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
The AVPA is the global trade body for the suppliers of privacy-preserving, independent age assurance (both age verification and age estimation) technology and associated businesses

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
In Para A1.10, the Call for Evidence refers to "age assurance or age verification." This phrase is repeated elsewhere several times. It suggests a misinterpretation of the language being adopted by the industry, platforms and in particular international standards bodies. Age assurance is the general term, within which both age estimation and age verification sit. So it is confusing to suggest a choice between age assurance and age verification. There is a choice between age estimation and age verification – but both are forms of age assurance. When first implementing the Age-Appropriate Design Code, the Information Commissioner re-named it as the Children's Code and applied it only to sites designed for children. The statute was clear that the legislation applied to sites "likely to be accessed by children". This position has now been rectified, but Ofcom should be careful to apply the legislation accurately however it is eventually passed. If the current phrasing is retained, then there is no shortage of evidence that children are likely to access services that are definitely not designed for them, such as pornographic websites, and sites not intended for them, such as dating sites which may operate a voluntary minimum age of 18 but are popular with those below that age.

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

Question 3: What information do services have about the age of users on different platforms (including children)?

• the methods used to gather any information that can assist in estimating or assuring a user's age, either at the point

1. user first accesses the service

When a user first accesses a service, they may not open an account or provide any personal details, but may simply be visiting the site to access its content. The platform will have no data beyond that which is inherent to any site visitor, such as their IP address and information of which hardware and software is being used. So, in general this will be insufficient to assist in estimating or assuring age. Proactive effort is required to obtain further data.

The user may choose or need to open an account. This process can provide an opportunity to gather more information on which to assess age.

The options for a new or existing user include:

1. Self-declaration – this could exclude very young children unable to tick a confirmation box or enter a date of birth, but generally provides little or no assurance.
2. Biometric age estimation – if the user is willing to share a facial image, a sample of their voice or other features that evolve with age, machine learning can create artificial intelligence to estimate the age of a user. Currently the best in class facial age estimation achieves this around key ages of 13 or 18 to within an average of +/- 1-1 ½ years error.
3. Identification documents – through reading either optically or electronically the data on ID such as a passport, driving licence or national ID, and for added security, face matching or comparing the image obtained with a live image of the user, and undertaking a document authenticity check
4. Reusable digital identity app or wallet - can enable the selective disclosure of verified details from a government issued identity document such as over 18, under 18, over 13
5. Open banking – a user can log into their bank and give consent for the bank to disclose their age
6. Authoritative databases – confirming details provided by the user with one or more databases such as the electoral roll, credit reference agency records or the Passport Office.
7. Mobile network operator records – another arguably authoritative database with the potential extra security of knowing the user in possession of the mobile phone.
8. Credit card usage – as in the UK only adults are allowed a credit card, a user can prove their age by demonstrating they have access to make payments with one.
9. Vouching – authoritative individuals can provide a reference as to a user's age

ii. or subsequently, for existing users only

Behavioural age estimation – similar to biometric, but based on action taken by a user, such as how they play a computer game, which connections they make to other online users, or what their interests are

Depending on which option is used, the platform will need to obtain the required data.

Question 3: What information do services have about the age of users on different platforms (including children)?

The user may also make a purchase. This process can provide an opportunity to gather more information on which to assess age; a payment method which may require a name and/or an address to be supplied, and for physical goods, a delivery address will be necessary (but it may not be that of the user unless this is required by the payment method).

• what, if any, mechanisms are available to enable services to identify children in different age groups (for example children below age 13, aged 13-15, or aged 15-17); and

Many of the methods above are of limited use to children because they do not have the required records or documents. Those which may be suitable include:

1. Self-declaration
2. Biometric age estimation
3. Identification documents – relatively fewer children than adults have passports; none under 15¾ has driving licences
4. Reusable digital identity apps, are available from 13 plus and can be set up with a Citizen Card, Passport or Driving License - these allow a data minimised attribute to be shared e.g. 13+, 15+, 16+, 18+ or 18-
5. Open banking – older children may have an online bank account
6. Authoritative databases – only the Passport Office is currently available but the UK Government is intending to allow certified ID providers to access more data such as education and benefits records which would include children .
7. Mobile network operator records – while many children have mobile phones, these are generally on accounts created by their parents
8. Vouching

iii. or subsequently, for existing users only

1. Behavioural age estimation

• how approaches to assessing the age of users are evolving.

