Department for
Science, Innovation
& Technology

Research and analysis

Age Assurance Data Access Study

Published 26 March 2026

Contents

Executive summary

This report presents findings from an exploratory study on data gaps affecting age assurance (AA) tools and potential solutions to improve their effectiveness. The primacy of issues related to training and testing data access led the study to concentrate on statistical age estimation (AE). The focus was further narrowed to facial age estimation (FAE) due to its status as the most widely adopted statistical AE method and the one with the clearest evidence base of data-related

performance gaps. The findings are based on a literature review, expert consultations across sectors, and legal input to assess the feasibility of proposed solutions.

Data access issues

A key challenge in advancing FAE tools is the limited availability of high-quality, representative training and testing datasets, especially for children. Despite significant progress in FAE accuracy, tools still struggle to distinguish users within narrow age bands (e.g. under 13 or under 18) and exhibit bias across gender, skin tone, ethnicity. These limitations are rooted in data gaps and compounded by ethical, legal, and financial challenges in collecting diverse facial images, with collecting images of children posing a particular challenge. Improved datasets could reduce bias and enhance accuracy, but some accuracy limitations may persist due to biological and technical constraints. There are evidence gaps around the real-world impact of limitations in FAE performance attributable to data gaps. This impact is likely to vary across sectors: in some (e.g., online gambling), regulatory requirements render FAE inapplicable, while in others (e.g., retail or online gaming), improved FAE could support safer, more inclusive AA provision and contribute to reducing online harms and improving digital access.

Potential solutions

Facial image database
The study explored the feasibility and design of a facial image database solution that enables access to high-quality, representative training and testing data. The market alone is unlikely to develop such a database solution. Ethical and legal constraints, along with substantial cost barriers, mean that most AA vendors, especially smaller or early-stage companies, lack the resources to collect facial images, in particular of children, at the scale and quality required. This creates a clear case for considering the desirability of public intervention. Three design options for a potential database solution were explored:

- **Option 1**: An external validation dataset focused on demographic groups underrepresented in the data, notably children.
- **Option 2**: An external validation dataset representative of the wider population.

- **Option 3**: A comprehensive solution including training, internal validation and external validation datasets.

While Option 3 could deliver the most far-reaching benefits, it would require significant investment, estimated in the millions, and faces major commercial viability challenges. There are also critical legal considerations applicable to the development and operationalisation of this database solution. Consent is the only viable legal basis for collecting and processing facial data, particularly when children are involved. Strong data minimisation practices, clear governance roles, and robust technical safeguards would be essential to meet UK GDPR requirements. A full Data Protection Impact Assessment (DPIA) would also be necessary before proceeding.

While, a facial image database could offer important benefits in improving the accuracy and fairness of FAE tools, particularly for safeguarding children online, these benefits are not well evidenced. A database solution would require careful design, phased development, strong safeguards and significant public leadership to overcome the many practical and legal hurdles. A roadmap for further action would involve:

- Coordinating with existing benchmarking efforts to avoid duplication
- Clearly defining the database's purpose, potential applications, governance structure and legal basis
- Refining data requirements and collection strategies
- Engaging with stakeholders to validate cost assumptions and expected benefits
- Undertaking a comprehensive legal assessment and DPIA

Funding in novation in AA
In parallel, the study considered a lower-cost alternative: an open innovation challenge aimed at supporting innovation in the development of age assurance tools. This approach offers a flexible and scalable way to foster innovation across a wider range of technical solutions. While promising, this type of intervention may not address efficiently and fully key research gaps, as there is no evidence to suggest that funding is the primary barrier to further research and solutions often still depend on access to high-quality, representative datasets for validation and refinement.

1. Introduction and overview of the study

The Department for Science, Innovation and Technology has taken a range of steps to support the widespread take up and development of age assurance technologies, particularly through provisions in the Online Safety Act (OSA). The OSA outlines that in-scope services must use highly effective age verification or age estimation to prevent children from accessing the most harmful kinds of content online ('primary priority content'). However, services are not required to use any specific age verification technology, reflecting the OSA's tech-neutral, outcomes-focused approach that allows for and encourages innovation in age assurance. Ofcom, the independent regulator for the OSA, has set out codes of practice and guidance on the actions services should take to comply with the Act. In January 2025 it published guidance on highly effective age assurance, which clarified which assurance solutions are and are not currently considered 'highly effective', to help in-scope services provide age-appropriate online experiences for their users. Under the OSA, a requirement on Ofcom means that, in recommending age assurance technologies, Ofcom must have due regard to a set of principles that include considerations of effectiveness, lack of bias, ease of use, interoperability and privacy. It is also required to produce a report assessing how services have used age assurance to comply with the Act, as well as its effectiveness. To support further innovation in age assurance, the Department for Science, Innovation and Technology has carried out research, of which this report is one example. This report presents the findings from an exploratory study investigating data access issues impacting the performance of age assurance (AA) tools and potential solutions to address these data issues and encourage the development of more effective AA tools.

The objectives of the study are to:

- Gain a better understanding of the issues related to data gaps and access that hinder the development, accuracy and performance of AA tools, and their impact on AA provision.
- Explore and identify key features of solutions that could help mitigate the identified data issues and contribute to the development of more effective AA tools, particularly by enabling secure access to appropriate training and testing data.

The focus on challenges related to training and testing datasets means that the study is focused on statistical age estimation (AE), rather than AA in general. The focus of the study was further refined to specifically address facial age estimation (FAE) for the following reasons:

- FAE is by far the most established statistical AE method, with the broadest market adoption. It is among the two AE techniques presently recognised by Ofcom as having the potential to effectively prevent children from accessing age-restricted

online content and services[footnote 1]

- The evidence analysed for this study was more informative about data-related challenges affecting the accuracy of FAE (compared with other AA methods), thus providing a stronger evidence base for identifying relevant and feasible solutions.

The study focuses exclusively on online cases with an application for online safety in the context of the Online Safety Act's implementation.

The study involved a literature review as well as consultations with AA solution providers and testing organisations, private sector organisations that have looked into/adopted AA solutions, public sector organisations, academics with expertise in AA and data access issues, and charities and not-for-profit organisations operating in the field of (child) online safety. The study also sought input from legal experts to assess the feasible legality of the features of the explored solutions.

