Call for evidence response form

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

Title
Second phase of online safety regulation: Protection of children

Full name
✂

Contact phone number
✂

Representing (select as appropriate)
Organisation

Organisation name
Reset

Email address
✂

Confidentiality

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

Your details: We will keep your contact number and email address confidential. Is there anything else you want to keep confidential? (select as appropriate)
No

Your response: Please indicate how much of your response you want to keep confidential (select as appropriate)
None

For confidential responses, can Ofcom publish a reference to the contents of your response? (select as appropriate)
Yes

Your response

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.
Is this a confidential response? (select as appropriate) *No, not confidential.*
Reset is a philanthropic initiative working to improve the health of democratic information ecosystems, so that Big Tech's business model serves the public good instead of purely corporate interests. We work with civil society and policymakers in the UK, US, Canada, Australia and the European Union.

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

Is this a confidential response? (select as appropriate)

No, not confidential.

The approach to cover services which are "likely to be accessed by children" is welcome because it avoids the known loophole and shortcomings of previous approaches that sought to cover "directed at children".

We would recommend the consideration of the following factors which might indicate that a service is likely to attract child users:

- The nature of the service including:
  - Features: such as **design features** and tools like **filters** and being able to "**like**" content
  - Functionalities: such as the affordances the products offer its users, like "**groups**", **livestream** or **search** for other users
  - Content: such as the **nature** of the content the product serves
- Whether children use other services with these **features, functionalities or content**
- Evidence about the actual user base of a service.

Question 3: What information do services have about the age of users on different platforms (including children)?
Is this a confidential response? (select as appropriate) *No, not confidential.*
● The status quo of opacity in the industry makes it hard to know what data points services have on the age of users. This is because there are currently few incentives or laws to force platforms to disclose this information. This highlights the importance of the forthcoming **transparency powers** in the Online Safety Bill, which Ofcom will have to deploy with regards to gathering information about these issues when the Bill is in effect. ● However, we note that where **investigations** have used simple techniques, like scanning biographies for age ranges, **researchers have easily been able to identify underage users.** A range of easy age estimating techniques are available but do not seem to be widely and consistently applied by platforms.[1] ● Platforms have a large amount of information about the age of users but consistently choose not to disclose these data points to the public and independent regulators.

Question 4: How can services ensure that children cannot access a service, or a part of it?
Is this a confidential response? (select as appropriate) *No, not confidential.*
We defer to evidence provided by the 5 Rights Foundation here.

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?
Is this a confidential response? (select as appropriate) *No, not confidential.*

[1] https://fairplayforkids.org/april-14-2022-new-meta-profits-from-pushing-pro-eating-disorder-content-to-children-on-instagram/

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

- We are not in a position to comment on the types of age assurance or verification technologies on the market. However some key principles should be applied when choosing if/which technologies should be adopted:
    - Any such technology should be **privacy respecting** given the sensitive user data involved
    - Such technologies should be **inclusive** and not **exclusionary.** The reality is that children use the internet and related services for social interaction. This will not change and therefore any steps to increase the safety and well-being of children should be cognizant of this.
    - Any technologies deployed should take into account General Comment 25 on UN Rights of the Child whereby, “promoting, respecting, protecting and fulfilling all children’s rights in the digital environment” is paramount.[2]

We defer to evidence provided by the 5 Right Foundation.

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Is this a confidential response? (select as appropriate)

No, not confidential.

There is demonstrable evidence that platforms routinely make harmful content accessible to young users. This includes:

- **Pro-eating disorder and dieting content:** Research has shown that it is both present on platforms and is often recommended to young people either as content to view[3] or as content creators to follow.[4] The *Facebook Files* indicated that Meta are aware of this problem on Instagram, but failed to take adequate actions.[5]
- **Manospehere and misogynistic content:** Research has identified a problem with both sheer volume of content—such as Andrew Tate content[6]—and the role of

[2] https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

[3] https://counterhate.com/research/deadly-by-design/

[4] https://fairplayforkids.org/wp-content/uploads/2022/04/designing_for_disorder.pdf

[5] https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp_lead_pos7&mod=article_inline

