Introduction

Ofcom recently published a consultation concerning how user-to-user and search service providers should approach their duty to prevent harm to children. The Online Safety Act (2023) makes these service providers responsible for keeping users, and especially children, safe from harm online. They must assess risks and take steps to address them. Ofcom is seeking feedback on their recommendations on how they intend to enforce service providers' obligations, which are provided across 5 published volumes and 15 annexes, as well as some further documentation. The 5 volumes are concerned with: 1 - overview; 2 - identifying services (children may access); 3 - causes and impacts (of online harms to children); 4 - assessing risks; 5 - mitigating risks.

We are members of a research group dedicated to the investigation of AI and Information Ethics at Northeastern University London, itself an institution of (higher) education. As such, we intend to address ourselves to issues the consultation raises in relation to three themes: AI and Information Ethics; Education; and Research.

We address some of the specific questions raised by Ofcom in the consultation below (though we notably omit any answers to the questions concerning volume 4 as beyond our immediate expertise). But we begin by framing some general, overarching points.

AI and Information Ethics

We have concerns about the proposals in relation to age verification. As discussed in the ICO's age appropriate design code of practice, there are many relevant rights of the child at play when considering access to online services by children. Perhaps most central here are the right to privacy, the right to access (appropriate) information whilst being protected from inappropriate information, and the right to protection from exploitation. Unfortunately, at times it may seem that there is a tradeoff between these rights. Our concern is that in verifying ages, solutions will be put into place that will not give enough consideration to the protection of children's data. Furthermore, given that adults will also need to have their age verified under the approach outlined by Ofcom, the widespread implementation of age verification methods will lead to impacts on privacy, appropriate access to information, and have the potential to be discriminatory. Adults in particular will face risks to their privacy, their ability to access services and information, and (depending on the service) their freedom of expression and association. The very services which pose a risk to children may be the services which are most likely to severely impact the lives of adults if data is not properly protected (see for example the Ashley Madison data leak, which is thought to have resulted in a considerable number of scams, and potential suicides).

We are particularly concerned with the suggested implementation of AI-based age estimation methods. AI-based methods have the additional risk of bias, which is a known concern with facial recognition systems to the extent that they have been found to be racially discriminatory (e.g. Uber's facial recognition tool). Humans are already known to be poor and biased estimators of age, and AI age estimation tools have been shown to exaggerate these biases, with facial expression, gender and age impacting accuracy.

Furthermore, we have concerns that alternative age verification methods (such as reliance on credit cards or identity documents) will further the digital divide, given that not all adults are able to access to the necessary documentation or acquire a credit card (e.g. they have poor credit, no fixed address or have limited access to banking). Similar concerns have been raised around the requirement of voter identification which is thought to disproportionately affect people from minority ethnic backgrounds, poorer people and those with disabilities, as warned by the electoral commission (with recent polls suggesting a real effect in the latest election).

The use of age verification methods also raises concerns for data protection. There is a risk that many services which previously held scant personal data would now have access to identity documentation, identifying photographs, or simply further personal information through requiring users to make a personal account. Whilst services will need to conform to existing data legislation, this will still provide companies with a large amount of additional personal data that will be required for accessing their service. Given these concerns, and those discussed above, we would advocate for the establishment of a robust, decentralised third party verification service (provided by a government body?) much like those advocated for during the COVID-19 pandemic in order to protect privacy and place control over data in the hands of individuals. In the development of this system, adequate thought should be given to the minimisation of impact on individuals who are less able to access identity documents or credit. We would strongly repudiate the recommendation to utilise AI-based methods due to concerns of discrimination.

We would also like to stress that, in general, issues of AI and information ethics are typically socio-technical in character, with multiple stakeholders beyond the service providers and their users typically relevant (see, e.g. Ball and Helliwell's 3D model of ethical AI practice). Indeed, research indicates that socio-economic status and parental education are extremely important in shaping children's digital lives. It is crucial to consider parental control, for example, as a multifaceted rights issue, encompassing both the rights of parents/carers and those of children. This is not merely about the application of technological means—though Ofcom reports that at least a third of all UK parents employ such methods to monitor their children's digital activities—but about the ethical dimensions of parental and child rights. This underscores the significant role

parents/carers play in mediating their children's online experiences. It involves a delicate balance between safeguarding children's welfare and honouring the emerging agency of young digital citizens, highlighting the need for policies that not only support and empower parents/carers in this critical responsibility but also respect the rights of children in the digital age.

