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?	N/A

Question	Your response
• **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?**	
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?**	Confidential? – No **All functionalities are designed to enable content to be disseminated easily, quickly and widely because the architecture, infrastructure and environment of user-to-user services are designed to keep users engaged and using the platform. Most user-to-user services do not create their own content and are reliant on users creating and sharing their own. The design of functionalities is to facilitate this and make it as easy and appealing as possible, which leads to quick and easy dissemination of content – whatever kind.** The business model of most user-to-user services is dependent upon advertising revenue to sustain and grow it, which means they have a strong commercial incentive to keep users engaged and using their platforms for as long as possible. The quantity of customers paying attention to the content on a product equates to the value of the business: the more people paying attention, means more eyeballs on adverts, which drives up revenue.[1] This business model drives the design decisions behind most functionalities, particularly on user-to-user services. Designers interviewed as part of a study commissioned by 5Rights told researchers that most design decisions are driven by three key objectives: to **maximise reach, maximise time and maximise activity.**[2] These key objectives mean most functionalities are designed to make using the services as frictionless or easy and appealing as possible to keep users on the services, the by-product of which is the easy, wide, and quick dissemination of content.

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

Question	Your response

	Functionalities which encourage content dissemination and consumption:
	
	When used in combination, which is the intention behind the design strategy of most services, functionalities can lead to the quick and easy dissemination of content. Once the user is online, strategies that make it easy and frictionless to keep consuming and sharing content are used to prolong their time online. From our research, the functionalities we have found are used to encourage the sharing of content include:[3]
	
	• Functionality which facilitates timed content: Features and functionality that makes content only temporarily available, or only viewable 'live', are used to encourage users to engage with it immediately, or on a regular basis. E.g.
	  ○ Functionalities which allow users to **post 'Stories'** – available on multiple social media apps and sites - are available for 24 hours only after being posted, with notifications reminding users when they are posted.
	  ○ **Live-streamed content** is presented, and often not available later 'on demand'.
	  ○ **Notifications that content is available,** new or about to expire increase this motivation to engage with it sooner.
	• Functionality intended to reduce friction: Functionality which minimises the need for users to make active choices and removes any distractions aid the continued consumption of content. E.g.
	  ○ **Auto-play** configures content across many social media apps to automatically play or refresh.
	  ○ **Endless scroll** feeds present more content.
	• Popularity metrics: Functionality that facilitates connection, interaction and creation between users helps to 'promote' content by associating it with positivity, popularity, and aspiration. E.g.
	  ○ Making the **'like' button** a pink heart or a 'thumbs up' icon, associates it with positive emotions and relationships, promoting these features as positive and valuable to the user.
	• Functionality that rewards users: Some user to user services actively design additional rewards and incentives - popularity is rewarded above all else. **Content that has received high volumes of engagement will be displayed more prominently**, be more likely to 'go viral' or be shown to greater numbers of other users. E.g.

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

Question	Your response
	○ Counting and prominently displaying quantified information about social activity is designed to draw user attention to it. Displaying how many **'friends', 'followers', 'following'** – each user has or displaying individual content alongside a count of how many **'likes' or 'shares'** it has received can encourage users to share the content. • Functionality that connects users known to them or strangers once removed: Social media companies design in features that facilitate the building of networks by enabling users to easily discover and connect with other users that know. E.g. ○ **Friend recommendations** based on the contact numbers or email addresses stored on the phone, or contacts brought across from other apps. ○ **Direct messaging** to other users that are not already connected to the user's account. ○ **Profile identifiers** such as usernames or QR codes enable users to share their profile with others on and off-app to make wider connections. ○ **Privacy settings 'off' by default**, meaning the user must go out of their way to make it harder for others to connect with them. • Functionality which streamlines validation and feedback: Companies simplify the channels for interacting by reducing the barriers to giving other users feedback. This then increases the expected or 'normal' volume of interaction, creating a self-reinforcing cycle. E.g. ○ **The 'like' button** is prominently displayed and positioned conveniently under the thumb (for a righthanded user) which increases the level of ease with which users can provide feedback. ○ **Emojis, stickers or comments** often available in comment boxes or instant messenger interfaces. These are almost always predominantly positive (smiling emojis, positive affirmations) as opposed to negative, nudging users towards positive feedback and validation towards other users. • Functionality that facilitates copying and content creation**:** Many social media products have features that enable the editing of content before it is posted. Children in particular experience validation and affirmation in the form of 'likes', comments and connections that shape what they do online. In order to receive this affirmation, they create mostly visual content, sharing them as widely as possible – copying trends and behaviours of others facilitates their desire to do this. ○ **Filters, lenses and photo-editing tools** are all designed to enable the user to 'improve' the aesthetics of images or videos that they might post as easily as possible. Some digital products 'beautify'

