solutions developed by larger platforms (if the cost permits it). This risks imposing uniformity of the online moderation landscape by requiring smaller platforms to turn to larger ones for moderation tools.[3]

How does the draft Bill differ to online safety legislation in other countries (e.g. Australia, Canada, Germany, Ireland, and the EU Digital Services Act) and what lessons can be learnt?

- In general, the draft Bill is similar in its key aims and provisions to other online regulations passed globally in the last four years, including in France and the EU.[4] However, the OSB also includes similar provisions to those passed in less democratic countries, such as India and Pakistan.[5] **We encourage the UK to reflect on the fact that several provisions of the OSB are similar to those being introduced in less democratic countries and criticised for undermining freedom of expression online.**

- **The duty to protect "democratic" content is also reminiscent of proposals in Brazil and Poland to prevent platforms from removing content based on their own Terms of Service.** Such provisions risk creating conflicting requirements for platforms to remove or not remove content, especially without a precise definition of what constitutes democratically permissible content, or of what constitutes harmful content.

---

[2] Namely those related to minimising the presence and dissemination of illegal content, and the length of time such content stays online in Section 21 (3).

[3] In July 2021, we released our Gap Analysis on Technical Approaches to Counter Terrorist Use of the Internet in partnership with the GIFCT. The report makes a variety of policy recommendations including the need to formulate a strategy that encourages stakeholders to work towards a common goal and to ensure that technical solutions are considered alongside policy responses. You can find the report here.

[4] Legislation increasingly requires platforms to publish transparency reports and act against illegal and terrorist content (for instance in France, the EU).

[5] Provisions similar to those in the Bill requiring platforms to publish detailed ToS and encouraging systems and processes to minimise the presence of illegal content have also been passed in India and Pakistan.

- **The OSB is also similar to the Protection of Online Falsehoods and Manipulation (POFMA) bill passed in Singapore in 2019, which applies indiscriminately to both public and private communication channels.** Legal provisions for platforms to monitor private communication channels, in particular end-to-end encrypted platforms, present major risks to online privacy and security, in addition to undermining the right to privacy online, and have negligible advantage in countering terrorist activities online.

  Tech Against Terrorism recently released an in-depth report assessing the risk of terrorist use of end-to-end encrypted services, outlining possible mitigation strategies that safeguard encryption and the right to privacy. You can find the report here.

- **The OSB differs from legislation passed in other countries[6] in being one of the rare laws aimed at countering illegal and harmful content without mandating a specific, and usually short, removal deadline for tech platforms.** This is commendable as the pressure put on tech companies to quickly remove content threatens freedom of expression. If tech platforms do not have the time necessary to adjudicate on the legality of content they are likely to err on the side of over-removal to avoid penalties. Removal deadlines also ignore the fact that most small and medium-size platforms do not have the capacity to comply.

- **The OSB also crucially differs from other online regulation by placing part of the legal liability on platforms' employees.** Such provisions, especially when platforms' employees can be sanctioned with jail term, as per 73(8(c)), would lead the UK down a similar path to that followed in Brazil were Facebook employees had been jailed and India were twitter offices had been raided by the police.[7]

- **The draft Bill also differs from other regulations by encouraging tech platforms to undertake risk assessments. Tech Against Terrorism welcomes this provision.** Understanding the risk is a crucial first step to effectively countering terrorist exploitation of the internet, and tech companies should to the best of their abilities consider how their platforms could be exploited and remain aware of adverse usage. However smaller platforms will need support in carrying out risk assessments, as they may not have the resources or capacity to conduct these. We encourage the UK government to support smaller tech platforms with this crucial task.

[6] Including in the EU, Australia, Germany, India, Pakistan, Indonesia, and Turkey

[7] In Brazil, Facebook employees were jailed to compel the platforms to remove content globally (rather than blocking access for users in Brazil). In India, Twitter offices in Delhi were raided by the police after the platform had labelled a tweet by a member of the ruling party as "manipulated media".

Does the proposed legislation represent a threat to freedom of expression, or are the protections for freedom of expression provided in the draft Bill sufficient?

We welcome the UK's commitment to safeguard freedom of expression online by outlining a duty for platforms to protect it. **However, we find that certain provisions of the Bill do threaten freedom of expression:**

- **The Bill broadly targets harmful content online, without specifying in clear terms what is considered to be harmful to adults or children** – beyond "a material risk of the content having, or indirectly having, a significant adverse physical or psychological impact". This, coupled with the circular definition of certain illegal content (for instance, terrorist content is defined as content that leads to a terrorist offence) effectively delegates the adjudication of what is harmful or illegal to private tech companies when **limits to freedom of expression should be adjudicated by independent judicial authorities in line with international human right standards.** This delegated responsibility means that platforms are likely to err on the side of caution by over-removing content, possibly removing legal and non-harmful material in the process, to avoid being sanctioned.