While Ofcom's report includes factors such as governance and media literacy, it could benefit from exploring how socio-economic and cultural factors influence children's online experiences and vulnerabilities. Socio-economic status impacts children's exposure to online harms and their ability to cope with them. For instance, children from lower-income families often have less access to digital literacy education and parental guidance, increasing their susceptibility to online risks such as cyberbullying and exposure to harmful content (Zhang & Livingstone, 2019). These children might also rely on less secure devices and public Wi-Fi, heightening their exposure to online threats.

Policies should ensure that digital literacy and support resources are universally accessible. Schools in disadvantaged areas require additional funding and resources to effectively address these disparities. A targeted approach, informed by socio-economic data, can help in the equitable distribution of educational and technological resources, thus safeguarding all children against online harms. For example, targeted digital literacy programmes and community outreach initiatives can help bridge the digital divide and provide essential support to vulnerable children. This targeted approach, informed by socio-economic data, can help in the equitable distribution of educational and technological resources, thereby safeguarding all children against online harms.

Education

We understand that Ofcom's remit is to implement the law passed by the UK parliament. Nevertheless, we are concerned that there is insufficient emphasis in the documentation on the role of education - and in particular, education in media and information literacy, and in AI ethics (which in turn involves AI literacy) - in preventing online harms. We would like to stress the recent publication by one of our members of a paper advocating AI ethics education in schools (Dabbagh et al., 2024); and to draw attention to the need, as well, for continuing (adult) education in this area (including for parents/carers).

NU London also views itself as 'a national leader in advancing AI literacy' (NU London AI Strategy 2023-24). From this perspective, we would like to suggest that, in order to implement the recommendation of Dabbagh et al. (2024), those in the education sector will themselves need to be trained in AI ethics, so as to be in a position to teach it to children. Mandatory digital citizenship education from an early age can help children

become responsible digital citizens. Schools should collaborate with technology companies to develop effective educational materials and training programmes.

While media literacy is emphasised in the report, integrating comprehensive digital citizenship education is vital. This involves teaching children ethical online behaviour, digital rights and responsibilities, and safe online practices. For example, a curriculum including empathy, respect, and critical thinking can help children navigate digital spaces responsibly. Programmes focusing on empathy, respect, and critical thinking can help children understand the consequences of their online actions and develop strategies to protect themselves and others. Educational curricula must evolve to address the complexities of the digital age, with collaborations between educators, companies, technologists, and policymakers creating comprehensive programmes that foster responsible digital citizenship.

The current age groupings in the Ofcom report might be too broad to capture the nuances within developmental stages. For instance, the experiences and vulnerabilities of a 13-year-old can differ significantly from those of a 17-year-old. Segmenting teenagers into narrower age groups, such as 13-15 and 16-17, can help tailor interventions more precisely, considering the varying levels of maturity and independence. Critical transition periods, such as the shift from primary to secondary school (around age 11-12), are times of significant social and psychological change, influencing online behaviour and risks. Tailoring interventions to these specific developmental stages can enhance the effectiveness of protective measures. For instance, targeted support during these transitions can help mitigate the risks associated with increased online activity and exposure to potential harms.

Research

Under the current heading, the first thing we would like to note is the ongoing need for research on the issues raised by the consultation: which services are children likely to access; what are the causes and impacts of online harms; which methods of mitigating risks are most effective in this specific context; and so on. We are aware that the UK is outside of the European Union, but in that context there is the Digital Services Act (2022), under which researchers are able to request and gain access to the data that various service providers have at their disposal. Here in the UK, the AI Safety Institute is able to support some research, which is currently focussed on so-called frontier models. In our view, much more access to data for research purposes is needed: such transparency will help to ensure that service providers - and Ofcom as their regulator - are able to protect children from online harms. Our colleagues at Northeastern University in Boston are currently building the US National Internet Observatory as well as a research community that can use it (as part of the Internet Democracy Initiative) to promote the digital good.