While new methods for estimating or verifying age are emerging all the time, the key evolution is towards **interoperability**. This allows a single age check to be re-used across multiple sites. It exists today to the extent several sites share the same age verification provider's service. The www.euCONSENT.eu project demonstrated successfully a proof of a concept where AV providers could rely on each other's age checks. The solution used a token on a user device to alert an AV provider that they need not re-check that user but could be redirected to another AV provider who has already completed a check, and get that provider to confirm the user is of the correct age to access the new site.

In the context of this question, a new user visiting a site for the first time could be recognised as having already completed an age check for some other purpose, and the provider of that check could confirm the user's age or age-range to the new site without any user action required. From time-to-time, the user may need to re-authenticate – to prove they are still the same person who did the original age check, but that could be as frequent or infrequent as regulators or the sites' own policies required.

Question 3: What information do services have about the age of users on different platforms (including children)?

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

Critically, this question asks how a service can prevent children accessing it. So this rules out any form of parental controls, including account confirmation, because the service has no control over what a parent decides to do. A parent may confirm their child is older than they really are to give them access to materials the site may have determined were unsuitable.

So services must use age assurance to ensure children cannot access them or parts of them. Both verification and estimation can be used, as estimation can be as effective as verification if the age the software tests for is sufficiently over (or under if applicable) the age limit. Some verification approaches rely on the possession of identity documents; for accessibility estimation approaches or verification from authoritative databases which include a large share of the population e.g. educational records, can promote societal inclusion.

• how age assurance policies have been developed to date and what age group(s) they are intended to protect;

To date, age assurance has mostly aimed to distinguish adults from minors at the age of 18. As this has generally been for compliance with legal minimum ages for sales of certain goods, it has typically been through age verification methods, rather than estimation.

In Germany over 90 approaches to assess age, have been approved by the German age regulator, the KJM, as detailed on the KJM Raster Page. The German FSM also accredits approaches with a Seal of Approval.

More recently, age assurance has been deployed to improve the accuracy of the recorded age of users of social media. So, for example, Instagram is using age estimation to confirm a user's age if they initially claim to be under 18 and then wish to amend their age to be 18+. Some social media platforms claim to use behavioural or facial age estimation to confirm the ages claimed by users and flag accounts for review where this is anomalous. Other sites use facial age estimation to assess if a teen is of the correct age to be on the 13-17 livestreaming platform, versus the 18 plus platform. Sites with younger children also deploy facial age estimation to assess if an adult is of age to give parental consent, e.g. over the age of 25.

Adoption of available age assurance solutions has been slow. In the UK, the ICO's reluctance to enforce the Age-Appropriate Design Code, and a delay of 2 years before Ofcom announced an

Question 4: How can services ensure that children cannot access a service, or a part of it?

enforcement strategy for video-sharing platforms, has left platforms weighing the costs and benefits of implementing age assurance and concluding it is not yet worth it – regardless of the existing legal requirements applying to them. And to date, most of the websites affected are large, well-known platforms with a UK presence who generally invest heavily in maintaining their reputation. Under the Online Safety Bill, up to 5 million new sites in the adult industry based in many offshore jurisdictions with a much lower level of concern for public relations will be in scope. Only firm regulation at scale is going to deliver the policy objectives of the new law

• if the service is tailored to meet age-appropriate needs (for example, by restricting specific content to specific users or part of a service), how this currently works or could work;

As far as pornography is concerned, there are already sophisticated and effective solutions available from the wider SafetyTech sector which can detect over 99% of adult content, and have done so in live operation reviewing thousands of items of content a week. The same technology can be re-purposed to identify other harmful content, although it will take time to train and trim the artificial intelligence, particularly for less easily defined harms (A reference to “diet pills” may be easier to spot than a harmful discussion about eating habits). Regulation should take into account the capabilities of technology to facilitate protections, and AI can be supplemented with reporting mechanisms of course.

As well as identifying content, particular users could be registered as suppliers of potentially harmful content and all their material, regardless of its content, put behind age assurance checks.

• how the efficacy of age assurance policies is or could be monitored; and

A number of platforms claim to monitor the effectiveness of their age assurance by the number of underage users detected. This is the equivalent of measuring the level of theft from a shop by the amount of goods in the possession of shoplifters when they are caught.

