Tech Against Terrorism
Submission to the consultation process for Ofcom's Consultation on guidance for VSP providers on measures to protect users from harmful material

Submitted 2 June 2021

1. BACKGROUND

Until the United Kingdom Online Safety Bill – aiming to counter harmful content online and announced in a White Paper in April 2019 – is implemented, UK-established Video-Sharing Platforms (VSPs) will have to comply with a dedicated VSP regulation meant at protecting users from "harmful" content.
The UK media regulator Ofcom will be given new powers to regulate VSPs in the UK. These include ensuring that VSPs have appropriate measures in place to protect users from illegal content as well as from incitement to hatred and violence. In line with this, Ofcom published a draft guidance for VSPs on measures to protect users from harmful content and compliance with the new regulation. The draft guidance was opened for public consultation from March to June 2021.

2. TECH AGAINST TERRORISM'S CORE ARGUMENTS AND RECOMMENDATIONS

Our arguments throughout our response can be summarised as follows:

- **Accountability and practical guidance** – Governments need to provide more leadership and practical guidance in tackling terrorist use of the internet and not place the onus on private companies. Ofcom should consider supporting organisations and mechanisms with the expertise to practically support smaller VSPs with complying with the new regulation.

- **Proportionality and consideration for smaller platforms** – Smaller platforms are the most exploited by criminal actors, however, they often lack the resources and capacity needed to counter the threat. Whilst we commend Ofcom's recognition of diversity in platforms' sizes, this should be further emphasised and Ofcom should specify which measures are recommended according to size and resources. A robust support programme for smaller VSPs, as well as practical Guidance and tools are needed to address the problem of terrorist and violent extremist online content.

- **Rule of Law** – Counterterrorism and tackling online harms need to be based on the rule of law and pay due regard for human rights, in particular freedom of expression. Lists and definitions of proscribed content and behaviours should be detailed, and inscribed in the rule of law by clearly referring to existing law on counterterrorism and acceptable limits to freedom of expression which lay out what is considered illegal speech in the UK.

3. GENERAL COMMENTS (question 16 of the consultation)

At Tech Against Terrorism, our expertise relates to terrorist and violent extremist material online. Therefore, throughout this response, our focus shall be on “relevant harmful material”.
Whilst the Guidance sets a commendable framework for platforms to comply with Part 4B of the Communications Act 2003, it lacks practical guidelines and recommendations (e.g., a checklist) and fall short of supporting platforms to operationalise the Guidance.

The Guidance also fails to lay out concrete recommendations adapted to the diverse VSP landscape and does not fully acknowledge that platform size and resources can impede practical measures, thus creating unrealistic expectations for platforms. Many of the recommendations, in particular regarding prohibited content and policy, included in the Guidance are also already implemented by most VSPs.

In our experience with smaller tech companies, Tech Against Terrorism finds that practical enforcement is where many smaller platforms struggle to follow up from policies and thus where most support is needed. Tech Against Terrorism recommends Ofcom to focus more on practical guidelines and tools on how to implement policy and enforce moderation, and on improving cooperation between law enforcement and tech companies.

- **Practical Guidance**

Unfortunately, the Guidance is neither practical in its advice nor realistic in its expectations. Assessing what constitutes harmful speech and thus a legitimate restriction to freedom of expression – including correctly assessing whether content is terrorist – is a complex and lengthy process. Although large platforms will have the resources to hire the necessary legal team to support such an assessment, smaller platforms simply won’t and the Guidance does little to support that.

In general, the Guidance is not needed for larger platforms, which have the resources and capacity to ensure they can comply with the 2020 VSP regulation (and which already have the necessary policy and mechanisms in place to comply). Smaller platforms are the ones that require support to comply with the VSP regulation. However, by focusing on broad guidelines and indications on what processes and mechanisms to deploy, with almost no Guidance on how to effectively operationalise them Ofcom falls short of providing the necessary support to smaller tech companies. Ofcom should consider supporting organisations and mechanisms with the expertise to practically support smaller tech companies with complying with the new regulation. Alternatively, Ofcom should develop the necessary support mechanisms, in particular regarding the dispute resolution mechanisms which set unrealistic expectations.

