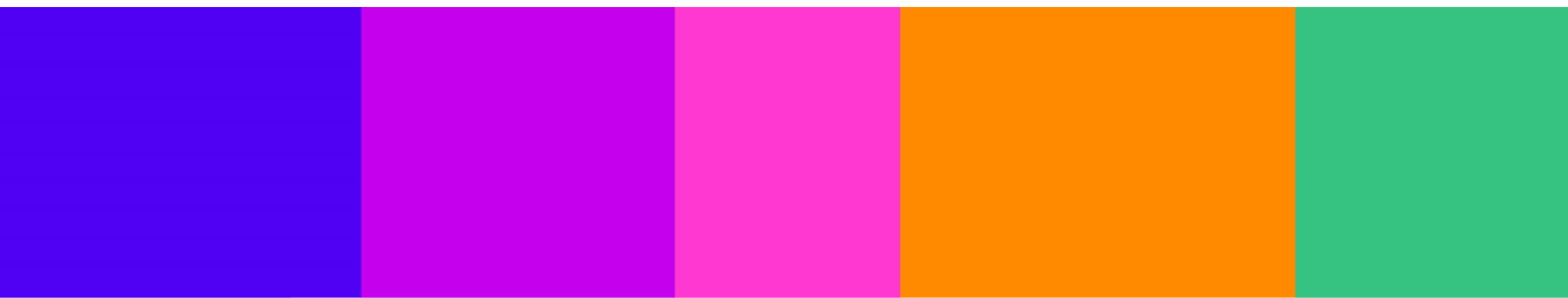




Consultation response form

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

Consultation title	Consultation: Protecting children from harms online
Organisation name	Integrity Institute



Your response

Question	Your response
Volume 2: Identifying the services children are using Children’s Access Assessments (Section 4).	
<p>Do you agree with our proposals in relation to children’s access assessments, in particular the aspects below. Please provide evidence to support your view.</p> <p>1. Our proposal that service providers should only conclude that children are not normally able to access a service where they are using highly effective age assurance?</p> <p>2. Our proposed approach to the child user condition, including our proposed interpretation of “significant number of users who are children” and the factors that service providers consider in assessing whether the child user condition is met?</p> <p>3. Our proposed approach to the process for children’s access assessments?</p>	<p>Confidential? – No</p> <p>1. Yes, it is reasonable to set a high bar for companies to meet in order to not have to assess risks for children.</p> <p>We would stress, however, that age assurance overall provides a poor grounding and foundation for child safety. We understand that Ofcom is NOT mandating across the board age verification, which we think is good. The following is encouragement of why not to go further:</p> <p>Grounding child safety approaches in age assurance practices, which will never achieve 100% accuracy, can draw company resources away from the teams that help design the systems to reduce the spread of harmful content, respond to negative user experiences, and improve trust and safety of all users on the platform (not just children).</p> <p>In addition, there are risks associated with mandating robust age assurance mechanisms. The process of age assurance is not yet mature enough to have well established best practices, and it may require a waterfall approach combining multiple methods. For instance, image recognition tools might have biases and systematically misclassify groups of adults as children or vice versa, and relying on government IDs can pose several challenges. Children may lack such documents, could potentially spoof them, and marginalized groups such as lower-income adults and undocumented immigrants might be unfairly blocked from services. Moreover, sensitive data used in age assurance processes could be susceptible to hacks and leaks, putting</p>

Question	Your response
	<p>children’s information at risk. Data minimization and security standards are essential for age assurance providers, especially in light of incidents like the AU10TIX hack.</p> <p>One proposed solution is on-device age assurance, where a “child mode” setting is retained on the user's device. However, this approach is vulnerable to malware, viruses, and smart youths creating apps to manipulate the age assurance system locally.</p> <p>Technical accuracy is often compromised because models are rarely trained on globally representative data. This lack of diverse training data leads to inequities, particularly in lower revenue-returning markets like APAC and EMEA, where people from different racial, ethnic, and cultural backgrounds might be inaccurately assessed. Therefore, age assurance models require much better training data, including a large and diverse sample of individuals below and above 18 years of age from various backgrounds to ensure an acceptable margin of error.</p> <p>Ultimately, while a "child mode" option on phones could be a useful feature communicated to various apps, it is crucial to recognize that completely preventing determined young users from accessing platforms may be unrealistic. To best protect children online, we need to ensure that platforms are designed in ways that are safe for them, which we currently do not see across all platforms.</p>
<p>Volume 3: The causes and impacts of online harm to children</p> <p>Draft Children’s Register of Risk (Section 7)</p>	
<p>Proposed approach:</p> <p>4. Do you have any views on Ofcom’s assessment of the causes and impacts of online harms? Please provide evidence to support your answer.</p>	<p>Confidential? – No</p> <p>4. One viewpoint is that the specific areas that should be given the highest priority are the ones that lead to child endangerment, such as self-injury (SSI), eating disorders (ED), cyberbullying, and dangerous stunts. While pornographic content is also important to monitor, it is less likely to lead</p>