- **The inclusion of private communication channels in the scope of the Bill is also of concern for freedom of expression and privacy.** The majority of online platforms providing private communication services now offer end-to-end encryption, and requiring such platforms to monitor their services to counter illegal content is effectively a mandate to break encryption and its inherent promise of privacy.[8] Asking platforms to monitor their services for illegal content, as the Bill suggests in Section 9 (3) on minimising the presence and dissemination of illegal content, further threatens freedom of expression given that users will self-censor if they know their online communications are systematically monitored. To avoid asking platforms to both protect and infringe on freedom of expression and privacy, we recommend that private communication is excluded from the scope of the Bill, or at least exempted from the duties to limit the presence of illegal content.

  On the question of the threat to freedom of expression, please see Tech Against Terrorism's written evidence to the House of Lords Communications and Digital Committee inquiry into Freedom of Expression Online.

[8] Tech Against Terrorism recently published an in-depth report assessing the risks of terrorist use of end-to-end encrypted services as well as possible mitigation strategies that do not infringe on the right to privacy nor create security risks. You can find our report and recommendations for governments and law enforcement here.

Will the regulatory approach in the Bill affect competition between different sizes and types of services?

Tech Against Terrorism welcomes the Bill's consideration for platforms' sizes as outlined by the provisions in Schedule 4 – which outline the different categories of platforms. **However, in the current absence of precise threshold conditions for the different categories of services, we are concerned with the negative impact the bill may have on smaller platforms:**

- **We regret that platforms' resources (human, technical and financial) are not listed in the threshold conditions for categories of regulated services**. Resources, or rather the lack therefore, are the factor most determinative of a platform's ability to counter terrorist use of its services. Platforms with limited resources will also struggle to comply with requirements that indiscriminately apply to platforms of all sizes, whereas larger platforms will have the capacity to allocate more resources to the systems and processes necessary for compliance. Indiscriminately requesting all platforms to comply with the same legal requirements will punish smaller platforms instead of providing them with the support needed to counter the dissemination of illegal content. To avoid this, HMG **should consider including platforms' financial, human, and technical resources in the criteria for categorising platforms, and consider how smaller platforms might be supported.**

Are there systems in place to promote transparency, accountability, and independence of the independent regulator?

- **Tech Against Terrorism welcomes the obligation for Ofcom to produce transparency reports. However, we would recommend that such reports be based not only on findings from the providers of regulated services, but that reporting also integrate Ofcom's own metrics and data relating to its practices of regulatory enforcement and provider collaboration.**

- **Tech Against Terrorism calls on governments to be more transparent about their collaboration with tech platforms to counter terrorist use of the internet**. In July 2021, we published our Guidelines on Transparency Reporting on Online Counterterrorism Efforts as a framework for governmental accountability both to their citizens and to internet users with regard to online counterterrorism measures. In particular, we recommend that Ofcom align with Part A of the Guidelines, as far as is possible in the lawful exercise of its powers, to explain the role played by the OSB in HMG's efforts to counter terrorism online, and equally with Part B, on process and systems, to explain the mechanisms that exist to support tech platforms in identifying terrorist content.

- **Beside the mandate to produce transparency reports and reports about researchers' access to information, the Bill contains limited information regarding the regulator's transparency and accountability.** Tech Against Terrorism recommends that Ofcom publish detailed and publicly available guidance for tech companies on how to comply with the Bill's provisions, as well as on the criteria Ofcom will use to categorise platforms (according to the categories outlined in the Bill) and assess their regulatory compliance.

Does the Bill deliver the intention to focus on systems and processes rather than content, and is this an effective approach for moderating content?

- **Tech Against Terrorism cautiously welcomes HMG's focus on risk assessments and consider this the first step in countering terrorist use of the internet.** Tech companies should to the best of their abilities consider how their platforms could be exploited and remain aware of hostile usage. Tech companies, especially smaller platforms, will need support in carrying out risk assessments, as they may not have the resources or capacity to conduct these. It is our hope that HMG will support smaller tech companies in responding to terrorist exploitation of the internet.

- We do warn HMG against **requiring** tech companies both big and small to carry out these processes and risk assessments. If the Bill does not adequately consider proportionality, it risks becoming ineffective, and the enforcement mechanisms in the Bill may risk putting smaller tech companies out of business. This is due to the fines and potential liability that await tech companies when they do not oblige with the sometimes stringent requirements that are put on tech companies, which smaller enterprises may not be able to meet. We also hope that HMG will support initiatives like Tech Against Terrorism to assist tech companies to conduct these risk assessments.