[6] https://www.theguardian.com/technology/2022/aug/06/revealed-how-tiktok-bombards-young-men-with-misogynistic-videos-andrew-tate

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

algorithmic amplification where manopshere content is pushed to boys accounts on YouTube for example[7]

- **Bullying content:** Research around the prevalence of online bullying indicates that bullying content must be prolific; 19% of children aged 10 - 15 in England and Wales reported experiencing online bullying behaviour in 2020[8]
- **Racist content:** Algorithms have been shown to increasingly push racist and stereotype content to children's accounts on TikTok, for example.[9] Research has also shown how harmful the content recommended to young people can be, including white supremacist and neo-Nazi content.[10]
- **Self harm and suicidal content:** Three quarters of Britons report seeing self harm content before the age of 14,[11] indicating a significant prevalence of this content. Investigations from the Center for Countering Digital Hate have demonstrated that new TikTok accounts were recommended self-harm and suicide content within 2.6 minutes of scrolling through content.[12]
- **Extreme violence:** A study of vulnerable young people outlined that 70% of young people reported having seen violent or extreme content online.[13]
- **Dangerous challenges:** Research has shown that dangerous and deadly challenges, such as train or car surfing, are readily available and easily recommended by search functions on social media platforms often without any warnings.[14]

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

Is this a confidential response? (select as appropriate)

[7] See for example, research that highlights the algorithmic recommendation of content to young boys accounts on YouTube https://au.reset.tech/news/algorithms-as-a-weapon-against-women-how-youtube-lures-boys-and-young-men-into-the-manosphere/,

[8] See a summary of the Crime Survey for England and Wales at https://anti-bullyingalliance.org.uk/tools-information/all-about-bullying/prevalence-and-impact-bullying/prevalence-online-bullying

[9] See for example, an investigation into TikTok's algorithm https://au.reset.tech/news/surveilling-young-people-online-an-investigation-into-tiktok-s-data-processing-practices/

[10] https://www.jstor.org/stable/27161413

[11] https://media.samaritans.org/documents/Samaritans_How_social_media_users_experience_self-harm_and_suicide_content_WEB_v3.pdf

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

[13] https://static1.squarespace.com/static/5d7a0e7cb86e30669b46b052/t/618b7cd8b5872f4721c9d59a/1636531420725/Online+Harms+Research+November+2021+-+Full+Report.pdf

[14] https://fairplayforkids.org/dared-by-algorithm/

Evidencing specific harms for children from exposure to content is still an emerging research area, but there are many studies and examples of self-reported evidence of young people explaining how services are harming their well-being. This includes:

- **Self harm and suicide content:** Research suggests that for some young people, time spent on social media—especially where they are exposed to self harm content—increases the prevalence of self harm.[15] Additionally, a Coroner recently declared that a young woman, Molly Russell, died while suffering the "negative effects of online content".[16]
- **Violent content:** There is evidence of a link in poor mental health outcomes and viewing extreme violence or terrorist content,[17] which social media platforms often allow young people to access or promote in their feeds.
- **Pro-eating disorder and dieting content**: There is a link between the use of more social media platforms in general and eating disorders, which is connected to the amount of dieting and 'skinny' content young people consume online.[18]

While social media companies routinely conduct research into the effects of their products on children the results of this research are more often than not, kept in-house. For example, internal research from Meta as seen in the *Facebook Files*, demonstrates:
"In one study of teens in the US and UK, Meta found that more than 40% of Instagram users who reported feeling 'unattractive' said the feeling began on the app. About 25% of teens who reported feeling 'not good enough' said the feeling started on Instagram."[19] Likewise, among teens who reported suicidal thoughts, 13% of British users and 6% of American users traced the desire to kill themselves to Instagram, one presentation from the *Facebook Files* showed.

Question 8: How do services currently assess the risk of harm to children in the UK from content that is harmful to them?

Is this a confidential response? (select as appropriate)

No, not confidential.