Question	Your response
<p>a. Do you think we have missed anything important in our analysis?</p> <p>5. 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.</p> <p>6. Do you have any views on the age groups we recommended for assessing risk by age? Please provide evidence to support your answer.</p> <p>7. 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.</p> <p>Evidence gathering for future work:</p> <p>8. Do you have any evidence relating to kinds of content that increase the risk of harm from Primary Priority, Priority or Non-designated Content, when viewed in combination (to be considered as part of cumulative harm)?</p> <p>9. Have you identified risks to children from GenAI content or applications on U2U or Search services?</p> <p>a) Please Provide any information about any risks identified</p> <p>10. 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:</p> <p>a) (i) specific examples of body image or depressive content linked to significant harms to children,</p>	<p>to death, suicide, self-injury, or eating disorders compared to the other areas mentioned. It is also important to look beyond content itself to the types of interactions users can take on the platform. For example, one important risk factor is the protections around interactions between children and adults on a platform, and the characteristics of communication that are allowed (image / text / ephemeral / E2EE).</p> <p>We understand that OFCOM is not enforcing detailed specific policies and definitions of content that platforms must adopt, which we think is good. In general, regulation should be careful about being overly prescriptive in terms of the exact types of content platforms should take action on. It is worth noting here that content-specific regulation (that is, regulation that mandates platforms implement detailed definitions of harmful content without consideration of the content and policy nuances within the platform) was identified by integrity workers as one of the least effective policy measures towards reducing harms to users on platforms. Harmful content changes and users adapt to find ways to get around platform (and policy) definitions.</p> <p>This is not to say that trying to reduce types of heinous content online is not a laudable goal, but it is an outcome that platforms are incentivized to seek regardless of regulation. This is why many platforms already have policies against the types of content outlined here. The problem is not that platforms aren't identifying and trying to remove harmful content (in most cases), the larger problem is that they are <i>not</i> incentivized to change the structure and design of their underlying content ranking systems that elevate harmful content.</p> <p>6.</p> <p>Having companies research the age appropriate experiences for more granular age buckets is reasonable. However, we would be hesitant towards going any further</p>

Question	Your response
<p>b. (ii) evidence distinguishing body image or depressive content from existing categories of priority or primary priority content.</p> <p>11. 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.</p>	<p>and mandating that platforms do age assurance on all users to tailor their user experience for many different, narrow age ranges.</p> <p>Any user experience designed for 18-21 could also be “attractive” for 16-17 year olds, and will lead to children trying to evade systems. How should we think about age appropriateness that’s more actionable?</p> <p>It is also unclear how platforms will reach this level of granularity in age assurance to act on assessments of risk and segment their services accordingly, as technology to identify children at that level of granularity (e.g., 5 years old vs. 9 years old) does not exist, and more children that age will not have official identification.</p> <p>10.</p> <p>A lot of the social media studies on this are non-experimental, correlational, and/or rely on self-reported social media use which has repeatedly been shown to not correlate with actual social media use. That said, there's a rich history of high-quality (often experimental) studies showing the effects of seeing particular body types on people's well-being.</p> <p>This meta-analysis of 25 studies shows that images of thin-ideal bodies increase body dissatisfaction (which correlates with outcomes like anxiety, depression, and self-esteem).</p> <p>This review paper provides a great overview of the relevant methods for how social media and body image are studied together, and highlights some of the current important shortcomings.</p> <p>This study shows that social media literacy may actual buffer the effects of thin-ideal images encountered on social media on the well-being of teenagers (particularly teenage girls).</p>