Efficacy can be assessed by

1. Assessing the theoretical effectiveness of an age assurance solution
2. Empirically testing the effectiveness of age assurance solutions with users whose real age is known to the auditor
3. Testing for circumvention through presentation attacks and other methods
4. Quantitative research surveying the population to estimate the number of underage users with access to a platform

• how services can identify users that do not meet any relevant age limits and how is this appropriately addressed?

See above for methods of age assurance.

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

• how accurate these technologies are in verifying the age of users,

Accuracy can be considered in a narrow interpretation, meaning how close to the real age of the user is the age determined by any given method of age assurance (including both estimation and verification techniques), or more widely.

For verification checks linked to documents to be deemed accurate, it is necessary to assess how well document checks are undertaken (e.g. with liveness detection, document authenticity checks and face matching to an acceptable level) and to also reauthenticate.

For estimation, the estimated ages will be within a range of tolerance They may be either under or overestimates, so "accuracy" can be measured first by considering the average error, and secondly by looking at how far wrong the most extreme errors are. This allows solutions to be described by how close to the truth they are most of the time, and how far from the truth they can be some of the time.

The impact of this distribution on the effectiveness of age assurance can be mitigated by testing for an age above (or in some cases below) the required age so the majority of those who pass the estimation test are actually above the required age, even if their age is overestimated.

With age verification, the outcome is generally binary – either the correct date of birth is obtained or it is not. So, the results in terms of age assurance can be true positives and true negatives, or false positives and false negatives.

The level of inaccurate results will be dependent on many factors: the quality of the authoritative data source; the capabilities of optical character recognition when documents are inspected remotely; the effectiveness of authentication to check the user is the person to whom the evidence of age applies.

The effectiveness of all methods is affected by the extent of anti-fraud measures applied across the solution; how often the check is made and how often the user is asked to re-authenticate if they wish to rely on a previously obtained age check.

So it is a multi-dimensional assessment of the reliability of any given method. In general, the reliability of all methods is converging as they are all constantly improving. And often providers will use more than one method to build an overall level of confidence in the age, where one contradictory data point may be outweighed by three other consistent data points. This implies that the level of accuracy can be selected as a matter of discretion, although this will often be weighed against cost and the impact on the user experience.

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 do know that age assurance solutions are in general far more accurate and effective than human estimates of age or processes that rely on humans to ask for proof, authenticate the owner and check the proof is not altered, borrowed or stolen.

Therefore, Ofcom should consider outcome based benchmarks for age assurance as an initial approach e.g. "Age assurance to prevent a child under 13 opening a social media account should prevent at least 50% of children under 13, 90% of children under 11 ½ and 99% of children under 9."

• whether accuracy varies based on any user characteristics, and how effective they are at preventing children from accessing harmful content;

Early attempts to use machine learning to develop artificial intelligence that could estimate gender or undertake facial recognition (not estimation) from facial images were found to be biased by skin tones. The effectiveness of AI is limited by the diversity of the training data used to develop it. These solutions were developed with a homogenous training data set so performed poorly on a diverse population.

This is a risk that has been addressed from the outset by those developing AI-based estimation solutions for age assurance. Testing of age estimation systems now includes testing for bias to ensure that it is minimised, and evidence has been published of rapid progress towards eliminating such bias. It is also key to be able to state the source of images and confirm that they have been collected in accordance with GDPR or other local laws..

• any potential unintended consequences of implementing age assurance (such as risk of bias or exclusion), and how these can be mitigated;

For age estimation - through diverse training data and testing to prevent unacceptable levels of bias.

In general, exclusion risks are mitigated by offering users a wide range of methods of age assurance.

This can be significantly enhanced through interoperability. This is where one age check can be used across multiple services, even if they are using different age assurance solutions individually. A user need only find one solution that they can access, and they can then re-use that check anywhere even if that service does not offer a usable solution for the particular user.

• the safeguards necessary to ensure users' privacy and access to information is protected, and over restriction is avoided;

While GDPR provides a general level of protection for privacy and data security, there is a strong argument for pre-emptive regulation to ensure high standards are achieved and maintained, given the sensitivity and volume of personally identifiable data that may be used as part of the age assurance process. We already recommend compliance with BSI PAS 1296:2018, and the forthcoming ISO and IEEE standards include extensive requirements in these areas too. Ofcom should consider requiring age assurance to be delivered to these standards or their equivalents,

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?

and incentivise this by recognising certification schemes, and taking the use of certified solutions into account in any enforcement action.