As an institution, we are therefore home to significant expertise in building the infrastructure needed to conduct research that can support Ofcom's efforts in this area - expertise which we are very willing to share, as appropriate.

One specific area where we would like to see additional research done concerns the causes and impacts of online harms. Not all children have the same access to the internet, or to parental or other (e.g. AI/tech literate peer or educator) social support: we accordingly feel that it is particularly crucial to note intersectionalities between socio-economic factors and the risks of online harms that children face.

Ofcom's report would benefit from exploring how various types of harm intersect and amplify each other, creating a compounded effect on children's well-being. For example, a child experiencing cyberbullying might also encounter self-harm content, where the bullying diminishes their self-esteem, making them more susceptible to harmful ideation encouraged by such content. A multi-faceted approach is necessary to understand these intersectionalities, requiring educational institutions to identify signs of compounded harm and provide integrated support services. Additionally, platforms should develop algorithms to recognise when minors are exposed to multiple risk factors, triggering interventions or providing resources.

Evidence supports that children experiencing cyberbullying are more likely to suffer from depression, and exposure to depressive content online exacerbates the risk of severe mental health issues (Bonanno & Hymel, 2013). Therefore, interventions should be holistic, addressing the entire online ecosystem rather than treating each harm in isolation.

Northeastern University London

Consultation response form

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

Consultation title	Consultation: Protecting children from harms online
Full name	Northeastern University London, AI and Information Ethics research cluster
Organisation name	Northeastern University London, AI and Information Ethics research cluster, including: **Brian Ball**, Associate Professor in Philosophy **Alex Cline**, Assistant Professor in Computing and Information Systems **Hossein Dabbagh**, Assistant Professor in Philosophy **David Freeborn**, Assistant Professor in Philosophy **Courtney Hagen Ford**, Assistant Professor in Marketing **Alice Helliwell**, Assistant Professor in Philosophy **Tom Williams**, Assistant Professor in Philosophy

Your response

Question	Your response
Volume 2: Identifying the services children are using Children's Access Assessments (Section 4).	
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. 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? 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? 3. Our proposed approach to the process for children's access assessments?	Confidential? – N Whilst it may seem clear that a service could not conclude that children are unable to access a service without highly effective age assurance, there is no exploration of other mechanisms to block access for children. It is a fallacy to conclude that children cannot normally access a service only through direct age checks. There may be other ways of providing assurance that children are not normally able to access a service without relying upon highly effective age assurance on the part of the service provider. Federated identity could be used to minimise data transmission (e.g. through single sign-in). It may also be possible to put in place barriers to access through an increase in obscurity, for example by making access to a site or the account creation process difficult, time consuming, or even requiring some kind of reasoning task for access (which may be effective particularly for younger children, though may cause concerns for adult accessibility). Ensuring that websites do not appear on safe searches could also reduce access by children.

	Given the privacy risks, and risks of utilising AI, alternatives such as these should be explored for feasibility as alternatives to the age assurance mechanisms put forward here. Tests for the effectiveness of these strategies could be conducted on a smaller scale rather than requiring widespread implementation. The proposed approach requires that services establish whether they meet the child user condition almost by default, with considerable evidence needed if they are to conclude that they do not meet the condition. We are concerned that this encourages either the collection of identity data to verify ages of users, or that it will involve the sort of analytical tracking that is discouraged by anti-cookie legislation. As discussed in (3) below, this will result in the majority of services collecting additional data on users, which carries with it associated risks (bias, privacy loss, risks to data security, increase of digital divide etc.), even when they are unlikely to have a significant number of child users. Acceptable evidence is vague, and the burden of proof lies with those who do not have significant child users, thus this policy will likely impact all relevant services, with knock-on effects for adult users. The first step in the proposed process relies upon highly effective age assurance. This has associated risks: as raised above, the implementation of several of the age assurance mechanisms designated as highly effective by OfCom would create considerable risk of bias, risk of privacy loss and risk of exacerbating the digital divide. Placing this step ahead of consideration of whether children are a target of the service increases the likelihood of services implementing age assurance mechanisms regardless of the necessity of this data collection. This is particularly likely as it is not clear how a service could establish it doesn't have a significant child user-base without age assurance (or the use of intrusive analytical tracking). OfCom could instead consider revising the process such that services which are certain they do not appeal to

	children could forgo age assurance (think, for example, of Linkedin, or similar work-focussed sites).