Question	Your response
	images by default through the camera function (e.g., smoothing skin, changing face shape) – subtly encouraging users to create content they may then feel comfortable sharing more widely. ○ **Tools and templates** for creating videos that fit into trends, such as using particular filters, sound-tracks or voice-overs. ○ Most apps make the **'re-share'** feature as frictionless as possible, e.g., with a 'one click' button. How functionality used in combination disseminates content: Misinformation case study Misinformation is a pervasive risk to all users, but children in particular. Ofcom research found that while 74% of children aged 12-17 felt confident in identifying misinformation, only 11% were able to correctly identify a genuine post, without making mistakes.[4] This is having real world consequences. During the COVID-19 pandemic, research carried out by Kings College University found one in five 16-24- year-olds believed there was no hard evidence coro-navirus actually exists.[5] Using the example of misinformation, the following features and functionalities used in combination can amplify its spread across services easily and quickly: • **Popularity metrics** (likes, shares and views), inform the recommendation algorithms that digital services use to promote content to users. Misinformation may attract thousands or millions of likes, shares and views, particularly when it is provocative, humorous or even just absurd.[6] Misinformation can seem more credible when it appears alongside visible popularity metrics or is shared by 'verified' accounts. Stemming the flow of misinformation is challenging when algorithms prioritise popularity metrics over the nature of the content itself, which can lead to misinformation being amplified and services profiting from its spread. A leaked in-ternal 2016 presentation from a major social media company revealed; "64% of all extremist group joins are due to our recommendation tools" and that most of the activity came from the platform's "Groups You Should Join" and "Discover" algorithms: "Our recommendation systems grow the problem."[7]

[4] https://www.ofcom.org.uk/news-centre/2022/one-in-three-internet-users-fail-to-question-misinformation
[5] https://www.kcl.ac.uk/policy-institute/assets/covid-conspiracies-and-confusions.pdf
[6] https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308
[7] https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/#:~:text=In%20an%20internal%20presenta-tion%20from,our%20recommendation%20tools%2C%E2%80%9D%20the%20presentation

Question	Your response
	• **Autoplay** is designed to prolong time spent on the service. Services that use autoplay expose users to recommended video or audio content that plays without initiation from the user. Autoplay risks taking users further into recommendation rabbit holes and exposing them to recommended video or audio content that can become increasingly more extreme.[8] Some services do not allow users to switch autoplay off. • **Trending lists** provide instant access to false information, particularly as popular **hashtags** are used to promote disinformation.[9] Trending lists are easily manipulated by fake accounts and some companies exploit this by offering the creation of a "**bot**" account for as little as £150 to make a hashtag trend for a few hours.[10] • **Fake accounts** include automated accounts or 'bots' and fake profiles created by users. Fake accounts can be created for malicious purposes, such as manipulating discussion online or spreading misinformation at scale.[11] Bots that use AI to appear more human-like are difficult to detect and can evade content moderation. Despite policies to tackle coordinated inauthentic behaviour, fake accounts have been used to create pages where fake accounts can generate fake engagement. A loophole in inauthentic behaviour policy means that, by using pages, malicious actors are able to exaggerate the credibility of information users see and ultimately influence the algorithms that recommend content to users. • **Ineffective content labelling** undermines efforts to identify misinformation or provide relevant information to users. Content labels that warn of inaccurate content or redirect users to credible sources of information are often too subtle and therefore ineffective. Labels have also inadvertently led to disputed content receiving more attention.[12] Visual warnings have been called for to overcome concerns about 'language and cultural barriers' that play a part in false information.