The report is structured as follows:

- Section 2 presents findings on the scale and nature of data access issues impacting the development of AE tools, and more specifically FAE tools
- Section 3 presents a potential solution in the form of a facial image database, including the likely challenges, limitations and costs involved with the solution
- Section 4 discusses the comparable benefits and limitations from an alternative solution involving grant-funding or open innovation challenges to supporting the development of more effective AA solutions
- Section 5 provides a conclusion to the report and study findings, as well as thoughts on potential next steps.

2. Scale and nature of the issue

This section discusses the scale and nature of the data access issues impacting AA tools, and FAE tools more specifically. The section covers:

- **Data limitations:** A major barrier to developing accurate FAE tools is the limited access to high-quality, representative datasets. While dataset size and diversity have improved, challenges remain, including ethnic imbalances, lack of children's data, limited live (video) data, and ethical

concerns about data sourcing and consent. Collecting large, diverse, and legally compliant datasets is costly and often commercially unviable, especially for underrepresented groups.

- **Impact on FAE tools:** Without high-quality and representative training and testing data, FAE can become biased, performing well for some groups but poorly for others. Despite progress, FAE tools still struggle with narrow age ranges, such as distinguishing 16- vs. 18-year-olds, and show lower accuracy for females and underrepresented groups.
- **Impact on users:** Bias in AE tools can exclude users without formal ID, especially in the UK. The real-world impact of AE is mixed: it is limited in some sectors (e.g. gambling) but promising in retail trials like “Challenge 25.” Access to better data could help, but some issues stem from biological limits. Still, accuracy gains may boost adoption, especially under new UK regulations.

The data requirement for automated AA solutions (including AE and age verification) can be divided into three types:

- Training data – used to improve the probabilistic accuracy of statistical/algorithmic tools
- Testing data - which should be similar to the training data but ideally independent from it, i.e. strictly segregated, to avoid risks of overfitting and data leakage. The testing data should be further segregated between that used for internal validation and that used for external validation.
- Enabling data – required to verify age directly (e.g. passport, credit check, school records) and enable age verification methods (based on hard identifiers or verified age information). A large proportion of interviewed stakeholders were interested in solutions to enhance access to this type of data, e.g. population-wide government datasets with verified age attributes. However, this type of solution was outside the scope of this study, which focused on challenges related to training and testing datasets.

Limitations in the quality and representativeness of training and testing datasets impact statistical age estimation (AE) tools, which can involve a wide range of methods drawing on the probabilistic analysis of biological or behavioural features of people that vary with age. Examples include facial and voice analysis, capacity testing (e.g. knowledge tests), movement analysis (e.g. keyboard dynamics), online content and behavioural pattern analysis, and others[footnote 2]. These techniques are still maturing, but facial age estimation (FAE) is the most widely adopted AE

technique in automated, digital, and consumer-facing contexts[footnote 3]. All interviewed stakeholders either mentioned FAE as one of the AA methods that they were aware of, reported selling or testing FAE solutions as a business, or mentioned having adopted FAE or considered adopting it:

- Of the 7 interviewed business users, 2 use FAE, 4 are looking into adopting FAE and 1 is familiar with FAE but found it to be insufficiently accurate (in particular against edge cases of young-looking adults) and has opted for ID-based age verification instead.
- Of the 6 interviewed entities operating in the AA industry, 4 are AA solution vendors, 1 is a conformity assessment body of ID and AA solutions (including FAE solutions), and 1 is a trade body representing AA solution providers (including FAE solution vendors). The 4 interviewed AA solution vendors all provide both FAE and ID-based age verification.
- Of the 6 interviewed public sector organisations and 11 interviewed third sector entities (incl. charities, NGOs and academics), all listed FAE as one of the AA methods that they were aware of. Some public sector stakeholders and academics mentioned emerging, more niche AE techniques (e.g. movement analysis of tendon movements using AI and wearable) but had limited expertise and/or experience with the application of these. Stakeholders also mentioned content and behavioural pattern analysis as a technique used more widely by large online platforms with access to significant behavioural user data.

2.1. Data limitations

Lack of access to high-quality and representative training and testing datasets has been reported as a key challenge and area of concern in the development of AE solutions (ACCS, 2023[footnote 4]; European Commission (EC), 2024[footnote 5]; Sas & Muelberg, 2024[footnote 6]; Unicef, 2021[footnote 7]). The level of accuracy achievable by statistical techniques depends on the quality of the datasets used to train them (EC, 2024). In a workshop held by the Age Check Certification Scheme (ACCS) in 2023, participants, including representatives from across the UK age assurance industry, statistical and measurement analysts, biometric scientists, and conformity assessment specialists highlighted the lack of access to appropriate datasets as a key issue for the AA sector (ACCS, 2023). Stakeholders interviewed as part of this study similarly noted that a lack of access to representative training and testing datasets impacts the accuracy of AE tools in general. The impact of data limitations on the performance of AE tools is discussed in section 2.

Limitations with facial image datasets available to train and test FAE tools
Interviewed FAE vendors and technical experts noted that there have been significant improvements in the volume and quality of facial image datasets available to develop FAE tools. These datasets include open-source facial image datasets, with 10,000+ to 100,000+ subject data points (e.g., MORPH, UTKFace, FACES, FG-Net, IMDB-Wiki, MEDS and Adience). However, the stakeholder consultations identified that there are still limitations with facial image datasets available to train and test FAE tools, including:

