Consultation response form

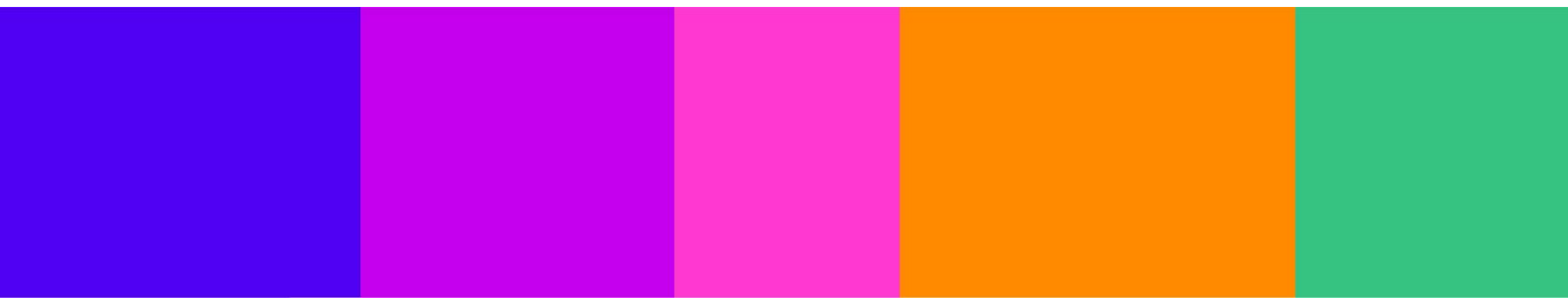Please complete this form in full and return to os-cfe@ofcom.org.uk.

Consultation title	Categorisation: research and advice Call for evidence
Full name	
Contact phone number	
Representing (delete as appropriate)	
Organisation name	Carnegie UK
Email address	

Confidentiality

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

Your details: We will keep your contact number and email address confidential. Is there anything else you want to keep confidential? Delete as appropriate.	Nothing
Your response: Please indicate how much of your response you want to keep confidential. Delete as appropriate.	None
For confidential responses, can Ofcom publish a reference to the contents of your response?	Yes

Your response

Question	Your response
Question 1: How do you measure the number of users on your service?	N/A
Question 2: If your service comprises a part on which user-generated content is present and a part on which such content is not present, are you able to distinguish between users of these different parts of the service? If so, how do you make that distinction (including over a given period of time)?	N/A
Question 3: Do you measure different segments of users on your service? • **Do you segment user measurement by different parts of your service? For example, by website vs app, by product, business unit.** • **Do you segment user measurement into different types of users? For example: creators, accounts holders, active users.** • **How much flexibility does your user measurement system have to define new or custom segments?**	n/A
Question 4: Do you publish any information about the number of users on your service?	N/A

Question	Your response
Question 5: Do you contribute any user number data to external sources/databases, or help industry measurements systems by tagging or sharing user measurement data? If not, what prevents you from doing so?	N/A
Question 6: Do you have evidence of functionalities that may affect how easily, quickly and widely content is disseminated on U2U services? • **Are there particular functionalities that enable content to be disseminated easily on U2U services?** • **Are there particular functionalities that enable content to be disseminated quickly on U2U services?** • **Are there particular functionalities that enable content to be disseminated widely on U2U services?** • **Are there particular functionalities that prevent content from being easily, quickly and widely disseminated on U2U services?**	Confidential? – N **Functionalities that allow the easy, quick and wide dissemination of content on U2U services include:** - Recommender systems and algorithmic promotion that reward users posting particular types of content (often this can be harmful, abusive, misogynistic, or false) and/or curate "who to follow" lists in users' feeds based on existing popularity and reach, regardless of whether or not those users follow the content creator or have engaged with their previous posts - Prominent upvote, resharing or like tools which allow users to instantly share posts with others who follow them. - Paid-for promotion of individual posts in users' feeds, again regardless of whether they have any prior interaction with the content creator or interest in the subject matter - Prominent indexes of popular or current hashtags enabling users to use these to make their content easily searchable, promotable or shared by users not known to them who are interested in that particular topic. - The ability to create and use hashtags in a coordinated way to orchestrate a pile-on on a particular user or to flood the particular service with posts on that content in short timescales - Large volume, synthetic features such as emojis or GIFs that encourage engagement - Ease of finding and tagging users with large numbers of followers that the individual user does not know - Unrestricted chatbots or bot networks that pick up and amplify posts relating to specific topics - Ease of embedding content from other platforms and use of third-party tools and, conversely,