Question	Your response
	<p>This study compares the effects of encountering images of different body types across different types of media (magazine, newspaper, TV, social media, etc), and looks at affective, cognitive, and behavioral outcomes. One notable finding is that exposure to thin-ideal body images increases risk for eating disorders.</p> <p>11.</p> <p>Areas to make sure are considered are health misinformation, sexual harassment or advancements, predatory behavior, grooming, and cyberbullying are critical areas to focus on, and should be included in the definitions of these categories. Bullying can be particularly challenging to identify and address effectively.</p>
Draft Guidance on Content Harmful to Children (Section 8)	
<p>12. 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?</p> <p>13. Do you have further evidence that can support the guidance provided on different kinds of content harmful to children?</p> <p>14. For each of the harms discussed, are there additional categories of content that Ofcom</p> <ul style="list-style-type: none"> a) should consider to be harmful or b) consider not to be harmful or c) where our current proposals should be reconsidered? 	<p>Confidential? – Y / N</p>
Volume 4: How should services assess the risk of online harms?	

Question	Your response
Governance and Accountability (Section 11)	
<p>15. Do you agree with the proposed governance measures to be included in the Children’s Safety Codes?</p> <p>a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence.</p> <p>b) 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.</p> <p>16. 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?</p>	<p>Confidential? – Y / N</p>
Children’s Risk Assessment Guidance and Children’s Risk Profiles’ (Section 12)	
<p>17. What do you think about our proposals in relation to the Children’s Risk Assessment Guidance?</p> <p>a) Please provide underlying arguments and evidence of efficacy or risks that support your view.</p> <p>18. What do you think about our proposals in relation to the Children’s Risk Profiles for Content Harmful to Children?</p> <p>a) Please provide underlying arguments and evidence of efficacy or risks that support your view.</p>	<p>Confidential? – Y / N</p>

Question	Your response
<p>Specifically, we welcome evidence from regulated services on the following:</p> <p>19. 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 their services pose to children and comply with their child risk assessment obligations under the Act?</p> <p>20. 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?</p> <p>21. 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?</p> <p>a) 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.</p>	
<p>Volume 5 – What should services do to mitigate the risk of online harms</p> <p>Our proposals for the Children’s Safety Codes (Section 13)</p>	
<p>Proposed measures</p> <p>22. Do you agree with our proposed package of measures for the first Children’s Safety Codes?</p> <p>a) If not, please explain why.</p> <p>Evidence gathering for future work.</p>	<p>Confidential? – No</p> <p>22.</p> <p>From our survey of Integrity workers in our community, the best tools to incentivize platforms towards the goal of creating safer online experiences is to mandate meaningful transparency about the scale, cause and nature of harms on the platforms, and utilize tools like risk assessments and audits to explore the ways these harms manifest and how</p>

Question	Your response
<p>23. Do you currently employ measures or have additional evidence in the areas we have set out for future consideration?</p> <p>a) If so, please provide evidence of the impact, effectiveness and cost of such measures, including any results from trialling or testing of measures.</p> <p>24. Are there other areas in which we should consider potential future measures for the Children’s Safety Codes?</p> <p>a) If so, please explain why and provide supporting evidence.</p>	<p>they can be mitigated. Our transparency framework is outlined in this blog post. Additional approaches include robust privacy settings on by default, limiting recommended content shown to children, limiting notifications, allowing parental controls, and limiting targeted advertising.</p> <p>An additional consideration is what content is considered to be harmful vs. controversial (ex. sexual health and gender identity may be classified as harmful) How can PC and PPC such as hate speech impact already marginalized communities such as the LGBTQ+ from discussing events that may trigger content moderation?</p> <p>24. The following is the Institutes general guidance on direction of focus for the framework:</p> <p>The primary focus should be on platform design and recommender/ranking systems. Platform design encompasses elements such as reshared content, access to content from accounts users don't follow, or content in groups where the user won't know everyone. Overall, it's crucial to limit content from accounts the user didn't specifically choose to follow and to be extremely cautious about recommending any accounts. Following that, strong privacy defaults and controls are essential, as well as focus on search, content moderation, and user reporting.</p> <p>Content moderation is often too slow because it requires sending posts off to moderators, and by the time a post has been labeled as harmful or not, 99% of exposures have already occurred. Age assurance is risky due to the collection of sensitive personal data on all users, and these companies have a poor track record for protecting user data. User reports are often not very useful because many platforms add so much friction to submitting user reports (e.g., to formally</p>