- **Representativeness of adult facial images:** There have been significant improvements in the quantity and diversity of adult facial image datasets, with datasets reaching millions of subject data points. These include government and institutional datasets (e.g. police mugshots), commercial and proprietary databases (e.g. user-submitted images), web-scraped images (e.g. celebrity databases like IMDB), and academic databases complied by universities or research labs. However, interviewed academics and technical experts report that the datasets remain unbalanced with the overrepresentation of certain demographics (e.g., Hispanic and Caucasian males) and underrepresentation of minorities, including people with disabilities impacting facial features. A survey of facial image datasets (including open-access ones) used in academic deep learning age estimation research published between 2015 and 2020 finds that most datasets are ethnically unbalanced, with Caucasian/White accounting for 80% of data subjects across all databases that include ethnic labels (Dahlan, 2021[footnote 8]).
- **Lack of facial images of children:** Large datasets for underage subjects that are ethically and legally sourced, with accurate labels are rare (Anda et al., 2020[footnote 9]). Interviewed stakeholders report that there is insufficient children's facial image data to train and test FAE tools, with existing datasets limited to tens or at most hundreds of subject data points.
- **Ethics and integrity concerns:** Stakeholders highlight ethical concerns with the use of facial image datasets sourced from jurisdictions without robust data protection. They note that some FAE solution vendors are not transparent about the data source used to develop their tools given the commercially sensitive nature of this information. This could impact the industry's reputation in the long term. They also note concerns around the accuracy and reliability of the age labels in publicly accessible datasets.
- **Live data:** The lack of access to live facial data (e.g. short-filmed clips) hinders the development of FAE technology that is effective against spoofing and other circumvention risks (ACCS, 2023). For example, a recent study finds that state-of-the-art FAE algorithms struggle to distinguish replayed images from genuine

images, making them vulnerable to replay attacks[footnote 10] (Korshunov et al., 2024[footnote 11]). Similarly, stakeholders reported limitations with live data resources available to test more advanced FAE spoofing techniques, including video injection attacks.

The lack of availability and access to appropriate datasets is driven by the significant practical challenges and financial resources involved in ethically and legally collecting large amounts of facial images, in particular of harder-to-reach groups such as children and demographic minorities (ACCS, 2023). The practical challenges include engaging with a large and diverse set of data subjects and obtaining their informed and comprehensive consent to collect and process facial images and other personal attributes (e.g. age, gender, skin tone, etc.) for the purpose of developing commercial FAE solutions. In the case of children under 13, consent would have to be obtained from the parent or legal guardian. One industry stakeholder noted that they investigated creating a sufficiently large and representative facial image dataset to test FAE tools in the past. However, the initiative was found to be too costly (over multiple millions of GBP) and commercially unviable in the medium term (3-5 years).

2.2. Impact on performance of AE tools

The accuracy of AE tools relies on access to high-quality, representative training data that captures real-world diversity in age, gender, ethnicity, physical features, and contextual factors. Without this, models may perform well for well-represented groups but poorly for others, resulting in bias, defined as the difference in average error between groups (ACCS, 2023). Bias can arise at various stages, particularly when models are trained on unbalanced datasets (Panic & Marjanovic, 2024[footnote 12]; Zheyan et al., 2021), or when data includes inaccurate age labels, which limits learning and reduces accuracy. Testing data must also be diverse and high-quality to assess real-world performance. Without it, results may overlook key limitations. Well-structured testing helps developers identify where tools struggle, enabling improvements in fairness and reliability.

Impact on the performance of FAE tools
The performance of FAE tools has greatly improved over recent years. The National Institute of Standards and Technology (NIST) Face Analysis Technology Evaluation (FATE) is the largest benchmarking exercise of FAE algorithms, using about 11.5 million photos of over 8.5 million individual data subjects from four databases. Its latest evaluation[footnote 13] findsthat five of six commercially developed algorithms

tested in 2024 outperformed the most accurate algorithm from the previous benchmarking evaluation in 2014. Specifically, using a common database of high-quality facial images sourced from visa applications, the best mean absolute error[footnote 14] (MAE) has reduced from 4.3 to 3.1 years. Despite these advances, FAE tools still struggle to accurately distinguish between children within narrow age bands (ACCS, 2023), as well as to differentiate young adults (18-22) from minors (under 18). For instance, FAE vendor Yoti's algorithm achieved a MAE of around 2.0 years for ages 6–17 in high-quality visa application photos – the best accuracy in the under-18 range across all FAE algorithms tested by NIST (2024). Yoti's internal evaluation (2024[footnote 15]) reports achieving about 1.2 years MAE for ages 6–12 and similarly 1.3 for ages 13–17 (two relevant age ranges to ensure that under 13s and under 18s cannot have access to age restricted goods and services). These results are encouraging, but even with 1–2 years average error, this means that there are cases where a 12-year-old could be mistaken for 14, or a 16-year-old for 18. This is problematic if the goal is to accurately assess whether a person is under the legal age threshold or not (such under 13 or 18) using FAE. While FAE tools still struggle to estimate ages within narrow bands, they have proven effective at performing the "Challenge 25" check, i.e. assessing whether someone appears to be under or over 25 (RASG, 2023[footnote 16]). The "Challenge 25" practice has been widely used in the retail and hospitality sector (performed by humans) and allows to request ID only from those who look younger than 25 to confirm they are over 18. FAE tools continue to show accuracy gaps also across demographic groups, including gender, ethnicity, and skin tone. NIST (2024) found all tested algorithms performed worse on females than males, with no clear explanation, though differences in training data representation and aging patterns may contribute. Algorithms also showed varied performance by region of birth (used as a proxy for ethnicity[footnote 17]), likely due to unbalanced datasets (Panic & Marjanovic, 2024). One FAE vendor noted poor performance for individuals with facial dysmorphisms, disabilities, or medical conditions (e.g., Down syndrome), largely due to their underrepresentation in training data. While some stakeholders believe that larger, more diverse datasets, especially including children, could improve fairness, others caution that the gains may be marginal compared to the significant costs and complexity of data collection. Moreover, some issues with accuracy and bias in AE tools may not be fixed with larger and more diverse datasets alone. For example, one interviewed conformity assessment body evaluating FAE tools said these tools perform less well in estimating the age of people with darker skin in low light. An academic added that there are still large unknowns around how skin tone and lighting interact and whether more training data alone can solve this. For example, poor lighting can make darker skin tones appear even darker, making it harder for the system to detect key facial features.

2.3. Impact on use cases and end-users

The inaccuracy and bias affecting AE tools have raised significant concerns as they carry the risk of fostering discrimination, hindering the access of underrepresented groups to age-appropriate and age-restricted services online, while insufficiently reducing the risk of online harms to children (Sas & Muelberg, 2024; 5Rights Foundation, 2021; EDRi, 2023[footnote 18]; Unicef, 2021). Moreover, the UK does not have a national ID system, and passport ownership is far from universal[footnote 19]. As a result, a significant percentage of UK children (and adults) could be prevented from engaging with use cases where age verification through hard identifiers (e.g. a passport) is required due to the insufficient certainty of age achievable by AE tools, or where age verification is required as a secondary method to back up and affirm AE processes. One way to help reduce this challenge is through the use of CitizenCards[footnote 20]. These Home Office endorsed PASS-accredited photo ID cards could offer a reliable and affordable means of proving age where more traditional forms of ID may not be available, in particular for younger individuals. Still, it remains important for the market to foster innovation and continue to develop and improve the effectiveness of a wide range of AA techniques, including AE techniques.