Question	Your response

	speed at which content can be shared across multiple platforms from the individual service.
	
	- Ease of direct public replies to users with large followings, whether or not a user follows them – these individual posts are often then picked up and further amplified by that user's followers.
	- Ability to create large private groups or linked groups (such as [WhatsApp's communities](#)) enabling individual users to share content directly with 100s of other users, who can then share further via multiple other groups or communities.
	
	Consideration might also be given here to whether any monetisation or revenue-sharing arrangements with content providers provide incentives for or provide financial support to the dissemination of harmful content, and take appropriate steps to mitigate any such risk.
	
	Some services have features that are inherently risky: a clear example of this is chat roulette, which is risky not just for children.
	
	Are there particular functionalities that prevent content from being easily, quickly and widely disseminated on U2U services?
	
	Turning this question round to ask which functionalities would prevent the easy, quick and wide dissemination of content is instructive – eg "what does good look like?". This might include:
	
	- Functionalities that create friction or mechanisms that provide a delay between a user writing or sharing a post or uploading visual or audio material and their posting of it. For example:
	    o prompts about harmful language used in a post;
	    o number of posts permitted over a given time period;
	    o provision of content wrapper features;
	    o more than one click required to repost content.
	- Prompts to ask whether a user wishes to share a post without engaging directly with the content (such as Twitter/X's prompt to users to read arti-

Question	Your response
	cles before sharing); or functions which alert users to the fact that a post has been shared multiple times or [restrict its further sharing]().
	- Restriction and/or identification of chatbots and bot networks
	- Functionalities that allow users to have more control over curating their own online environment (this goes beyond the scope of current user empowerment tools; it could for example also include the option to choose chronological feeds over personalised feeds)
	- Policy and capability to monitor traffic on the site for early signs of virality for particular posts or hashtags and a means by which a brake can be applied by the service provider to allow a review of the content for harm. (eg criticism of Twitter for slow reaction to Wiley's antisemitic tweets)
	- Design to support uptake of user support mechanisms
	- Processes to incorporate understanding of viewpoints from different societal groups
	In relation to content that is deemed to be harmful, services could deploy any of the following tools or policies – though we note that some different design choices affecting earlier stages in the communications process could also have the result of reducing prevalence and virality of content harmful for the purposes of the online safety regime:
	- Demonetising content
	- Suppressing content in recommender tools and/or search engines;
	- Geo-blocking of content;
	- Suspension of content;
	- Removal of content;
	- Non-recommendation of user and/or group as person to follow;
	- The existence of a strike system;
	- Geo-blocking of account;
	- Suspension of account;
	- Termination of account

