Consultation response form

Please complete this form in full and return to IHconsultation@ofcom.org.uk.

Consultation Title	Protecting people from illegal harms online
Your Full Name	Ayesha Bhatti
Your Contact Phone Number	(+44) 07598279179
Representing (Self or Organisation only)	Organisation
Organisation Name (if applicable)	Center for Data Innovation
Email Address	Abhatti@itif.org

Confidentiality

Is your name confidential? (please enter **yes** or **no** only)	No
Is your organisation name confidential? (please enter **yes** or **no** only)	No
Can Ofcom publish a reference to the contents of your response? (please enter **yes** or **no** only)	Yes
Please indicate if your full response is confidential. Partly confidential responses can be indicated under each question. (please enter **yes** or **no** only)	No

We ask for your contact details along with your response so that we can engage with you on this consultation. We will keep your contact number and email address confidential. For further information about how Ofcom handles your personal information and your corresponding rights, see Ofcom's General Privacy Statement.

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?
Response: We are concerned about the list of functionalities, including end-to-end encryption, pseudonymity and anonymity, livestreaming, and recommender systems, that Ofcom considers as creating risk for online harm. Ofcom is correct that these functionalities are not inherently bad and add value to users. Yet by listing these functionalities as risk factors that online services must consider when conducting risk assessments, Ofcom appears to suggest services should consider limiting use of such functionalities across their platforms. Even if such restrictions would offer some benefits, limiting the use of these functionalities may leave consumers worse off. For example, limiting the use of end-to-end encryption may allow for additional content moderation, but it may also expose users, including children and vulnerable adults, to additional privacy and security risks. Likewise, limiting online anonymity may result in an increase in certain online harmful behaviour, but the costs to privacy and free expression of eliminating this feature may create significantly more harm. Given that online services design these features primarily to benefit users, Ofcom should reconsider their inclusion. As an analogy, a risk factor for bank robberies is having cash on site, but the solution to this problem is not to suggest that banks stop handling money. We are also concerned that Ofcom identifies a service's "revenue model" and "growth strategy" as risk factors. Ofcom writes that "revenue models may create financial incentives that – intentionally or unintentionally – lead to business decisions that promote or tolerate illegal activity." In essence, Ofcom suggests online services will put profits over user safety and argues "the same logic applies across all kinds of illegal harms." However, online services, especially large ones, have invested millions in online safety and many have robust user safety teams. The marginal benefit an online service might receive from additional users engaging in illegal activity on their platform is likely offset by the potential negative harm to their reputation. Indeed, many businesses in the offline world face similar trade-offs, but most businesses, even though they strive to generate profits, do not promote illegal activity even if they have some financial incentives. Ofcom also specifically mentions advertising business models as a potential risk, which negatively portrays one of the most popular business models for online services and suggests businesses should consider alternatives. Reducing the number of ad-supported online services would negatively impact Internet users who often prefer these to paid alternatives.
ii) Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.
Response: Ofcom's analysis of the causes of online harm focuses too much on the factors that may lead bad actors to use a particular service for illicit purposes, rather than factors that increase the overall prevalence of online harms. For example, it is very likely that illegal firearms and drug dealers use mainstream online services, such as marketplaces to surreptitiously list their products or direct messaging to communicate with customers. Online services should make efforts to stop

those bad actors from using their services. However, Ofcom has presented no evidence that stopping these bad actors from using mainstream services will decrease the overall prevalence of illegal sales of firearms and narcotics. Instead, this activity will likely migrate to non-covered services or bad actors will further disguise their activities. Ofcom should be careful that its recommendations do not simply move online harms from online services that take these issues seriously to other services which may not or that it claims an artificial victory because reported online harms on major platforms decreases even as the underlying harmful behaviour does not change.
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.
Response:
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

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: We agree with the concept of governance and accountability to oversee the performance of illegal content duties. However, the proposed approach fails to consider a rapidly growing subset of online services built around decentralised architectures. As a result, the proposed governance and accountability are ill-suited for the wide range of online services that UK users may encounter.	
ii)	Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.
Response: The proposed approach fails to consider and address decentralised online services. Under this approach, centralised governance would be ineffectual as the host of any decentralised service would have little to no autonomy over them. The micro-blogging service Mastodon is an example of a decentralised online service, offering users the opportunity to either join or host their own servers, which act as micro-environments for content sharing. Any control over the operation and moderation of a server would fall on the server owner, which can range from organisations to individuals. The German non-profit organization that develops the Mastodon software is clear that it is unable to remove servers that share harmful content given the nature of the service, instead choosing a de-prioritisation/prioritisation approach to signpost users to servers that have undergone a vetting process. This vetting process involves a review of the server against the Mastodon Server Covenant. Decentralised online services do not yet have the level of popularity as centralised ones, but the situation could change quickly with the right service. The EU Blockchain Forum's recent report states that "[a]lthough the adoption of decentralised social media applications has not yet reached massive levels, it is evident that user awareness and migration from centralised to decentralised platforms are noteworthy. Decentralised social media has firmly established itself as a legitimate and valid alternative to the heavily centralised incumbent social media landscape. As these platforms mature, their appeal and user base are likely to grow, further reinforcing their credibility as a viable option in the digital space." If Ofcom wants to create future forward proposals, it should fully consider the governance and accountability of decentralised systems.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

Response: No.