Impact of limitations with FAE accuracy and fairness attributable to data gaps on real-world use cases and end-users

The existing research does not provide significant evidence of the exact impact of FAE accuracy gaps and bias attributable to data access issues on real-world harms accruing to children accessing inappropriate content, nor to the detriment accruing to age-appropriate users being denied access to age-prohibited content. Assessing this impact involves understanding the practical implications of data issues and corresponding limitations of FAE tools on the use cases that they can effectively serve. Interviewed business users from various sectors, each aiming to address different use cases, reported a range of practical limitations when implementing age assurance (AA) using facial age estimation (FAE). Most of these issues cannot be addressed by improving FAE performance with better training and testing datasets alone:

** - Online gambling:** Interviewed stakeholders involved in the online gambling industry explained that the industry is legally required to rely on age verification using hard identifiers for AA. They note that improving the accuracy of FAE (and other AE tools) would likely have limited practical implications in this sector.

- User-to-user: A private image sharing platform provider finds that FAE tools will

inherently exhibit inaccuracies in the 17- to 25-year-old age range and has opted for ID-based age verification instead. Access to better training and testing data, beyond what is currently available, could help reduce these inaccuracies. However, as was noted by this respondent and interviewed academics, even with perfect training and testing datasets, inherent accuracy limitations with facial age estimation (FAE) are likely to persist. This is due to several fundamental reasons including biological variability in aging, visual data limitations, and the subjective nature of age perception.

- **Online gaming:** One interviewed stakeholder in the online gaming industry reports that while industry leaders have started running pilots using facial AE tools, the wider industry has so far been resisted adopting AA solutions, as they are costly and introduce user friction.
- **Adult content:** One interviewed stakeholder in the adult content industry notes that a fragmented global regulatory landscape for AA requirements, enforcement issues and user resistance continue to be key limitations in wider industry adoption of AA, including FAE.
- **Retail:** Stakeholders in the retail industry noted positive outcomes from a Home Office sandbox trialling the use of facial AE tools to undertake the "Challenge 25" check in the sale of alcohol and other age-prohibited items (RASG, 2023[footnote 21]). While this is not an online age assurance use case, stakeholders report that the key obstacle to the wider adoption of this use case rests on government/legal approval. It is difficult to assess the extent to which accuracy gains and reduced bias in FAE would result in real-world benefits through the provision of better AA by online service providers and an increase in the number of use cases adopting FAE. The enforcement of the Online Safety Act and regulatory guidance from Ofcom and the ICO will lead to an increasing number and variety of online service providers assessing their need to provide AA and coming to decide to implement AA systems. FAE is listed by Ofcom as one of the methods having the potential to provide highly effective AA[footnote 22], alongside photo-ID matching and reusable digital identity services among others. There remains a strong possibility that FAE will continue to gain widespread adoption[footnote 23], with further improvements in accuracy and fairness having a material positive impact on end-users. Additionally, there is a feedback loop at play: as FAE performance improves, it is likely to drive increased adoption.

3. Facial image database solution

The study explored the key features and requirements of a viable facial image database solution. This section starts by discussing the rationale for exploring a facial image database solution specifically, followed by the key steps and questions required to be addressed to robustly assess the viability of a facial image database solution. It then presents the corresponding viability assessment and exploration of key features of a potential facial image database solution that were undertaken in this study, based on the evidence and insights collected through the literature review and stakeholder consultations. Lastly, it discusses key challenges and limitations with the explored database solution.

3.1. Rationale

The study initially set out to explore a solution providing statistical AE tools with access to better training and testing data. The scope of the solution was further narrowed down to addressing data issues impacting FAE tools for the following reasons:

- **Proportionality and viability**: A solution targeting data gaps impacting FAE tools was identified as most likely to optimise the real-world benefits achievable from improving the performance of FAE tools by enabling them to access better training and testing datasets. FAE is by far the most established statistical AE method, with the highest market adoption so far and a strong possibility that it will continue to increase in importance. This means that it is more likely for:

    - Improved FAE performance to lead to improved AA provision through online service providers and use cases adopting FAE as part of their AA provision, which in turn leads to;
    - Reduced incidence of online harms accruing to children accessing age-restricted content and services and reduced detriment accruing age-appropriate users being unable to access age-restricted content and services.

- **Stronger evidence base on data access issues:** Data gaps affecting the performance of FAE were more extensively discussed in the reviewed literature and stakeholder interviews, providing a stronger evidence base for identifying a relevant and viable database solution. In particular, study findings suggest that there is scope for improved accuracy and fairness of FAE tools vis-à-vis demographic groups underrepresented in existing datasets, in particular children, for which available datasets are limited in scale and representativeness and the

data is very hard to obtain.

- **Scope for government intervention:** There are significant practical and financial barriers to ethically and legally collecting large, diverse facial image datasets, especially those including children (ACCS, 2023). The FAE market, primarily composed of startups and small companies, lacks the resources to build and maintain such comprehensive databases. One industry stakeholder noted that from a commercial perspective, the costs of creating a high-quality, representative database are unlikely to be recouped through licensing, making it an unviable investment for individual businesses. This data gap not only creates entry barriers but also risks stifling competition and innovation in the FAE market. A government-led initiative with public funding could help address this potential market failure by prioritising broader social benefits such as improved child online safety and reduced discrimination, over purely commercial returns. This could make such an initiative more viable from a cost-benefit perspective than one driven by the private sector.