[8] https://www.technologyreview.com/2020/01/29/276000/a-study-of-youtube-comments-shows-how-its-turning-people-onto-the-alt-right/
[9] https://www.isdglobal.org/digital_dispatches/hashtag-pairing-is-being-used-on-twitter-to-facilitate-soviet-propaganda-tactic-whataboutism/
[10] https://www.bbc.co.uk/news/blogs-trending-43218939
[11] https://www.cits.ucsb.edu/fake-news/spread
[12] https://misinforeview.hks.harvard.edu/article/twitter-flagged-donald-trumps-tweets-with-election-misinformation-they-continued-to-spread-both-on-and-off-the-platform/

Question	Your response
	• **Disappearing content** promotes disinhibited behaviour and 'consequence-free' content sharing. This type of content can only be viewed and shared during a certain time period and often 'disappears' before it can be fact-checked. Features that allow users to create disappearing content are popular among young people, but are also more difficult to report. 86% of 13-17 year olds use disappearing content to interact with their friends.[13]
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? – No **Categorisation of regulated services must be risk-based, taking into consideration the higher risk that some smaller services can pose due to the topic or content they discuss on the service. Ofcom must look to measure the risk of functionalities and features which can cause children harm as set out in the 4C's framework of online risk to children (Annex A).[14]** While user to user services with large user bases and high functionality can lead to fast dissemination of harmful content, smaller platforms with fewer users and less functionality are not by definition less risky. Categorising services based on functionality and user base must account for: ○ '**Popular by surprise' services:** Categorising services by the size of its user base does not take into consideration services which can grow very quickly overnight. For example, the tech company Meta recently launched a new user to user service, Threads, which attracted 2 million sign ups within two hours of it launching in July. By the next day, users had already posted 95 million posts, and 190 million likes.[15] The subscription-based service OnlyFans, where users can pay for sexually explicit content, had only 120,000 users in 2019. By December 2020, the service had more than 90 million users and over one million content creators.[16] Children are often the first adopters of new technologies and services, including those that are not intended for them and , placing them at a greater risk of harms which could arise in the case of popular by surprise services.

[13] https://www.childnet.com/wp-content/uploads/2021/11/Youth-perspectives-on-expiring-content-new-youth-research-by-Childnet.pdf
[14] https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf
[15] https://time.com/6292957/threads-fastest-growing-apps/
[16] https://www.nytimes.com/2021/01/13/business/onlyfans-pandemic-users.html?smtyp=cur&smid=tw-nytimes

Question	Your response
	○ Smaller does not mean safer: ○ **Small services can promote the most extreme content:** While high functionality can facilitate quick dissemination of content, a service with low functionality can also serve as a basis from which extreme content is spread. Research from the Centre for Countering Digital Hate (CCDH) found that a forum maintained and used to promote incel propaganda has around 4000 active users, but around 2.6 million visits a month. During the time period of the research, over a fifth of posts in the incel forum feature a misogynist, racist or homophobic slur.[17] • **Small services can lack robust safety features:** Smaller services often have weak content moderation, which means content that the service says it does not allow on the service is left on the platform. For example, Clapper a video-sharing platform has under 1 million downloads on the Google Play store. Despite a minimum user age of 17, the service's weak age assurance means a child can log into Clapper via their Google account, even if they are underage. The service was found to harbour misinformation and took steps to announce that it does not allow conspiracy 'QAnon' content[18] however, this 'ban' not featured anywhere in the services community guidelines. Indeed, the service's terms of service also explicitly state "we cannot ensure the prompt removal of objectionable material as it is transmitted or after it has been posted.[19]" **The approach to categorisation should be risk-based** In setting the threshold for Category 1 services, Ofcom should assess each functionality against the known range of harms - content, contact, conduct and contract harms[20] - children experience online (Annex A). Many of these harms are created and facilitated by the design features and commercial decisions of platforms, which are often the underlying drivers of harm to children online.[21]