• which methods of age assurance users prefer, when offered a number of ways to verify their age;

This can vary by use-case. It is likely to be different for situations where estimation is sufficient, than when users must choose a form of verification. There is some evidence users prefer verification methods they consider more anonymous, such as supplying a credit card number.

• the cost of implementing and operating such technologies;

The UK Government impact assessment for the Bill estimated age checks would cost 10 pence each, but was heavily caveated.

Our members have been asked to assist with the creation of such estimates, but it is not generally possible to provide specific figures. It will depend on the method, with some incurring third party costs such as credit reference agency fees, and others with much lower marginal costs, such as estimation techniques (although it would be wrong to assume these have no marginal cost attached to them, and there will be investment or licensing costs). Pricing will be dependent on the opportunity; some sales may be strategically more advantageous than others.

The market has grown significantly – the AVPA had just 6 members in 2018 and now has 26 and is still expanding. This has put inevitable downward pressure on pricing.

Interoperability could also significantly impact pricing, with checks which can currently only be re-re-used when a single provider supplies multiple services, being reapplied across far more services through competing providers.

But the discussion about pricing is moot because it would clearly be wrong to absolve a service carrying content harmful for children from a duty to prevent minors accessing it on the grounds of cost. Even if every age check cost it £10 per user, if a site's business model is unsustainable if it has to pay that level of fee, then it should not be operating because it cannot do so without causing harm to children. We do not excuse construction companies from health and safety legislation because it is too expensive to implement.

• how age assurance and age verification or related technologies may be circumvented; and

There is a balance to be struck between the level of effectiveness and the impact on the user experience. In theory, you could design an age assurance solution (either through verification or estimation) that always ensured with 99.9% certainty that a user was 18+. This would entail the highest level of assurance with the original age check – perhaps reading a Passport's data from its chip, confirming the data with the Passport Office, checking the document was not lost or stolen, and then comparing the image from the passport with a live video of the user being instructed to repeat randomised phrases while carrying out equally random movements. You could then confirm that it is the same user accessing the device every five minutes by asking for a PIN, and perhaps a fresh selfie image every fifteen minutes or at random intervals to ensure the PIN was not given to a younger person in the intervening period. An extra check could be triggered if the

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?

user switched from one game to another, or a slightly longer period was detected from finishing one game to starting the next, or the way the controls were being used was noticeably different. But it would be highly disruptive to the user.

At the other end of the spectrum, a single age check could be relied upon for a year, with a password entered once a month to confirm it is the same user whom you trust not to share that password with someone to young, and to keep their device locked or log out of their user account when finished playing.

More technical circumvention is also possible, but the effectiveness of this is a function of time and money. An individual user may not have a lot of time and cash to try and get round an age check. There will be better funded attempts with more time invested by those who seek to beat the technology and then offer that at scale – but they will also want to be paid for their work by users who then take advantage of such mechanisms, thus again limiting the use of the evasion through its cost.

• what mitigations exist to reduce circumvention among users.

As with all technical security, it is a cat and mouse game, and providers and any interoperability network will need to be vigilant for signs of largescale attacks, and respond to mitigate the impact on the overall integrity of the age assurance ecosystem.

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)

[Please select]

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)

[Please select]

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

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) *[Please select]*

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) *[Please select]*

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) *[Please select]*

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

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
Where age assurance solutions are used, services should be transparent about how these are provided, and if they are audited and certified to deliver high standards of privacy and data security. Users should be advised how to rectify inaccurate decisions either through the service or directly with the provider, depending on what is agreed between the service and the AV provider if this is a third party.

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) *[Please select]*

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) *[Please select]*

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

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

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) *[Please select]*

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]*

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) *[Please select]*

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.

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)

[Please select]

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)

[Please select]

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)

[Please select]

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)?

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]*

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 22: How are human moderators used to identify and assess content that is harmful to children?
Is this a confidential response? (select as appropriate) *[Please select]*

Question 22: How are human moderators used to identify and assess content that is harmful to children?

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

Is this a confidential response? (select as appropriate)

[Please select]

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)

[Please select]

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]

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?

Question 26: What other mitigations do services currently have to protect children from harmful content?
Is this a confidential response? (select as appropriate) *[Please select]*

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]*

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) *[Please select]*

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?