making communications work
for everyone

Your response

Please refer to the sub-questions or prompts in the annex to our call for evidence.

Question	Your response
Question 1: Please provide a description introducing your organisation, service or interest in Online Safety.	*Is this response confidential? – N* Tech Against Terrorism is an initiative established in 2017 by the UK-based Online Harms Foundation under the aegis of the United Nations Counter Terrorism Executive Directorate (UN CTED). Tech Against Terrorism supports the global technology sector to respond to terrorist use of the internet whilst respecting human rights and promotes public-private partnerships to fortify this response. Our focus is on smaller and newer tech platforms which are at higher risks of terrorist and violent extremist exploitation but often lack the necessary knowledge and resources to tackle the threat. Tech Against Terrorism's research shows that terrorist groups - both Islamist and far-right in orientation - consistently exploit smaller tech platforms when disseminating propaganda. Tech Against Terrorism's mission is to support smaller tech companies to tackle this threat whilst respecting human rights and to provide companies with the practical tools to do so. As a public-private partnership itself, the initiative is supported by the Global Internet Forum to Counter Terrorism (GIFCT) and by the governments of Spain, Switzerland, the Republic of Korea, and Canada. We were recommended as an industry partner in the UK

Government's Interim Code of Practice on terrorist content and activity online (published in December 2020).

We support the global tech sector across three pillars:

- Analysis of the threat and original research: Our open-source intelligence (OSINT) team monitors the evolution of the terrorist and violent extremist online threat landscape to identify at-risks platforms and monitor terrorist use of the internet trends. Our research team works to produce in-depth analysis of terrorist use of the internet, based on data from our Terrorist Content Analytics Platform and insights from our OSINT team. You can find our latest OSINT report on the threat of terrorist and violent extremist operated websites here.
- Policy Advisory and Response: we work in direct collaboration with 40+ platforms across the tech sector to have a positive and sustainable impact on global online counterterrorism efforts, supporting resilient yet flexible online counterterrorism policy responses via knowledge-sharing and capacity building programmes. Our key policy support programme is our Mentorship, which helps tech companies in strengthening their online counterterrorism response in a rights-respecting and transparent manner. Our policy team also supports tech companies in understanding online regulation and how legal requirements on moderating online content may impact their counterterrorism approach.
- Technical support: We build tools and technical approaches to support platforms in swiftly and accurately removing terrorist content on their

	services. Our Terrorist Content Analytics Platform (TCAP) alerts over 70 tech platforms when we locate terrorist content on their services. As of September 2022, the TCAP alerts tech companies of content produced by 34 terrorist entities, in line with terrorist designation processes by democratic countries. To date, the TCAP has sent over 19,000 alerts, with 92% of the content now offline.
Question 2: Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services? **IMPORTANT: Under this question, we are not seeking links to or copies/screenshots of content that is illegal to hold, such as child sexual abuse. Deliberately viewing such images may be a criminal offence and will be reported to the police.**	Tech Against Terrorism research shows that terrorist and violent extremist use of the internet is increasingly concentrated on smaller platforms, who struggle to action terrorist and violent extremist content due to a relative lack of resources to understand and tackle the threat. Our analysis of TCAP data further shows that material produced by designated terrorist entities is particularly prevalent on small file-hosting / sharing, archiving, pasting, and video-sharing services which act as content stores. The increased exploitation of smaller and newer platforms by terrorist and violent extremist actors can be partly explained by the broad improvements in moderation of the most agregious terrorist material on the most mainstream and large platforms. The increasingly hostile environments terrorist actors face on mainstream platforms has forced a broad migration to smaller and newer platforms which typically lack the capacity or resources to effectively tackle the threat. Far-right terrorist and violent extremist actors in particular tend to congregate on platforms where they believe content is less likely to be moderated. • Terrorists and violent extremists actors also exploit smaller platforms as part of a multi- platform strategy meant to ensure an online presence that is as stable and wide-reaching as possible. Within this approach, multiple smaller

platforms are targeted simultaneously to host and archive content, which is then disseminated via aggregated links on beacon platforms, ensuring that the material remains available for as long as it takes the slowest platform to take it down.