[17] https://counterhate.com/wp-content/uploads/2023/08/CCDH-The-Incelosphere-FINAL.pdf
[18] https://www.theverge.com/2021/2/11/22278480/clapper-tiktok-clone-bans-qanon-content-parler-deplatforming-capitol-riot
[19] https://www.clapperapp.com/terms
[20] https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf
[21] https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf

Question	Your response
Question 8: Do you have evidence of other objective and measurable factors or characteristics that may be relevant to category 1 threshold conditions?	N/A
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?***	Confidential? – No In setting category 1 threshold conditions, Ofcom should assess services and their functionalities against the risks outlined in the 4Cs framework (Annex A)[22] which provides a non-exhaustive list of known online harms to children and how they manifest. The risk register of commons features (Annex B) is a non-exhaustive list informed by this framework and sets out how some features commonly used by children can cause harm. Some functionality can be harmful to children when used in combination and can also impact different groups of children in different ways. 5Rights commissioned research, *Pathways*[23] and *Just One Click*[24] provides evidence of how functionality leads children to harmful content online. The 4C's: Annex A is the 4C's risk framework[25] - a classification of risks to children online - content, contact, conduct and contract (sometimes referred to as commercial) risks, or cross-cutting risks for those that fall into more than one category. The framework provides a guide to help identify different kinds of risks, whether acute or severe, immediate or cumulative, individual or multiple. The examples in Annex A include risks from *Children Online: Research and Evidence*, the ICO's *Children's Code Harms Framework*, the Australian eSafety Commissioner's *Safety by Design* work and the OECD's *Revised Typology of Risks for children in the digital environment*. They are

[22] https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf
[23] https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf
[24] Annex C
[25] Ibid

Question	Your response

	indicative of the types of risks children may be exposed to online. The lists are not exhaustive, but they give a good indication of the breadth of risks that should be considered currently.
	
	Risk register of common features: Annex B sets out common features and functionalities of services children use. In assessing the risk level of functionality, attention should be paid to how certain features might impact different groups of children.
	
	In addition to assessing risk of functionality to children using the above frameworks, Ofcom should also consider the following:
	
	• **Using functionalities in combination:** risks created by individual features can increase when they are used in combination with other features. For instance, a service which makes use of algorithmic friend recommendations that recommend child accounts to adults and make a child's location discoverable by other users would make the potential for grooming and sharing CSAM more likely.
	• **Misuse of features:** functionalities can be misused by actors with malign intent. Such risks may arise through:
	  • Inauthentic use of the service, such as the creation of fake accounts .
	  • The use of bots or deceptive use of the service .
	  • Other automated or partially automated behaviours.
	  • Coordinated manipulation and use of their services .
	  • Systemic infringement of their terms of service .
	  • Providers should pay particular attention to how their services, and any use of algorithmic amplification, may contribute to these systemic risks.
	• **Risks over time:** Certain risks may expose children in particular to low levels of immediate harm but increase in severity over time. A single notification may momentarily distract a child, for instance, but over time may have a more serious impact on their sleep, schoolwork and ability to concentrate.
	  • Functionalities can expose children to risks over time in many ways, including:
	  ▪ Isolated exposure to risks that cause immediate harm, such as seeing a violent, sexual or otherwise developmentally inappropriate content .
	  ▪ Cumulative passive exposure to risks over time, such as seeing the same narrow ideals of beauty consistently promoted in newsfeeds or timelines .

Question	Your response
	▪ Cumulative active engagement with risks, such as participating in pro-anorexia or self-harm groups . • Similarly, the impact of harm can be either: ▪ Immediate or delayed – whether the impact of the experience occurred immediately after exposure or manifested at a later point . ▪ Direct or indirect – whether the impact of the hazard occurred through direct exposure to the child who was harmed or indirectly through exposure . <u>How different functionality can be harmful:</u> In 2021, 5Rights commissioned research from Revealing Reality to explore how the design of digital products, in particular social media, shape the experiences and behaviours of children. The research found that features and functionalities found commonly on user-to-user services are driving and nudging children towards harmful content. These findings formed the *Pathways: how digital design puts children at risk*[26] report which drew out how functionalities such as direct messaging, hashtags, recommended accounts, and search functionalities within user to user services led children to harmful content: **Direct messaging:** • Children were directly messaged and directed to pornographic content : • All of the child avatars were directly messaged by accounts on Instagram they did not follow. This included being added to group chats by strangers with other adults. The apparent motives behind these messages varied, but they included promoting websites with paid-for porn content, promoting brands or pages as well as offers to 'collaborate' in promoting products. • All four male child avatars and two female child avatars on Instagram were added to group chats by people they didn't know, in which there were multiple other strangers with links to paid-for porn sites or pornographic dating sites. **Recommendation systems (nudges, hashtags):**

