Your response

Section 1.01 Volume 2: The causes and impacts of online harm

(a) Ofcom's Register of Risks

Question 1:	
i)	Do you have any comments on Ofcom's assessment of the causes and impacts of online harms?
Response:	
ii)	Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 2:	
i)	Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer.

Response:

The following responses will cover the spread of terrorist content on user to user (U2U) and search services, based on Tech Against Terrorism's expertise. You can access all our reports via our Public Resources page on our Knowledge Sharing Platform.

Generative artificial intelligence (AI)

1. Terrorist actors are resilient, and their frequent tactical adaptations can involve a range of emerging technologies. Ofcom should acknowledge the risk posed by terrorist exploitation of generative AI in its register of risk.
2. Tech Against Terrorism has identified users exploiting generative AI tools to bolster the creation and dissemination of AI-generated propaganda in support of both violent Islamist and neo-Nazi ideologies.[1]
3. **[CONFIDENTIAL✂]**
4. **[CONFIDENTIAL✂]**

[1] Tech Against Terrorism, Early Terrorist Adoption of Generative AI (2023).

5. **[CONFIDENTIAL✂]**
6. Whilst we have found little evidence that generative AI services are being systematically exploited by terrorists - and the engagement with generative AI is likely to be in the experimental phase - these experiments do indicate an emerging threat of terrorist exploitation of generative AI, in the medium to long term; as such, this risk should be reflected in Ofcom's risk registry.

File-sharing services

7. Ofcom rightly emphasises that file-sharing services are amongst the most exploited tech platform types for terrorist purposes. File-sharing services act as content stores - hosting text, audio, images, and video - and are used as online libraries of content. Terrorists rely on these to ensure that their content maintains a stable presence online and often place outlinks to these sites on larger platforms with a wider reach (beacon platforms).
8. However, when discussing the risks posed by terrorist exploitation of file-sharing services, Ofcom could further explore the unique ways in which these platforms are used in a crisis.
9. Tech Against Terrorism has identified an increased reliance on file-sharing platforms in the immediate aftermath of crisis events.[2] In particular, Tech Against Terrorism found that the use of outlinks from larger beacon platforms to smaller file-sharing sites to disseminate the Bratislava attack perpetrator's manifesto reflected a new typology of behaviour within a subset of far-right terrorist networks online: namely, lone-actor attackers who seek to exploit file-sharing platforms to host their most operationally important material, including manifestos and livestreams.[3]
10. As demonstrated by the Bratislava attack, the perpetrators of far-right terrorist attacks and their supporter networks increasingly rely on small file-sharing platforms in anticipation of being moderated by larger social media and messaging platforms. By exploiting small file-sharing platforms, these actors are likely to be attempting to ensure the longevity of crisis-related content online by targeting a diverse range of platforms with limited moderation abilities. It is likely that future attack perpetrators will learn from what has worked in past attacks and will similarly adopt a multi-platform dissemination strategy to counteract expected content moderation.

Archiving and paste platforms

11. When identifying types of services that can be used to commit or facilitate offences related to terrorism, Ofcom should also include archiving platforms and paste platforms.
12. Archiving platforms enable the storage of public information from defunct webpages or documents for anyone to view publicly. Paste sites enable the uploading of text online and are often utilised for sharing source code.
13. In 2023, Tech Against Terrorism found archiving platforms to be the second most heavily exploited platform type amongst those alerted via the Terrorist Content Analytics Platform

[2] Tech Against Terrorism, State of Play (2022).

[3] Tech Against Terrorism, Far-Right Lone-Actor Terrorist Attacks and Violent Extremist Use of File-Sharing Platforms (2023).

(TCAP).[4] Of all platform types, only archiving platforms were exploited to a significant extent by terrorist actors across the ideological spectrum, including both Islamist and far-right actors.

14. Similarly, Tech Against Terrorism found a high concentration of terrorist content on a small number of paste sites.
15. Archiving and paste platforms are likely to be popular with terrorist actors due to their multifunctional nature. Archiving sites are used to aggregate outlinks to content stores, as well as to act as content stores in their own right to maintain a stable presence for terrorist content online. It is likely that the purpose of terrorist actors in exploiting archiving sites to function as content preservers is to host terrorist content for indefinite periods of time.
16. Similarly, paste sites are also used to store content and to aggregate information, such as lists of URLs which link to further content and are not immediately identifiable as necessarily terrorist in nature.
17. Given the potential utility of these platforms to terrorist actors, and the high concentration of terrorist content found on these platforms, Ofcom should consider categorising archiving and paste platforms as types of services used to commit or facilitate offences related to terrorism.