- **Proportionality**

We recommend that Ofcom provides a clear framework indicating what baseline measures a platform should consider depending on their size. This framework should be drafted in consultation with tech companies of all sizes, and civil society organisations supporting the tech sector such as Tech Against Terrorism and IWF to ensure that the measures suggested are reasonable.

In section 5, Ofcom commendably notes that various metrics can be used to determine platform size, in particular resources and capacity. However, the Guidance does not specify a threshold for any of these metrics, and instead suggests that platforms themselves should consider what measures are proportionate. Whilst we welcome the focus on proportionality and Ofcom's acknowledgement that platforms' sizes and resources impact content moderation enforcement, the Guidance is asking platforms to adjudicate on what they should do to comply with legal requirements without providing a practical framework for platforms to operationalise the Guidance.

In section 5.14 of the Guidance, Ofcom further notes that "cost and resources cannot be considered in isolation when determining whether a measure is practicable and proportionate. What is practical and proportionate must be considered in the round and weighed against the risk of harm to users on a platform." We are concerned that Ofcom may have overlooked overwhelming evidence that smaller platforms are the most exploited by criminal actors, and that a strict penalty regime without a robust support programme or practical Guidance and tools will not address the problems of terrorist and violent extremist online content.

- **Public Interest**

Section 5.35 of the Guidance states that "In designing and implementing protection measures, VSP providers should also take into account the impact such measures may have on the general public. For example, some content which might initially seem harmful, may actually be in the public interest."

This is important – especially in the context of reporting on human rights violations and war crimes – however quite complicated in practice. Ofcom should consider, for example, terrorist and violent extremist propaganda content produced by terrorist organisations with the aim of being reshared by media organisations and journalists, thereby bypassing content moderation, and provide Guidance for platforms on how to address this.

4. CONSULTATION RESPONSE

Question 1: Do you have any comments on Section 3 of the draft Guidance on harmful material and related definitions?

- We commend the Guidance for providing an initial framework and regulatory basis for tech companies to action this content. However, the Guidance remains vague in specifying what content is covered by the regulation, and provides little practical Guidance for platforms to properly assess online content.
  Section 3.24 of the Guidance refers to "material the inclusion of which would be a criminal offence." This presents a circular problem: terrorist content is already illegal under UK law, and the VSP Guidance does not provide any further indications for platforms on how to assess what is terrorist content. Platforms are therefore left to adjudicate on what constitutes a criminal offence, when it should be the role of the government to provide clear and precise Guidance on what exactly constitutes terrorist content.

  Section 3.22 of the Guidance defines "incitement to hatred" as " having its usual meaning in everyday language." Beyond that the Guidance is vague and recommends platforms to refer to case law as well as consider context when actioning content. This does little to help platforms assess whether a content falls into incitement to hatred, and thus does not support platforms in operationalising the 2020 VSP regulatory framework on how to correctly identify and action "grey area" content.

  The lack of a practical framework on how to assess incitement to hatred and violence will also complicate platforms' capacity to correctly identify terrorist and violent extremist content, as the two types of content often overlap in practice.

  Tech Against Terrorism recommends Ofcom to clarify and detail its definition of harmful material. Listing and definitions of proscribed content and behaviours should be: detailed; inscribed in the rule of law by clearly referring to existing law on counterterrorism and acceptable limits to freedom of expression which lay out what is considered illegal speech; providing examples of the type of proscribed content. it. Overall, Ofcom's definition should practically support tech companies in building a framework to classify content.

  These recommendations are similar to Tech Against Terrorism key recommendations for tech companies on prohibiting terrorism and violent extremism on their platforms. Tech Against Terrorism recommends tech companies to be as clear and detailed as possible in their Community Guidelines, and to refer to national and international designation lists to inscribe their prohibition of terrorism in the rule of law.