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

Question	Your response

	• Children were nudged towards harmful and more extreme content:
	  ○ Ciara's (age 15) avatar searched for '#skinny' on Instagram, which led to an account titled 'skinny.quick' promoting a website selling 'fat burner' and 'breast enhancer' gummy bears.
	  ○ Charlotte's (age 15) avatar was recommended accounts and posts relating to weight loss and fitness. After 'following' and 'liking' a selection of these, the 'explore' feed filled with more similar content promoting weight loss journeys, fitness 'before and after' comparisons, dieting tips or photos of women emphasizing their slimness or low weight.
	
	Search functions: *Children were easily able to search for and access harmful content*
	• Child avatars searched terms that aligned with content that children in the research had told us they had seen on social media apps and sites, including 'thin', 'bodygoals', 'porn', 'darkmemes', 'suicide' and 'proanaa' (proana with one 'a' is blocked by Instagram, but adding a second 'a' unlocks access to pro-anorexia content).
	• Laura's (age 13) avatar searching 'suicide' on Instagram was recommended images of graphic self-harm injuries
	• Ciara's (age 15) avatar searching for 'proanna' on Instagram was recommended posts titled "no food all week"
	• Jordan's (age 14) avatar was served up sexual content on Instagram alongside adverts for Roblox and a school revision study app
	• Owen's (age 15) avatar was served sexual content on Instagram alongside adverts for T-levels and a Home Office campaign for recognising and reporting child abuse online
	• Laura's (age 13) avatar was able to search for 'depressed' theme content on Instagram, while also being served adverts for a sweetshop, Nintendo Switch and a teen targeted-tampon advert. Laura's avatar was also recommended a post of pro-suicide material saying "it's so easy to end it all."
	
	<u>Functionality used in combination: Livestreaming and video sharing case study</u>
	
	Livestreaming or sharing videos can be an engaging way for a young person to express themselves and connect with peers however it can also expose them to inappropriate and avoidable contact, conduct and content risks. For example:

Question	Your response
	• Services which encourage users to share videos from personal spaces create a window into a young person's life, which groomers can exploit in an attempt to build trust.[27] • Services often set a young person's livestreams to **public by default**, making them visible to millions of users, including unknown adults. A service that randomly pairs children to unknown adults has been used by perpetrators of CSEA[28] to gather material. • Many livestreaming and video-sharing services enable viewers to **move from a public interaction to private messaging**. In 74% of cases[29], when children are contacted by someone they don't know online, this first happens via private messaging. • Design choices such as hearts that visualise **'likes' enable others to exploit the desire for social affirmation** which is strong in children and young people. Children as young as 7-10 years old have been pressured into performing sexual acts on livestreams in exchange for likes.[30] • This enables viewers to interact with **live-streaming young people in real-time**. 6% of children who livestreamed have been asked by viewers to change or remove their clothes on camera.[31] • Services that **do not uphold their own age restrictions** undermine protections for children. One popular app that enables video chat between random users automatically prefills a user's age as 18 on registration. Underage users are able to access the app without further verification.[32] • **Adult and harmful content** (such as self-harm and suicide content) is routinely promoted to children by recommendation algorithms, even when expressly forbidden by community guidelines.[33] • Videos of children are served up to users who have shown a prior interest in young people, or to users who had previously watched sexually themed videos.[34]