- By disrupting terrorist and violent extremist use of smaller platforms on which content is hosted, we can effectively disrupt the entire terrorist ecosystem used to disseminate propaganda by targeting the source. You can read Tech Against Terrorism’s assessments of how smaller platforms are used for terrorist purposes here and here.
- Tech Against Terrorism is happy to share further analysis of TCAP data on the exploitation of smaller and newer platforms with Ofcom on request.
- Search engines represent a key element of terrorist and violent extremist use of the internet, notably for strategic purposes as they can provide a bridge between material and users. Rather than hosting TVE content, search engines risk facilitating the discovery or promotion of TVE networks and their material on the indexed web, including on Terrorist Operated Websites (TOWs), via search results. Most online counterterrorism efforts are aimed at disrupting the dissemination of terrorist content. However, less attention is paid to disrupting the visibility of material on search engine results. Cooperation between search-engines and user-to-user services is particularly important to ensure that moderation enforcement taken on a platform cannot be bypassed by attempting to

	search for / access the content via a search engine.
Question 3: How do you currently assess the risk of harm to individuals in the UK from illegal content presented by your service?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 4: What are your governance, accountability and decision-making structures for user and platform safety?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 5: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?	*Is this response confidential? – Y / N (delete as appropriate)* • Tech Against Terrorism, via its Policy Advisory and Response work, advises 40+ tech companies on counterterrorism and related content moderation policies. We do so directly via our Mentorship programme, and indirectly via our Knowledge Sharing Platform which is accessible to all platforms and free to use and hosts all of our key policy recommendations. • Our key policy recommendations for tech companies on online counterterrorism include: ○ Including an explicit prohibition of terrorism (or violent extremism) in content standards. This prohibition should be inclusive, covering for the different services offered by the platform, and based in the rule of law by explicitly referring to national

🟥	and international designation lists or other applicable counterterrorism law and online regulation. Designation of terrorist groups can be a useful tool to guide tech companies in assessing what material to remove, since it removes a layer of uncertainty for tech platforms in assessing what material should be considered terrorist. ○ Preferring the Community Guidelines or Acceptable Use Policy to detail what behaviour and content are prohibited on the service. Community Guidelines are often less legal in terms and more accessible to the users. ○ Using a language in the Community Guidelines adapted to the user-base principally used by young users. ○ Detailing the prohibitions listed in the Community Guidelines, including by explaining what content the prohibitions apply to (e.g., content, comments, usernames, profile pictures), and in detailing what each specific prohibition covers (for instance are all references to terrorist activities covered, including imagery and symbols, and are there exception for journalistic purposes?). Examples should be included as much as possible. ○ Providing a clear explanation of what is covered by a specific prohibition, referring to the legal basis for actioning specific

types of content whenever possible.

  - Laying out what actions can be taken by the platform in response to a violation and how users can report violating content. In doing so, the Community Guidelines should clearly lay out what users can expect of the platform to keep them safe, as well as what is expected of users to maintain a safe online environment.
  - Publishing regular blogposts that detail the platform's approach to content moderation and counterterrorism, including on the evolution of moderation policies and enforcement practices, how moderation is enforced, or on the ratio of human vs. automated moderation.
  - Publishing regular transparency reports that include information about policy and enforcement evolution over the coverage period beyond simple metrics. For instance, detailing a policy change, or explaining why the numbers are significantly different from the previous reporting period.
  - Publishing detailed blogposts and reports on compliance with specific regulation that impacts how moderation is conducted on the service.

- Tech Against Terrorism is happy to provide Ofcom with its complete list of policy recommendations if requested.

