Your response

Volume 2: The causes and impacts of online harm

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?

Background

My response focuses on (the risk of) harms that is/are associated with terrorism offences, notably, encountering terrorism content online. I draw on original research that my colleagues and I conducted in the UK. Specifically, I refer to Schumann, Clemmow, Rottweiler, and Gill (2024), a paper that presents a representative survey study (N = 1495) that captured whether participants had actively sought or had been passively exposed to information (online and offline) that incited or facilitated the use of violence to pursue ideological goals as well as individuals' violent extremist attitudes and violent extremist behavioural intentions. The peer-reviewed paper is available (open-access) here: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0293810. The data that the paper refers to is accessible here: https://osf.io/jxsva/. Based on the insights gained in the aforementioned survey study, I submit five comments as a response to the consultation.

Comment 1 – Rates of encountering terrorism content online are likely higher than stated and should be revised.

6B.11 highlights that "3% of UK internet users aged 13 and above claimed they had experienced content that encouraged extremism, radicalisation or terrorism in the past four weeks". 6B.12 further notes that "the risk of coming across terrorism content is relatively small".

Based on our research, the rate of exposure to terrorism content online is likely higher than stated in the proposal document. In a slightly older sample (> 18 years) and assessing exposure over the lifetime, we showed that approximately 23% of participants displayed information use behaviour that reflected a moderate or high likelihood of actively seeking information that incited/facilitated terrorism. Additionally, approximately 31% of participants exhibited information use behaviour defined by a moderate or high likelihood of being incidentally (i.e., passively) exposed to content that promoted/facilitated terrorism. Examples of rates of specific online activities also point to higher base rates of encountering terrorism content online. For instance, 16.1% of participants indicated that they had ever (i.e., over the lifetime) searched for images or videos portraying violence to achieve political, religious, or social goals; 30% reported having received, without having asked for it, memes or videos that promoted terrorism.

The higher rates of both actively seeking and being incidentally exposed to terrorist content found in our research might be due to differences in the data collection procedure (i.e., we referred to lifetime exposure) as well as possibly lower effects of social desirability in a study that was not commissioned by a government agency. *Overall, I am of the opinion that the risk of harm*

associated with terrorism content online is currently underestimated and that the proposal document should be revised to reflect our evidence that points to higher exposure rates.

Reference

Schumann, S., Clemmow, C., Rottweiler, B., & Gill, P. (2024). Distinct patterns of incidental exposure to and active selection of radicalizing information indicate varying levels of support for violent extremism. *Plos one*, *19*(2), e0293810.

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

Comment 2 – The term 'encounter' should be specified to distinguish the risks of harm of active selection of and incidental exposure to terrorism content.

The proposal document references throughout that individuals 'encounter' illegal harms online. The term 'encounter', however, is too ambiguous and does not specify the underlying processes of the information use behaviour. More precisely, it is unclear whether the 'encounter' is due to individuals having actively sought the content (i.e., active selection) or because they were incidentally exposed, without having searched or asked for it. Drawing this distinction, in the context of harms related to terrorism offences but also with regards to other illegal harms, is crucial. The potential risk of harm as well as the type/degree of impact that are to be expected differ for individuals who are actively seeking terrorism content and those who are incidentally exposed.

Notably, following Slater's (2015) reinforcing spirals model, individuals who actively search for terrorism content online would do so because their beliefs are (at least to some degree) already aligned with the material and its source/author. In other words, actively seeking out terrorism content can serve as an indicator for a high(er) risk of radicalisation. Moreover, if those who actively seek terrorism content are successful and find the information, the respective materials will further strengthen their ideology, violent extremist attitudes and behavioural intentions. By contrast, and following the Political Incidental News Exposure Model (Matthes et al., 2020), incidental exposure to terrorism content could imply two types of risk of harms: individuals who don't agree with the material will experience negative affective responses (i.e., anger, frustration, fear); individuals who endorse the message, may be at risk of being enticed to actively pursue more similar material, which could, ultimately, drive their radicalisation.

Our research confirms the aforementioned argument and highlights that the term 'encounter' should be further specified in the proposal document to distinguish between the active selection of and incidental exposure to terrorism content. Moreover, the risks of harm of each type of information use behaviour should be outlined. We demonstrated (Schumann et al., 2024) that individuals who displayed an information use behaviour that reflects at least a moderate probability of actively seeking material that incited or facilitated terrorism reported stronger violent extremist attitudes and behavioural intentions than those with a near-zero likelihood of active selection (but perhaps a high likelihood of incidental exposure). In other words, the risk of harm of terrorism content (on radicalisation) is expected to be higher for individuals who actively look for the content rather than being only incidentally exposed.