[27] https://www.sciencedirect.com/science/article/abs/pii/S2211695816300095?via%3Dihub
[28] https://www.bbc.co.uk/news/av/technology-56103351
[29] https://www.nspcc.org.uk/globalassets/documents/online-safety/delivering-a-duty-of-care.pdf
[30] https://www.iwf.org.uk/about-us/who-we-are/annual-report-archive/
[31] https://learning.nspcc.org.uk/media/1559/livestreaming-video-chatting-nspcc-snapshot-2.pdf?_ga=2.195183788.1083629909.1589538845-336995373.1589375216
[32] https://smartsocial.com/post/antiland-app
[33] https://www.wired.com/story/when-algorithms-think-you-want-to-die/
[34] https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html

Question	Your response

Question 10: Do you have evidence of other objective and measurable characteristics that may be relevant to category 2B threshold conditions?	Confidential? – No
	
	Services that pose less of a risk to children, based on the 4C's of online risk and the risk register of common features, which have no or limited functionality, or limit the spread of content could potentially have characteristics relevant to 2B threshold conditions.
	
	• **Services which pose less of a risk to children**: Using the harm and risk registers outlined in response to Question 9, services with fewer functionalities or functionalities which pose less of a risk of harm to children could be a characteristic of 2b thresholds. However, as outlined in response to Question 7, services with low functionality are not necessarily by virtue, safer. For example:
	  ○ Forums that promote and encourage disordered eating, for example pro-anorexia and pro-bulimia blogs or websites with content which includes fasting tips, statements such as "being thin is more important than being healthy" and "eat and you'll never be skinny. Starve and forever be pretty[35]", 'thinspiration' images of emaciated bodies and sections on how to manage cravings.
	  ○ Online games which promote violence and financial harms such as gambling. For example, Crazy Games where children can "practice a little violence and get covered in blood and gore!."[36]
	• **Services with functionality which limits the sharing of content:** Some services have introduced features which place limits on how content is shared. While this does not prevent dissemination of harmful content entirely, it does slow down the speed.
	  ○ **Limits on resharing content:** For example, WhatsApp has placed forwarding limits on its service.[37] WhatsApp users can forward a message or a channel update with up to five chats at one time. However, if you're forwarding a message or update that was forwarded to you, it can only be shared with one group chat.

[35] https://proanalifestyledot.wordpress.com/
[36] https://www.crazygames.com/t/bloody
[37] https://faq.whatsapp.com/1053543185312573

Question	Your response
	○ **Warning messages or prompts:** The user to user service X (formerly Twitter) introduced a feature which prompts users to read an article they are trying to 'retweet' or 'quote tweet'.[38] The service says this is to encourage more informed discussion on its platform. • **Services which practice data minimisation:** Most services profile users for commercial gain by collecting and data gathering and profiling them. Algorithms follow user behaviour patterns on such tight loops that they know the 'exact' mix of ingredients that will appeal to each individual user. This data is often used in persuasive design strategies, intended to keep users using the service for as long as possible. There are examples of large platforms with a high user base which only collect limited data and present less of a risk to users. For example, Wikipedia does not profile users on its service, and collects only very limited data which it uses to improve the service and not chase users with adverts or other content.[39]
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	Confidential? – No **Search services are driven by the same commercial objectives as user-to-user services, which means much of their functionality is driven to keeping users on the service. This means that many of the risks associated with user-to-user services are relevant for search services.** In addition to the known harms and risks of features functionality listed above, matters that can impact the prevalence of harmful content on search services include: • **Autocomplete:** Where some search engines suggest possible search terms based on the first few letters entered by the user, these can interrupt, misinterpret and possibly redirect a child's thought process. This can steer the child towards sometimes towards extreme, stereotypical or unwelcome views. Search functionalities within user to user services also pose this risk. • **Bypassing content moderation:** Words and hashtags associated with trying to bypass content moderation often result in harmful content appearing prominently on search services. Research from the Centre of

[38] https://www.theguardian.com/technology/2020/jun/11/twitter-aims-to-limit-people-sharing-articles-they-have-not-read#:~:text=In%20the%20test%2C%20pushed%20to,Twitter%20said%20in%20a%20statement.
[39] https://foundation.wikimedia.org/wiki/Policy:Privacy_policy