  Tech Against Terrorism also support platforms in understanding the terrorist and violent threat and identifying related content:

- *Platforms using our Knowledge Sharing Platforms (KSP) will have access to information on the terrorist and violent extremist threat landscape to strengthen enforcement mechanisms and content moderation, including a compendium of symbols associated with designated terrorist groups and violent extremist groups, including logos as well as other visual identifiers such as visual imagery, flags, and tattoos; as well as a terminology dataset, containing key terms and phrases used by terrorists and violent extremists of different ideologies and groups. These visuals and terminology datasets can be used by platforms to inform their moderation enforcement.*

- *Tech Against Terrorist also alerts content linked to designated terrorist groups to tech companies via our Terrorist Content Analytics Platform (TCAP). We also alert users of content that does not fall within the TCAP's Group Inclusion Policy, via email and share detailed threat assessments with concerned platforms.*

Question 2: Do you have any comments on the draft Guidance about measures which relate to terms and conditions, including how they can be implemented?

- Section 2.33 of the Guidance requires platforms to "include terms and conditions to the effect that a person must not upload to the service a video containing relevant harmful material." This is already common practice for most tech companies which delineates what is acceptable on their platforms in their Content Standards (Community Guidelines and Terms of Services). These usually include provisions on illegal content, and most platforms would have baseline prohibition of harmful material.

  The VSP Legislation and Guidance are commendable in requiring all VSPs to include a prohibition of harmful material. However, the Guidance should be developed to provide practical guidelines. Additional Guidance could be provided regarding what form this prohibition should take, recommending platforms to inscribe this prohibition in the rule of law by referring to international and national designation lists of terrorist organisations, or to the European Court of Human Rights fact sheet on incitement to hatred for instance.

  There is a lot of focus in the Guidance on Terms and Conditions, rather than on Community Guidelines. Whilst this is mostly a difference of terminology, in practice Community Guidelines are where tech companies comprehensively delineate what content and behaviour is prohibited on their platforms, and how they respond to a violation. Community Guidelines are also more commonly presented in a user-friendly format and adapted to the user-base.
  Tech Against Terrorism recommends Ofcom to review existing Community Guidelines from VSPs to ensure its recommendations are informed by what is already practiced by tech companies.

  Tech Against Terrorism conducts in-depth policy reviews of platforms' content standards and provide bespoke recommendations on how to strengthen and future-proof counter terrorism and violent extremism whilst safeguarding human rights and freedom of expression, as part of our Mentorship programme for smaller platforms. All of our policy support is accompanied by practical assistance, including via our KSP and Terrorist Content Analytics Platform.

The KSP is a collection of interactive tools and resources designed to support the operational needs of smaller tech platforms. It is a "one stop shop" for companies to access practical resources to support their counterterrorism and transparency efforts. Our resources include research and analysis on terrorist use of the internet, such as the threat landscape and proscribed organisations, on global online regulation, as well as guidelines and recommendations on content standards and transparency reporting.

- Sections 4.27 through 4.30 address the length and readability of terms and conditions. This inclusion is commendable – Tech Against Terrorism already encourages platforms to consider ease of use when conducting policy review.
  In general, the Guidance included is commendable and interesting. However, it would make more sense for the Guidance to recommend platforms to focus on the demographics of the user base to tailor the Community Guidelines . Some platforms already do that well.

  Sections 4.50 - 4.53 recommend VSPs to regularly review their terms and conditions and amend when necessary. This is a good recommendation given the fast-changing terrorist and violent extremist online threat landscape.

  Tech Against Terrorism recommends developing this section to include specific factors that platforms should consider when reviewing their content standards. For instance, the Guidance could underline the risks of malevolent actors adapting their use of the platform to content moderation policy and practices to avoid having their content removed or account banned.

