Snap Inc.

SNAP INC RESPONSE: OFCOM PROTECTING CHILDREN FROM HARMS ONLINE

26 July 2024

CONTENTS

Introduction

Thank you for the opportunity to respond to Ofcom's consultation on 'Protecting children from harms online' as part of Ofcom's work to implement the UK's online safety regime. We are pleased to see that many of the draft measures published in the consultation are steps that Snap already takes to protect our younger users.

We recognise that there is crossover in this consultation with some of the measures and guidance included in the Illegal Harms consultation, which we responded to in March. On that basis, we recommend Ofcom reads this response in conjunction with our submission on the Illegal Harms consultation (Please see Annex A), as we have avoided repeating our positions where there is no or limited differentiation in the recommended measures between the two consultations. Our response to the Illegal Harms consultation also provides a short introduction on how Snapchat works and we recommend you refer to this if you are unfamiliar with the platform.

To support this consultation submission, we have provided the following information on our approach to children on Snapchat to help contextualise our commentary in the detailed section of the response.

[REDACTED/CONFIDENTIAL] We take our responsibilities seriously to help protect our community from harm and ensure they have a safe positive experience on Snapchat – especially for our younger users.

This begins with our Terms of Service, which is clear that a user must confirm that they are aged 13 or above to join Snapchat, and Community Guidelines which prohibit the use of Snapchat for any illegal or harmful activity. We have made deliberate design choices from the outset (guided by safety- and privacy-by-design principles) to ensure the platform supports the safety of our users, including to help connect people to real friends rather than strangers and prevent the spread of harmful and other violating content. For example:

- We do not have a public newsfeed or the ability to live stream;
- The public areas of Snapchat – Spotlight and Discover – prevent widespread, unvetted distribution. Content is curated or pre-moderated before it can reach a large audience.
- 1:1 communication can only commence once there has been a mutual friend request.
- Friends lists are private to each user (and their parents via our Family Centre).
- Blocked users are not able to send new friend requests sent from other accounts created on the same device.
- Location sharing is off-by-default and location can only be shared between friends on Snapchat (rather than the public at large).

We also offer additional protections for our younger users aged between 13-17. For example:

- Pop-up warning messages to a teen to inform them of the potential risk when connecting with someone:
    - They don't share any mutual friends with or have in their contacts;
    - Who has been blocked or reported by others; or
    - Is from a region where the teen's network is not typically located.
- Preventing a delivery of a friend request altogether when a teen sends or receives a friend request from someone they don't have any mutual friends with, and that person has a history of accessing Snapchat in locations often associated with scamming activity.
- Parental tools - Family Centre - to give parents insights on who their teens are talking to and prompt conversations about healthy online habits.

We welcome Ofcom's public recognition of Snap's measures to protect young users and support the Online Safety Act (OSA) as well as adding some of these to your proposed codes relating to illegal harms (e.g. grooming and CSAM).

Detailed response

[REDACTED/CONFIDENTIAL]

NOTE: A person must confirm they are 13 or above to join Snapchat. Therefore, our references to 'children', 'minors', 'teen' in this response refer to those users who are aged between 13-17.

1. Draft Guidance on Content Harmful to Children

Do you agree with our proposed approach, including the level of specificity of examples given and the proposal to include contextual information for services to consider? Do you have further evidence that can support the guidance provided on different kinds of content harmful to children?

For each of the harms discussed, are there additional categories of content that Ofcom

a. should consider to be harmful or
b. consider not to be harmful or
c. where our current proposals should be reconsidered?

Snap is generally aligned with Ofcom's approach, including the level of specificity of examples given and the proposal to include contextual information for services to consider. It would be helpful, however, if Ofcom's definitions of harm incorporated some of the elements of the examples provided to ensure services have the level of clarity required to identify and enforce against content most likely to cause harm to children as part of its operations (and any scaling-up). For example, Ofcom's definition of "pornography" *(content that is not text-only; produced solely or principally for the purpose of sexual arousal)* is extremely broad and will likely not match the definition of pornography used by online services.

Please see below Snap's response to each proposed category of harm.

Pornographic/Sexual Content

Ofcom's line between sexually suggestive content and pornographic content should be made clearer to services. Snap agrees that *both* sexually suggestive and pornographic content are inappropriate for minors, but platforms' treatment of each type of content is likely to be very different. In Snap's case, sexually suggestive content might be filtered from minor accounts, limited to an adult audience, and contain no penalty for the posting user (e.g. Snap has filters in place to prevent minors from viewing sexually suggestive content, but we would not "punish" a user for posting it). In contrast, posting pornographic content, even to an adult audience, would likely result in a more serious consequence to a violating user given our clear prohibition for this type of content in our Community Guidelines, and the content would be removed from the platform.

Suicide/Self-Harm

We are generally aligned with Ofcom's guidance on suicide and self harm content. However, we would welcome further clarity on *"recovery content"* or suicide or self harm content that is not *"intentionally or deliberately"* encouraging suicide or self harm. In our experience, we find evidence of intent to be helpful in making policy and enforcement decisions. However, where there is no evidence of intent, we rely on other critical criteria, such as detailed instruction or description of methods or depiction of inherently dangerous activities, to avoid over-enforcement. It would be helpful if Ofcom could draw on some of this criteria in its own proposed guidance to ensure a fully considered approach.

While we recognise Ofcom's concern that recovery content may have a negative effect on minors, we think, on balance, removing such content shared by a user, especially a user who is a minor, could cause more harm than good. For example, it could be a cry for help or derail an individual's recovery if part of their journey has been to feel empowered to share their story and inspire others. Ofcom's own research (s21.226, page 402, Volume 5) as part of its proposal to signpost children to support (US5) recognises that posting or reposting recovery content is a way for some children to indicate their desire for and gain access to support.

We believe it is also important for services like ourselves to work directly with experts on paediatric mental health and wellness and engage teens directly (e.g. Snap has a teen council) to inform our policies and keep up with trends based on our platform specificities. In Snap's case, we have been advised by experts not to remove such content – for example, where a user is posting what may be a cry for help, as long as it is not graphic and not violating any other parts of our policies (e.g. no glorification or encouragement of suicide).

Turning to suicide or self harm content that is not *"intentionally or deliberately"* encouraging suicide or self harm, there is a case to consider the benefits of social media to highlight that help is out there, especially for marginalised or deprived youth. For example, the US Surgeon General issued an advisory in 2023: *"research suggests that social media-based and other digitally-based mental health interventions may also be helpful for some children and adolescents by promoting help-seeking behaviours and serving as a gateway to initiating mental health care."* Help-seeking behaviours amongst teens may not be as straightforward or obvious as those of adults. In many cases, it might mean sharing content that is provocative, but not intentionally or deliberately encouraging self harm or suicide. This may be an attempt to connect with others going through similar feelings and/or seek help. We have noted an article published by the Journal of child psychology and psychiatry, and allied disciplines in 2020 regarding the role of digital technology in children and young people's mental health, which said:

> *"... the digital space is increasingly successfully being harnessed for the identification and treatment of mental health problems."*
> …
> *"While the instinctive reaction of adults, from parents to politicians, is to try to shut down online sites where self-harm is discussed (fearing social contagion or copycat problems), online ethnographic methods and qualitative interviews reveal that the young people who engage with these discussions not only tend to be already self-harming (implying that the causes lie elsewhere), but also that online peers are primarily supportive."*[1]

Regarding Ofcom's advice that services should take care when assessing artistic or fictional representations of suicide and self-harm that *"may romanticise or glamourise suicide in a way that is*

[1] Hollis, C., Livingstone, S., & Sonuga-Barke, E. (2020). Editorial: The role of digital technology in children and young people's mental health - a triple-edged sword?. Journal of child psychology and psychiatry, and allied disciplines, 61(8), 837–841.

harmful to children," further clarification would be helpful. We would caution against such an interpretation. Snap does not generally prohibit artistic works, such as music, poetry, artwork or fictional works, drawings, stories, paintings and other art and we pride ourselves on being a platform that celebrates self-expression and creativity. We would not generally prohibit our users from making references to popular culture on Snapchat of this nature – even if it may invoke a sense of negativity. As noted above, this could be vital to users getting the support they need; or conversely, could lead to unintended censorship and restricting freedom of expression.

We agree that it is important to provide minors (and all users) with genuine support resources, and err on the side of over-communicating these resources to users, even in instances where there may not be a serious threat of self-harm or suicide (please see our response to User Support below).

Eating Disorder Content

We are aligned with Ofcom's guidance on eating disorder content and take a strict approach on such content, given the demographic of our users.

Abusive/Hateful Content

We are generally aligned with Ofcom's proposed guidance on abusive and hateful content, and its recommendation of taking a victim-centric approach to evaluating these harms. We also agree that there is often overlap between abusive and hate content and other harms, such as bullying, harassment, and incitement of violence, and that context is critical. We have found working with third party experts, such as anti-hate NGOs, to be helpful, as trends around hate content can evolve quickly.

Bullying

Snap is generally aligned with Ofcom's proposed guidance on bullying, including taking a victim-centric approach and considering all the context available. We also agree with taking a stricter approach to bullying content when it comes to minors, such as images shared without consent that were intended to humiliate, or modified images, including images applying augmented reality effects or AI effects intended to embarrass or humiliate. We have stricter penalties for content that is intended to bully minors.