Question	Your response
harmful to children, but we welcome suggestions on additional definitions. • **Do you have evidence relating to the measurement of the prevalence of content that is illegal or harmful to children on search services?**	Countering Digital Hate found forums in the "incelosphere" feature on the first page of Google search results for terms associated with body image and unemployment. Researchers examined the term "looksmaxxing" is a verb meaning 'to improve your appearance' that is popular in incel communities and has seen wider use in health and fitness communities. The term "NEETs" is often used by incel communities to refer to young men who are 'not in education, employment or training'. Searches for the name of the suicide forum returned results in which the forum ranked top in UK, and second in the US. This is despite reporting of the forum's role in the suicides of dozens of young people, leading suicide experts to brand the site as "extremely dangerous".[40]
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**	Many of the risks associated with large search services remain for smaller services. Weaker content moderation and safety tools pose a high risk to children with regards to being expose to harmful content. See response to Question 7.

[40] https://counterhate.com/wp-content/uploads/2023/08/CCDH-The-Incelosphere-FINAL.pdf

Question	Your response
illegal or harmful to children?	
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

Annex A: 4C's online risk to children

Content risks A child or young person experiences content risks when they are exposed to harmful material. They include:	Contact risks Contact risks arise when a child or young person participates in an activity with a malign actor, often, but not always, an adult. They include:	Conduct risks A child or young person encounters conduct risks when they are involved in an exchange, often, but not always, peer-to-peer, as either a perpetrator, victim, or sometimes both. They include:	Contract risks (or commercial) Contract or commercial risks occur when a child or young person is exposed to inappropriate commercial contractual relationships or pressures. They include:
Violent material Child Sexual Abuse Material (CSAM)	Child Sexual Exploitation and Abuse (CSEA), including grooming Developmentally inappropriate activity Catfishing (targeting a victim by using a fake identity)	Trolling	Loss of digital footprint Hidden costs Compulsive use Identity theft
Developmentally inappropriate content Extremism	Scams Blackmail	Cumulative or volumetric attacks	Fraud Phishing Scams Gambling
Eating disorder promotion Disinformation/ misinformation	Stalking, unwanted surveillance	(‘pile-ons’)	Inaccurate profiling Bias in automated decision-making

Scams		Sexual extortion (‘sextortion’) Non-consensual sharing of intimate material or image-based abuse	Excessive data collection, sharing
Body image pressures		Bullying, abuse, insults, rumors, social exclusion Individual identity attacks Dehumanization	
		Hate speech Sexual harassment/ aggression	
		Doxing (publishing private information)	
		Extremism	
		Direct and indirect threats of violence, intimidation and harassment	

		Doctored images (including deepfakes and shallow fakes) Scams	
		Stalking, unwanted surveillance	
		Chilling effects on free expression	
		Over-exposure, over-sharing	
Cross-cutting risks	A number of risks to children online cut across some or all of the categories of risk, and result in children being exposed to infringements of their privacy, threats to their health or unfair treatment. Cross-cutting risks include:		
Infringement of privacy Adverse effect on data rights Restriction of access to services	Discrimination Risks to physical and mental health	Interference with sleep or schoolwork Security risks	Addiction, compulsive use Loss of non-financial resources (e.g. time, sleep)

Annex B: Risk register of common features

Features	Risk Category	Potential Harms
Friend recommendations that introduce adults to children	Contact, Conduct	Child Sexual Exploitation and Abuse (CSEA) Developmentally inappropriate activity Phishing and catfishing
Notifications on by default	Conduct, Contract	Loss of non-financial resources (e.g. time, sleep) Unwarranted intrusion
Discoverable location	Contact, Conduct, Contract	Bodily harm Unwarranted intrusion
Targeted advertising on by default	Content, Contract	Unwarranted intrusion Undue commercial pressure Financial harm
		Manipulation and exploitation
Lootboxes	Contract	Hidden costs Compulsive use Gambling
In-service 'gifts'	Contract	Child Sexual Exploitation and Abuse (CSEA) Developmentally inappropriate activity Catfishing
		Scams
End-to-end encryption	Content, Contact	Child Sexual Abuse Material (CSAM) Developmentally inappropriate content Extremism
		Child Sexual Exploitation and Abuse (CSEA), Developmentally inappropriate activity