Gaming platforms

18. When considering the risks posed by the exploitation of gaming services, Ofcom should also emphasise that these platforms are becoming increasingly important to far-right terrorist networks as a means of providing ideological support to attack perpetrators.
19. Of particular concern, gaming and gaming-adjacent platforms are increasingly being used by terrorist supporters to recreate gamified versions of terrorist attacks.
20. In 2022, Tech Against Terrorism identified at least 40 computer-generated versions of offline attacks on one platform, including, but not limited to, the attacks in Christchurch (2019), Buffalo (2022), Oslo and Utøya (2011), Bataclan, Paris (2015), and the Nairobi Westgate Mall attack (2013).[5] The computer-generated versions of offline attacks all allowed users to take part in a simulated game and most of them allowed users to play as the original attack perpetrators. Our research shows that the Buffalo (2022) and Christchurch (2019) attacks are those most frequently recreated on gaming platforms. Both these attacks were themselves filmed using helmet-mounted cameras in deliberate emulation of popular first-person-shooter games.

Platform features

21. When identifying those platform features that may pose a greater risk of terrorist exploitation, Ofcom could also include lack of registration, on-platform search, and channels with large or unlimited participants as high-risk features.
22. Platforms with minimal sign-up requirements are likely to attract terrorist actors because the online personal security of such actors is increased when their identities are harder to trace.
23. Platforms that are easy to navigate are typically preferred by all users, including terrorist actors. The availability of on-platform search functions makes a platform simple to navigate and content, groups, users, and other information easy to find. With on-platform search features, terrorist actors can more easily find terrorist content utilising keyword and phrase searches.

[4] Tech Against Terrorism, TCAP Insights: Patterns of Online Terrorist Exploitation (2023).

[5] Tech Against Terrorism, State of Play (2022).

24. Leadership and organisational messages from terrorist entities are frequently posted through a one-way messaging channel which does not allow users to directly respond to messages, or post in the channel, and which have a very large or unlimited audience. Channels like this are highly advantageous to terrorist entities as they allow complete control over the content in a channel and an uncomplicated means of providing clear direction to other platforms or channels. It is likely that terrorist entities are drawn to platforms that enable this type of one-way communication. Terrorist entities will almost certainly continue to exploit these channels to redirect users elsewhere throughout the short- to longer-term future, both to circumvent moderation and to ensure content stability.

Search engines

25. Deliberate terrorist exploitation of search engines is low. However, search engines can be used to circumvent content moderation efforts undertaken elsewhere online and can contribute to the longevity and discoverability of terrorist content.
26. Search engines can greatly increase the reach of terrorist content, primarily by redirecting users to terrorist-operated websites (TOWs), and thus pose a risk of harm to the public. TOWs and other pages hosting terrorist content are often indexed in mainstream search engine results, allowing them to be accessible and easily discoverable for vulnerable individuals who are seeking out the content.[6]

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Yes: paragraphs 3-5 above.

Section 1.02 Volume 3: How should services assess the risk of online harms?

(a) Governance and accountability

Question 3:

i) Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice?

Response:

ii) Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

Response:

[6] Tech Against Terrorism, State of Play (2022).

iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 4:	
i)	Do you agree with the types of services that we propose the governance and accountability measures should apply to?
Response:	
ii)	Please explain your answer.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 5:	
i)	Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 6:	
i)	Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to tie remuneration for senior managers to positive online safety outcomes?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(b) Service’s risk assessment

Question 7:	

i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Specifically, we would also appreciate evidence from regulated services on the following:

Question 8:	
i)	Do you think the four-step risk assessment process and the Risk Profiles are useful models to help services navigate and comply with their wider obligations under the Act?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 9:	
i)	Are the Risk Profiles sufficiently clear?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Do you think the information provided on risk factors will help you understand the risks on your service?
Response:	
iv)	Please provide the underlying arguments and evidence that support your views.
Response:	
v)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(c) Record keeping and review guidance

Question 10:	
i)	Do you have any comments on our draft record keeping and review guidance?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 11:	
i)	Do you agree with our proposal not to exercise our power to exempt specified descriptions of services from the record keeping and review duty for the moment?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Section 1.03 Volume 4: What should services do to mitigate the risk of online harms

(a) Our approach to the Illegal content Codes of Practice

Question 12:	
i)	Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 13:

i)	Do you agree that in general we should apply the most onerous measures in our Codes only to services which are large and/or medium or high risk?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 14:	
i)	Do you agree with our definition of large services?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 15:	
i)	Do you agree with our definition of multi-risk services?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 16:	
i)	Do you have any comments on the draft Codes of Practice themselves?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 17:	

i) Do you have any comments on the costs assumptions set out in Annex 14, which we used for calculating the costs of various measures?