Violent Content

Snap is generally aligned with Ofcom's proposed guidance on violent content, including the potential overlapping with other harms, such as hate, violence, threats. In some cases, such as denial of tragedies or tragic events, i.e. *"content which trivialises or misrepresents violent acts, where the purpose is to normalise or discount the impact of violent behaviour,"* we also see potential overlap with harmful misinformation.

With respect to gaming content, it would be helpful to understand what Ofcom means by "realistic." On the whole, we agree that realistic, graphic violence should not be accessible by minors, but Snap does not generally prohibit content that is fantasy-based, or references popular video games or pop culture. As a creative-centric platform, we try to allow as much self-expression and creativity as possible while still keeping users safe. As an example of content that encourages, promotes or provides instructions for an act of serious violence against a person, Ofcom includes *"A music video that uses incendiary, boastful or taunting lyrics about an incident such as the injury of another individual to encourage violence."* What is not clear is whether Ofcom would expect services to remove content posted by a minor that merely references or quotes popular songs and movies. We recognise that intent and context are important as part of the consideration but further clarity from Ofcom would be helpful to avoid any unintended consequences of stifling creativity and freedom of expression.

Harmful Substances Content

Snap agrees with Ofcom's guidance on harmful substances, and has stricter penalties for glorification of the use of these substances where minors appear to be involved.

Dangerous Stunts and Challenges Content

We are generally aligned with Ofcom's guidance on dangerous stunts and challenges content. One area of concern, however, is *"Encouraging others to emulate stunts carried out by professionals that could cause serious injury if emulated,"* and the example,*"Content encouraging individuals to replicate dangerous stunts in the context of extreme sports, (e.g. snow sports, climbing, skate-boarding, or parkour), when the action is likely to cause serious harm if attempted by someone that has not been trained, and the post does not contain sufficient commentary relating to safety.*" For the encouragement of certain dangerous challenges, such as jumping out of a moving car or ingesting poison, it is clear to us that such content should not be allowed on Snapchat. With extreme sports, however, it might be difficult to differentiate admiration versus encouragement or promotion. In addition, snow sports, climbing, skate-boarding, and parkour are all sports that minors are likely to engage in, and are not inherently dangerous. Further guidance around certain contexts would be welcomed to make it clearer when Ofcom would expect services to take action on this type of content (e.g. if a user was climbing at a certain altitude with no safety harness).

Non-designated content: Body Image and Depressive Content

We agree with Ofcom that further research is needed to assess non-designated content relating to *"depressive content"* and *"body image content"* and we would urge Ofcom to develop clear guidelines on this grey area of policy.

We already have strict policies which prohibit content glorifying or promoting suicide/self harm, eating disorders or extreme dieting. However, creativity and self-expression are core tenets to Snapchat and we would not want to take an overly-zealous approach to Non-Designated Content (NDC) when its definition is sufficiently broad. For example, in the context of positive body image or teens celebrating their bodies. Furthermore, Snapchat offers Augmented Reality (AR) experiences with partners as a shopping utility that can focus on the body or parts of the body depending on the product – e.g. allowing users to try on trainers or a lipstick shade in an AR format. Our data shows that 85% of Snapchatters feel more confident in their purchases as a result of using AR.[2] Therefore, we would want to ensure that Ofcom's requirements do not unduly hamper the user experience or Snap's business interests.

While Ofcom, at a meeting with Snap, clarified that non-designated content referred to depressive content (i.e. content glorifying or encouraging feeling negative or depressed) rather than depressing content; we note Ofcom's consultation relied on research which states: *"images, video clips and text concerning or*

[2] 2022 Alter agents study commissioned by Snap Inc and Publicis Media I Base: UK Shoppers (n=1000), Base: UK Snapchatters (n=526), UK Non-Snapchatters (n=152) I Q: How much do you agree with the following statements about AR technology and shopping?

concerned with self-harm, suicide or that were otherwise negative or ***depressing*** *in nature". Molly Russell Foundation Page 204, Volume 3.*

Irrespective of the definition, we remain concerned that depressive content could inadvertently capture content that celebrates self or artistic expression (as mentioned above). For example, many of today's popular music artists have content (such as dark lyrics in their music) which could emote (or even be considered to glorify and encourage) negative or depressing feelings. However, we do not believe that this in its own right justifies the prohibition of such content. To do so, creates a severe risk of censorship and hampering freedom of expression. It is also important that users have access to support resources and we do not want to curtail any potential cry for help (subject to the content violating our Community Guidelines – see our expert advice above). This is why Ofcom must create clear definitions and guidelines to ensure services can make the right call that is balanced and proportionate. Without such clear definitions and guidelines, responsible services faced with the threat of heavy fines are likely to err on the side of caution – which may similarly lead to unintended censorship and restrictions on freedom of expression.

2. Risk Assessment Guidance

Draft Children's Register of Risk

Proposed approach

Do you have any views on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

We are broadly supportive of Ofcom's assessment of the causes and impacts of online harms. That said, we have the following feedback:

- The draft risk register refers to eight primary priority and priority content risks as well as two NDC risks (sub-sections 7.1-7.9, Volume 3). With regards to these content risks, please see our feedback on the Guidance on Content Harmful to Children above.
- With regards to search services (subsection 7.10), currently, we would not consider Snapchat to be a search service. We note in 7.10.17 that Ofcom states that a user-to-user service that includes a public search engine which operates as a general or vertical search service would be considered a 'combined' service - quoting Article 4(7) of the UK OSA. In Article 229 (2) of the UK OSA, however, it is clear that a search engine is not to be taken to be "included" in an internet service or user-to-user service if the search engine is controlled by a person who does not control other parts of the service. We think it would be helpful to be clearer in the draft risk register that when a service that includes a search service that is controlled by a different provider, e.g. Google Search or Bing, that service is not considered to be a combined service.
- Regarding the final sections (sub-sections 7.11-7.15):
    - We agree these are important topics. However, it would be worth stressing further that these are not risks themselves, but factors that may increase or decrease the likelihood of the risk of exposure to harmful content occurring. While this is explained in the introduction in Section 6.13, as the document progresses, it becomes less clear and somewhat confusing that Ofcom is referring to these matters as 'factors' that must be considered together to determine the residual risk of a service. A service may be low risk

overall, despite the presence of one or more of these factors, and they shouldn't be assessed independently. For example:

- Regarding governance, systems and processes (subsection 7.11), a service deploying fewer content moderation resources may increase likelihood of exposure to harmful content but this might be mitigated by other factors like service design. The fewer content moderation resources is not a risk in and of itself.
- Regarding 7.12 Business models and commercial profiles, we agree these models/profiles could be aggravating factors and increase the likelihood a specific company does not appropriately mitigate risks. But it is important not to demonise or conflate industry models/profiles vs specific company decisions – it is reasonable to expect that many of the user-to-user services in-scope of the Act will predicate their business growth strategy on advertising revenue and increasing user engagement. At Snap, brand trust is a critical component to our business strategy which bakes in effective and proportionate measures to mitigate the risk of harm. For example, we note that section 7.12.8 suggests that children contribute to a significant share of advertising revenues of many services. In our experience, while certain advertisers have products and services that are most relevant to teenage users, these do not account for a significant share of advertising revenue. Although we do not consider proportionate targeted advertising to be harmful, Snap has chosen not to display ads based on profiles to our teenage users in the UK.
- Regarding 7.13, we agree that features and functionalities affecting time spent using services may increase the likelihood of exposure to harmful content but only if the service does not have effective mitigations against harmful content appearing on the service in the first place.
- Regarding 7.14 wider context, similarly, we agree that having a system that recommends similar content to what a user (or similar users) have watched could increase the likelihood and/or severity of the risk minors being exposed to harmful content. However, whether this means the overall risk is high / medium depends on many other factors, such as the nature of the content they are recommending, how they are designed, the presence of effective content moderation, etc. It is important that the words 'recommender systems', 'algorithms' and 'content tagging' are not demonised by suggesting they are risks rather than factors that may influence the risk depending on the context. The same is true of user base demographics, media literacy and the rise of GenAI.
- Regarding 7.15 recommended age groups, here actually it is stressed that this factor is only one of many factors affecting risk of harmful content to children which is helpful.

- We note the reference to user access and age assurance in subsection 7.11.19 (page 225, Volume 3). We feel it is important for Ofcom to stress that Ofcom's own evidence shows this risk applies to everything from devices, to app stores to apps to websites - in other words - this is an issue that is prevalent throughout the digital ecosystem and therefore extensive research on the type and impact of various solutions need to be properly considered at the ecosystem level. In this respect, please see our feedback on Age Assurance measures in the Draft Code of Practice.
- Regarding section 7.15.3 and age groups, see our additional comment below.

Do you have any views about our interpretation of the links between risk factors and different kinds of content harmful to children? Please provide evidence to support your answer.

We do not have any additional comments here. We broadly support Ofcom's interpretation.

Do you have any views on the age groups we recommended for assessing risk by age? Please provide evidence to support your answer.