Volume 3: The causes and impacts of online harm to children

Draft Children's Register of Risk (Section 7)

Proposed approach: 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. a. Do you think we have missed anything important in our analysis? 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. 6. Do you have any views on the age groups we recommended for assessing risk by age? Please provide evidence to support your answer. 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. **Evidence gathering for future work:** 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)? 9. Have you identified risks to children from GenAI content or applications on U2U or Search services? a) Please Provide any information about any risks identified 10. Do you have any specific evidence relevant to our assessment of body image content and depressive content	Confidential? – N The report addresses the concept of cumulative harm, where different types of harmful content interact to exacerbate the impact on children. However, the complexity of these interactions and their long-term effects might require further in-depth studies. Ofcom's report could benefit from exploring how multiple types of harm intersect and amplify each other, creating a compounded effect on children's well-being. It's crucial to recognise and address the compounded effects of multiple harms. Interventions should be holistic, considering the entire online ecosystem a child is exposed to, rather than treating each harm in isolation. Ofcom's broad definition of NDC as content that is not explicitly categorised but still presents a material risk of significant harm to an appreciable number of children is a prudent approach. This flexibility allows for the inclusion of emerging and unforeseen types of harmful content. However, while the broad definition is a strength, the dynamic nature of online content means that NDC categories must be continuously updated. Emerging technologies and trends can quickly create new forms of harmful content that may not fit neatly into existing categories. The rise of new social media challenges or trends that promote dangerous behaviour, such as the "Tide Pod Challenge" or the more recent "Benadryl Challenge," exemplifies how quickly new harmful content can emerge. Additionally, NDC often intersects with designated content, creating compounded risks. For example, depressive content may intersect with bullying or self-harm content, amplifying the overall harm. Ofcom's approach should consider the intersectionality of harms and develop strategies that address the compounded effects of multiple types of harmful content. This could involve integrated risk assessments that look at how different types of content interact and affect children.

as kinds of non-designated content? Specifically, we are interested in: a) (i) specific examples of body image or depressive content linked to significant harms to children, b. (ii) evidence distinguishing body image or depressive content from existing categories of priority or primary priority content. 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.	Additionally, while the report includes contextual factors such as governance, business models, and media literacy, it might benefit from exploring further how cultural factors influence children's online experiences and vulnerabilities. The report could also analyse further how socioeconomic status affects children's exposure to online harms and their ability to cope with them. Socioeconomic status significantly influences a child's online experience. Children from lower-income families might have less access to digital literacy education and parental guidance, increasing their vulnerability to online harms. The report primarily focuses on immediate and short-term impacts. Longitudinal studies tracking the long-term effects of exposure to harmful content on children's development and mental health would provide a more comprehensive understanding. For example, a child exposed to cyberbullying may develop long-term issues such as chronic anxiety, depression, or social withdrawal, which can persist into adulthood and affect their overall quality of life. While the report emphasises media literacy, integrating comprehensive digital citizenship education is vital. This involves teaching children about ethical online behaviour, digital rights and responsibilities, and safe online practices. Digital citizenship education should be mandatory in schools from an early age. Such education helps children become not only savvy consumers of digital content but also responsible digital citizens. Schools should collaborate with tech companies to develop effective educational materials and training programs. Programmes in schools that focus on empathy, respect, and critical thinking can help children understand the consequences of their online actions and develop strategies to protect themselves and others. Educational curricula should evolve to address the complexities of the digital age. Collaborations between educators, technologists, and policymakers can create comprehensive programs that foster responsible digital citizenship.