Response:

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

(b) Content moderation (User to User)

Question 18:

i) Do you agree with our proposals?

Response: The below response provides our feedback on the proposed mitigations regarding accounts operated by or on behalf of proscribed organisations, as per Ofcom's request for further evidence in the consultation documents.

ii) Please provide the underlying arguments and evidence that support your views.

Response:

Proposed mitigations regarding accounts operated by or on behalf of proscribed organisations

1. Ofcom's recommendation that U2U services should employ a strikes and blocking system against users where they are found to have posted or shared illegal content, or committed or facilitated illegal behaviour, is a proportionate mitigation and well-evidenced as a deterrent of such behaviour.
2. We also agree with the assessment that there is insufficient evidence to include broad measures in the proposed Codes that would recommend that services **immediately block** accounts which post any illegal content. Beyond the technical and rights-based challenges highlighted, more research is needed on the unintended effects of 'deplatforming' extremist actors and on measuring overall harm reduction. These risks include migration to unmonitored spaces, radicalising effects on communities, and driving innovative moderation evasion.
3. We make the following observations concerning Ofcom's recommendation that U2U services should immediately remove accounts run by or on behalf of proscribed organisations.

   - Tech Against Terrorism agrees this is a proportionate mitigation given the high likelihood that content generated by these accounts would amount to a priority illegal offence.
   - However, there is an inconsistency between this (sound) logic and its practical application as is recognised by Ofcom in the drafting. Namely, it is often impossible for service providers to verify whether an account is run **by** or **on behalf of** a proscribed organisation, and therefore the stated justification for removal cannot be met to any degree of confidence.
   - In practice, the guidance provided for service providers to support assessments of whether an account is operated by or on behalf of a proscribed groups, including considering the username, profile images and information, is insufficient.

• Based on our experience of monitoring online terrorist content, there are relatively few 'official' accounts that are verifiably run by proscribed organisations, and they tend to be on stable platforms resistant to takedown such as terrorist-operated websites or servers, or on encrypted messaging apps. In most cases, accounts professing to be official or that include symbols or terminology affiliated with a proscribed group are **likely** to be run by supporters of that organisation. • In addition, terrorist actors and their supporter networks will often use content moderation avoidance techniques, especially on larger platforms, to avoid the suggestion or mask the fact of any affiliation with proscribed groups. • Therefore, service providers are unlikely to have the necessary tools to distinguish between 'official' and 'supporter-run' accounts, with smaller platforms even more disadvantaged in their capacity to make such judgements. **Ofcom should consider reducing this burden on service providers by reformulating the "reasonable grounds" threshold for inferring an account's affiliation with a proscribed organisation.** • Ofcom should consider expanding their interpretation of being "run by or on behalf of" to include accounts that **profess to be a** proscribed organisation either through explicit statement or through any use of official symbology or terminology. This can be justified given affiliation declared in this way, although unverified, will nonetheless benefit a proscribed organisation if the account is run by supporters rather than members.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No

(c)

(d) Content moderation (Search)

Question 19:
i) Do you agree with our proposals?
Response:
ii) Please provide the underlying arguments and evidence that support your views.
Response:
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

(e) Automated content moderation (User to User)

Question 20:
i) Do you agree with our proposals?

Response: The below response provides arguments for and suggestions on how to use URL detection and key word detection for terrorism content, as per Ofcom's request for further evidence in the consultation documents.

ii) Please provide the underlying arguments and evidence that support your views.