Low-privacy profiles by default	Contact	Child Sexual Exploitation and Abuse (CSEA) Stalking, unwanted surveillance
		Identity theft Catfishing
Direct messaging of children by unknown adults	Contact	Child Sexual Exploitation and Abuse (CSEA) Developmentally inappropriate activity Catfishing
		Stalking, unwanted surveillance
Livestreaming / video chat	Contact, Conduct	Child Sexual Exploitation and Abuse (CSEA) Developmentally inappropriate activity Blackmail
		Stalking, unwanted surveillance Over-exposure, over-sharing
Video-sharing	Content, Conduct	Child Sexual Exploitation and Abuse (CSEA)
		Developmentally inappropriate activity Blackmail,
		Stalking, unwanted surveillance Over-exposure, over-sharing
Image-sharing	Content, Conduct	Stalking, unwanted surveillance Over-exposure, over-sharing Body image pressures
Anonymity	Contact, Conduct	Child Sexual Exploitation and Abuse (CSEA) Catfishing
		Stalking, unwanted surveillance Trolling
		Bullying, abuse, insults, rumors, social exclusion
Search functions	Content, Contract	Violent material

		Developmentally inappropriate content e.g. eating disorder promotion or disinformation/misinformation
Engagement ‘streaks’	Contract	Addiction, compulsive use Over-exposure, over-sharing Excessive data collection
Algorithmic curation of feeds	Content, Contract	Violent material
		Developmentally inappropriate content: Extremism
		Eating disorder promotion, Disinformation/misinformation Scams
		Body image pressures Inaccurate profiling
		Bias in automated decision-making Excessive data collection, sharing
Virality	Content, Conduct	Developmentally inappropriate content Body image pressures Developmentally inappropriate activity Over-exposure, over-sharing
Endless scroll	Content Contract	Compulsive use
		Loss of non-financial resources (e.g. time, sleep)
Popularity metrics	Contact, Conduct	Child Sexual Exploitation and Abuse (CSEA), including grooming

		Developmentally inappropriate activity Stalking, unwanted surveillance
		Over-exposure, over-sharing
Autoplay	Content, Contract	Addiction, compulsive use
Trending lists	Content, Conduct	Developmentally inappropriate content Disinformation/misinformation Addiction, compulsive use
Disappearing/ time-limited content	Content,	Violent material
	Contract	Child Sexual Abuse Material (CSAM) Developmentally inappropriate content Extremism
		Scams
		Addiction, compulsive use
Disappearing/time- limited messages	Content, Contact	Child Sexual Exploitation and Abuse (CSEA) Developmentally inappropriate activity Scams
		Blackmail
		Sexual extortion ('sextortion')
		Non-consensual sharing of intimate material or image-based abuse
		Bullying, abuse, insults, rumors, social exclusion
		Direct and indirect threats of violence, intimidation and harassment

		Doctored images (including deepfakes and shallow fakes)
Groups	Content, Contact, Conduct	Violent material, Child Sexual Abuse Material (CSAM)
		Developmentally inappropriate content Extremism
		Eating disorder promotion Disinformation/misinformation Developmentally inappropriate activity Hate speech
		Chilling effects on free expression Over-exposure, over-sharing
Pay-to-play	Contract	Hidden costs
		Restriction of access to services
Autocomplete	Content, Contract	Developmentally inappropriate content Inaccurate profiling
People also liked…	Content, Contract	Developmentally inappropriate content Disinformation/misinformation Inaccurate profiling
Image altering (filters)	Conduct	Body image pressures
		Doctored images (including deepfakes and shallow fakes)
'Creator' accounts	Content, Conduct, Contract	Over-exposure, over-sharing Undue commercial pressure
Public comments	Contact, Conduct	Child Sexual Exploitation and Abuse (CSEA) Stalking, unwanted surveillance
		Trolling

		Cumulative or volumetric attacks (‘pile-ons’) Bullying, abuse, insults, rumors, social exclusion Individual identity attacks
		Hate speech
		Chilling effects on free expression
‘Quick add’ features that promote frequent and frictionless adding of contacts	Contact,	Developmentally inappropriate activity
	Conduct	Phishing and catfishing
		Overexposure, oversharing