Draft Guidance on Content Harmful to Children (Section 8)	
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? 13. Do you have further evidence that can support the guidance provided on different kinds of content harmful to children? 14. 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?	Confidential – N In many cases, there are risks that beneficial, helpful or essential content could plausibly wrongly flagged as harmful. Examples could include essential material for children and teenagers under the age of 16, which has some potential to be confused with harmful content, such as LGBT-related material, sex education resources, recovery content and advice dealing with overdoses, mental health problems etc. Recovery-oriented material might use similar language or imagery as harmful content, making it challenging for automated systems, as well as human-recommendation systems to differentiate. Indeed, human-recommendation systems can, as Ofcom notes, make considerable mistakes and exhibit systematic biases -- the same can hold for automated systems as well. Distinguishing these beneficial materials from harmful requires context-sensitivity, a context which automated systems are likely to lack. For example, content discussing body image could either aim to promote eating disorders or encourage recovery from them. Similarly, mental health content intended to offer support and advice on coping with depression or suicidal thoughts might be mistakenly flagged as promoting self-harm. The context in which this content is presented is crucial, yet it is often subtle and complex, requiring sophisticated understanding that current technology may not consistently achieve. The moderation of online content is inherently fraught with the risk of false positives and false negatives. False positives occur when beneficial content is incorrectly flagged as harmful, while false negatives happen when harmful content is not detected and remains accessible. Both scenarios present significant risks. False positives can deprive children of essential information and support, as discussed. Conversely, false negatives can expose children to content that can cause psychological or physical harm. Recommender

	systems, as considered by Ofcom, play a significant role in shaping children's online experiences. These systems often rely on engagement metrics to suggest content, which can inadvertently lead to the recommendation of harmful content. For example, children seeking recovery content for eating disorders might be recommended harmful content that promotes disordered eating behaviours due to the similarities in the keywords and themes. NLP-based recommender systems, while not explicitly discussed in the Ofcom report, are increasingly used to automate content moderation. However, these systems have limitations. NLP algorithms can struggle with the nuanced understanding required to distinguish between harmful and beneficial content, especially when the context and intent are subtle. They can generate both false positives and false negatives, leading to inconsistent content moderation outcomes. Moreover, the lack of explicit discussion of NLP in the Ofcom report suggests a gap that needs to be addressed. Incorporating advanced NLP techniques requires continuous updates and refinements, based on a robust and diverse dataset, to improve accuracy and effectiveness. Evaluating the effectiveness of content moderation is a complex task. Metrics such as precision, recall, and F1 scores are typically used to assess the performance of NLP systems. However, these metrics are inherently limited when we consider context-sensitive information. The real-world effectiveness of content moderation must be evaluated by its ability to protect children from harm while allowing access to beneficial content. Human moderators play a crucial role in this process, but they also face challenges such as inconsistency and scalability issues. A combination of automated tools and human oversight, supported by clear guidelines and continuous training. Additionally, transparency in reporting the outcomes of content moderation, including rates of false positives and negatives, can help build trust and ensure accountability. Ideally the report should pay

	consideration to some of these limitations and address some of these considerations. While Ofcom's guidance rightly highlights the risk of harmful content being presented as recovery content, it is also important to recognize the risk in the opposite direction: beneficial recovery content being misclassified as harmful. A purely safety-first approach, which prioritises the prevention of false negatives, may lead to an overabundance of false positives. This approach could limit access to critical support systems, in areas related to mental health, sexual health, and substance abuse recovery. A more holistic approach is necessary, one that carefully weighs the risks of both excluding beneficial content and including harmful content. Ideally, any approach will need to recognize a balance, between protecting children from potential harms whilst also ensuring they have access to beneficial content, such as supportive and educational resources. This misclassification can have significant negative consequences. For example, children in need of support and guidance may be deprived of critical resources that help them cope with mental health issues, substance abuse, or sexual health concerns. A purely safety-first approach, which errs on the side of caution by excluding potentially harmful content, runs the risk of inadvertently silencing valuable information and support. Given that children primarily now access such material via online resources, the risks of incorrectly flagging content could be considerable, effectively blocking access to beneficial content. Thus it is essential to balance the need for safety with the recognition that some content, while sensitive, is beneficial and necessary for children's development and well-being.
Volume 4: How should services assess the risk of online harms? **Governance and Accountability (Section 11)**	

15. Do you agree with the proposed governance measures to be included in the 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. 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?	Confidential? – Y / N
Children's Risk Assessment Guidance and Children's Risk Profiles' (Section 12)	