[15] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6278213/

[16] https://www.independent.co.uk/news/uk/molly-russell-north-london-pinterest-instagram-b2183199.html

[17] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4803729/

[18] https://onlinelibrary.wiley.com/doi/abs/10.1002/eat.23198

[19] https://www.wsj.com/articles/facebook-instagram-kids-tweens-attract-11632849667?mod=article_inline

Question 8: How do services currently assess the risk of harm to children in the UK from content that is harmful to them?

- Currently, the only entities able to answer this question accurately are companies themselves. Under the Online Safety regime Ofcom will have increased transparency tools at its disposal to learn more about this process. But the era of self-regulation has demonstrated there is not a shortage of risk assessments done internally at companies. But without **transparency requirements** there are **no incentives for disclosing this data.**
- Harms arising to children from data processing might be covered by DPIA as required by the **Age Appropriate Design Code**, but these are not always publicly available nor comprehensive. *(NB any proposed changes by HMG to the UK Data Protection Act need to ensure the Age Appropriate Design Code remains untouched).*
- Carnegie UK have drafted a “Model Code” for social media regulation which outlines in great detail best practice for risk mitigation and risk assessment.

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

Is this a confidential response? (select as appropriate)

No, not confidential.

- **Attention optimization:** services’ primary motive rests in profit, not safety. The structure of the business model as it currently stands is to optimise for attention in order to maximise advertising revenue. Engagement-based rankings, which are the basis of many news feeds provided by services, are not optimised to show users content from their social circles or that might be relevant. Rather the goal is to ensure that users “engage” (whether through positive or negative emotions) via clicks, likes, shares or other forms of engagement. The calculus is simple from the platforms’ perspective: the more engagement with content, the more advertising revenue.
- **Algorithmic amplification:** many news feeds designed by services rank content based on engagement based rankings which optimises for attention, amplification and profits rather than safety.
- **Behavioural advertising:** The impact of behavioural advertising on young people is poorly understood, but evidence suggests that young people may be particularly affected by it. For example, research has shown that despite young people’s privacy concerns, they do not appear to be able to effectively safeguard themselves from the persuasiveness of this advertising.[20] Other research shows that when teenagers are provided with more information and ‘debriefed' about

[20] Specifically, higher levels of targeting using more personalised data generates stronger responses among teens *regardless* of their concerns about privacy. Michel Walrave, Karolien Poels, Marjolijn L. Antheunis, Evert Van den Broeck & Guda van Noort 2018 “Like or dislike? Adolescents’ responses to personalised social network site advertising,” *Journal of Marketing Communications*, https://doi.org/10.1080/13527266.2016.1182938.

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

how behavioural advertising works, any initially strong intentions to make purchases are moderated.[21] Research on younger children have also found that “children seem to process targeted online advertising in a noncritical manner”[22] *vis a vis* adults. This leaves young people vulnerable to economic harm.

- **Extended use designs:** Young people can be vulnerable to extended use designs or ‘addictive’ design features that attempt to keep young people ‘hooked’ on a digital product. These include push notification designed to pull young people back into an app,[23] endless scroll, content recommender algorithms that are “optimised for addiction”[24] (i.e., “trained” to maximise the amount of time young people spend watching videos)[25] to removing video time markers[26] or other features that might remind young people to log off and take a break.[27] In rare cases, this extends to a medical addiction, called Internet gaming disorder.[28] More commonly, extended use design causes constant relationship harm. Intrafamily conflict around screen time is rife,[29] and many teachers report conflict in the classroom over the use of digital devices.[30] These can also cause physical harm, because they can lead to a loss of sleep.[31]

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

Is this a confidential response? (select as appropriate)

No, not confidential.

[21] http://dx.doi.org/10.1016/j.chb.2016.11.050.

[22] https://doi-org.ezproxy-b.deakin.edu.au/10.1080/02650487.2016.1196904.