- Regarding algorithmic recommendations, our position paper on algorithmic amplification and terrorist use of the internet demonstrates that algorithms often play a lesser part in the promotion of terrorist content than is commonly thought.[9] **We advise HMG to focus on the removal of terrorist content, rather than on interference with the algorithms that promote content.** Through Tech Against Terrorism's Terrorist Content Analytics Platform (TCAP) we found that the majority of tech companies where most terrorist content is found do not use promotion algorithms.

- While we recognise that HMG wants to go beyond the prohibition of content, **we advise that more definitional clarity be provided when detailing the different types of content.** Even if the focus is on processes and risk assessments, it is the existence and prevalence of certain content on a given tech platform that remains the central problem. To commit to processes for removing this content, a definition must be the starting point.

- Tech Against Terrorism welcomes the fact that the OSB recognises that in moderating online speech, journalistic and otherwise legitimate material may be erroneously removed. This has negative consequences for human rights and freedom of speech. **However, we caution against the way in which the UK government endeavours to protect such material online**. The Bill specifies "democratic content" as any content that furthers democratic debate in the United Kingdom. Tech Against Terrorism is concerned that tech companies will struggle to adjudicate on what content does not fall within the remit of this vague definition. **Without clear definitions and detailed guidance, individual tech companies have to interpret this definition for themselves, which in turn risks inconsistent application of the law**. Likewise, journalistic content, according to section 14.8, is defined in the Bill as content produced by a news publisher, content that is UK-linked,[10] or generated for the "purposes of journalism". Tech

[9] This is based on our monitoring of online content dissemination by designated terrorist groups included in our Terrorist Content Analytic Platform's group inclusion policy.

[10] This means whether the content is likely to be of interest to a significant number of UK users.

companies are likely to struggle to determine what constitutes terrorist content shared for terrorist purposes as opposed to terrorist content shared for journalistic objectives. HMG should provide more context and guidance to tech platforms on how to determine what constitutes journalistic content.

- **To circumvent platforms' content moderation systems, terrorists and violent extremists use a variety of avoidance techniques. We worry that the broad definitions of journalistic and democratic content can be exploited for this purpose**. In our research we note that both Islamist and far-right terrorist and violent extremists avoid content moderation by posing as news agencies or journalists. This frustrates the task of identifying whether content is journalistic or terrorist in nature. Furthermore, violent extremists or terrorists may deem that their online content furthers the democratic debate (see definition of "democratic content" above) and this could provide a justification for them to appeal the removal of their content. Further detail and guidance on how to differentiate between legitimate journalistic and democratic content and terrorist content is required to support platforms in rendering effective compliance.

- **Beyond this, we recommend that, if HMG wants to keep the provision around journalistic content, the Independent Press Standards Organisation (IPSO) provide more guidance on how journalists should report on terrorist content**. Tech Against Terrorism warns against journalists sharing terrorist content without any contextual relevance as this furthers the dissemination of terrorist material, which is exactly what terrorists aim for. Therefore, we advise HMG to encourage IPSO to publish guidance on how to report on terrorist content responsibly lest the provision on journalistic content in the Bill indirectly promote the further dissemination of terrorist content.

What would be a suitable threshold for significant physical or psychological harm, and what would be a suitable way for service providers to determine whether this threshold had been met?

- **Tech Against Terrorism argues that adjudication of the legality or level of harm of content should be the role of governments, not tech platforms.** We also argue that if the Bill is going to establish a threshold for significant physical or psychological harm, this should also be decided by governments, not tech companies. This is because physical or psychological harm is a term that is ill-defined and open to interpretation and would therefore leave too much power to tech companies having to decide on what falls under this. An overly expansive interpretation could lead to over-removal of content, and negatively impact human rights. An excessively narrow interpretation would not accurately protect the United Kingdom's internet users from harm. Whatever definition of harm is adopted by HMG, it must be clearly stated in the Bill, with supplementary guidance made available.

Are the definitions in the draft Bill suitable for service providers to accurately identify and reduce the presence of legal but harmful content, whilst preserving the presence of legitimate content?

- **Tech Against Terrorism cautions that the OSB's definitions for both illegal content and terrorist content are impractically broad and circular.** Illegal content is, in the draft Bill, defined as content that leads to an offence, and terrorism content is defined as "terrorism content that leads to a terrorist offence". Whereas terrorist offence has been defined in other relevant legislation, this definition does little to inform a tech company about what content falls under these definitions, nor does it inform how tech companies should operationalise this definition when acting against terrorist exploitation of their services. **This leaves tech companies to adjudicate on what constitutes terrorist content.** Whilst terrorist content that clearly depicts violence and incites violence might be easy for platforms to detect, in reality most terrorist groups frequently share "grey area" content, which is generally difficult to identify. The current definition means that tech companies will need to make difficult decisions in correctly assessing whether content is terrorist or not.