We agree that the age groups proposed, particularly as it aligns with the ICO Age Appropriate Design Code. However, we note that Section 7.15.3 states that services must consider risk of harm across these age groups to ensure appropriate mitigations. We believe it should be recognised that an alternative approach would be to consider the risk of harm for the lowest age group and apply those mitigations to all minors or all users. It is not always practical or proportionate to tailor mitigations to every age group individually. This section should clarify that increased risk to children in different age groups will normally only arise if a service does the opposite i.e. considering the risk to the oldest age group and applying those limitations to all minors or all users.

Do you have any views on our interpretation of non-designated content or our approach to identifying non-designated content? Please provide evidence to support your answer.

Please see our response above.

Consultation Questions (specifically for NDC)

Do you have any views on our interpretation of non-designated content or our approach to identifying non-designated content? Please provide evidence to support your answer.

We support the intent but are concerned about the risk of censorship and hampering freedom of expression if non-designated content categories are drafted too broadly. It is important to remember that every new category will have a significant impact on platforms in terms of cost and engineering work to modify systems, training for moderators, etc. Please also see other feedback with regard to NDC in this response.

Do you have any specific evidence relevant to our assessment of body image content and depressive content as kinds of non-designated content? Specifically, we are interested in:

- **specific examples of body image or depressive content linked to significant harms to children,**
- **evidence distinguishing body image or depressive content from existing categories of priority or primary priority content.**

Please see our response above.

Do you propose any other category of content that could meet the definition of NDC under the Act at this stage? Please provide evidence to support your answer.
No, we do not believe this is necessary at this stage for Ofcom's Risk Register. Snap has for a long time managed its own categories, which are prohibited by our Terms of Service (including our Community Guidelines). While our prohibited categories go beyond the PPC and PC categories, they reflect Snap's

experience of operating Snapchat and the sort of community that we wish to flourish on Snapchat and may not be relevant to other services.

For example, we are very wary of the risk of harmful false information. We disallow all political content from Spotlight (our broadcast platform for User Generated Content) unless it's from trusted news partners and creators, and pre-moderate that surface to ensure that other such political content is not distributed. This safeguard ensures that Snap is not algorithmically promoting political statements from unvetted sources, and generally reflects Spotlight's function as an entertainment platform. This will not be suitable for other platforms that are intended to distribute an unvetted feed of algorithmically curated political information.

Children's Risk Assessment Guidance and Children's Risk Profiles

NOTE: there is an Error in Figure 12.1 (page 46) - should be children's risk assessment in the third bubble.

What do you think about our proposals in relation to the Children's Risk Assessment Guidance? Please provide underlying arguments and evidence of efficacy or risks that support your view.

What do you think about our proposals in relation to the Children's Risk Profiles for Content Harmful to Children? Please provide underlying arguments and evidence of efficacy or risks that support your view.

Specifically, we welcome evidence from regulated services on the following:

- **Do you think the four-step risk assessment process and the Children's Risk Profiles are useful models to help services understand the risks that its services pose to children and comply with its child risk assessment obligations under the Act?**
- **Are there any specific aspects of the children's risk assessment duties that you consider need additional guidance beyond what we have proposed in our draft?**
- **Are the Children's Risk Profiles sufficiently clear and do you think the information provided on risk factors will help you understand the risks on your service?**
- **If you have comments or input related to the links between different kinds of content harmful to children and risk factors, please refer to Volume 3: Causes and Impacts of Harms to Children Online which includes the draft Children's Register of Risks.**

We broadly agree with Ofcom's proposals and are pleased Ofcom has closely aligned its approach with that proposed for the Illegal Harms Code, which we also support.

Please see the areas where we disagreed or sought further clarification in our Illegal Harms Code consultation response as we have similar feedback for the Content Harmful to Children's Code - in particular with regards to: (1) a materiality / frequency threshold for updates, (2) changes which should not be considered significant; (3) mapping of content harmful to children against categories in other similar legislation; and (4) the extent to which enhanced evidence is required.

Our comments on the Draft Guidance on Content Harmful to Children and on the Draft Code should also be applied to the risk assessment. For example, we have expressed concerns about the need for clear guidance on NDC if included by Ofcom in the Children's Register of Risks and this concern is also relevant to the assessment of risk from NDC under the Risk Assessment.

3. General questions

Our proposals for the Children's Safety Codes (Section 13)

Proposed measures

Do you agree with our proposed package of measures for the first Children's Safety Codes? If not, please explain why.

Yes, while there are certain areas that we disagree with or have made requests for clarification, overall we support the package of measures that Ofcom has proposed for the first Children's Safety Codes. An evidence based approach is critical to objectively set an appropriate set of mitigation measures to protect children from harmful content and in this respect Ofcom work is impressive and world-leading.

Evidence gathering for future work

Do you currently employ measures or have additional evidence in the areas we have set out for future consideration? If so, please provide evidence of the impact, effectiveness and cost of such measures, including any results from trialling or testing of measures.

Yes, Snap currently employs:

- **Automated content moderation** - We are conscious of the same issues that Ofcom has identified: accuracy, effectiveness and lack of bias and conduct careful assessment of any new automated content moderation tools to determine whether and how to deploy such measures in a manner that appropriately balances privacy, safety and security. Please see our European transparency report for more information regarding our automated content moderation.

- **Generative Artificial Intelligence** - Snapchat has a number of services that may handle generative AI content, which Snap divides into: (a) generative services that may give rise to content creation risks, such as our My AI chatbot service, which are generally not user-to-user services; and (b) dissemination services that allow users to share content they have created using generative AI with other users, such as our Spotlight service, which are generally user-to-user services. It is important to recognise that user-to-user services i.e. dissemination services can be handling generative AI content that can come from any source, including potentially disreputable services that have not implemented any measures that are often proposed for generative AI services, such as watermarking and labelling. It is also significant to note that the risks associated with generative AI are not new; online services have been dealing with the challenges of 'photoshopping' for many years. That said, Snap is aware that there is intense interest and concern surrounding the ways in which advancements in generative AI technologies are impacting online platforms. In this respect, Snap's focus with regards to its user-to-user services and dissemination risk associated with generative AI content is also 'tech neutral'. Our mitigation measures are focused on preventing illegal content or content that otherwise violates Snap terms whether or not generative AI technology has been used. We are reviewing Ofcom's recently published research on the effectiveness of measures to identify and address harmful AI-generative content on online platforms with interest.

- **Impact of choice architecture** - As noted above in the draft Children's risk register section, we agree that features and functionalities affecting time spent using services and engagement with services is a factor that might increase the likelihood of exposure to harmful content but only if the service does not have effective mitigations against harmful content appearing on the service in the first place. Screentime is not proven to be a risk in and of itself[3]. What is most important is what a user is doing during the time they spend online (and offline). For example, while Snap does not currently use it, the BBC makes use of autoplay functionality in the iPlayer, similar to many other Video on Demand (VOD) services. Does that make those VOD services a risk? Existing research shows that what is most important is whether a service has the potential for harmful content and if so effective mitigations to prevent exposure to that content.

- **Children of different ages** - We believe the focus on setting expectations on protections for all children under the age of 18 is the right approach at this stage. As noted in our response above on the draft Children's risk register, while we agree that risk assessments must consider risk of harm across these age groups to ensure appropriate mitigation, we believe it is still important to recognise that services may choose to apply the same set of mitigations to all age groups provided they address the risks across all of those age groups. It is not often practical or proportionate to implement tailored mitigations for every age group individually, and if Ofcom proposes further research in this area, we would recommend taking that into account.

- **Parental Controls** - We agree with Ofcom's inclusion of parental controls as a future area of focus and we note that Ofcom has already considered the use of these tools for those services in-scope of the Video Sharing Platforms Regulation (including Snap). We continue to receive positive feedback from parents who use our Family Centre and the tools it offers, and periodically seek feedback on our approach from our Safety Advisory Board and Teen Council. However, one area that remains a challenge is awareness, despite our awareness raising initiatives. We would like to encourage Ofcom (including as part of its legal duties to consider the role of app stores and device operating systems (OS)) to consider how services can work together to create interoperable parental tools that might allow parents to gain the visibility and use of each service's parental tools in one place to easily provide support to their children and encourage healthy online habits. For example, parents may already have access to accounts for certain services (such as those provided by companies that also provide device OS and app stores) and can easily access the centralised parental control tools of those services from those accounts. If Snap's Family Centre could be accessed through the centralised parental tools of those services, we believe that would radically improve usability and awareness for parents (and better protect children accessing a wide range of services). We recognise that this is a topical issue for the British public which continues to resurface amongst parental community groups (e.g. Smartphone Free Childhood) and civil society.

Are there other areas in which we should consider potential future measures for the Children's Safety Codes? If so, please explain why and provide supporting evidence.

Yes. Ofcom has an obligation under Section 161 of Online Safety Act to assess and report on the use of app stores (and related device OS accounts) by children. This is critical for establishing a more proportionate, interoperable age assurance framework that will properly protect children across the entire online ecosystem, which the current focus on individual services alone cannot achieve. The Government

[3] For example, see the work of Professor Pete Etchell (https://www.peteetchells.com/)

cannot use its power under section 215 of the Online Safety Act to specifically regulate app stores until this report is produced and this seems more important and more impactful than some of the potential future measures that Ofcom is proposing to consider. See also our similar feedback in response to the age assurance mitigation measures in the Children's Safety Codes.