Response:

URL detection

1. URL detection for terrorist content should be focused on the following platform types:
   - Beacons: platforms used by terrorists to project their content to the widest audience possible. Beacons act as both a centrally located lighthouse and signpost to where the content can be found. Through beacons, terrorists redirect their target audience to the platforms on which content is hosted. These platforms include social media, messaging platforms, gaming platforms, and video-sharing platforms.
   - Aggregators: centralised databases of where content can be found online. These collate a wide range of URLs to content-hosting platforms. They thereby facilitate the diffusion of content. Aggregators include paste sites, social media, and forums.

Terrorist-operated websites (TOWs)

2. URL detection should also focus on disrupting outlinks that lead users to TOWs, which pose one of the most significant threats to the global effort to tackle terrorist use of the internet.
3. These websites are used to disseminate and archive propaganda material, as well as to recruit and communicate internally, and are easily accessible on the surface web. Since TOWs are difficult to remove at the infrastructural level, disrupting the dissemination of outlinks to TOWs could aid in mitigating the threat these websites pose.

URL shorteners and content moderation evasion

4. Any consideration of URL detection to tackle the harm created by the dissemination of links should include measures to mitigate the risk of utilising URL shorteners to bypass the detection and moderation of identifiably problematic URLs.
5. For instance, if a specific URL has been identified and blocked by a platform for directing users to terrorist content, a modified URL obtained via a URL shortener service can help bypass these blocks.
6. This strategy can be countered by platforms blocking online domains for terrorist-operated websites entirely, rather than blocking specific URLs to individual pieces of content.

Tech Against Terrorism resources

7. Tech Against Terrorism's Knowledge Sharing Platform contains resources that platforms could utilise for the URL detection of terrorism content, namely:

- The Terrorist-Operated Websites Database. This database is designed to support the regular monitoring and disruption of terrorist-operated websites (TOWs), and covers violent Islamist and far-right terrorist websites.[7]
- The Beacon Channels Tracker: This tracker is designed to support the regular monitoring and disruption of terrorist activity online. The database contains links leading to verified, active accounts and channels that are used by terrorist entities to reach a wide audience – we consider these Beacon Channels. All groups and entities included in the tracker have been designated as terrorist and are included within our Terrorist Content Analytics Platform inclusion policy. The tracker can be filtered by platform name and group/entity affiliation. It also contains a brief description with information about the type of channel or content to which it leads. The tracker is updated regularly by Tech Against Terrorism's Open-Source Intelligence (OSINT) Team.

Further recommendations

8. As discussed in a recent Resolve Network research report,[8] pro-IS users are likely to operate on the assumption that the outlinks they share to TVE content will be deactivated. Therefore, their strategy depends on volume and speed with considerable reliance on the use of automation for rapid generation and dissemination of banks of URLs (such as through Telegram bots).
9. Platforms could, therefore, make use of behaviour-based cues, such as abnormal posting volume, which can be picked up more easily by automated systems. Similarly, platforms could also make use of moderation mechanisms already in place for spam behaviour to disrupt terrorist use of their services.[9]

Keyword detection

Tech Against Terrorism resources

10. Tech Against Terrorism's Knowledge-Sharing Platform contains resources that platforms could use to augment keyword detection to counter terrorism content, namely the Terminology Database. The database is designed to support the monitoring and disruption of terrorist activity online. It contains keywords, terms, and phrases relating to terrorist groups and identities that have been identified in the context of OSINT research.

Safeguarding human rights

11. An important caveat for the use of keyword detection for terrorism content is that – whilst terms may be observed in the context of terrorism – they may also have relevance to non-extremist ideologies.
12. Automated moderation can heighten the risk of interference with human rights, and in particular with freedom of expression, because it lacks human context and can amplify mistakes or bias in the data which informs it.

[7] Some of the websites are operated by or in support of entities that are not designated, but warrant our monitoring and analysis as violent extremist entities.

[8] Stuart Macdonald, Connor Rees, and Joost S. (2022), Remove, Impede, Disrupt, Redirect: Understanding & Combating Pro-Islamic State Use of File-Sharing Platforms, Resolve Network.

[9] Tech Against Terrorism, TCAP Insights: Patterns of Online Terrorist Exploitation (2023).

13. To promote the safeguarding of human rights when recommending automated moderation measures, particularly key word detection, Ofcom could:
    - Encourage companies to ensure a balance between human and automated moderation to ensure the importing of necessary context and nuance into decision making.
    - Encourage companies not to remove content or accounts on the basis of one key word alone, but rather to also consider other key terms used or other metrics of identification, such as key images or symbols used (see below).