Question	Your response
	<p>report a post on Facebook, one has to click through seven screens) that very few people actually submit reports, and so relying on them should also come with incentives for the companies to design them in easy to use ways. Reports are often denied because they lack the context of the entire encounter and the reported content on its own may not be policy-violating. Additionally, the reports that are submitted tend to come from accounts that are more inclined to be abusive themselves, suggesting that they may be weaponizing reports against benign content or well-behaving users.</p> <p>The most impactful lever is improving the recommendation systems, which determine what shows up in the feed, in search results, at the tops of comment threads, etc. Some straightforward approaches have been developed, tested, and shown to be effective across a number of platforms, but many of these platforms opt not to implement these interventions because they tend to reduce time spent on the platform in the short term. Additionally, creating a classifier seems to be proposing a solution rather than addressing the problem, incentivizing companies to try to meet the bare minimum classifier requirements rather than addressing how harmful content is amplified by their systems.</p> <p>For further insights, we recommend our algorithmic risk assessment report, Instagram report, and child safety briefings.</p>

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

25. Do you agree with our approach to developing the proposed measures for the

Children’s Safety Codes?

a) If not, please explain why.

26. 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?

a) Please explain your views.

27. 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?

28. Do you agree with our definition of ‘large’ and with how we apply this in our recommendations?

29. Do you agree with our definition of ‘multi-risk’ and with how we apply this in our recommendations?

30. Do you agree with the proposed measures that we recommend for all services, even those that are small and low-risk?

Confidential? – Y / N

Age assurance measures (Section 15)

31. Do you agree with our proposal to recommend the use of highly effective age assurance to support Measures

Confidential? – N

33. Please see the comments above under Volume 2 for our general feedback on grounding the approach to child safety in age assurance measures.

AA1-6? Please provide any information or evidence to support your views.

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

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

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

33. 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?

34. Do you have any comments on our assessment of the implications of the proposed Measures AA1-6 on children, adults or services?

a) Please provide any supporting information or evidence in support of your views.

35. 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?

Many children are honest about their age on platforms, yet still face unsafe experiences. Therefore, it is more important to create accountability and better incentives for companies to think comprehensively about all the methods available to reduce harms to children, rather than mandating specific age assurance measures. While age assurance may be necessary in some discrete cases listed (Platforms with the goal of disseminating inappropriate content for children), historically, methods similar to age assurance have consistently failed because children are adept at finding ways around verification systems. The least effective way to protect children is through **age assurance alone**. If platforms took more responsibility and seriously committed to removing and reducing harmful content, age assurance and verification would be largely unnecessary, with a few exceptions.

Demanding that platforms implement age assurance for all users will likely undermine privacy and take resources away from the teams within the platforms that can understand the ways bad actors are abusing the platform and tailor responses to each platform's shape and design. There are also concerns about age assurance technology's ability to identify people who don't fit into clear categories (e.g., individuals going through gender transitions or whose names may not match their IDs). This introduces serious concerns about the ability of all individuals to access online platforms.

Rather than rely on systems to accurately identify if a user is a minor and then classifiers to identify if certain content is harmful (both of which will never be perfect and come with false positives and negatives that have undesired impacts), it is much more important for companies to create ranking and recommendation systems that do not amplify harmful content. This is key because classifiers for down-ranking harmful content will not be effective if the **underlying systems are still promoting** harmful content. Without additional language specifying that platforms should test

	<p>how their ranking systems respond to harmful content without classifiers, this measure would be inadequate. There are also dangers of inadvertently preventing minors from accessing mental health support content, as classifiers are not always good at identifying content in context.</p> <p>35. The technology is not advanced enough to group kids into age groups as specific as 13-15, 16-17, etc. based on classifiers alone. And as noted above, many children will not have official IDs, so it is questionable whether or not such a granular approach is feasible.</p> <p>However, in assessing age assurance systems, best practices could include metrics such as precision and recall to evaluate the system's effectiveness. Precision measures whether the system accurately identifies individuals as underage when they are, while recall assesses the percentage of underage individuals correctly identified. Testing should be focused on individuals near the threshold age, ensuring the system does not simply default to assuming everyone is 18 or older. It's important to acknowledge that no system can achieve 100% accuracy, and there will always be a trade-off between privacy, user burden, and accuracy. Therefore, rather than focusing on specific methods, it is crucial to establish desired precision and recall levels.</p> <p>Finally, there should be an established appeal process for those who have been identified incorrectly as children.</p>
<p>Content moderation U2U (Section 16)</p>	
<p>36. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.</p>	<p>Confidential? – Y / N</p>