3.2. Key steps to assess the viability and proportionality of a facial image database solution

Relevance, proportionality and value for money are key considerations for the assessment of an appropriate facial image database solution. The appropriate solution should both minimise costs of development, operation and maintenance and maximise benefits in terms of improving the performance of FAE tools and AA provision across online product and service offerings, reduce online harms to children and increase benefits to adults and children from being able to seamlessly and safely participate in the digital space. The key steps and questions to assess the viability and proportionality of a database solution and come to a decision on whether or not to proceed with development and implementation of the solution are captured in the diagram below. This study has gathered evidence on some of the steps and questions, but open questions remain.

Figure 1: Decision tree to support the viability and proportionality of explored facial database solutions (text version)

- What data gaps would the solution help fill?
- How would the database be used (e.g. external validation, training,…)?
- Is there no duplication of efforts with existing interventions and initiatives (e.g. ensure that there is no overlap with NIST’s facial AE validation and benchmarking work, and existing datasets)

- Can we expect buy-in from the relevant population of tool developers?
- What would be the expected improvement in tool performance?
- What is the expected improvement in age assurance provision across online services and use cases?
- What would be the expected benefits in terms of reduced incidence of online harms for children and increased utility for consumers, both adults and children from an improved online experience?
- Can the solution be used for other purposes and by other population of users? If so, what is the corresponding expected benefit stream?

Additional stakeholder engagement and collaborative exploration (beyond the scope of this study) will be needed to complete the viability assessment. The types of stakeholders to further engage with to address key assessment questions are suggested in the table below.

Table 1 Stakeholder type and relevant assessment questions

Stakeholder types to engage/collaborate with	Assessment questions
FAE tool developers	1 to 12, 14 to 15 and 19 to 21
FAE testing organisations, technical experts and academics	1 to 15 and 19 to 21
Database infrastructure technical experts	13 to 14, 18 and 21
Businesses and public sector organisations using/looking to adopt FAE solutions	1 to 10, 14 to 15
Academics in fields using large amount of facial image databases (e.g. medicine, healthcare, psychology, criminal justice, computer science & AI, cybersecurity)	10 to 12
Charities and NFP organisations	17
Data ethics experts and academics	17
Data protection legal experts	16

Regulators and public bodies (ICO, Ofcom, Children's Commissioner)	10, 13 to 14, 16 to 17

3.3 Exploration of the benefit delivery mechanism

A key element of the viability and proportionality assessment of the database solution involves understanding and evidencing the benefits achievable from improving the performance of FAE tools beyond what they deliver now by providing them with access to a database solution that addresses existing data gaps. The literature review and stakeholder consultations suggest that providing FAE tools with access to more representative and high-quality datasets could help further improve their accuracy, in particular with AE of demographic groups that tend to be underrepresented in existing datasets, notably children.

A hypothesis of the impact pathway enabled by FAE tools having access to better datasets is captured as a high-level Theory of Change (ToC) in the diagram below. This ToC rests on some key assumptions, including the starting assumption that a high-quality and representative database relevant to developing FAE tools is successfully created, made accessible and known to FAE tool developers.

Figure 2: High-level Theory of Change of facial image database solution

This is a high-level ToC; some benefits are not captured. For instance, the adoption of FAE by an online service provider could replace or provide an alternative to an existing AA solution, e.g. using hard identifiers, already offered by the online service provider to end-users. If end-users find the FAE solution less costly or invasive, this

creates an additional stream of benefits from the solution that accrues to consumers. Moreover, if the database solution encourages innovation and competition in the market for FAE solutions, and by extension AA solutions, this could drive down prices and improve the overall AA service offering, benefitting consumers. Equally, the potential benefits of better FAE tools may be tempered by the emergence and broader adoption of the government-led Digital ID initiative[footnote 24] and other digital identity solutions, such as digital wallets. As these systems become more widely implemented, they may offer alternative and more effective means of age verification that reduce reliance on FAE.

As discussed in section 2.4, evidence gaps remain along the hypothesised benefit pathway, making it difficult to quantify the magnitude of benefits that can be achieved from enabling FAE tools to be trained and tested on representative and high-quality datasets.

These include:

1) The extent to which existing shortcomings with the accessible data are impacting the performance of FAE tools and, relatedly, how much FAE accuracy could be improved from having access to appropriate datasets at the development stage.

2) The extent to which FAE limitations (attributable to gaps with the existing datasets) are hindering access or making it more difficult for age-appropriate users to online products and services as well as leading to an increased incidence of online harms to children.

3.4 Exploration of key features of the solution

Phased approach to developing the database solution
Given the evidence gaps and corresponding uncertainty with the size of the benefits achievable from enabling FAE tools to access representative and high-quality datasets, a phased approach is most appropriate if looking to further explore a facial image database solution, a phased approach is most appropriate. This would involve starting with a relatively small database that minimises development costs. The database could be further expanded, e.g. if there is emerging evidence that the intervention is leading to positive results.

Explored options
The set of database options explored as part of this study as well as relative advantages and disadvantages are captured in the table below.

Table 2 Explored facial image database solution options

	1: External validation (EV) database focused on key demographics currently underrepresented in the data (e.g., children)	2: Representative EV database	3: Representative EV and internal validation (IV) database combined with training (T) database focused on underrepresented demographics
Pros	Lower data collection costs, less complex consent management Lower risk of data misuse and corresponding costs of technical and governance setup and safeguard requirements		Higher potential for vendors to improve FAE accuracy vis-à-vis key demographics currently underrepresented in training data
Pros	Minimised data collection costs Better understanding of FAE performance vis-à-vis key demographics currently underrepresented in the data	Better understanding of FAE performance vis-à-vis representative sample	
Cons	Potential limitations in comparability of validation results	Remaining uncertainty around NIST validation results benchmarking datasets needs resolving to avoid duplication	High data collection costs More complex consent

	of efforts; potential gaps with NIST data, mobile phone camera images, better representation of children in the dataset	management Higher risk of data misuse and corresponding costs of technical and governance set-up and safeguard requirements

Option 1: The study initially explored developing option 1 in the table above, i.e. a database solution for external validation of FAE tools only. External validation is the process of evaluating the performance of a FAE tool using data that was not used in the original training and development of the model. An initial outline of a potential solution that was developed and discussed with key stakeholders is shown below.