Developing the Children's Safety Codes: Our framework (Section 14)

Do you agree with our approach to developing the proposed measures for the Children's Safety Codes? If not, please explain why.

Yes, we agree with Ofcom's approach.

Do you agree with our approach and proposed changes to the draft Illegal Content Codes to further protect children and accommodate for potential synergies in how systems and processes manage both content harmful to children and illegal content? Please explain your views.

Yes, we agree with Ofcom's approach. Aligning systems and processes as far as possible, and only building or adding additional measures when necessary to specifically protect children, must be more efficient, effective and reduce costs, resulting in better protection of minors, and reflects how responsible companies operate in practice.

Do you agree that most measures should apply to services that are either large services or smaller services that present a medium or high level of risk to children? Do you agree with our definition of 'large' and with how we apply this in our recommendations? Do you agree with our definition of 'multi-risk' and with how we apply this in our recommendations? Do you agree with the proposed measures that we recommend for all services, even those that are small and low-risk?

No, we do not fully agree with this approach. We would reiterate the point that we made in the Illegal Harms consultation, that we do not consider it appropriate to apply measures to large services regardless of the level of risk. We believe that all user-to-user services should face the same measures in accordance with their risk profile. The number of users, particularly such a large threshold of 7 million, should not be the qualifying criteria for the reasons we have explained in our Illegal Harms consultation submission (see Annex A for more details):

- There is evidence for this in the General Risk Factors (Annex 5) to the Illegal Harms consultation, which states that:
    - *Low capacity or early-stage services may increase the likelihood of different illegal harms as they may have limited technical skills and financial resources to introduce effective risk management.*
    - *A fast-growing user base may negatively affect effective risk management, given the increased scale and sophistication of the moderation technologies and processes required to keep track of a fast-growing user base (particularly since the sources of risk can change quickly as the user base develops).*

  We agree with these general risk factors and they particularly arise with smaller services.

- It is also important to recognise that the requirements of the OSA will inevitably have an impact on the speed of online service innovation and development, as companies will be concerned to ensure they have met all the accountability steps before they make changes or release new products. In the online world, where popularity of services can change very suddenly, this could be a competitive advantage where a company with a sizable user base is able to gain a significant foothold in the market before having to put in place equivalent governance and accountability measures as a 'large service'. We have also raised these concerns in respect of the DSA and its thresholds in the EU.

4. Governance and accountability

Do you agree with the proposed governance measures to be included in the Children's Safety Codes? Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.

Yes, we generally agree with Ofcom's proposals in relation to governance and accountability. As with our Illegal Harms consultation response, there are some areas, however, where we would like further clarification. In particular, it would be useful for Ofcom to clarify who it would envisage would be the person accountable to the governance body for compliance and the senior members of staff who make decisions. As in our previous response, we believe the intention is that: (a) the 'person accountable to the governance body for compliance' is equivalent to the EU's DSA Head of Compliance role (and could be performed by the same individual); and (b) the 'staff who make decisions' are the most senior Product/Engineering and Operations managers responsible for deciding on and implementing the risk mitigation measures. We would be grateful if Ofcom could clarify further whether our understanding of the intention for these measures in the Code is correct in this respect.

We note that some of the measures in the draft Child Protection code are limited to large and multi-risk services. We would reiterate the point that we made in our response to the Illegal Harms consultation, that we do not consider it appropriate to limit governance and accountability measures to large and multi-risk services. We have set out our reasons again in our response to the General Questions above.

Do you agree with our assumption that the proposed governance measures for Children's Safety Codes could be implemented through the same process as the equivalent draft Illegal Content Codes?

Yes, this would also be our assumption. Separate processes for each code would not be effective, reasonable or proportionate.

5. Age assurance measures

Do you agree with our proposal to recommend the use of highly effective age assurance to support Measures AA1-6? Please provide any information or evidence to support your views.

Snapchat does not have as its principal purpose the hosting PPC or PC content and PPC or PC content is prohibited on Snapchat. Snap also does not believe that the user-to-user services we offer on Snapchat for which we have a significant recommender system are high or medium risk for PPC or PC content, given the nature of our service and the extensive mitigations we already have in place. [REDACTED/CONFIDENTIAL]

That said, we fully support Ofcom's proposal to recommend the use of Highly Effective Age Assurance (HEAA) to support Measures AA1-AA4. Services whose principal purpose is hosting PPC or PC content, or do not prohibit PPC or PC content, and may be accessible to children should be applying service-wide or content control measures supported by the use of highly effective age assurance. We expect that these measures will only impact a relatively small, niche set of service providers on the basis of Ofcom's criteria.

Are there any cases in which HEAA may not be appropriate and proportionate?

While we broadly support the use of HEAA to support Measures AA5-6 (i.e. to support recommendation systems on services that are at a high or medium risk for PPC or PC content), we have several concerns about whether the proposal as currently drafted would be appropriate and proportionate:

- *Multi-Service Apps:* Measures AA5-6 appear to be written based on a simple single purpose service, with the assumption that if a recommender system is being used, it applies to the entire service. Some apps like Snapchat are, however, more complex and made up of multiple services, such as messaging and stories, camera, lens, map and video sharing, and only some of these services use recommender systems. The thresholds for the application of AA5-6 are not clear whether the measures apply if the entire multi-service app is assessed to be high/medium risk of PPC or PC content, or if it refers only to the user-to-user services within the app that are using the recommender system. We would interpret the requirement to mean the latter, as AA5-6 is clearly focused on the potential increased risk arising from recommender systems. However, to ensure legal certainty, the Code should be written so that the threshold for the AA5-6 measures is very clear to all stakeholders.

- *Varying Sophistication:* Annex 5 makes clear that there are different types of "recommender systems" and that Ofcom is focused on 'content recommender systems' only. Measures AA5-6 are not therefore intended to include systems that underpin search functionalities or network recommender systems. However, there are also significant differences between the level of sophistication and the breadth of the source content between different content recommender systems. For example, Snap uses the type of 'content recommender system' that we believe AA5-6 is intended to capture for its broadcast content services. [REDACTED/CONFIDENTIAL] We believe this is already the intention given the explanation of recommender systems in 20.1 of Annex 5 ("Recommended content is sourced widely") but would welcome further clarification that Measures AA5-6 would not apply to recommender systems with: (1) a narrow range of UGC source material; (2) source material from only vetted publishers; and/or (3) a low level of capability i.e. the results are not highly personalised due to limitations in the algorithm or number of signals being relied on to deliver content recommendations.

- *Interoperable HEAA:* While Measures AA1-4 will apply to a relatively small, niche set of service providers, we expect Measures AA5-6 will impact a much more significant number of services. As a result, we do not understand how Ofcom can require Measures AA5-6 to be supported by HEAA when it notes itself in AA10.67 that the development of interoperable solutions is at an early stage and it cannot make any specific recommendation about the principle of interoperability. None of the kinds of age assurance that Ofcom has explicitly identified as being highly effective in its 'draft guidance on highly effective age assurance' in A10 are in fact interoperable. Ofcom has not at all considered the cumulative cost of requiring each affected service provider to individually carry out non-interoperable age assurance checks and whether this is an appropriate and proportionate approach as opposed to proposing that affected service providers work together to implement interoperable age assurance. An approach that shares the results of age assurance checks across all of the affected services may only need one check per user and inevitably result in a much lower cumulative cost and so would be a more appropriate and proportionate approach. Ofcom must have proper regard to interoperability under Section 82(3)(c) and Paragraph 12(2)(g) of Schedule 4 of the OSA and we do not believe it will be able to demonstrate that it is appropriate and proportionate to support Measures AA 5-6 with HEAA based on its current limited analysis of interoperability.

In this case, are there alternative approaches to age assurance which would be better suited?

Yes, we believe there are alternative approaches to age assurance that would be better suited:

- *Application of Recommendation System Measures to All Users:* In paragraph 15.305, Ofcom recognises that the OSA provides flexibility for user-to-user services in meeting the duty to protect children from encountering PPC and PC content (except where PPC is not prohibited). Paragraph A12.47 also recognises the possibility that some services may offer users access to child appropriate services where users do not have to confirm their age. However, in paragraphs 15.304-15.311, Ofcom has not explicitly referenced the possibility that service providers may deploy other mitigations to avoid the need to apply HEAA to its recommender systems. Snap applies the majority of its privacy, safety and security mitigation measures to all of its users of Snapchat, regardless of age. For example, we take down illegal content and content that is otherwise violating our terms (including PPC and PC content) so that it can no longer be accessed by any of our users. We do not seek to suppress it for certain age groups but continue to allow access for other age groups. Ofcom should be clear that if mitigations measures in the Code that are required to be supported by age assurance (i.e. AA5-6 in the current draft) are in fact applied to all users, regardless of their age, then this would serve as an appropriate alternative approach to age assurance.