17. What do you think about our proposals in relation to the Children's Risk Assessment Guidance? a) Please provide underlying arguments and evidence of efficacy or risks that support your view. 18. What do you think about our proposals in relation to the Children's Risk Profiles for Content Harmful to Children? a) Please provide underlying arguments and evidence of efficacy or risks that support your view. Specifically, we welcome evidence from regulated services on the following: 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? 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? 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? 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.	Confidential? – Y / N

Volume 5 – What should services do to mitigate the risk of online harms **Our proposals for the Children's Safety Codes (Section 13)**	
Proposed measures 22. Do you agree with our proposed package of measures for the first Children's Safety Codes? a) If not, please explain why. **Evidence gathering for future work.** 23. Do you currently employ measures or have additional evidence in the areas we have set out for future consideration? a) If so, please provide evidence of the impact, effectiveness and cost of such measures, including any results from trialling or testing of measures. 24. Are there other areas in which we should consider potential future measures for the Children's Safety Codes? a) If so, please explain why and provide supporting evidence.	Confidential? – Y / N

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 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?	Confidential? – N We have considerable concerns regarding the proposals in relation to age verification. As discussed in the ICO's age appropriate design code of practice, there are many relevant rights of the child at play when considering access to online services by children. Perhaps most central here are the right to privacy, the right to access (appropriate) information whilst being protected from inappropriate information, and the right to protection from exploitation. Unfortunately, at times it may seem that there is a tradeoff between these rights. Our concern is that in verifying ages, solutions will be put into place that will not give enough consideration to the protection of children's data. Furthermore, given that adults will also need to have their age verified under the approach outlined by Ofcom, the widespread implementation of age verification methods will lead to impacts on privacy, appropriate access to information, and have the potential to be discriminatory. Adults in particular will face risks to their privacy, their ability to access services and information, and (depending on the service) their freedom of expression and association. The very services which pose a risk to children may be the services which are most likely to severely impact the lives of adults if data is not properly protected (see for example the Ashley Madison data leak, which is thought to have resulted in a considerable number of scams, and potential suicides). We are particularly concerned with the suggested implementation of AI-based age estimation methods. AI-based methods have the additional risk of bias, which is a known concern with facial recognition systems to the extent that they have been found to be racially discriminatory (e.g. Uber's facial recognition tool). Humans are already known to be poor and biased estimators of age, and AI age estimation tools have been shown to exaggerate these biases, with facial expression, gender and age impacting accuracy. Furthermore, we have concerns that alternative age verification methods (such as reliance on credit cards or identity documents) will further the digital divide, given

	that not all adults are able to access to the necessary documentation or acquire a credit card (e.g. they have poor credit, no fixed address or have limited access to banking). Similar concerns have been raised around the requirement of voter identification which is thought to disproportionately affect people from minority ethnic backgrounds, poorer people and those with disabilities, as warned by the electoral commission (with recent polls suggesting a real effect in the latest election). The use of age verification methods also raises concerns for data protection. There is a risk that many services which previously held scant personal data would now have access to identity documentation, identifying photographs, or simply further personal information through requiring users to make a personal account. Whilst services will need to conform to existing data legislation, this will still provide companies with a large amount of additional personal data that will be required for accessing their service. Given these concerns, and those discussed above, we would advocate for the establishment of a robust, decentralised third party verification service (provided by a government body?) much like those advocated for during the COVID-19 pandemic in order to protect privacy and place control over data in the hands of individuals. In the development of this system, adequate thought should be given to the minimisation of impact on individuals who are less able to access identity documents or credit. We would strongly repudiate the recommendation to utilise AI-based methods due to concerns of discrimination.
Content moderation U2U (Section 16)	

36. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views. 37. Do you agree with the proposed addition of Measure 4G to the Illegal Content Codes? a) Please provide any arguments and supporting evidence.	Confidential? – Y / N
Search moderation (Section 17)	

38. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views. 39. Are there additional steps that services take to protect children from the harms set out in the Act? a) If so, how effective are they? 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? 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 provider arguments and evidence to support your views: 41. Do you consider that it is technically feasible to apply the proposed code measures in respect of GenAI 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?	Confidential? – Y / N

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. 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.	Confidential? – Y / N

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 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.	Confidential? – Y / N

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 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.	Confidential? – Y / N

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.