Figure 3 - Outline of potential external validation facial imagery database solution

This external validation database solution would test FAE tools against a dataset that is representative of key demographic groups currently underrepresented in existing data, particularly children. The dataset would need to include accurately labelled images with details such as age, gender, and ethnicity, along with variations in lighting, resolution, and potentially live images. Creating such a dataset would likely require primary data collection, as this study did not identify existing datasets or technical solution (including synthetic data[footnote 25]) that could fill these gaps. A more thorough assessment involving broader engagement with FAE developers, testing bodies, technical experts, and researchers (including those in medical or biometric fields) is recommended to ensure no duplication with existing resources.

This solution was initially considered the most cost-effective and lowest-risk option and has the potential to provide a robust understanding of FAE limitations for underrepresented groups. It could also serve as a first step in a phased approach toward developing a larger dataset to support training and internal validation by FAE tool developers.

Option 2: During the workshop with FAE vendors and a technical expert from a conformity assessment body, participants emphasised that an external validation database representative of the wider population, rather than only underrepresented groups, would be useful. Maintaining the same data quality and image variation standards as in option 1, this broader approach would improve the comparability and interpretability of validation results across different tools. This is reflected in option 2 in the table above. Option 2 would still provide the same following relative advantages (as option 1) over a larger database solution available for training and internal validation:

- **Lower data collection costs:** Insights from individual interviews and the workshop with FAE industry suggest that the number of individual data subjects needed for external validation of FAE tools is roughly one-tenth of what would be required to also support training and internal validation. However, discussions on the exact data size requirements, specifically the number of unique facial images, were inconclusive. One stakeholder suggested that a minimum of 15,000 data subjects is required to create a representative dataset for external validation, with an indicative average data collection cost of £50 per subject, bringing the total to approximately £750,000. One stakeholder at the industry workshop suggested a £50 per-unit data collection cost, based on their own experience, and that of others, in gathering facial images. While no other industry stakeholder challenged or offered specific input regarding this estimate during the workshop, one indicated that the figure seemed sensible in a separate discussion. Expanding the database to also support training and internal validation would increase data requirements and costs tenfold[footnote 26]. While there are fixed costs associated with setting up and maintaining the database, these are relatively small compared to data collection costs, meaning that scaling the solution would offer limited potential for cost savings.
- **Less complex consent management framework:** Using the database exclusively for external validation would likely allow access to the database to be limited to a few entities involved in managing the database and undertaking the external validation. Consent from participating individuals (giving their data) can be obtained upfront and cover pre-set entities involved in the solution. If the database were to be accessed for training and/or internal validation, various FAE tool developers would require access to the database. There might be a need to go

back each time a new entity requires access to the database to obtain explicit consent from participating individuals. Any consent given would have to be specific to the purpose: i.e. if the database expanded on its stated purpose, then additional consents would be required. This could be mitigated by setting a strict application process and criteria through which tool developers can qualify for access to the database and covering access by all qualified tool developers through upfront consent.

- **Reduced risks of data misuse:** Enabling tool developers to train and test their tools using the information held in the database increases risks of data misuse and thus could require a more involved technical and governance set-up and safeguards to ensure the data is kept safe and used appropriately.

The proposed external validation solution may risk some duplication with existing benchmarking initiatives, particularly those led by NIST. NIST conducts the largest benchmarking evaluation of FAE tools, utilising approximately 11.5 million photos from over 8.5 million individuals, sourced from four databases[footnote 27]. These include data from US government and law enforcement records as well as visa applications through the US embassy in Mexico. However, as NIST does not publicly disclose the full details of these datasets, there is uncertainty regarding their representativeness of the UK population, especially UK children. This may limit the ability of NIST's evaluations to provide a detailed understanding of FAE performance within the UK context. The proposed solution could still serve as a valuable complement to NIST's benchmarking efforts by offering insights more tailored to the UK demographic.

Option 3:FAE vendors participating in the workshop noted that expanding the database to support training and internal validation would be beneficial for improving the accuracy of FAE tools. However, the high data collection costs required to build a representative and sufficiently large facial image database for these purposes are a key barrier to feasibility. Additionally, participants did not offer specific or quantified estimates of the accuracy gains that might result from access to such a dataset.

Governance and responsible entities
This study explored the potential governance structure and entities that could be involved in creating and operating the database solution:

- **Database management:** The option suggested by a larger number of stakeholders is to have a trusted organisation (e.g. NGO) managing the database with oversight from government (e.g. specific task unit commissioning the work).
- **Data collection:** could be subcontracted to a specialised firm under strict contractual obligations.

- **External validation:** could be managed by an independent testing organisation. This organisation could also be responsible for managing the database (e.g., instead of an NGO). There was no consensus on the exact entity that should be responsible for the external validation (and manging the database). Alternatively, the solution could explore undertaking the external validation using a 'black box' whereby tool developers submit their models and are fed back the validation results. Both seem to be feasible and ensure the confidentiality and security of the data.

A key next step is to clearly identify and define the roles of all parties involved in the solution from a data protection perspective, such as determining whether they act as data controllers (joint or independent) or data processors. While the study considered this aspect, it was challenging to draw definitive conclusions due to the early stage of the solution's development and the current uncertainties around its intended use and data requirements.

Legal assessment

The study also explored the legal feasibility and compliance of a facial image database solution. The processing of facial images is governed by the UK GDPR. A full Data Protection Impact Assessment (DPIA) cannot yet be completed due to the early stage of the project and key uncertainties, particularly around its purpose and governance. However, an initial legal review of the external validation-focused solution[footnote 28] identified several key features that should be taken forward, including proactive data minimisation practices (e.g. excluding names and ID documents), appropriate privacy safeguards (e.g. controlled access via testing bodies or APIs), and the implementation of a clearly limited purpose to support informed consent. The legal basis for collecting and processing this data would be consent, with parental or legal guardian consent required for data subjects under the age of 13Still, unresolved legal questions remain, including the need for clear consent withdrawal processes, verified parental consent, defined roles and responsibilities, robust security measures, and appropriate retention policies. If the initiative is progressed further, these legal questions need to be addressed, after which a DPIA should be completed. The legal considerations and challenges involved with the database solution are discussed further in section 3.5.2.

Roadmap to delivery

Once the scope, objective and key features of the database solution are granularly defined and a robust cost-benefit assessment indicates sufficient proportionality, DSIT can consider commissioning the development of the database solution. The diagram below captures the key stages to solution delivery and operationalisation.