Question 3: Regarding terms and conditions which prohibit relevant harmful material, do you have any comments on Ofcom's view that effective protection of users is unlikely to be achieved without having this measure in place and it being implemented effectively?

- We welcome the enforcement and sanctions regime as set out in the Guidance. We recommend Ofcom provide greater detail and clarity on the nature of enforcement, e.g., the thresholds for sanctions.
  Tech Against Terrorism agrees with Ofcom that policies and guidelines should be followed by effective measures of implementation, which is why Tech Against Terrorism's support to tech companies include both a policy and practical aspect. Our Knowledge Sharing Platform notably includes a guide on content moderation, outlining the positives and limitations of different content moderation strategies for tech companies to inform their moderation enforcement.

  However, we are concerned that the Guidance provided by Ofcom is too limited to general comments and broad policy guidelines without providing enough indications on how to practically operationalise those.

  We recommend Ofcom to conduct a more in-depth assessment of what moderation policy and practice are already used by VSPs to inform practical recommendations to ensure that platforms can be supported in implementing the necessary moderation measures.

- Section 4.40 states: "Effective action in response to violations might include warnings; temporary bans on posting content; bans on interacting with the content of others; demonetisation; temporary account restrictions; and permanent removal or deletion of accounts. We are aware that some VSPs also block IP addresses."

  We regret that Ofcom's recommendations on content moderation enforcement strategies is limited to the above listing of a few examples of moderation strategies.

  Tech Against Terrorism recommends Ofcom to develop on the sections related to "Enforcement and Sanctions" and "Moderation" to ensure that it provides tech companies with a complete and detailed guides on possible enforcement strategies, adapted to the nature of VSPs' products offering, for VSPs to inform their response in line with Ofcom's Guidance. Ideally, this should take the form of a broad landscape review of content moderation strategies in use by VSPs, the positives and the negatives of each as well as the resources they require.

  Tech Against Terrorism's Knowledge Sharing Platform includes a section on "Alternative Content Moderation Solutions" which provides a similar landscape review of moderation enforcement strategies to inform smaller platforms' understanding of the enforcement mechanisms available to them beyond content removal. Tech Against Terrorism could collaborate with Ofcom to publish a similar handbook for VSPs based on the Guidance.

Question 5: Do you have any comments on the draft Guidance about reporting or flagging mechanisms, including on Ofcom's view that reports and flagging mechanisms are central to protecting users?

- We highly commend sections 4.54 through 4.72 of the Guidance for paying attention to user reporting and flagging. These functions are particularly important for smaller platforms which lack the technical tools to proactively monitor content and prevent upload. For smaller platforms, user reporting can be a crucial means of identifying violating activity and content.

  We also commend section 4.68 in particular on notifying users on the progress of their report. At Tech Against Terrorism, we have made similar recommendations.

  To complement this, Tech Against Terrorism suggests recommending platforms to notify users when their content or account is removed, or otherwise actioned, and explain then the moderation decision in full. This is part of Tech Against Terrorism's key recommendations for tech companies, as it allows users to better understand moderation policy and enforcement.

  We would also recommend a tiered approach to user reporting, allowing not just logged-in users, but all users to report content; a 'trusted flagger' system in order to prioritise content reports; and the inclusion of all prohibited content and behaviour in the reasons given for reporting moving content across all reporting functions, in order to assist prioritising terrorist and violent extremist content in the review process.

Question 6: Do you have any comments on the draft Guidance about systems for viewers to rate harmful material, or on other tagging or rating mechanisms?

- We are concerned that the rating system in sections 4.73 through 4.86 of the Guidance and the reporting system of sections 4.54 through 4.72 may contradict each other.
  According to the reporting system outlined in the Guidance, users are to be asked to report any content that may be harmful. Under the rating system however, the decision of whether content is indeed harmful is to be "crowd-sourced". Whether platforms should favour user reporting or rating should be clarified.