Question	Your response
Question 7: Do you have evidence relating to the relationship between user numbers, functionalities and how easily, quickly and widely content is disseminated on U2U services?	Confidential? – N ISD's research demonstrates the powerful effects of algorithmic ranking practices on shaping online discourses for users. A growing number of users will increase the volume of content generated on social media platforms. With vast numbers of content, platforms aim to keep attention span through algorithmic ranking for users. However, this research demonstrates how algorithmic ranking on a large scale shapes the online discourse users are exposed to collectively, and how they behave online[1]. The Centre for Countering Digital Hate research shows similar findings on users' recommendation algorithms, demonstrating how vulnerable users are targeted with self-harm and suicide content in very high quantities on TikTok[2] Avaaz have demonstrated how platforms like Facebook, without effective regulation of content, facilitate the spread of harmful content to millions of users. Their research focused on the spread of climate misinformation, with the 163 posts they analyzed accumulating an estimated total of 25,102,970 views in the research period[3]
Question 8: Do you have evidence of other objective and measurable factors or characteristics that may be relevant to category 1 threshold conditions?	There are a number of systemic factors relating to the design of individual services that should be taken into account here, beyond the means by which users can create or share content or how content from other users is promoted to them in their feeds. We set out here some of the aspects we cover in our work on codes of practice: these include a model code of practice for regulatory or self-regulatory approaches to online harm reduction which includes sections on some of the design-based features and functions that services should consider when taking decisions on how to mitigate the risk of harm occurring[4], and prior to this a hate crime code of practice and a VAWG code of practice.

---

[1] https://www.isdglobal.org/isd-publications/suggested-for-you-understanding-how-algorithmic-ranking-practices-affect-online-discourses-and-assessing-proposed-alternatives/

[2] https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design_120922.pdf

[3] https://secure.avaaz.org/campaign/en/facebook_climate_misinformation/

[4] https://d1ssu070pg2v9i.cloudfront.net/pex/pex_carnegie2021/2023/01/16173402/MODEL-CODE-A-reference-model-for-regulatory-or-self-regulatory-approaches-to-harm-reduction-on-social-media.pdf

Question	Your response
	- The service provider should consider the risks of tools/features used for organising content (eg hashtags) and what safeguards should surround their use, for example to prevent terms inciting violence against minoritised groups being used. - The service provider should consider the impact of autoplay functions, especially in the context of content curated or recommended by the provider. When a service provider seeks to take control of content input away from the person in this way the provider should consider how this feature might affect a person's right to receive or impart ideas. - The service provider should consider whether to provide appropriate information to its users about the accuracy (or otherwise) of information (eg flagging content that has been fact-checked) and should make its policies in this regard available. - The service provider must consider how its advertising delivery systems affect content seen by users. In particular, it must consider the circumstances in which targeted advertising may be used and managerial oversight over the characteristics by which audiences are segmented where those segments might be computer or user - generated. - The service provider must have terms of service and/or community standards in respect of its advertisers that are fit for purpose taken against its values, local laws and international human rights and should have processes in place to enforce that policy consistently. - The service provider must consider the need for explainability or interpretability, accountability and auditability in designing AI/ML systems. - The service provider should consider the speed and ease of content transmission. This could include, for example, methods to reduce the velocity of forwarding and therefore the occurrence of harm cross-platform. - The service provider should assess the risks posed by any features/tools (eg upvote/down

Question	Your response
	vote; like buttons) provided that encourage users to respond and/or to engage with other user's content.
	Other factors to consider here include the degree of personalisation deployed by the service's systems, including adequate oversight over the segments used for personalisation and policies.
Question 9: Do you have evidence of factors that may affect how content that is illegal or harmful to children is disseminated on U2U services? • **Are there particular functionalities that play a key role in enabling content that is illegal or harmful to children to be disseminated on U2U services?** • **Do you have evidence relating to the relationship between user numbers, functionalities and how content that is illegal or harmful to children is disseminated on U2U services?**	**Are there particular functionalities that play a key role in enabling content that is illegal or harmful to children to be disseminated on U2U services?** There are particular risks, with regard to illegal content, from embedded content from other services and click through to external sites. The use of breadcrumbing, which will be addressed by provisions in the OSB, allows child abusers to use phrases, keywords, or other hints that signpost to illegal content or to introduce abusers to networks of others. Closed groups with large numbers of members are a common way of facilitating or enabling the sharing of illegal or CSEA content. A service that has no means by which to carry out "know-your-customer" checks, or which enables users to create multiple, false or anonymous identities for account creation might also be of note. The use of autoprompts in search functions and the oversight of the services' algorithm is also important, to ensure that content is not recommended or shared in contravention of the site's own terms and conditions. For children, age-appropriate barriers must be introduced to stop harmful contact with minors. All account settings for children should be set to the safest level by default. As with question 6, it is important here to consider the functionalities and tools required to help users prevent exposure to illegal content and/or for children to manage their online experience in a way that is as safety-protecting as possible, which is the corollary of the functionality and tools that would help facilitate its dissemination. Accessible and transparent user mechanisms must be in place for adult users to also implement such features that protect them from exposure to harm. This could include:

Question	Your response
	- features to prevent the direct messaging of accounts that do not follow a user;
	- messages from unknown contacts reviewed by moderators; and
	- control features around who can search for a profile, what content is visible for example features which filter harmful content and words appearing, and how personal content can be shared or re-distributed online.
	- controls over recommendation tools, so a user could choose for example to reject personalisation;
	- user-set filters (over words, images, sound, videos or topics);
	- tools to limit who can contact/follow a user, or to see a user's posts;
	- tools to allow users to block or mute users, or categories of user (eg blocking anonymous and/or unverified accounts);
	- controls for the user over who can and cannot redistribute their content or user name/identity in real time.
	Do you have evidence relating to the relationship between user numbers, functionalities and how content that is illegal or harmful to children is disseminated on U2U services?
	- Research from 5Rights Foundation demonstrates how the functionalities of several user2user services are designed to maximise reach, and draw as many people as possible into their services. This includes increasing their reach to child users, for the ultimate purpose of increased revenue generation. Features fulfilling this objective include making it easy to share content, quantifying/displaying popularity, and making it easy to connect with friend or follower suggestions. 5Rights research demonstrated that such features also make social media platforms risky to children by design, demonstrating how algorithmic targetting also pushes harmful content onto children, including self-harm and suicide material, extreme body image content and sexual content[5]

[5] https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf

Question	Your response
	- The NSPCC reiterates that the platforms' design choices, and functionalities are used for the grooming of children and production of first generation child abuse images[6]. They also allow the sharing of such materials across the web, for example, through messaging platforms[7] - Note that IWF research on the occurrence of child sexual abuse imagery across platforms also demonstrates that the majority of child sexual abuse imagery is hosted on platforms 'not commonly known about', and that 'it's a bit of a myth that all this content is hosted on social networks' highlighting the complexity of the relationship between user numbers and illegal/harmful content to children is disseminated[8] -
Question 10: Do you have evidence of other objective and measurable characteristics that may be relevant to category 2B threshold conditions?	Confidential? – N We do not have evidence on this but would flag that there may well be a point of principle here that some functionalities or characteristics are inherently so risk that, no matter how small the service, they should be within scope. Chat roulette is one such example.
Question 11: Do you have evidence of matters that affect the prevalence of content that (once the Bill takes effect) will count as search content that is illegal or harmful to children on particular search services or types of search service? For example, prevalence could refer to the proportion of content surfaced against each search term 16 that is illegal or harmful to children, but we welcome suggestions on additional definitions.	N/A

---

[6] https://www.nspcc.org.uk/globalassets/documents/online-safety/duty-to-protect---nspcc-report---sept-2021.pdf

[7] https://www.nspcc.org.uk/globalassets/documents/online-safety/duty-to-protect---nspcc-report---sept-2021.pdf

[8] https://www.iwf.org.uk/news-media/news/iwf-publishes-platform-specific-data-for-child-sexual-abuse-imagery/

Question	Your response
• Do you have evidence relating to the measurement of the prevalence of content that is illegal or harmful to children on search services?	
Question 12: Do you have evidence relating to the number of users on search services and the level of risk of harm to individuals from search content that is illegal or harmful to children? • Do you have evidence regarding the relationship between user numbers on search services and the prevalence of search content that is illegal or harmful to children?	N/A
Question 13: Do you have evidence of other objective and measurable characteristics that may be relevant to category 2A threshold conditions?	N/A

Please complete this form in full and return to os-cfe@ofcom.org.uk.