Figure 4 - High-level roadmap to development and operationalisation of the

solution

Particular focus in delivery should be given to the **data collection and consent management framework**.

This stage involves establishing the data collection protocol, consent framework, engagement strategy, and a finalised DPIA. It will be essential to involve charities and associations representing the targeted demographics to ensure inclusivity and ethical considerations.

Consent management framework
A structured consent management framework should be developed, considering the following key factors:

- Due to the sensitive nature of the data being collected, consent must be obtained from each data subject, with a narrowly defined purpose for data collection.
- Parental or legal guardian consent is required for data subjects under the age of 13.
- There must be provisions for revisiting consent, allowing individuals who reach the age of consent to withdraw their data if they choose.
- Data subjects should provide informed consent, which requires a clear and accessible briefing on how their data will be processed. This should be designed to be inclusive, e.g. for individuals with disabilities.
- Consent for collecting special category data could be gathered through self-declaration, using “yes/no” tick boxes for ethnicity and whether the individual identifies as having a disability that affects their facial features. This information

will likely be based on self-reporting rather than verification by the data collector. There would be no practical or non-intrusive way for the data controller to independently verify this information. It is entirely reasonable to rely on the data subject's self-declared responses, provided there is confidence that they understand what is being asked. This approach is widely used in contexts such as diversity questionnaires.

- Age could be self-reported, but this carries the risk of misrepresentation. Alternatively, age verification could be carried out by requiring the submission of a hard identifier document (e.g., passport or birth certificate). This approach would limit participation to children who possess such documents, so it is important to assess whether this would affect the representativeness of the sample in relation to the broader UK child population. Once the age is verified, any hard identifier evidence must be securely discarded to protect privacy. Still, this could risk capturing excessive information.

Data collection strategy for hard-to-reach demographic groups
There are no inherent data protection concerns in collecting a greater number of images from underrepresented groups to ensure balanced dataset representation, and there is a clear need to collect information from underrepresented groups to address bias in FAE systems. However, it should be acknowledged that any attempt to collect information from underrepresented groups could be open to accusations of profiling. This can be addressed by ensuring that participation is entirely voluntary and that individuals are not treated differently based on their characteristics. This would likely be acceptable given that the purpose of collecting this information is to help reduce discrimination and improve fairness in facial age estimation.

Collecting data from "harder-to-reach" communities, such as children, will be a key challenge in the delivery of this solution and require a strategic, ethical, and culturally/socially sensitive approach. Key considerations involved in the data collection strategy are discussed in section 4.3.

3.5 Costs, challenges and limitations of the database solution

3.5.1 Proportionality of costs vs. benefits

A key challenge with the explored database solution options is their limited commercial viability. Data collection costs are expected to be substantial and are unlikely to be recovered through fees charged to FAE vendors for dataset access or external validation services. One interviewed conformity assessment body

evaluating FAE tools explained that they had also investigated developing a sufficiently large and representative facial image database for external validation of FAE tools. However, they did not proceed due to lack of commercially viability in the medium term (3-5 years) given the sizeable costs involved (> £1 million) and limitations in how much of the costs could be reasonably recouped by charging FAE tool developers for the validation and certification service. There may be an opportunity to leverage existing facial image datasets held by His Majesty's Passport Office or the Driver and Vehicle Licensing Agency to reduce the data collection costs associated with developing FAE tools. However, these images may lack the diversity in lighting, angles, and expressions typically required to train robust models. A further challenge is that these images were not originally collected for this purpose, raising questions around consent, data protection, and the suitability of repurposing them for AI model training. While such limitations could potentially be mitigated through synthetic augmentation techniques, the challenge remains in securing appropriate consent from data subjects for their images to be used in this context.

If public funding was available, from a societal perspective, value for money and proportionality of costs vs. benefits of a database solution could still be achieved if the wider socio-economic benefits associated with improving the accuracy of FAE tools outweigh the costs. Key benefits include the reduced incidence of online harms accruing to children from accessing age-restricted content and services as well as improved access of age-appropriate users to age-restricted content and services. A robust cost-benefit analysis is required to determine the proportionality of the explored database solutions. However, this is challenging to undertake because of limited evidence on the magnitude of achievable benefits and lack of consensus on the costs. Key areas for further research are:

3) Examining how potential improvements in the scale and representativeness of training datasets relate to marginal accuracy gains in facial age estimation (FAE) tools. The relationship is expected to show diminishing returns. This study reviewed existing literature and engaged with FAE vendors to explore this relationship in more detail but was unable to uncover substantial insights. A deeper understanding would require access to results from controlled experiments assessing how incremental improvements in data quality and quantity translate to performance gains.

4) Identifying the specific use cases, sectors, and demographic groups where limitations in FAE accuracy have real-world implications and quantifying the scale of those effects to better understand the potential benefits of FAE improvement. For example, FAE performance would impact AA provision in sectors such as online gaming, user-to-user services, social media platforms, and adult content that could

see broader adoption of FAE as a means to meet online safety regulatory requirements. Conversely, improvements in FAE accuracy are unlikely to influence the online gambling sector, where ID-based age verification remains a legal requirement.

3.5.2 Legal considerations and challenges

The explored database solution is at a very early stage, leaving many unknowns, particularly in relation to the precise purpose of the database solution and the roles of those parties who engage with it. Given these unknowns, a comprehensive and detailed legal assessment of the database solution cannot be undertaken. The question of the focus and purpose of the solution needs to be addressed as a priority to progress the development of the solution and ensure that corresponding legal questions are addressed.

The legal assessment of the external validation database solution that was developed and discussed with key stakeholders[footnote 29] has indicated the following positive suggestions:

- A consent-based approach is most suitable.
- Effort to be taken to ensure data minimisation, particularly in respect of removing names from the database and the exclusion of any ID documents
- Privacy measures to restrict access of those validating against the database, which would help to achieve minimisation as well as security and fairness (including an independent testing entity accessing database for external validation or through API black box[footnote 30].
- Collection of ID documents for the sole purpose of verifying the age of the relevant data subjects is a good and likely necessary measure.
- Adopting a limited purpose (even if this is not yet fully known) should assist in ensuring that any consent is ‘informed’ and make the overall process easier to understand for those data subjects wishing to take part.

