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
YourDataKey provides a real-time, verified Digital Identity controlled and used by citizens in all virtual and most physical situations – included in this verified Digital Identity are 'characteristics' such as DOB, Residency status, home address, gender, etc., – the Digital Identity provides the appropriate response to threshold challenges (such as age, location etc) from organisations, platforms, networks and e-tailers

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
1. Promotion by parent/guardian 2. Self-initiated access 3. Peer group recommendations 4. 'nudges' from multiple (social) media platforms

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
1. Not all platforms hold age related data (nor other threshold characteristics) 2. User provided/self-certified – therefore subject to wilful inaccuracy 3. Binary 'over 18?' rather than actual age via DOB

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

1. Establish a universal digital ID service(s) that validates and verified the individual's characteristics such as age
2. As part of that ID service ensure that parent/guardian 'vouching' for under-age individuals is in place
3. Platforms to require certified digital ID before account set-up and guest access

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

Mostly self-certification is in place today; some legacy document scanning and some biometric/AI filters are in current deployment, but with the advent of deepfake AI their ongoing effectiveness is being questioned

Newly emergent solution using real-time verification of an individual's ID and characteristics like age, from multiple trusted data-sources deals with 2 sides of the issue –

- Illegal content posting
- Underage access to legal content

Cost to providers depends upon the service deployed and the service provider – the real-time data verification is fractions of a penny at scale – so irrelevant

Costs to individuals for real-time verification is free at the point of entry; scaling to <£10 per annum for a fully inclusive premium family subscription

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
No

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) No
The evidence is daily in the media and increasing

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
Differently, on a scale of ‘they don’t’ to ‘they do X&Y, because it keeps the authorities happy’

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

1. UK specific Laws
2. Accessibility to content sourced from outside of the UK
3. Integrity of the process establishing the age/age-threshold of the user
4. Integrity of the process establishing a traceable ID of the person posting content

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

A clear set of laws that the platform must meet and be seen to meet; civil and criminal (for Directors) consequences – enforced without undue delay and published

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)?

Is this a confidential response? (select as appropriate)

No

Terms of Service and Policy statements are 'get out clauses' for the provider and unless wilfully ignored by a parent or guardian of a child, should not be considered in mitigation.

Age of user is easily verified by the provider, there is no longer a need to 'estimate' age when it can be definitively established in real-time

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?[1]
Is this a confidential response? (select as appropriate) No
See response to Q 11 above – in terms of real child safety this is irrelevant

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
Simple pop-up screen on exit/log-off asking a binary question and if response is negative – 1. Recording the response (for statistical analysis) and 2. taking the user to the reporting/complaints screen(s) 3. on completion thanking the user for their help

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
No

[1] See A1.2 to A1.3 of the call for evidence for more information on the indicative list of harms to children.

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)

No

On the basis that the report/complaint comes from a traceable, verified user –

1. Immediate systemic suspension of access to the content identified
2. Followed asap by human review
    a. If complaint is upheld; internal QC review or identification of individual posting the content
    b. Action against the Individual posting the content
    c. If appropriate authorities advised
    d. Metrics provided to Governing body at agreed frequency

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

1. Legacy processes associated with account holder/user age – mostly easy to work around
    a. Self-certification at initial account set-up or service sign-on
    b. Official document scan
    c. Facial scan & AI age estimation – now at risk with Deepfake type AI capabilities
2. New to market, real time dynamic identity and age verification from multiple trusted sources
3. AI scans searching out illegal content

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

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?

1. Stop guessing at age qualification and put in place 100% definitive ID & Age verification
2. Continue to develop the AI content monitoring capability removing illegal content
3. See reporting/complaint response Q13

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

1. By shielding the child from the legal but age-inappropriate content = definitive ID & Age verification

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

1. Continuously improve the AI categorisation of content/ content-links by age threshold
2. Promptly action user complaints/reports (see exit pop-ups Q 15)
3. Work with sources of linked content to the common good (they are all under the aegis of current UK laws)
4. Sever relationships with unreliable sources of linked content; report to the Governing body and publicise

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)

No

1. Ensure that content is provided by identified sources, subject to UK legal scope – it is their continued anonymity that is the core of illegal content provision
2. Continuously improve the AI categorisation of content/ content-links by age threshold

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)

No

Am not in place to advise

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

In different ways by different platforms – content posting is on such a scale that human moderation needs to be reserved for post AI content review and user complaints/reports

Statistics based minimum investment and standards for human moderation could be established as a standard by the Governing body

Question 23: What training and support is or should be provided to moderators?

Is this a confidential response? (select as appropriate)

No

See response to Q22 above

Human Moderators will be/are affected by what they see; impacting on their performance and own health; They require

1. Regular mental health support sessions
2. A work cycle that includes scheduled 'down-time' allowing them to recover their mental health and functional sharpness
3. Belief that steps are in place to deal with illegal content providers …….. and that they are helping to win the battle

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

Unable to provide a response to scale.

Regarding automation bias – a process of prompt human review of automatically banned content that is the subject of an appeal is an appropriate solution and a viable use of human review

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)

No

Unable to provide a response as we are not a platform provider

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

Is this a confidential response? (select as appropriate)

No

Unable to provide a response as we are not a platform provider

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)

No

1. All actions are upon the platform/provider
    a. Remove the easy work arounds available on most platforms. Insist upon definitive (not estimated) age identification at account set-up, guest access log-on and other appropriate points in the provision of the service
    b. Hold providers criminally and legally liable – unless they can evidence their actions under point a) above
    c. Fast-track early prosecutions and publicise outcomes
2. Consider placing children who bypass any age thresholds, on a temporary 'stop' list for non-standard access; this would involve their parents/guardians knowing of the issue and could use the existing 'stop-list' capabilities of the Disclosure and Barring Service

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

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?

Ensuring that all users are real people (albeit some allowance for alias usage) by real-time Digital Identity verification including characteristics such as age thresholds and more; and that these real people know that in the event of an illegal act they can be/will be traced and prosecuted; will see a huge reduction in harmful content being posted.

That same discipline with the platform verifying identity & Age, will remove the easy access of under-age users to adult categorised content

A similar approach to that of prosecution under Health & Safety legislation (guilty until proven innocent) could be adopted in the early days to ensure that platforms/content providers engage with the required culture change.

Because platforms can deploy verified Digital ID and age verification now, such deployment should be a viable defence in the event of prosecution.