Image and symbol detection

14. To complement keyword detection, Ofcom could also similarly consider recommending image and symbol detection for terrorism content.
15. Tech Against Terrorism's Knowledge-Sharing Platform contains an Image Compendium that platforms could use to detect images and symbols relating to terrorism. The Image Compendium is designed to support the monitoring and disruption of terrorist activity online and contains images associated with designated terrorist groups, specifically: group-affiliating logos and posters (including tattoos and flags), official media outlet 'branding', and meme content.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response: No

Question 21:

i) Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated 'publicly' or 'privately'?

Response:

ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Do you have any relevant evidence on:

Question 22:

i) Accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services;

Response:

ii) Please provide the underlying arguments and evidence that support your views.

Response:

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Question 23:

i)	Ability of services in scope of the CSAM hash matching measure to access hash databases/services, with respect to access criteria or requirements set by database and/or hash matching service providers;
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 24:	
i)	Costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of fuzzy matching for CSAM URL detection;;
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 25:	
i)	Costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services;
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 26:	
i)	An effective application of hash matching and/or URL detection for terrorism content, including how such measures could address concerns around ‘context’ and freedom of expression, and any information you have on the costs and efficacy of applying hash matching and URL detection for terrorism content to a range of services.
Response:	

The below response summarises the utility and efficacy of hash-matching for terrorism content addressing concerns around 'context' and freedom of expression, as per Ofcom's request for further evidence in the consultation documents.

ii) Please provide the underlying arguments and evidence that support your views.

Response:

Consideration of hash-matching for terrorism content

1. Hashes are unique digital identifiers assigned to individual pieces of content. These are subsequently shared to recipients who can input hashes into automated detection mechanisms so that they can be matched against content found on their platforms. Hashes can be stored in databases or tables designed to easily look up the original content using the hash value. Hashes are designed to be irreversible. Hash sharing allows for the creation of large-scale datasets, providing value not only to content moderators but also to academics and researchers for future analysis. We assess below the advantages and disadvantages of sharing and matching hashes.

2. Advantages
   - Hash-sharing provides access to large-scale databases of hashed terrorist content.
   - Hash-sharing allows for enhancement of automated detection mechanisms, as platforms can train algorithms to match hashed content with those in hash-sharing databases.

3. Disadvantages
   - To fully exploit the value of hash-sharing databases, tech platforms need access to them. Tech platforms currently sign up on a voluntary basis, which limits the success of hash-sharing to those that are willing to use them.
   - Tech platforms require the capability to hash content that reflects the cryptographic methods used by the hash-sharing consortiums.
   - The process requires consistent maintenance of databases, including adding new hashes, which can take time; tech platforms are dependent on those adding hashed content to the databases.
   - Hashed content does not always cover edited versions of content. Since a hash is unique to an individual file, the slightest change to a piece of content may present as a minor iteration of the original but would theoretically generate a new hash. Edited or new versions of terrorist content will, therefore, need to be hashed.
   - Hash-sharing databases can be criticised for lack of transparency, particularly if access to hash sharing consortiums is tightly regulated.

In response to Ofcom's call for further evidence on specific areas relating to hash-matching:

The accuracy and effectiveness of hashing solutions for terrorism content

4. The accuracy and effectiveness of hashing solutions for terrorism content varies greatly and is dependent on multiple factors including the type of hashing protocol used and the type of content.

5. Basic cryptographic hashing protocols, such as MD5, require a file (such as a propaganda image) to match exactly against the hashed file. Even the smallest edit, crop or colourisation will change the MD5 hash value and prevent a match. This ensures high accuracy and prevents any false negatives but limits effectiveness at catching any edited content (low 'recall').
6. Locality-sensitive hashing (perceptual), such as TMK or PDQ, is more sophisticated and enables the identification of roughly similar content even if not identically accurate (higher recall). This increases the chance of false positives but improves the effectiveness of hash-matching, enabling it to capture some edited versions of terrorist content. However, these algorithms are typically more resource-intensive to produce, store, and use than cryptographic hashes.[10]
7. Video presents significant technical challenges for hash-matching. Any edits to the video, such as shortening the length or overlaying with different effects, will reduce the reliability of perceptual hashing. This is especially problematic in relation to terrorist livestreams, which are often shortened or edited.
8. Additionally, as online actors seeking to disseminate terrorist propaganda begin adopting advanced generative AI tools to produce new content and edit existing content, the opportunities for hash-matching against known material are likely to correspondingly decrease.