I further recommend that this latter insight is also added to section 6B.16. Doing so will provide a more nuanced description of how (or rather when) terrorism content can enhance the risk of radicalisation.

Comment 3 – It should be acknowledged explicitly that the risk of encountering terrorism material is increased if individuals actively seek out such information.

6B.35 stipulates demographic characteristics that are associated with an increased risk of encountering "content that encouraged extremism, radicalisation or terrorism".

However, one of the strongest predictors of individuals encountering (more) terrorism content is whether they actively look for it. Specifically, in our research (Schumann et al., 2024), we showed that individuals whose information use behaviour displays even a moderate likelihood of active selection of terrorism material are also exhibiting a moderate (or high) likelihood of incidental exposure. That is, in addition to their active pursuit of terrorism content, those individuals are then also incidentally exposed to terrorism content when they are not looking for it. *I am, therefore, of the opinion that individuals' information use behaviour, that is, distinct patterns of the active selection of or incidental exposure to terrorism content, should be added to the description of user base risk factors.*

References

Matthes, J., Nanz, A., Stubenvoll, M., & Heiss, R. (2020). Processing news on social media. The political incidental news exposure model (PINE). Journalism, 21(8), 1031-1048.

Slater, M. D. (2015). Reinforcing spirals model: Conceptualizing the relationship between media content exposure and the development and maintenance of attitudes. Media psychology, 18(3), 370-395.

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

Response: no

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.

Comment 4 – U2U services risk profiles should include their search function.

As outlined in Comment 2, our research (Schumann et al., 2024) provides evidence that actively seeking for terrorism content is associated with a higher risk of radicalisation (i.e., stronger support for violent extremist attitudes and behavioural intentions). In turn, it can be concluded that the ability to search for such content enhances the risk of harms of terrorism materials as well as the severity of the potential impact associated with terrorism content online. U2U services do typically enable users to search for accounts or content based on specific keywords. *I propose, therefore, that the search function is added as a relevant functionality to the risk profile of U2U services.*

Comment 5 – Mitigating measures required from search services should be extended.

Moreover, I recommend that as it pertains to terrorism content search services should be required to take more conservative mitigating measures. As documented in our research (Schumann et al.,

2024), those who exhibited information use behaviour characterised by even a moderate likelihood of active selection of materials that incite/facilitate terrorism endorsed stronger violent extremist attitudes and behavioural intentions (i.e., this is a correlation not a causal effect). Reiterating a point made earlier, this finding suggests that individuals who actively seek out terrorism content are at a higher risk of radicalisation.

It is noted in the proposal document that "Services should provide crisis prevention information in response to search requests that contain general queries regarding suicide and queries seeking specific, practical or instructive information regarding suicide methods. This information should include a helpline and links to freely available supportive information provided by a reputable mental health or suicide prevention organisation. It should also be prominently displayed to users in the search results.

Services should employ means to detect and provide warnings in response to search requests the wording of which clearly suggests that the user may be seeking to encounter CSAM. This warning should include information about the illegality of CSAM and links to resources provided by a reputable child sexual abuse organisation to help users refrain from committing CSEA offences. It should also be prominently displayed to users in the search results."

Similarly, search services should be required to provide additional information for individuals who search for terrorism content. This information may include alternative (or counter) narratives or links to EXIT organisations that support individuals to leave terrorist groups. In fact, this approach has been successfully implemented in the so-called redirect method that has been pioneered by Jigsaw and Moonshot. An evaluation by Facebook in 2019 demonstrated that the redirect method is promising, although it should not be considered as the sole measure to prevent and counter radicalisation online (see https://moonshotteam.com/resource/facebook-redirect-programme-moonshot-evaluation/).

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

Response: no

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

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:	
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:	

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:	

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:	

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

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:	

Content moderation (User to User)

Question 18:	
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:	

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:	

Automated content moderation (User to User)

Question 20:	
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 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:	
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:	

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:	

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:	

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:	

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:	

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:	

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:	

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:	

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:	

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:	

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:	

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

The Illegal Content Judgements Guidance (ICJG)

Question 49:	
i)	Do you agree with our proposals, including the detail of the drafting?
Response:	
ii)	What are the underlying arguments and evidence that inform your view?
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 50:	
i)	Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?
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 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:	

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

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:	

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:	

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:	