- We understand that Ofcom will provide further guidance to tech companies on how they need to uphold the Bill's obligations on illegal and terrorist content. **However, we still deem that an unambiguous definition should be inserted into the Bill to relieve Ofcom and tech companies of the full burden of implementation**. For these regulations to target what they intend to, we advise more clear and detailed definitions that can be operationalised by tech platforms.

- This also applies to democratic and journalistic content, which we further elaborate on in the response pertaining to that provision.

What role do algorithms currently play in influencing the presence of certain types of content online?

- We welcome the increased focus amongst policymakers on the role played by content personalisation and other algorithmic recommendation systems on online platforms. Such scrutiny is warranted. Terrorist groups exploit platforms that make use of recommendation algorithms, and there are examples of individuals encountering terrorist and violent extremist content via platforms using content personalisation. **However, we are concerned that the current debate is, on a policy level, based on an incomplete understanding of terrorist use of the internet, and that a focus on content personalisation is a distraction from more important steps that should be taken to tackle terrorist use of the internet.**

- **There is very limited evidence that content personalisation leads to terrorism.** There are cases in which an individual's radicalisation process has partly consisted of visiting platforms using content personalisation, but there is no conclusive evidence to suggest that algorithmic recommendation systems lead individuals to terrorism. Even one of the studies most critical of YouTube (and the platform's role in radicalisation) does not highlight the platform's recommendation algorithm as the most significant reason behind exploitation of the platform.[11] In addition, a more recent study that compares YouTube, Reddit, and Gab's recommendation

[11] Lewis Rebecca (2018), "Alternative Influence: Broadcasting the Reactionary Right on Youtube", published by Data&Society.

algorithms finds that far-right content is promoted on YouTube, but not on Gab or Reddit.[12] We encourage policymakers to look beyond the disproven 'conveyor-belt' theory of radicalisation, which suggests that people are radicalised by simply viewing terrorist content online. It is well-established that radicalisation is much more complex, and most academic research argues that individuals are largely radicalised offline and subsequently seek content online proactively. In our view, it is crucial that any policy or regulation affecting content personalisation is underpinned by evidence and extensive research.

Are there any foreseeable problems that could arise if service providers increased their use of algorithms to fulfil their safety duties?

- **Service providers' increased reliance on automated content moderation tools can lead to wrongful removal of content, threatening human rights and freedom of speech in particular.** We recommend tech platforms to rely on both proactive removals as well as human moderation to ensure that as far as possible they mitigate this risk. It would contravene the stated purpose of the Bill, to keep users from online harm and at the same time protect legitimate speech online, if tech companies are forced to rely on proactive algorithmic removal alone. It will also undermine an open and diverse internet and risk forming "content cartels",[13] online speech expert Evelyn Douek. Douek uses this term to describe collaborative tech industry arrangements which harmonise their practices in removing illegal and harmful content online, including child sexual abuse material and terrorist content. We therefore recommend that HMG take this into account when taking the Bill forward.

Is Ofcom suitable for and capable of undertaking the role proposed for it in the draft Bill?

- T**he Bill remits the implementation of many provisions to Ofcom, which therefore faces a significant challenge in providing all the codes of practice and guidance which will be required by tech companies to ensure their compliance**. In addition, Ofcom would also have to work with a very wide range of service providers, whether search engines or user-to-user services where proportionality would have to be considered. We therefore assess that Ofcom would need a significant increase in resources to undertake this. In addition, we advise Ofcom and HMG to establish public-private-partnerships which could help Ofcom provide these codes of practice, and to build the constructive relationships which are the Bill's best hope of fulfilling its stated objective.

[12] Whittaker Joe, Looney Sean, Reed Alastair and Votta Fabio (2021), "Recommender systems and the amplification of extremist content" in *Internet Policy Review,* Vol.10, Issue 2.
[13] Evelyn Douek. The Rise of Content Cartels (Colombia University, 2020).

How will Ofcom interact with the police in relation to illegal content, and do the police have the necessary resources (including knowledge and skills) for enforcement online?

- **We argue that terrorists and violent extremists should ultimately be the ones who should be held accountable for posting terrorist content on the internet, not tech companies or their employees**. Therefore, we advise the UK government to work with the police to identify those responsible and prosecute them accordingly.

- Concerning **the police and their ability to enforce accountability online, we argue that the UK has existing legal measures in place to do this.** We recommend against expanding the scope of police powers, and in this we agree with the previous Independent Reviewers of Terrorism Legislation that the existing measures are sufficient.

- **Finally, we would urge HMG not to consider any powers that disproportionately target users' right to privacy in an attempt to moderate terrorist content of the internet or hold to account those who disseminate such content.**