[23] https://www.dmu.ac.uk/research/research-news/2022/dmu-research-suggests-10-year-olds-lose-sleep-to-check-social-media.aspx#:~:text=Research%20support-,DMU%20research%20suggests%2010%2Dyear%2Dolds%20lose%20sleep%20to%20check,up%20to%20use%20social%20media

[24] https://ssrn.com/abstract=3682048

[25] https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

[26] https://www.wired.com/story/tiktok-time/

[27] For example, Instagram allows users to set daily time limits to prevent overuse. Consumer’s used to be able to self define their daily limit, including setting limits at 10 or 15 min. Earlier this year, Meta set a new ‘limit’ to these daily limits. Consumers can only now set a daily limit of 30 minutes or more (See Natash Lomas 2022 ‘Instagram quietly limits ‘daily time limit’ option’ *TechCrunch*)

[28] As defined in DSM5 onwards (See American Psychiatric Association 2013 *Diagnostic and Statistical Manual of Mental Disorders. 5th edn*. American Psychiatric Publishing Arlington). See also Cecilie Andreassen 2015 ‘Online social network site addiction: A comprehensive review’ *Current Addiction Reports* doi:10.1007/s40429-015-0056-9, who explores the potential for social networking sites to be addictive

[29] Sarah Domoff, Aubrey Borgen, Sunny Jung Kim, Jennifer Emond 2021 ‘Prevalence and predictors of children's persistent screen time requests: A national sample of parents’ Human Behavior and Emerging Tech doi.org/10.1002/hbe2.322

[30] https://www.cnbc.com/2019/01/18/research-shows-that-cell-phones-distract-students--so-france-banned-them-in-school--.html

[31] See De Montfort University 2022 as above

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

- **Terms of service:** currently these are the only rules that govern what content is or isn't allowed on platforms, in most jurisdictions. However, these are written, enforced and evaluated by the platforms themselves. If a platform decides to weaken their terms of service there is currently no accountability or recourse for users.
- In Australia there is a **public facing complaints mechanism** where young people have recourse under the Online Safety Act directly to the Office of the eSafety Commission. Similar infrastructure does not exist in the UK but might be worth contemplating.
- The current lack of governance oversight and accountability as it relates to child users and platform safety is why the **transparency measures** and **information gathering powers** in the **Online Safety Bill** will be so important. Ofcom will need to exercise the **full extent of its powers** under the new regime in order to increase governance and oversight of platforms and services.

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

N/A

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[32]

Is this a confidential response? (select as appropriate)

No, not confidential.

[32]

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[32]

- Terms of service are platform-specific. These terms are written, enforced (or not) and evaluated by platforms themselves. There is currently no external oversight of a services' enforcement of their own terms of service. Similarly, the status quo provides no external oversight over internal decision making or the prioritisation of content. Without increased **transparency**, we are unable to evaluate the decision-making of platforms.

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Is this a confidential response? (select as appropriate)

No, not confidential.

N/A

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Is this a confidential response? (select as appropriate)

No, not confidential.

- Australia's Online Safety Act provides a **public facing complaint mechanism** for children and families **harmed by legal content** that falls into the definition of 'bullying', where it is directed at a child. Under this regime, complaints are investigated by an **independent regulator** and if they are found to be 'bullying', the regulator can make a number of recommendations, including demanding that platforms remove the content within either 24 or 48 hours depending on the nature of the content.

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

Is this a confidential response? (select as appropriate)

[Please select]

N/A

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Is this a confidential response? (select as appropriate)

No, not confidential.

The functionalities and features that are designed to prevent or mitigate harm to children are fundamentally lacking. There has been a significant under-investment by tech companies in safety features for young people, especially where they might impact their profits. However, a number of products or features do exist:

- Privacy respecting age verification technologies
- Privacy risk assessments as mandated by the Age Appropriate Design Code
- While functionalities or features for child safety might exist, platforms and services need to be able to demonstrate to users and regulators that these measures are actually effective. **Transparency measures** will be a key mechanism for measuring efficacy, under the online safety regime.
- Providing **data access** for **researchers and civil society** organisations via a **privacy-respecting mechanism** is an important avenue for evaluating the efficacy of such features or functionalities. This mechanism should be prioritised by Ofcom under the Online Safety regime.
- **Where functionalities or features exist that *can* mitigate risk of harm to children, these features should be turned-on by default for everyone under-18.**

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Is this a confidential response? (select as appropriate)

No, not confidential.