Question 6: How do your terms of service or public policy statements treat illegal content? How are these terms of service maintained and how much resource is dedicated to this?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 7: What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?	*Is this response confidential? – Y / N (delete as appropriate)* • Tech Against Terrorism's Policy Advisory and Response team advises tech platforms on transparency and accountability on online counterterrorism efforts. Core to this are our Transparency Reporting Guidelines. We also advise tech companies on general transparency and accountability beyond moderation enforcement. • Our key transparency and accountability recommendations for tech platforms include: ○ Having a page dedicated to user reporting and user appeal in the help and/or safety centre, which details a step-by-step process on how to report content that is in violation of the Content Standards, and on how to submit an appeal if one believes that their content or account was erroneously actioned ○ Having on-platform reporting features, in addition to a dedicated email address or online form to report content. If the user reporting function is only accessible to logged in users, another public reporting function should be made available to all users ○ Ensuring that user reporting features should offer users the

	possibility to select a precise reason for reporting, aligned with the prohibitions listed in the community guidelines. This should include a dedicated category for terrorist and for violent extremist content. A dedicated reporting category helps prioritising and segmenting reviews of user reports by the Trust & Safety team, ensuring that terrorist content is dealt with swiftly ○ Ensuring that user reporting features are available for different types of content on a platform (including comments and user profiles) ○ Publishing a detailed explanation of the moderation workflow, detailing what happens once a user has submitted a report, and the key elements considered by the platform when reviewing the report ○ Including sections on user reporting and user appeal in their transparency report ○ Notifying users when acting on their content or account and including a detailed explanation of the prohibition violated to improve users understanding of what is acceptable on the service. Platforms should also notify users of the results of their report and explain why the content was not found in violation of the Community Guidelines if applicable ○ Sending users a detailed explanation of the decision-

	process when responding to an appeal
Question 8: If your service has *reporting or flagging* mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 9: If your service has a *complaints* mechanism in place, how are these processes designed and maintained?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 10: What action does your service take in response to *reports* or *complaints*?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 11: Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?	*Is this response confidential? – Y / N (delete as appropriate)* • Tech Against Terrorism assesses moderating content at scale and adjudicating on what constitutes terrorist and violent extremist content to be the main two challenges faces by tech platforms when moderating content on their services whilst ensuring that human rights are safeguarded, in particular the rights to freedom of expression, access to information, and political opinion. • To ensure that content moderation adequately balances between user safety and user activity, Tech Against

Terrorism recommends significant human oversight over moderation enforcement. Both over selected moderation cases to ensure adequate context understanding, and over the general moderation process to prevent systematic infringement on user activity and rights. An adequate balance between human and automated moderation is thus crucial to limit content moderation’s impact on user rights.

- Related to the need for regular oversight, automated content moderation solutions should be developed with input from subject matter experts and with a human rights impact assessment at its core. They should also be developed with sufficient transparency and accountability. If solutions are developed without these, they risk being inaccurate and have adverse impact on human rights.
- Platforms should ensure easy to use appeal processes. User appeals are a key safeguard for freedom of speech online and increase accountability towards users by ensuring the possibility to contest moderation.
- Platforms should also consider using content moderation tactics other than content removal for content that is not strictly illegal (“grey area content”, or content that is produced by a violent extremist organisation that has not been designated as a terrorist organisation by democratic governments. This allows for platforms (particularly small or micro platforms) to deincentivise harmful content whilst limiting impact on freedom of speech. These can include hiding content, disengagement, educational or comms-

	based tactics, and community empowerment. • Before considering moderation tools that risk infringing upon freedom of expression, such as pre-upload filter or systematic monitoring of private online discussions, platforms should improve mitigation strategies to counter content moderation evasion techniques used by terrorists and violent extremist actors. Including by strengthening key texts and imagery techniques to prevent the use of alteration techniques to bypass content moderation. • Given the flexibility of terrorists and violent extremist actors to use a wide range of online platforms to disseminate content in a fast and resilient manner, strengthened cross-platforms collaboration is necessary to effectively tackle terrorist use of the internet and to ensure the effectiveness of the moderation measures taken by each platform. This is also needed to mitigate against the effect of platform migration (when terrorist and violent extremist actors migrate en masse to smaller and newer smaller platforms following their deplatforming from larger platforms).
Question 12: What automated moderation systems do you have in place around illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 13: How do you use human moderators to identify and assess illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*