<p>37. Do you agree with the proposed addition of Measure 4G to the Illegal Content Codes?</p> <p>a) Please provide any arguments and supporting evidence.</p>	
--	--

Search moderation (Section 17)

<p>38. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.</p> <p>39. Are there additional steps that services take to protect children from the harms set out in the Act?</p> <p>a) If so, how effective are they?</p> <p>40. Regarding Measure SM2, do you agree that it is proportionate to preclude users believed to be a child from turning the safe search settings off?</p> <p>The use of Generative AI (GenAI), see Introduction to Volume 5, to facilitate search is an emerging development, which may include where search services have integrated GenAI into their functionalities, as well as where standalone GenAI services perform search functions. There is currently limited evidence on how the use of GenAI in search services may affect the implementation of the safety measures as set out in this code. We welcome further evidence from stakeholders on the following questions and please provide arguments and evidence to support your views:</p> <p>41. Do you consider that it is technically feasible to apply the proposed code measures in respect of GenAI</p>	<p>Confidential? – No</p> <p>40. The effectiveness of "safe search" as a strategy depends on its specific implementation. It is reasonable for regulators to ask that companies determine content that is inappropriate for children and to take steps to ensure it cannot be seen by them. However, the intent behind existing "safe search" settings might differ, such as focusing on returning results appropriate for workplaces or other settings with different content appropriateness levels. These may not align perfectly with "child safe search."</p> <p>Additionally, "safe search" alone will not comprehensively solve the problem of children being exposed to harmful content online. It would be more effective to create broader incentives for companies to minimize harms, rather than mandating a specific design practice. "Safe search" might not have any significant impact on children seeing harmful content, so companies should not be forced or rewarded for implementing an ineffective design. Implementing "safe search" could be beneficial if companies have high confidence and low error in identifying children.</p> <p>There is also the risk, assuming "safe search" results rely on classifiers to filter out content deemed harmful to children, that a child searching for resources on mental health, sexual identity or eating disorder support may be prevented from finding those results. The risk of "safe search" being wrong leads to stifled speech, frustrating user experience, and lower quality results. As we discuss in this blog post on Instagram turning off search results for sensitive queries, effectively turning off search results gives companies an opportunity to claim they can't design better content ranking systems, rather than incentivizing them to do away with systems that reward low quality, often harmful content and</p>
---	--

functionalities which are likely to perform or be integrated into search functions?

42. What additional search moderation measures might be applicable where GenAI performs or is integrated into search functions?

build systems that would surface high quality, educational or authoritative content on sensitive topics. The example provided in our blog post is as follows: Searching for “bulimia” on Instagram returns harmful results: three of the top 15 results contain images that violate the National Eating Disorder Association’s guidelines for sharing content online. Whereas on Google Search, the top results are from the Mayo Clinic, National Eating Disorder Association, the NHS in the UK, the Cleveland Clinic and so on. The difference in results is due to the difference in optimization of content algorithms (engagement on Instagram vs quality and relevance on Google). If the idea of “safe search” is that such a mode *would* return higher quality results, it again begs the question of why this setting is reserved for children and not applied to enhance the online experience of all users.

User reporting and complaints (Section 18)

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

a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence.

b) 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.

44. 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?

a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence.

Confidential? – Y / N

b) 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.

45. Do you agree with the inclusion of the proposed changes to Measures UR2 and UR3 in the Illegal Content Codes (Measures 5B and 5C)?

a) Please provide any arguments and supporting evidence.

Terms of service and publicly available statements (Section 19)

46. Do you agree with the proposed Terms of Service / Publicly Available Statements measures to be included in the Children’s Safety Codes?

a) Please confirm which proposed measures your views relate to and provide any arguments and supporting evidence.

b) 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.

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

48. Do you agree with the proposed addition of Measure 6AA to the Illegal Content Codes?

a) Please provide any arguments and supporting evidence.

Confidential? – Y / N

Recommender systems (Section 20)

49. Do you agree with the proposed recommender systems measures to be included in the Children’s Safety Codes?

a) Please confirm which proposed measure your views relate to and provide any arguments and supporting evidence.

b) If you responded to our illegal harms consultation and this is relevant to your response here, please signpost

Confidential? – No

50. Regulation of recommender systems should focus primarily on requiring platforms to analyze and report why people, particularly children, are exposed to violating content, thereby creating broad accountability (ex. % exposures to harmful content due to recommendations, % exposures from reshares, etc.).