  Rating can be useful to crowdsource the moderation of content that is not harmful but may be offensive or should not be viewed by children, and is already practiced by certain VSPs offering users the possibility to tag "Not safe for work", or in certain instances, "Not safe for life" content.

  However, such rating systems would need to be complemented by some form of proactive moderation based on ratings to limit the risks of malevolent actors using such ratings in the hope of circumventing content moderation. Tech Against Terrorism can provide Ofcom with documentations on instances of violent extremist groups found to be labelling their content on VSPs with "Not Safe for Work/Life" ratings.

Question 7: Do you have any comments on the draft Guidance about age assurance and age verification, including Ofcom's interpretation of the VSP Framework that VSPs containing pornographic material and material unsuitable for classification must have robust age verification in place?

- We are concerned that the age assurance and age verification systems recommended in the Guidance would not only fail to address the identified problem, but also be impossible to implement without some sort of mandatory legal user registration.

  For age verification to be effective, it would require the deanonymization of the services, as platforms need to verify the identity of users. This is not a proportional response and violates existing norms.

  We strongly advise against such registration: similar policy proposals in other jurisdictions have been criticised for seriously endangering undermined fundamental freedoms and human rights (see the Brazilian Senate, PLS 2630/2020, Articles 7 and 8).

Question 10: Do you have any comments on the draft Guidance about the measure regarding complaints processes or on the regulatory requirement to provide for an impartial dispute resolution procedure?

- Tech Against Terrorism commends the importance given to redress mechanisms for users to contest content moderation decisions. User appeals are a key component of a platform's accountability towards its users as it ensures the possibility to contest a decision and be more informed about the thought process behind takedowns. This is an important safeguard for freedom of speech online, and it increases accountability towards users.

  Overall, the Guidance is focused on how easy the complaint process should be for users. However, redress processes can be challenging for platforms to manage and respond to swiftly, especially for smaller platforms.

  For example, sections 4.125 and 4.126 on analysing the effectiveness of complaints process is highly demanding of tech companies and potentially costly. Although some platforms are attempting to replicate the automation and processes of user reports for appeals, this requires significant resources which small platforms do not have. More consideration needs to be given to how this process can be facilitated for tech companies.

  The dispute resolution procedure in sections 4.128 through 4.146 is also unrealistic and impractical: the Guidance essentially suggests that all platforms, regardless of size, should establish the equivalent of the Facebook Oversight Board. Requiring all VSPs to set up a similar oversight body is unrealistic even for large and long-established platforms.

  It would require platforms to dedicate significant time and resources to establishing and running this body, and most platforms, in particular smaller ones, will not be able to do soThe Facebook Oversight Board offers an example of the complexity of dispute resolution mechanisms for tech companies. Despite Facebook's resources, the Oversight Board took two years to be set up (the creation of the Board was announced by Mark Zuckerberg in November 2018, the Board began receiving cases in November 2020), and still has limited capacity with the Board having to make decisions on which cases to accept. To this day, Facebook is the only online platform to have established such a dispute resolution mechanism, no other platforms has attempted to do so. This is easily explained by the significant resources necessary to set up and run such a mechanism even for large platforms.

  Tech Against Terrorism notes that the Guidance state the possibility of this role being delegated to a third-party. However, this would still be costly and inaccessible to smaller platforms. There is also no existing organisation that provides such services, and the responsibility to ensure a mechanism is in place still falls on the platforms.

  We recommend that if Ofcom wishes to see effective and impartial dispute resolution mechanisms, this should be the role of Ofcom itself; we cannot reasonably expect such mechanisms to be implemented by every single VSP.

Question 11: Do you have any comments on the draft Guidance about media literacy tools and information?

- We commend the focus on media literacy in sections 4.147 through 4.160. However, we are conscious that such Programmes will be costly and time-consuming to develop for most platforms.

  Improving media literacy should be the educational responsibility of the government not private organisations, particularly if universal access to media literacy is a priority – ideally this should be part of a school programme).