- *Interoperable Age Assurance:* As we note above, in AA10.67 Ofcom has said that the development of interoperable solutions is at an early stage and it cannot make any specific recommendation about the principle of interoperability. That is not the case. Under paragraph 14 of Schedule 4, Ofcom has the power to make different provision for different purposes within the Code and may, in particular, make different provision with regard to user-to-user services of different kinds and otherwise differentiate between Part 3 services, ***and between providers of such services***, in such manner as Ofcom consider appropriate. Some providers ("Gatekeeper U2U Service Providers") of user to user services with recommender systems are also providers of 'core platform services' like app stores and device operating systems, as well as for profit digital wallet solutions. Those providers generally require users to access their user-to-user services using the same account as their core platform services i.e. the same account is used for device OS, app store and user-to-user services. Where those Gatekeeper U2U Service Providers are required to implement HEAA for their user-to-user services under this Draft Code, it would be appropriate and proportionate under the principle of interoperability for Ofcom to require such Gatekeeper U2U Service Providers to share their user account's assured age with other, third party user-to-user service providers (and provide means for those third party user-to-user services to share an assured age signal with the Gatekeeper U2U Service Providers). Those Gatekeeper U2U Service Providers already provide core platform services, and often also already provide interoperable identity and wallet services for user-to-user services. As they are Gatekeeper U2U Service Providers, they already have the capability, infrastructure and means to meet this requirement. The additional cost to them would be limited and if necessary could be claimed back via the existing fee mechanisms that Gatekeeper U2U Service Providers have to share infrastructure costs with user-to-user services.

Ofcom could therefore include "HEAA signal shared by other user-to-user service providers for the same user" in the list of "Kinds of age assurance that could be highly effective" so all affected user-to-user services would feel comfortable relying on the shared signal (unless they have other reasons to believe it is inaccurate - which could then also be shared between user-to-user services). This direction from Ofcom would rapidly result in a HEAA solution for all user-to-user services that is fully interoperable, consistent, has the smallest impact on users but with the greatest impact on online safety and a vastly reduced cumulative cost across all affected user-to-user service providers. It would then allow all user-to-user service providers to apply the assured age to support all of the mitigation measures Ofcom has proposed in the Code, and any other mitigation measures that each user-to-user service has chosen to implement to protect the privacy, safety and security of its users. We understand that Ofcom has refrained from adding this option to the consultation on the basis that Gatekeeper/App Store providers are not in-scope of the OSA (see below). However, this should not impede on the inclusion of this approach. We note that Ofcom has suggested a similar approach of relying on mobile-network operators' (voluntary) age checks and other methods which rely on third party organisations that are not in-scope of their regulatory remit (Annex 10).

- *Privacy and Security:* We appreciate the collaboration between Ofcom and the ICO to produce the advice for service providers on the application of data protection principles to HEAA. We agree that many HEAA mechanisms raise privacy concerns and service providers will only be able to implement such measures if they can be satisfied data protection requirements will be met. Two of the most challenging privacy aspects of implementing HEAA in practice will be: (1) minimising the extent of the sensitive data that may need to be collected and checked, for example, biometric data, important ID documents or authentication details for digital wallets; and (2) ensuring appropriate security, particularly when third party vendors are involved as any such engagement increases risk and requires extensive, ongoing due diligence assessments. Although we have started to see sophisticated solutions to address privacy concerns, including double anonymisation systems and on-device mechanisms, these are still relatively new and we have also seen prominent identity verification solutions suffer high profile security breaches. The best way to minimise this risk is to minimise the number of age checks and companies involved, and ensure that where sensitive data is being processed, it is managed by sophisticated companies with proven track records. This is another reason supporting the interoperable age assurance approach, centrally managed by sophisticated Gatekeeper U2U Services Providers, that we have outlined above. Ofcom and the ICO must consider this as a potential means for service providers to collectively reduce privacy and security risks.

- *App Stores:* We note that Ofcom has an obligation to consult and report on the potential role of App Stores under Section 161 of the OSA. Ofcom should accelerate its analysis and production of this report. In doing so, we believe it highly likely to receive sufficient evidence to recommend to the Secretary of State to bring app stores (and associated device OS/App Store accounts) into the scope of the Act. This will give Ofcom the formal powers to ensure it can implement an alternative approach to age assurance that would be fully interoperable and better suited for all the reasons set out above (*Interoperable Age Assurance*).

- *Inferred age models:* [REDACTED/CONFIDENTIAL] We would like to see clear recognition that such inferred age models can be considered a HEAA means, subject to meeting the criteria that age assurance should be highly effective and any of the relevant constraints set out in the OSA. Where inferred age is conducted in-house by the service provider, this helps limit privacy and data security concerns and is less susceptible to circumvention. This is because it is based on a factual reflection of the user's actual conduct and characteristics on the platform and is an indicator of their probable age (versus reliance strictly on static external documentation that can be falsified, or on estimation technology that carries an increasing margin for error for younger and diverse audiences).
- *Parent Certification:* We note that Ofcom has not explicitly discussed the use of parental certification of a self-declared age as a HEAA mechanism. If the age assurance measures proposed by Ofcom in the Codes would result in a significant proportion of the users of a user-to-user services to be age checked, we believe parental certification could be an appropriate and proportionate alternative to age assurance. This would involve a parent, guardian or other responsible adult, who has verified their age as over 18, to provide certification for the users' self-declared age. This would allow adults that do not have any ID or do not want to share personal data with the service provider (or their HEAA vendor) to still have a proportionate means of accessing the service. While this method of age verification would be subject to circumvention risk, we believe that is true of most, if not all HEAA methods (where a person could stand in for the real user). We suggest that this is considered and included in Ofcom's HEAA recommendations.

Finally, we would just note that there are other initiatives looking at enabling interoperable age assurance than the three initiatives mentioned in A10.69. Ofcom is also aware, for example, of the CIPL / We Protect initiative which has several workgroups looking to establish a holistic, interoperable and principles-based approach to age assurance.

Do you agree with the scope of the services captured by AA1-6?

We support Ofcom's proposals that the age assurance mitigation measures must apply to all user-to-user services that meet the relevant thresholds. If that were not to be the case, age assurance measures would have very significant market impacts as Ofcom identities in paragraphs A12.46-47.

Do you have any information or evidence on different ways that services could use highly effective age assurance to meet the outcome that children are prevented from encountering identified PPC, or protected from encountering identified PC under Measures AA3 and AA4, respectively?

[REDACTED/CONFIDENTIAL]

Do you have any comments on our assessment of the implications of the proposed Measures AA1-6 on children, adults or services? Please provide any supporting information or evidence in support of your views.

Yes, we have several comments on Ofcom's assessment of the implications where we have contrary views and/or conflicting evidence, or where we believe there are additional aspects that Ofcom should consider:

- *Third party Age Assurance Costs:* In our experience, the costs associated with third-party age assurance methods are significantly higher than those outlined by Ofcom in Annex 12. In particular, the per-check unit cost. We understand Ofcom used the UK Government's impact assessment analysis to support its conclusion, and based on this Ofcom has estimated the following per-check unit costs: (a) a low estimate of £0.05 per check (approx $0.06) and (b) a high estimate of £0.20 (approx $0.26) (see paragraph A12.29). Based on this, Ofcom calculated the following illustrative cost estimates:

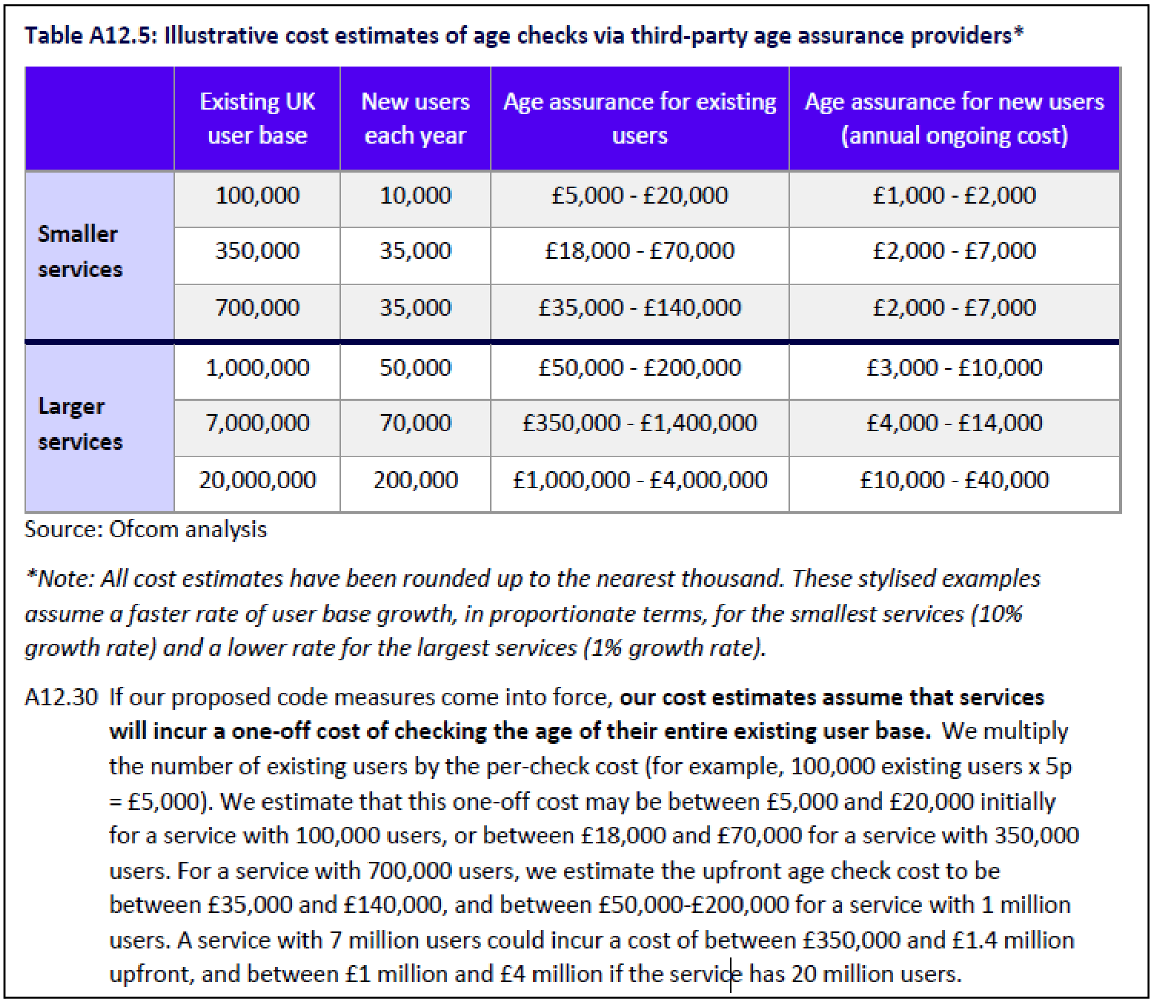


Table A12.5: Illustrative cost estimates of age checks via third-party age assurance providers*

	Existing UK user base	New users each year	Age assurance for existing users	Age assurance for new users (annual ongoing cost)
Smaller services	100,000	10,000	£5,000 - £20,000	£1,000 - £2,000
	350,000	35,000	£18,000 - £70,000	£2,000 - £7,000
	700,000	35,000	£35,000 - £140,000	£2,000 - £7,000
Larger services	1,000,000	50,000	£50,000 - £200,000	£3,000 - £10,000
	7,000,000	70,000	£350,000 - £1,400,000	£4,000 - £14,000
	20,000,000	200,000	£1,000,000 - £4,000,000	£10,000 - £40,000

Source: Ofcom analysis

**Note: All cost estimates have been rounded up to the nearest thousand. These stylised examples assume a faster rate of user base growth, in proportionate terms, for the smallest services (10% growth rate) and a lower rate for the largest services (1% growth rate).*

A12.30 If our proposed code measures come into force, **our cost estimates assume that services will incur a one-off cost of checking the age of their entire existing user base.** We multiply the number of existing users by the per-check cost (for example, 100,000 existing users x 5p = £5,000). We estimate that this one-off cost may be between £5,000 and £20,000 initially for a service with 100,000 users, or between £18,000 and £70,000 for a service with 350,000 users. For a service with 700,000 users, we estimate the upfront age check cost to be between £35,000 and £140,000, and between £50,000-£200,000 for a service with 1 million users. A service with 7 million users could incur a cost of between £350,000 and £1.4 million upfront, and between £1 million and £4 million if the service has 20 million users.

We believe the costs are significantly higher in practice. [REDACTED/CONFIDENTIAL]

- *Indirect Costs* - We agree with the issues that Ofcom has highlighted with regards to Indirect Costs in paragraphs A12.46-47. However, we are concerned that this analysis is brief and as Ofcom notes, it does not appear to have specific evidence on how users will behave when faced with age assurance checks other than two examples relating to Aylo and a video-sharing platform. [REDACTED/CONFIDENTIAL]
- We also think that Ofcom should take account of:
  - [REDACTED/CONFIDENTIAL]
  - *Multi-service apps:* Some multi-service apps have services that use recommender systems but those services are not the primary purpose or use of the app. [REDACTED/CONFIDENTIAL]
  - We do not believe Ofcom's cost analysis in Annex 12 has properly taken this into account when considering the indirect costs on services. [REDACTED/CONFIDENTIAL]
- *Minimum age restrictions* - We note that Ofcom has concluded that it is not proportionate to require services to implement HEAA mechanisms to enforce minimum age. We fully agree with Ofcom's conclusions in this respect for the reasons set out by Ofcom in paragraphs 15.312-315, as well as those we have outlined in point 2 (Indirect costs) above.

Do you have any information or evidence on other ways that services could consider different age groups when using age assurance to protect children in age groups judged to be at risk of harm from encountering PC?

No. In general, Snap's primary approach is to apply the mitigations that we believe are necessary for our youngest users (i.e. 13 year old) to everyone using our platform and ensure that all of our users benefit from our privacy, safety and security measures. In cases where we do apply additional measures for specific age groups in the UK, these are currently applied to 13-17 together, based on what is suitable for the youngest age group in that bracket (i.e. 13 years old).

6. Recommender systems

Do you agree with the proposed recommender systems measures to be included in the Children's Safety Codes? Please confirm which proposed measure your views relate to and provide any arguments and supporting evidence.

Yes, broadly speaking, we agree with the three measures proposed for recommender systems. It is important that content recommendation systems on services that are a medium or high risk of PPC and PC are designed not to recommend that content to minors, and minors have a means to inform the service when they do not like the content they are being recommended.

However, there are a number of areas where we believe the measures would benefit from greater clarity and/or flexibility:

1. In respect of RS2, while we support the intent behind the inclusion of the two additional NDC categories, we have doubts about how such content could be easily and consistently identified in

practice, and we are concerned that seeking to do so would unduly impact fundamental rights of freedom of speech and access to information. Please see our comments on this above.

2. In respect of RS1 and RS2, while we support the built-in flexibility in allowing platforms to use “all relevant available information” to identify content likely to be PPC and PC, rather than a rigid standard, it would be helpful to further guidance on the extent to which services need to take into account the relevant available information identified in paragraph 20.7:

    a. *User Feedback*: User complaints and trusted flaggers are an invaluable source of information regarding illegal and harmful content. However, they are not always correct. As a result, we do not believe it is appropriate or proportionate for the Code to be recommending that we consider this feedback to be ‘relevant available information’ until it has been reviewed by our content moderation processes and confirmed to be illegal or otherwise violating our terms as harmful content.

    b. *Negative Feedback:* Similarly, in paragraphs 20.7 and 20.158, Ofcom suggests that negative feedback could be used to provide relevant information to the service provider about which content could be NDC. In our experience, when users select our existing ‘I don’t like’ feature to effectively down vote content, this can be for many different reasons. These may not be reasons that have anything to do with the privacy, safety or security of our teenage users. It may simply be that they just disagree with something in the content e.g. if a person in the video says they like cats more than dogs. As such, we would not recommend that negative feedback is considered ‘relevant available information’ unless content is confirmed to be illegal or otherwise harmful by our content moderation processes.

3. In respect of RS2, it would be helpful if the recommendations articulated what *“reducing the prominence of"* means from a practical and operational standpoint. If content is confirmed to be PC (or PPC) following reports from users, trusted flaggers and other trusted sources, Snap’s approach is simply to remove the content for all users and it would not therefore be eligible for recommendation. We assume approaches that go above and beyond merely reducing the prominence of such content would be acceptable under the Code - but we believe this should be made clear to take into account the rights for freedom of expression. However, given the harmful nature of PC content, we are unclear why Ofcom is not proposing the same measure for RS2 as it is for RS1. Content that has been made available to minors on one platform due to less stringent approaches can more easily find its way onto another platform, which places additional strain on content moderation processes. Unless the service applies different rules for different minor age groups, we believe such content should not simply be ‘down ranked’, it should be filtered out entirely for all minors.

If you responded to our illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.

Yes. Please refer to Annex A (section 7 Recommender Systems, p.23). Our comments apply equally with respect to Measures R1 and R2.

Are there any intervention points in the design of recommender systems that we have not considered here that could effectively prevent children from being recommended Primary Priority Content and protect children from encountering priority and non-designated content?

Our experience with our own platform indicates there is a very low prevalence of content associated with PPC and PC on Snapchat's public surfaces, and our recommender systems are not a material risk factor for this type of harm. See also our response above with regards to NDC.

Is there any evidence that suggests recommender systems are a risk factor associated with bullying? If so, please provide this in response to Measures RS2 and RS3 proposed in this chapter.

Our experience with our own platform indicates there is a very low prevalence of content associated with bullying and harassment on Snapchat's public surfaces, and our recommender systems are not a material risk factor for this type of harm.

We plan to include in our RS2 and RS3, that services limit the prominence of content that we are proposing to be classified as non-designated content (NDC), namely depressive content and body image content. This is subject to our consultation on the classification of these content categories as NDC. Do you agree with this proposal? Please provide the underlying arguments and evidence of the relevance of this content to Measures RS2 and RS3.

Self-harm and suicide content are generally well-known categories for which services should already have well established practices for identify and moderating. As noted above in the Draft Guidance on Content Harmful to Children section, depressive and body image content are broader, more subjective and therefore far more difficult to detect. We therefore do not believe depressive and body image content should be included in RS2 as, without very clear, universal guidance on what these categories are intended to cover, it would be extremely difficult for content review and Trust & Safety teams to consistently identify and/or take appropriate moderation actions. This could have a chilling effect on freedom of expression and access to important information (such as support resources) that, while suitable for most, could conceivably exacerbate underlying mental health issues in a few individuals which platforms are unaware of.