The extent to which a hashing solution can identify terrorism content accurately in different contexts

9. There is certainly the risk that non-terrorist content could be captured through hash-matching due to it being so perceptually similar to terrorist content that it gets wrongly removed.
10. Tech Against Terrorism is not in a position to assess the accuracy of existing hash-matching solutions given we do not have access to hash-sharing databases or tech company hashing systems. More information on GIFCT's hash-sharing database, including how they measure their effectiveness, can be found here.
11. However, there is evidence that hash-matching offers an effective solution for moderating terrorist content in many contexts, particularly where the content is stable and unlikely to be changed as it proliferates. This is especially so when used in tandem with access to a database of known terrorist material, such as that currently being built as part of the Terrorist Content Analytics Platform (TCAP).
12. For the reasons stated above, it is likely that hash-matching for terrorism content is far more accurate and effective for PDFs, text, and images than for video and audio content.
13. However, there is no particular distinction between terrorism material and CSAM in terms of the difficulties and variations between or within types of files (images/videos/documents) or the relevant hashing protocol. No hash-matching solution can ever be fully accurate.

The degree of human oversight necessary to ensure the technology is sufficiently accurate in identifying terrorism content

Hash-matching ambitions for the Terrorist Content Analytics Platform (TCAP) Archive

14. **[CONFIDENTIAL✂]**

[10] Tom Thorley, Advances in Hashing for Counterterrorism (2023)

GIFCT's Hash-Sharing Database (HSDB)

15. **[CONFIDENTIAL✂]**
16. **[CONFIDENTIAL✂]**
17. **[CONFIDENTIAL✂]**
18. **[CONFIDENTIAL✂]**
19. **[CONFIDENTIAL✂]**

The potential costs associated with hash matching for terrorism content

20. **[CONFIDENTIAL✂]**
21. **[CONFIDENTIAL✂]**

iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: Yes: paragraphs 14-21.	

(f) Automated content moderation (Search)

Question 27:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(g) User reporting and complaints (U2U and search)

Question 28:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(h)

(i) Terms of service and Publicly Available Statements

Question 29:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 30:	
i)	Do you have any evidence, in particular on the use of prompts, to guide further work in this area?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(j) Default settings and user support for child users (U2U)

Question 31:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 32:	
i)	Are there functionalities outside of the ones listed in our proposals, that should explicitly inform users around changing default settings?

Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 33:	
i)	Are there other points within the user journey where under 18s should be informed of the risk of illegal content?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(k) Recommender system testing (U2U)

Question 34:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 35:	
i)	What evaluation methods might be suitable for smaller services that do not have the capacity to perform on-platform testing?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

We are aware of design features and parameters that can be used in recommender system to minimise the distribution of illegal content, e.g. ensuring content/network balance and low/neutral weightings on content labelled as sensitive.

Question 36:	
i)	Are you aware of any other design parameters and choices that are proven to improve user safety?

Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(l) Enhanced user control (U2U)

Question 37:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 38:	
i)	Do you think the first two proposed measures should include requirements for how these controls are made known to users?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 39:	
i)	Do you think there are situations where the labelling of accounts through voluntary verification schemes has particular value or risks?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(m) User access to services (U2U)

Question 40:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.

Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Do you have any supporting information and evidence to inform any recommendations we may make on blocking sharers of CSAM content? Specifically:

Question 41:	
i)	What are the options available to block and prevent a user from returning to a service (e.g. blocking by username, email or IP address, or a combination of factors)?
Response:	
ii)	What are the advantages and disadvantages of the different options, including any potential impact on other users?
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 42:	
i)	How long should a user be blocked for sharing known CSAM, and should the period vary depending on the nature of the offence committed?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

There is a risk that lawful content is erroneously classified as CSAM by automated systems, which may impact on the rights of law-abiding users.

Question 43:	
i)	What steps can services take to manage this risk? For example, are there alternative options to immediate blocking (such as a strikes system) that might help mitigate some of the risks and impacts on user rights?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(n)

(o) Service design and user support (Search)

Question 44:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(p) Cumulative Assessment

Question 45:	
i)	Do you agree that the overall burden of our measures on low risk small and micro businesses is proportionate?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 46:	
i)	Do you agree that the overall burden is proportionate for those small and micro businesses that find they have significant risks of illegal content and for whom we propose to recommend more measures?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 47:	

i)	We are applying more measures to large services. Do you agree that the overall burden on large services proportionate?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(q) Statutory Tests

Question 48:	
i)	Do you agree that Ofcom's proposed recommendations for the Codes are appropriate in the light of the matters to which Ofcom must have regard?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Section 1.04 Volume 5: How to judge whether content is illegal or not?