  Potentially, we recommend Ofcom to create a microsite containing all media literacy Guidance it wishes platforms to convey for all VSPs would have to refer or hyperlink to on their platforms.

  We also emphasise that media literacy should not be considered as a panacea: we need to address the root causes of terrorism and violent extremism, and this goes beyond online activities and the capacity of tech platforms.

Question 12: Do you have any comments on the draft Guidance provided about the practicable and proportionate criteria VSP providers must have regard to when determining which measures are appropriate to take to protect users from harm?

- The practical and proportionate criteria are generally appropriate. However, they are also quite complicated to assess for tech companies, in particular with regard to the nature of the material and the harms it may cause in sections 5.22 through 5.25.
  Most platforms do not have a deep understanding of the terrorist and violent extremist threat online, and any assessment of these criteria would have to be delegated to a third party.

  We encourage Ofcom and the UK government to support cross-sector initiatives providing supports capacity-building and training for smaller tech companies to fully understand the threat they face and how to efficiently counter it, such as Tech Against Terrorism. We already provide policy and practical support to smaller platforms on how to counter terrorism whilst respecting human rights, and we could adapt our recommendations to assist VSPs in identifying proportional counterterrorism and moderation in line with the criteria set out in the Guidance.

  These trainings should be regular to ensure that policy and enforcement actions are adapted to the evolving threat, in line with the Guidance on reviewing terms and enforcement mechanisms.

  As part of our practical support to tech companies, Tech Against Terrorism supports tech companies in understanding the threat to their platforms and draw recommendations on how to best counter it bespoke to the service and user base, including via our Mentorship Programme and Knowledge Sharing Platform.

- Section 5.4 states that "Decisions about which measures to take and how to implement them are related to the risk of harm to users on a platform. A low risk of harm is, generally, likely to require fewer or less sophisticated protection measures compared to a VSP with a greater risk of harm." The Guidance also notes that platforms' resources should be considered in assessing what measures are practicable and proportionate. Both are good inclusions, and Tech Against Terrorism commends Ofcom for acknowledging the limitations posed by platforms' resources.

  Ofcom is right to stress that different VSP will present different risks of harms to users, and that platforms' resources should be considered in how VSPs are to respond to harms. However, there is an inherent contradiction in those two considerations.

  Tech Against Terrorism research has shown that the smaller platforms with fewer resources and capacity are most exploited by terrorist groups. According to Ofcom's "risk of harms to users" criteria, smaller platforms heavily exploited by terrorists and violent extremists would thus represent a high "risk of harms to users" and would be expected to deploy the most sophisticated protection measures without receiving the practical support they need to do so. Tech Against Terrorism recommends a greater consideration to be given to supporting smaller platforms with practical Guidance and tools on which measures to implement, and how to implement them despite limited resources.

Question 13: Do you have any comments on the draft Guidance about assessing and managing risk?

- We commend the Guidance for underlining that measures taken by platforms should be reviewed regularly in consideration with the evolution of harmful material. We do however have some concerns regarding operationalisation and practicality. Smaller platforms are likely to struggle to conduct meaningful risk assessments as this would require significant understanding not only of their services, but also the overall threat landscape.

  In our experience many tech platforms lack the resources to build an understanding of the terrorist and violent extremist online landscape, which also requires understanding the terrorist and violent extremist offline landscape, radicalisation factors, and the risk of violence.

  Platforms will not only have to understand their own risk of exploitation, but how they fit in the broader terrorist and violent extremist ecosystem, e.g., how other platforms respond to terrorist and violent extremist content and how that influences activity on one's platform. This is a complex threat picture, and impossible to comprehend without knowledge-sharing and OSINT support, especially for small and medium platforms.
  Tech Against Terrorism recommends that Ofcom supports knowledge sharing for smaller tech companies to understand and efficiently respond to the threat. For instance, by supporting Tech Against Terrorism's OSINT monitoring capacity and Knowledge Sharing Platform, which both provide practical support to tech companies in understanding the evolving threat and which measures are required.