We believe a better approach would be to apply RS3 to all services with a recommender system, regardless of whether they are high or medium risk of PPC, PC or NDC content. This would allow services to monitor content that is repeatedly negatively flagged by minors and ensure this signal is appropriately incorporated into recommendation systems to limit the prominence of such content.

7. Terms of service

Do you agree with the proposed Terms of Service/Publicly Available Statements measures to be included in the Children's Safety Codes? Please confirm which proposed measures your views relate to and provide any arguments and supporting evidence. If you responded to our illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.

Can you identify any further characteristics that may improve the clarity and accessibility of terms and statements for children?

Much of our response on the measures relating to Terms of Service has already been covered by our previous submission on Ofcom's Illegal Harms consultation (see Annex A). However, we have provided further commentary on TS1: *Terms and statements regarding the protection of children should contain all information mandated by the Act* as follows*:*

We are generally aligned that user-to-user services likely to be accessed by children should include in their Terms of Service (or Community Guidelines) how they will prevent or protect children from encountering PPC and PC, including what proactive technology the service uses to safeguard children and how it works. We also agree with Ofcom that services should apply such measures consistently, and that they should disclose what policies and processes govern the handling and resolution of complaints related to such content.

However, for the reasons we have set out in detail above, we are concerned on how services are expected to address NDC in our Terms of Service as part of these measures, given the broad and subjective interpretation of depressive and body image content; and the legitimate exposure to it in many cases (e.g. songs containing depressive lyrics). Ofcom mentions in its Rights Assessment that it does not find that these measures would constitute an interference with a user's freedom of expression or privacy rights. On the contrary, we are concerned that services may be forced to err on the side of caution without clear definitions in place to appease regulatory bodies (e.g. heavy fines). This could lead to unintended consequences of over-enforcement, censorship and encroachment on freedoms of expression and information.

We recognise that this is an area that Ofcom has identified for further research. We urge Ofcom to develop clear definitions and guidelines on NDC to support services in their explanation and handling of it in a way that balances safety while preserving the fundamental rights of expression and information of users.

Do you agree with the proposed addition of Measure 6AA to the Illegal Content Codes? Please provide any arguments and supporting evidence.

We note that the categorisation thresholds are yet to be agreed by the Secretary of State and put onto a statutory footing. Ofcom is also yet to publish their consultation on the additional duties that will apply to categorised services. Given this measure will fall within those additional duties, we think it would be prudent of us to comment on it as part of this consultation response to ensure a full consideration of all the measures that would apply to categorised services under the additional duties.

8. Content moderation

Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.

Much of our response on the measures relating to Content Moderation has already been covered by our previous submission on Ofcom's Illegal Harms consultation (see Annex A). We agree with the majority of Ofcom's recommendations set out in Section 3B of Annex 7 and we have provided the following supplementary response:

CM1: Services should have in place content moderation systems and processes designed to swiftly take action against content that is harmful to children.

Ofcom specifically states that *"[s]ervice providers may set about making judgements as to whether individual pieces of content should be classified as content that is harmful to children (by reference to Ofcom's Guidance on Content Harmful to Children if they wish), for the express purpose of complying with the children's safety duties;* ***OR*** *[i]f service providers are satisfied that their terms of service are cast broadly enough to necessarily cover PPC, PC and NDC content, and secure that appropriate action is taken when that content is identified, service providers may choose to apply those when moderating content to secure compliance with the children's safety duties."*

We note that this measure has been adapted for content that is harmful to children based on the measure outlined in the Illegal Harms consultation. Our response very much echoes our original consideration as part of the Illegal Harms submission whereby we recommend that Ofcom not adopt the first option as an appropriate course of action and our preferred approach follows the second option.

However, on NDC we have set out our concerns on how we would be expected to address this in our Terms of Service above without clear definitions and guidelines from Ofcom. We broadly agree with Ofcom's recommendations that where services do not include NDC within their Terms of Service, they can rely on their **discretion** to choose what type of "further content moderation" to employ (downranking, overlays/labels, or restricting access to certain content for children). We suggest this discretion extends to whether the content requires any moderation action in the first place – as we have noted above, we believe that some of the definitions of NDC (e.g. artistic content; popular culture references) should not fall within scope, as they risk a moderation culture of censorship and one which restricts freedom of expression and information. We recommend that Ofcom provides robust guidelines around NDC to support any subsequent moderation practices, which do not penalise the user for content that, while may arouse negative feelings towards body image and suicide, may not actually be harmful.

It is also worth noting that employing an additional layer of content moderation especially for children (which will require specialised internal policies and guidelines, training for moderators, and resources to implement) will result in increased costs for services.

9. User support

Do you agree with the proposed user support measures to be included in the Children's Safety Codes? Please confirm which proposed measure your views relate to and provide any arguments and supporting evidence. If you responded to our Illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.

US2: User blocking and muting and US3: Disabling comments

We have already responded to these measures as part of our submission on the Illegal Harms consultation, and we agree these measures can help reduce the risk of children encountering PPC and PC. As the Protection of Children consultation recognises, Snap already deploys user blocking and muting options. As noted in our Illegal Harms consultation, we believe there is opportunity to enhance the 'disabling comments' measure: in the case of Snapchat, by default, the user must manually approve any inbound comments before they can appear publicly. We believe this default achieves a stronger protection than that proposed in the Code, which proposes a reactive disabling after the comment has already been posted. In contrast, our approach effectively prevents any users from making comments that contain harmful content posted by another user.

US1: Group chats

[REDACTED/CONFIDENTIAL]

While we recognise that Ofcom's measure will give children more control and choice over their online interactions, we do not believe that this measure alone will have the desired effect to help prevent children from encountering PPC or illegal harms content, such as child sexual abuse material, in group chat contexts.

In the case of Snapchat, which is predominantly used as a messaging platform [REDACTED/CONFIDENTIAL], there is a real risk of adding friction to the user experience and creating decision fatigue for young users by having to opt-in/out of every group chat. Alongside other important safety prompts they receive on Snapchat (and other recommended prompts by Ofcom), teens could simply tune out of making an informed decision; instead choosing to 'accept' all prompts for group chats (as seen in the case of cookie consent/fatigue).

On that basis, we recommend that Ofcom moves away from a blanket permission on all group chats and considers additional measures to group chat safety. For example, we already provide a number of safety measures to group chat functionality on Snapchat, which go beyond the code's recommendations and what other messenger services offer; and could be adopted in the code:

- a user on Snapchat can only be invited to a group chat by someone they are already friends with;
- If they are invited to a group chat where a participant is someone they have blocked, they receive a notification and have the option to decline the invitation before joining the group;
- The size is limited to a maximum of 200 people;
- Group chat reporting tools are available to support users in raising quick and confidential reports which are reviewed and, if appropriate, enforced, by our global Trust & Safety team that work 24/7; and
- a 'leave' functionality exists within the group chat should the user wish to exit at any time.

We believe that these measures, alongside a proportionate use of the group chat permissions, could have an instrumental impact in helping to prevent young people from encountering harm and restrict the ability of perpetrators contacting children via group chats.

[REDACTED/CONFIDENTIAL]

US6: Provide age-appropriate user support materials for children

We agree with this measure and are already in the process of adopting it through our new Policy Centre to ensure age-appropriate accessibility and understandability of our support materials for children (e.g. our Explainer series which sit beneath our Community Guidelines).

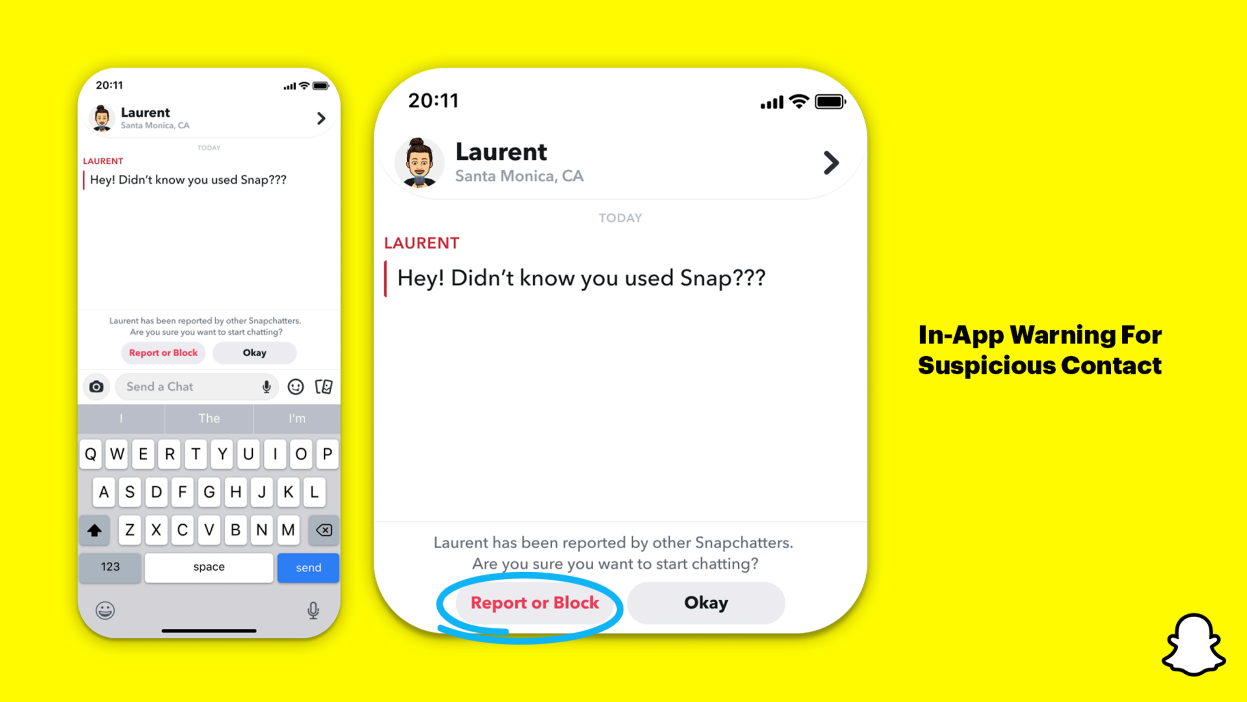
Of note is Ofcom's guidance which suggests that these materials should also target the adults who care for the child. We understand that Parental Controls is a future area of focus for Ofcom but it is remiss of Ofcom not to include or reference such tools in its proposal to help ensure parents have the necessary tools and resources at their disposal to protect teens online. While Snap has a dedicated parents site and