Question 14: How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?	*Is this response confidential? – Y / N (delete as appropriate)* • Tech Against Terrorism finds that there is generally a lack of transparency about which moderation enforcement actions is applied to which violation, and whether a strike or any other form of proportionality process is applied on the service. • There is also a lack of information shared regarding how platforms are preventing banned users from re-accessing the service.
Question 15: In what instances is illegal content removed from your service?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 16: Do you use other tools to reduce the visibility and impact of illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 17: What other sanctions or disincentives do you employ against users who post illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*

Question 18: Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed more widely by industry?	*Is this response confidential? – Y / N (delete as appropriate)* • User reporting is the easiest and least privacy-intrusive solution to moderate online services. • Metadata analysis also offers a viable and less privacy-intrusive alternative to identify terrorist use of online services. Metadata analysis can be particularly useful to detect terrorist use of private online spaces without infringing on users' privacy as it does not rely on accessing the content itself. • Crisis response protocols to prevent the livestream of or manifestos related to a terrorist or violent extremist attack from rapidly circulating. Tech Against Terrorism particularly advocate for smaller platforms to be involved in crisis protocols to prevent their service to be exploited to bypass the action taken by larger tech platforms with the resources to constantly monitor their services and rapidly remove content. • Hashing technology can also help significantly to disrupt terrorist use of the internet by allowing tech companies to preemptively ban verified terrorist content before it is viewed by any user. To support smaller tech platforms in benefiting from hashing technology, Tech Against Terrorism has begun hashing URLs submitted to the TCAP and sharing them with the GIFCT's hash-sharing consortium. You can find more about hashing of TCAP URLs here.

Question 19: To what extent does your service encompass functionalities or features designed to mitigate the risk or impact of harm from illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 20: How do you support the safety and wellbeing of your users as regards illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 21: How do you mitigate any risks posed by the design of algorithms that support the function of your service (e.g. search engines, or social and content recommender systems), with reference to illegal content specifically?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 22: What age assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 23: Can you identify factors which might indicate that a service is likely to attract child users?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 24: Does your service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 25: If it is not possible for children to access your service, or a part of it, how do you ensure this?	*Is this response confidential? – Y / N (delete as appropriate)*

Question 26: What information do you have about the age of your users?	*Is this response confidential? – Y / N (delete as appropriate)*
Question 27: For purposes of transparency, what type of information is useful/not useful? Why?	*Is this response confidential? – Y / N (delete as appropriate)* • Tech Against Terrorism recommends all tech companies and governments to follow our [Guidelines on Transparency Reporting on Online Counterterrorism Efforts](). These Guidelines seek to improve transparency and accountability from governments and tech companies around online counterterrorism activities. The Guidelines serve as a starting point for increased transparency, and it is our aim that all governments and companies will report on the baseline set out in the Guidelines. • In general, we recommend tech companies to include the following information, beyond metrics, in their transparency reports: ○ Introduction to content moderation on the services and overview of the Community Guidelines, including a link. ○ Platform's explicit commitment to human rights and freedom of expression when conducting content moderation ○ Total number of content removed, or otherwise actioned. ○ Metrics on user reports that did not lead to content being

	removed, or otherwise actioned ○ Metrics on user appeal ○ For each category of violating content: include a “mini-report” covering the metrics mentioned above ○ Overall yearly or quarterly comparison ○ “Behind the numbers” explanation, detailing why some numbers might low or others high, and the general context for content moderation. For more information, see Tech Against Terrorism’s recommendations for transparency reporting.
Question 28: Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?	*Is this response confidential? – Y / N (delete as appropriate)*

Please complete this form in full and return to OS-CFE@ofcom.org.uk