- Without commenting on the technical feasibility or effectiveness of system functionalities, it is clear that there are many changes or mitigation measures that platforms could make in order to reduce both risk and impact of content that is harmful to children on their service. This will not happen without regulation or a shift in the incentives structures, whereby profit is not the only outcome companies are optimising for. Hopefully a more robust regulatory framework and compliance regime in the UK will result in further uptake of mitigation features and measures.

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Is this a confidential response? (select as appropriate)

No, not confidential.

- **Terms of Service:** Platforms can enforce their existing terms of service which often include robust protections of children. But enforcement of these terms of service needs to be increased in effectiveness; and platforms need to be more transparent about what measures they take to enforce these terms of service and where violations of said terms are happening.
- **Privacy respecting age assurance:** a privacy respecting age assurance regime would ensure users are protected from age inappropriate material and services without restricting the ability of adult users to access content and information.
- **Age Appropriate Design Code:** adhering to the existing legislation relating to childrens’ use of online platforms and services, such as the Data Protection Act, the Age Appropriate Design Code and the forthcoming Online Safety Bill. Independent regulators in the UK such as the ICO and Ofcom need to be robust in their enforcement of existing legislation in order to support the safety and wellbeing of UK child users.
- **Transparency:** By providing researchers and civil society access to data relating to their services’, platforms would be increasing the body of research and public scrutiny about their products and services. This helps informed decision-making amongst regulators, parents and young people, while also helping to highlight where platforms are falling short in their commitments to the safety and wellbeing of UK child users.

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

- **Accountability:** Providing easy to access means to report harm and seek help where issues arise. These need to be more effective than computer-based 'flagging' tools for content that causes harm.

We also defer to evidence provided by the 5 Rights Foundation here.

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

Is this a confidential response? (select as appropriate)

No, not confidential.

See previous responses relating to algorithmic accountability, transparency, risk assessments and mitigation measures.

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Is this a confidential response? (select as appropriate)

[Please select]

N/A

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Is this a confidential response? (select as appropriate)

[Please select]

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?
N/A

Question 22: How are human moderators used to identify and assess content that is harmful to children?
Is this a confidential response? (select as appropriate) *No, not confidential.*
Since most platforms and services outsource the work of human content moderation to third-party service providers, there is little to no transparency about the following key metrics: ● The number of human moderators working in a specific language; ● Geographic location of moderators; ● Working conditions of moderators; ● Incentive structure surrounding priority content moderation

Question 23: What training and support is or should be provided to moderators?
Is this a confidential response? (select as appropriate) *No, not confidential.*
● Human moderators are often outsourced by big tech companies to third-party service providers. A number of high-profile cases have highlighted the challenges faced by human moderators such as: severe understaffing, little to no mental health support, PTSD, and labour exploitation. The work of content moderation is psychologically and emotionally taxing. Moderators should be compensated appropriately, provided with adequate psychological support and providers should have robust safeguarding policies in-place. ● Reporting done by Billy Perrigio at *TIME*, as well as the excellent work done by Foxglove and the Signals Network starkly highlight these issues.

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

Is this a confidential response? (select as appropriate)

No, not confidential.

N/A

Question 25: In what instances is content that is harmful to children, that is in contravention of terms and conditions, removed from a service or the part of a service that children can access?

Is this a confidential response? (select as appropriate)

[Please select]

N/A

Question 26: What other mitigations do services currently have to protect children from harmful content?

Is this a confidential response? (select as appropriate)

No, not confidential.

Much of the risks young people face and the harms they experience remain hidden and known only to platforms. Increasing transparency around these may help drive positive changes. Measures such as data access for researchers, or requirements for public transparency around risk assessments or DPIAs, might help further mitigate harms.

Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?

Is this a confidential response? (select as appropriate)

[Please select]

N/A

Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?

Is this a confidential response? (select as appropriate)

No, not confidential.

Many questions in this consultation could be better answered with **increased access to data for researchers and civil society.** Consequently, under the Online Safety regime Ofcom should **prioritise** establishing a **privacy-respecting mechanism** to allow for such access. Through independent research of service design, product changes and platforms' business models, Ofcom will have a dramatically increased evidence-base on which to make regulatory interventions.