YouTube series to support understanding of Snapchat and our safety measures (recognising that most parents are not on Snapchat), our Family Centre is our main resource, which not only gives parents the tools to be assured of their teens safety on Snapchat but is also designed to instigate conversations between parents and teens about healthy online safety habits and learning together. We suggest that there is some recognition in Ofcom's guidance for parental controls in this measure given the number of services that do adopt such tools to effectively support parents or trusted adults. Arguably, engagement with these types of tools can be more effective in building awareness and understanding of the support or safety mechanisms available given the active engagement with the product. See also our comments in the General Comments on interoperable parental controls which would enhance awareness (recognising that most parents are not on Snapchat).

US4: Provision of information to child users when they restrict interactions with other accounts or content

We agree with Ofcom's proposed measure and already implement this at Snap. For example, teens receive a pop-up warning if they receive a message from someone they don't already share mutual friends with or have in their phone contacts book. The message informs teens of potential risk, so they can carefully consider if they want to be in contact and reminds them to only connect with people they trust.

We have recently expanded our in-app warnings to include those chats **from someone who has been blocked or reported by others, or is from a region where the teen's network isn't typically located**:



Laurent
Santa Monica, CA

TODAY

LAURENT

Hey! Didn't know you used Snap???

Laurent has been reported by other Snapchatters.
Are you sure you want to start chatting?

In-App Warning For Suspicious Contact

We believe these advanced and additional signals as part of the information we provide to teens will significantly increase their safety, helping to prevent unwanted contact and potential harm. We therefore recommend Ofcom goes further in its proposed measure to reflect on Snap's approach on flagging risks or restrictions to interactions based on other factors beyond those taken only by the individual user; noting Ofcom's public recognition for this.

US5: Signpost children to support at key points in the user journey

We agree with Ofcom's research and findings that online self-help tools and support resources may be immensely helpful to young people to get the support they need if they are experiencing harm, including suicide, self-harm, eating disorders or bullying. As recognised in the consultation, Snap already offers users support resources in various parts of the app (both in our user-to-user services and other services like My AI); these not only ensure users are directed to third party organisations to receive expert help but also aim to educate and empower the user to spot signs of harm and what they can do about it on Snapchat (e.g. report it).

However, while we support this measure in principle, we are concerned by how Ofcom proposes to implement it. There is a risk that intervention at some of the points identified by Ofcom in the user journey may lead to excessive or frequent notifications which could cause 'alert fatigue' i.e. where the user fails to engage with the information presented and support offered due to having seen the notification too many times. We consider it is possible that repeatedly displaying a notification would be considered an annoyance and a degradation of the user experience. For example, intervention point 2 could see a user post multiple pieces of content related to suicidal feelings in one go and consequently receive multiple support notifications (as well as potential take-down notices), which they may simply ignore and feel like they have been 'spammed'.

This measure should offer some flexibility and be considered in the round when Ofcom is also proposing the use of other notifications (e.g. group chats – see above) and services may use prompts for important safety information. For example, beyond the recommendations made by Ofcom in their draft codes, Snap provides safety updates to its users to check their location settings or friends lists. We would therefore urge a balanced approach to the use of information provisions, signposts, notifications etc that allows service providers to reflect on the cumulative effect of the support and information they provide in order to determine the how best to ensure users are informed in an impactful way, engage with the information provided and make an informed decision.

10. <u>User reporting and complaints</u>

Do you agree with the proposed user reporting measures to be included in the draft Children's Safety Codes?

Do you agree with our proposals to apply each of Measures UR2 (e) and UR3 (b) to all services likely to be accessed by children for all types of complaints? Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.

The majority of our response in relation to the proposed measures for user reporting and complaints have already been covered by our previous submission on the Illegal Harms consultation (see Annex A). However, we have provided specific responses in relation to complaints about an incorrect assessment of a user's age as follows:

Prioritisation

Subject to our considerations on the use of HEAA methods (as set out above), we agree with Ofcom's proposal that services in scope of Recommendations PCU H4 to PCU H7 should prioritise complaints about an incorrect assessment of a UK user's age based on the seriousness of the associated restriction applied to the user's ability to access content on the service, and whether the age assessment was made without human oversight.

However, as explained in our previous response to the Illegal Harms consultation, we have concerns about relying on the accuracy of automatic detection technologies and the service's past error rate as criteria for prioritising the review of complaints that challenge such assessments. This would require services to constantly update their operational processes in real time, which would be extremely resource-intensive and may present particular capacity challenges for a company of Snap's size.

Moreover, using representations made by the complainant on the impact of the decision on their livelihood as prioritisation criteria could be difficult to verify and may differ from platform to platform. Ofcom's consultation gives the example of an adult performer who may be unable to access earnings on an adult website if incorrectly assessed to be under 18. However, adult content is not allowed on Snapchat and so we would welcome clear guidance and indicative examples on how services like Snap can sufficiently make this assessment to ensure rightful prioritisation.

Performance targets

We agree that all providers of services likely to be accessed by children should determine complaints of incorrect age assessment promptly. We further agree that services in scope of Recommendations PCU H4 to PCU H7 should monitor their performance against performance targets they themselves set with respect to turnaround times and the accuracy of their decision-making. Subject to our position on HEAA, such providers should further monitor trends in complaints about incorrect age assessments to guide improvements in their age assurance processes.

Action after determination

We agree that services should, to the extent possible, aim to restore users who they incorrectly assessed are a child to the position they were in prior to the incorrect assessment. However, Ofcom's one-size-fits all approach – i.e. that services reverse the restriction that they applied to the user's ability to access content on the service as a result of the incorrect age assessment – does not sufficiently take into account platform differentiation. Content on Snapchat is available for only a short period of time before it deletes by default. As a result, Snap will not be able to restore a complainant's access to restricted content that was deleted while the complainant was incorrectly subject to age restrictions.

Moreover, Snap prohibits children who declare they are under the age of 13 from creating Snapchat accounts. If Snap obtains knowledge that a Snapchat user is under the age of 13, Snap terminates the user's account and deletes the user's data, which is consistent with our obligations under applicable law. This makes it impossible for Snap to restore accounts that were incorrectly determined to be held by a person under the age of 13 to the position they were in prior to the incorrect determination.

For the reasons noted above, we recommend that Ofcom offers services greater flexibility in determining the best course of action following a determination that restrictions were applied to a user's ability to access content pursuant to an incorrect age assessment. This would enable providers to adapt their corrective measures to the design of their service as well as their legal obligations globally.

Do you agree with the inclusion of the proposed changes to Measures UR2 and UR3 in the Illegal Content Codes (Measures 5B and 5C)? Please provide any arguments and supporting evidence.

We have no concerns regarding the proposed changes to Measures UR2 and UR3 in the Illegal Content Codes (Measures 5B and 5C) – these are measures we already provide on Snapchat (Explainer on Snapchat Moderation, Enforcement and Appeals; Transparency Reports; Quick Guide to Snapchat Reporting). We note, however, that these changes only apply to user-to-user services and search services likely to be accessed by children. We would urge Ofcom to apply these measures to all user-to-user and search services in-scope of the OSA, as arguably, children are not the only cohort that under-report due to anonymity concerns and/or a lack of understanding of reporting processes, including what happens to their complaint. This is also true of other vulnerable groups, such as women who are victims of violence or coercive control[4]. As such, we recommend this approach is extended beyond those services likely to be accessed by children to help support a systemic shift in the culture of reporting for all users.

Conclusion

We hope this response is helpful to Ofcom's consideration when finalising the codes and guidance on protecting children from harm under the OSA. Please let us know if you have any questions or require additional information and we would be happy to discuss in further detail if required. [REDACTED/CONFIDENTIAL]

[4] Estimates suggest that one in three UK women have experienced online abuse or harassment online, and research from the Victim's Commissioner found that 35% of women surveyed who experienced online abuse thought the matter was too trivial to report to either the police, internet companies, or both.