Question 4:	
i)	Do you agree with the types of services that we propose the governance and accountability measures should apply to?
Response: Yes, but only for centralised services.	
ii)	Please explain your answer.
Response: Given that the proposals assume a centralised governance scheme, we agree with the types of services, provided that these types of services are limited to centralised systems. Ofcom would need to develop other proposals to cover the unique framework of decentralised architectures.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.	

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: Partially.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: We believe Ofcom should apply the most onerous measures to services that pose a high risk of causing harm, regardless of their size. As Ofcom correctly notes, "Where risks are very high, it is important that people are afforded protection even when the services they are using are relatively small." However, Ofcom should not apply the most onerous measures to services simply because they are large. Just because a service is large that does not mean that it has the resources to cover costs for onerous measures. For example, Twitter (now X) famously operated with massive losses for several years.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.	

Question 14:	
i)	Do you agree with our definition of large services?
Response: Partially.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: We agree that Ofcom should use a simple definition for large based on the number of users (ideally active monthly users), not metrics such as number of employees or revenue. However, counting the number of users of decentralised services is not straightforward. For example, the total number of users of an interoperable social network may exceed the threshold for a large service, but these users may be distributed across different service providers. Ofcom should provide clear guidance on how this threshold would apply to decentralised services.	

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

Question 15:
i) Do you agree with our definition of multi-risk services?
Response: No.
ii) Please provide the underlying arguments and evidence that support your views.
Response: We believe this definition does not allow for any meaningful distinction between different types of online services. The definition of multi-risk services will likely apply to most covered online services because bad actors can use these services for more than one type of illegal offense and Ofcom has set the threshold at two potential online harms. For example, any online service that has direct messaging capabilities poses a risk for drug offences and intimate image abuse. It would therefore be more meaningful to use a higher threshold or remove the distinction altogether.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

Question 16:
i) Do you have any comments on the draft Codes of Practice themselves?
Response: We agree with not taking a “one size fits all” approach. However, Ofcom only applies this principle to user-to-user services and search services. But its definitions of these services are not future proof. For example, Ofcom does not consider other methods of information retrieval, such as AI chatbots, that consumers may substitute for traditional search engines. Use of AI chatbots such as ChatGPT as the first source for information retrieval is increasing, as well as the integration of generative AI solutions into existing search engines such as Copilot in Bing. The current proposals split online services into either user-to-user or search services, for which neither appropriately capture the use of generative AI-as-a-service for gathering information. The taxonomy of search services looks at either general or vertical search services, and Ofcom does not consider this new form of information retrieval in its definitions. Ofcom should strive to create technology-neutral rules that neither favour nor penalize the use of emerging technology like AI. The taxonomy of search services should consider a third category for systems that are unable to provide real-time information based on indexing but can nonetheless offer a source of information similar to a search engine, either based on historical training data and/or Retrieval-Augmented Generation (RAG).
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

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: We agree with some proposals and agree with approach 3 as being the best approach for content moderation, however we have concerns about the enforcement of the Codes (dealt with in Question 53), and the establishment of Key Performance Indicators (KPIs)	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: Content moderation involves both reprioritisation of the recommender algorithm, as well as the take down of illegal content under the Act. Therefore, Ofcom should bear in mind that the KPIs will differ depending on how a service under approach 3 develops its content moderation system. Whilst the main KPIs in content moderation tend to be speed of take down and accuracy of identifying illegal content, the proposals should allow other KPIs for alternative content moderation systems, which may prefer prioritisation and the efficacy of the recommender algorithm to de-prioritise potentially illegal content.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.	

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: No.	

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: We have concerns over the enforcement and, in particular, penalisation of false negatives.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: The proposals should be clearer about when Ofcom will impose penalties, as there is a threat of increased censorship and moderation if Ofcom penalise services for not taking down content later deemed illegal by regulators (a false negative). This problem would also include situations where online services take down illegal content, but Ofcom does not deem the removal swift enough. Because services must employ a subjective and contextual "reasonable grounds to infer" approach, services will likely err on the side of caution which will result in over enforcement of the rules, thereby negatively impacting freedom of speech and expression. This risk is particularly acute where there is no penalty for false positives i.e., removing content that Ofcom deems legal. We therefore welcome greater clarity on when penalties will be enforced, such as an expansion on what Ofcom deems "appropriate and proportionate."	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.	

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:	