In addition, companies should study how key components of the recommender system, such as machine learning clas-

to the relevant parts of your prior response.

50. 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?

51. 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.

52. 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.

- Please provide the underlying arguments and evidence of the relevance of this content to Measures RS2 and RS3.

sifiers that play a role in the final ranking of content, respond to violating content. For example, studies done inside of companies have shown that many engagement based classifiers return higher engagement probabilities for violating content, which will lead to violating content being amplified on the platform.

Following that, it is essential to ensure that the various signals boosting content, such as classifiers predicting engagement or quality, do not amplify harmful content, as outlined in our blog post about the [safety of Instagram search](#).

Limiting content from accounts that the user did not proactively follow and ensuring recommended content systems amplify high-quality, child-appropriate content across all topic areas, including sensitive ones, are also crucial steps.

Content identification and classification are important areas to intervene in addition to lightweight negative feedback mechanisms. If platforms made it easier to provide negative feedback (i.e., not burying it under many screens of friction), the quality of that feedback would improve. Currently, most platforms offer users many more opportunities to provide positive feedback than negative feedback. This imbalance makes it harder to understand the meaning of positive feedback (e.g., likes, reactions, comments, re-shares) because it lacks a relative measure against negative feedback. Platforms like Reddit perform better in this context because users can upvote or downvote content, creating an even playing field where people can just as easily indicate liking or disliking. This balanced feedback and easier methods of providing feedback would benefit all platforms.

Incorporating negative feedback from users is one of the most effective measures observed in the proposed code. If these systems are built and maintained sufficiently, and designed in ways that are easy to use, their output can be applied in many ways to reduce harms. For instance, if a piece

	<p>of content scores above a classifier's threshold, it could be demoted in ranking or filtered out from search entirely. If someone frequently posts content that scores high for potential harms, content from that user could be slowed in its distribution until verified as non-harmful.</p>
--	---

User support (Section 21)

53. Do you agree with the proposed user support measures to be included in the Children’s Safety Codes?

a) Please confirm which proposed measure your views relate to and provide any arguments and supporting evidence.

b) 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.

Confidential? – Y / N

Search features, functionalities and user support (Section 22)

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

55. Do you have additional evidence relating to children’s use of search services and the impact of search functionalities on children’s behaviour?

56. Are there additional steps that you take to protect children from harms as set out in the Act?

a) If so, how effective are they?

As referenced in the Overview of Codes, Section 13 and Section 17, the use of GenAI to facilitate search is an emerging development and there is currently limited evidence on how the

Confidential? – No

56.

Other important controls to give to users include: limiting content from accounts they don't proactively follow, enhanced privacy controls and defaults to limit, for example, show or hide friends/connections list, and limited direct messages or comments from unknown or anonymous accounts. We need interventions to teach and remind users about the tools as just having them in menus isn't enough. The "inform" tools (informing users about rules and restricted interactions) tend to be on the weak side, and don't actually stop people from viewing the potentially harmful content much of the time. Additionally, making all these settings private by default ensures **all users** have a safer experience.

use of GenAI in search services may affect the implementation of the safety measures as set out in this section. We welcome further evidence from stakeholders on the following questions and please provide arguments and evidence to support your views:

57. Do you consider that it is technically feasible to apply the proposed codes measures in respect of GenAI functionalities which are likely to perform or be integrated into search functions? Please provide arguments and evidence to support your views.

Combined Impact Assessment (Section 23)

58. Do you agree that our package of proposed measures is proportionate, taking into account the impact on children's safety online as well as the implications on different kinds of services?

Confidential? – Y / N

Statutory tests (Section 24)

59. Do you agree that our proposals, in particular our proposed recommendations for the draft Children's Safety Codes, are appropriate in the light of the matters to which we must have regard?

a) If not, please explain why.

Confidential? – Y / N

Annexes

Impact Assessments (Annex A14)

60. In relation to our equality impact assessment, do you agree that some of our proposals would have a positive impact on certain groups?

61. In relation to our Welsh language assessment, do you agree that our proposals are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English?

a) If you disagree, please explain why, including how you consider these proposals could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.

Confidential? – Y / N

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