(a) The Illegal Content Judgements Guidance (ICJG)

Question 49:	
i)	Do you agree with our proposals, including the detail of the drafting?
Response: Please see the below response which is based on Tech Against Terrorism's practical experience of identifying and verifying terrorist content online, as well as the challenges of assessing the legality of that content. You can access all our reports via our Public Resources page on the Knowledge-Sharing Platform.	
ii)	What are the underlying arguments and evidence that inform your view?
Response: 1. The drafting of the Illegal Content Judgements Guidance (ICJG) is comprehensive and provides detailed guidance for tech platforms on how to interpret the priority terrorism offences set out	

in the Online Safety Act. The layout of the guidance is usefully structured to address the least subjective forms of terrorist content first.

2. The usage examples provided alongside each offence are critically important for interpreting the legislation in practical terms. These examples are likely to be especially useful for tech companies in relation to the more subjective offences. Further examples on different types of platforms would be beneficial.

Proscribed organisations offences

3. While perhaps self-evident, it would be useful to specify that **all** official propaganda produced by proscribed terrorist organisations will be priority illegal content. This would be based on the logic that official propaganda inherently "invites support for a proscribed organisation" even if the content itself does not explicitly demonstrate support for the group, for example a cooking recipe from an IS magazine. This clarification would simplify decisions for tech platform moderators who could refer to official branding and logos used by proscribed groups. To operationalise the removal of this illegal content, it is necessary for Ofcom to provide practical guidance to tech platforms on how official content can be identified. This might include a database of official media outlets affiliated with proscribed Islamist terrorist organisations such as Islamic State or the official branding used by proscribed far-right groups such as Atomwaffen Division.
4. Unofficial or 'supporter-generated' content relating to proscribed groups is more difficult to identify and assess. The guidance clarifies that content does not have to be officially produced, that is originating from a member or entity of a proscribed organisation, to amount to an offence. This approach is more practically applicable given that, in Tech Against Terrorism's experience, it is sometimes difficult to assess whether content is produced by a member of a proscribed organisation.
5. Therefore, our interpretation is that unofficial propaganda ***indicating support*** for a proscribed organisation, such as a poster with the proscribed organisation's logo, will be in scope. Beyond clear statements of support which are easier to assess, moderators will need to interpret what content meets the threshold of 'support' and therefore require guidance to understand media logos, symbols, and terminology connected with proscribed groups.

Instructional material

6. Material that provides instructions or training is more difficult to assess in relation to the intent and purpose of the material. The guidance for "information likely to be of use to a terrorist" highlights the importance of the terrorist purpose with the examples provided illuminating as to the more obvious content in scope such as the Anarchists Cookbook or Al-Qaeda's *Inspire* guides.
7. The "terrorist training offences" are more difficult to assess. According to the guidance, the burden is on the user to show that the purpose of the instructions or training is for a purpose other than terrorism. Based on our monitoring of instructional content, the purpose is often ambiguous and therefore would need to be explicitly stated by the user.
8. This burden on the user and ambiguity of ascertainable purpose may result in the removal of legal content for legitimate uses such as for hunting, legal military training, or scientific purposes. One illustrative example is the instructional guidance on making Molotov cocktails spreading online in the context of the Ukrainian defence against the Russian invasion.

9. In relation to 3D firearms, it is possible that blueprints available online for making these weapons could equally be used for purposes which are legitimate (such as hunting) or terrorist in nature (as in the case of Halle terrorist attack). Given the risk of use for malevolent purposes, the burden on the user to demonstrate legitimate intent is proportional.
10. Given that platform moderators are likely to lack expertise in instructional material, any practical guidance that Ofcom could provide in relation to specific instructional material would be useful.

Dissemination of terrorist publications

11. In simpler terms, a terrorist publication is one that directly or indirectly encourages acts of terrorism OR the main purpose of which is providing information that could be useful in the commission or preparation of terrorism acts. In regard to the latter, there is clear overlap with "information likely to be useful to a terrorist." While some guidance is shared as to what this type of content looks like in practice (such as glorification of terrorism or terrorists), there are likely to be many examples of grey area content requiring subjective interpretation.
12. An important distinction between the offence of disseminating 'terrorist publications' and the substance of other offences is the form of content it covers, namely that it must be a self-contained piece of content in the form of text, video, audio, or images rather than a user post for example. While the term 'publication' in this regard may be misleading, the guidance reduces the scope for subjective interpretation of the criteria to classify a piece of content as a 'terrorist publication.' *We provide additional suggestions on practical guidance in our response to question 26.2.*

Other terrorism offences

13. The other terrorism offences require greater subjectivity in the assessment of their constituent elements and stray away from Tech Against Terrorism's expertise in online propaganda.
14. In relation to the "encouraging terrorism offence," the definition is too broad and thus operationally impracticable at scale for tech platforms. The selected user examples, such as calling on others to emulate the acts of historical figures who used violence for political ends, also seem too broad and is likely to negatively impact freedom of expression given the tendency of platforms to over-remove content to avoid penalties.
15. The challenge in relation to the offence of 'inciting terrorism overseas' will be establishing the requisite intent given the defence of humour which is often employed by the violent far right.
16. One of Tech Against Terrorism's main strategic priorities is the disruption of terrorist-operated websites (TOWs). Of relevance to TOWs is the Section 5 offence concerning the preparation of terrorist acts. A website that appears to be run for and on behalf of a proscribed organisation falls under the definition of terrorism as it is an act that benefits that organisation. While the Online Safety Act does not regulate websites, it has powers over U2U services which host links to websites and search services that index websites. Ofcom should consider what mitigations might disrupt access to TOWs from U2U and search services including through removing those links to verified TOWs. Tech Against Terrorism maintains a database of the domains of verified terrorist-operated websites relating to proscribed terrorist organisations that could support the exercise of the statutory powers in question.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response: No

Question 50:

i) Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?

Response:

While the guidance is accessible, it still places a significant burden of responsibility on services and their moderators to make appropriate judgements. Find our full response below.

ii) Please provide the underlying arguments and evidence that support your views.

Response:

1. While the guidance is accessible, it still places a significant burden of responsibility on services and their moderators to make appropriate judgements. The guidance notes that "appropriately trained and culturally aware content moderators will need to be empowered to make sensible judgements having regard to the information they have." This may overlook the limited capacity of medium- and smaller-sized service providers to hire and train moderators to make those informed decisions. Ofcom should consider providing additional practical resources to aide and complement these guidelines and simplifying the prioritisation of content for platforms to target.
2. **[CONFIDENTIAL✂]**
3. In volume 5 of the ICJG, Ofcom recognises the need for suitable resources to guide tech companies in their assessments relating to membership or support for proscribed organisations. Tech Against Terrorism's Knowledge Sharing Platform (KSP) contains a comprehensive and up-to-date resource that includes an image compendium and terminology database relating to designated terrorist organisations. This resource would allow tech companies to make more informed assessments of whether a piece of content amounts to support for a proscribed organisation. For example, the image compendium contains the media logos for unofficial media outlets supportive of proscribed groups such as al-Shabaab while the terminology database includes phrases that indicate coded support to proscribed groups such as Atomwaffen Division. The combined visual and lexical resources currently cover 14 organisations proscribed by the UK, totalling over 160 images and 520 phrases.
4. **[CONFIDENTIAL✂]**
5. **[CONFIDENTIAL✂]**
6. **[CONFIDENTIAL✂]**

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

Response:

Yes: paragraphs 2, 4-6.

Question 51:

i)	What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Section 1.05

Section 1.06

Section 1.07 Volume 6: Information gathering and enforcement powers, and approach to supervision.

(a) Information powers

Question 52:	
i)	Do you have any comments on our proposed approach to information gathering powers under the Online Safety Act?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

(b) Enforcement powers

Question 53:	
i)	Do you have any comments on our draft Online Safety Enforcement Guidance?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Section 1.08

Section 1.09 Annex 13: Impact Assessments

Question 54:	
i)	Do you agree that our proposals as set out in Chapter 16 (reporting and complaints), and Chapter 10 and Annex 6 (record keeping) are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English?
Response:	
ii)	If you disagree, please explain why, including how you consider these proposals could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	