In addition to the areas of uncertainty where further consideration is needed, and potential areas of challenge from a legal perspective, include:

- There will need to be an easy and accessible way for data subjects to withdraw their consent and to exercise their data subject rights.
- As facial images of children under 13 will be included in the database solution, significant thought will need to be given to verified parental consent and how this will be achieved.
- The purpose of the database solution will need to be clarified. At present the

explored solution seems to have covered a number of areas in which it could operate, but ultimately it will need to identify the specific area of facial age estimation that it seeks to address and how this will operate against / alongside other services.

- The proposed solution will need to consider further whether special categories of personal data are necessary to achieve its purpose.
- If special categories of personal data are concluded to be necessary (and therefore collected), the proposed solution will need to consider how it can ensure that this information is accurate.
- The ownership, operation and governance of the database will need to be finalised in order to understand the roles (and the responsibilities) of the relevant parties. Thought will also need to be given to those developers looking to validate against the database and their role.
- It will be vital that the database (as well as any other systems used in relation to data collection and storage) has robust technical and organisational measures to ensure that all personal data is kept secure.
- The proposed solution will need to consider how long it will retain personal data and the criteria on which it will determine that personal data no longer needs to be retained.

Once the proposed solution has been finalised and the challenges above have been addressed, a DPIA should be undertaken. Additionally, thought should be given to whether the finalised proposal should seek to participate in the ICO's regulatory sandbox to obtain further support and assistance from the ICO.

A more extensive legal assessment of the external validation facial imagery database solution that was developed and discussed with key stakeholders is provided in Annex 1.

3.5.3 Practical challenges with data subject engagement and data collection

A key challenge with developing the solution resides in identifying and engaging with a large number (possibly thousands to tens of thousands) of data subjects, including from harder-to-reach demographic groups such as children and ethnic minorities, to collect their facial images. It will be key to partner with relevant organisations, e.g. charities, advocacy groups, health clinics and schools serving and/or representing

these demographic groups. This might lead to engaging directly with the targeted communities, undertaking public awareness campaigns around the initiative, explaining the scope of social benefits and building trust. Consulting with representative organisations and targeted communities (if possible), will be important in helping assess whether providing incentives for participation would improve engagement. Incentives could involve cash, vouchers, mobile credit, transport reimbursement, free health check-ups or other useful services. Financial incentives offered to individuals for contributing their facial images vary across prior study initiatives. A US study focusing on public perceptions of sharing images for AI in dermatology found that participants required, on average, USD 18.25 to share facial images, which was the highest among different body parts considered (Ly et al, 2023[footnote 31]. In a UK study examining the effect of facial ageing on forensic facial image comparison, 106 adult participants were recruited and compensated £5 each for donating their facial images (Sexton et al, 2023[footnote 32]).

To maximise participation while minimising costs, the data collection process should be designed for seamless, remote, and individual participation by data subjects whenever possible. For example, this could involve enabling mobile data collection through offline-capable apps as well as deploying pop-up booths in schools, clinics, local event, etc. for voluntary participations. If needed, assistance should be available for specific groups, such as young children and individuals with disabilities. This is likely to introduce variation in image resolution, lighting, and angles, depending on how each participant takes their photo. To address this, the app could include clear instructions and real-time feedback to guide users in capturing images from different perspectives and under varying lighting conditions.

A key challenge in the data collection process is verifying the parental or legal guardian relationship of an adult providing consent on behalf of a child. This process could leverage assistance from passport offices or other government agencies that have already verified such relationships.

The data collection should involve a pilot stage before scaling, with collection of feedback from the relevant communities to adjust the data collection methods accordingly.

4. Alternative solution: funding innovation in AA

The study also considered alternative approaches to supporting the development of more effective AA solutions beyond improving the performance of FAE tools with a database solution. However, the literature review and stakeholder interviews yielded limited suggestions for viable alternatives. Most stakeholders pointed to broader interventions outside the study's scope, such as enhancing the standardisation of AA tools and providing more detailed government or regulatory guidance on which methods are appropriate for different use cases and sectors. A few stakeholders proposed that government support could instead focus on stimulating innovation in the AA market through grant funding or an open innovation challenge. In response, the study explored what the grant-funding intervention could address as well as the potential benefits, costs, and limitations of directing resources toward this type of alternative rather than a database solution.

4.1 What innovation gaps would the grant-funding intervention try to address?

A grant-funding intervention requires identifying and defining a specific innovation gap or problem and inviting innovators such as startups, researchers, developers, or the public to propose and develop creative solutions to the problem. It is a competitive process, so funding or support is awarded to the most promising applicant or applicants depending on the amount of funding available.

Age assurance is a broad and rapidly evolving field of research, with state-of-the-art academic work spanning a wide range of technical approaches, from advanced AI-based AE using biometric, behavioural, and digital activity data, to privacy-preserving digital identity systems such as zero-knowledge proofs and on-device age checks. The following examples, drawn from the literature review and stakeholder consultations, illustrate specific areas where further innovation could enhance the performance of AA tools:

- **Improving the accuracy of FAE, in particular with darker skin tones, under poor lighting conditions:** Interviews with technical experts reveal that there are still large unknowns around the complex interplay between skin tone and lighting. This performance gap might require technical innovation beyond training FAE tools on larger amounts of facial images with high variation in data subject skin tone and lighting conditions to be fully resolved. **-Improving behavioural proxies for age:** Collaborating with child development experts could lead to the identification of new behavioural proxies for age, for example, cognitive games or puzzles that adapt in difficulty and subtly to gauge a user's probable age based on

performance (i.e. similar to capacity or knowledge testing). This could lead to AA techniques that, if designed with care, would help determine maturity level in a child-rights-respecting way. There is potential for innovation in this area, as many existing profiling techniques, including keystroke dynamics, response speed, and emoji usage, have yet to be adapted or applied to capacity testing-based age estimation (5Rights Foundation, 2021[footnote 33]).

While the study did not find clear evidence of a current lack of funding across all areas of AA research, these highlighted areas represent promising opportunities where targeted government funding could accelerate progress.

4.2 What are the benefits in comparison to a database solution?

The main benefits of a grant-funding/open innovation intervention over a database solution would be:

- **Innovation across a broader range of solutions beyond FAE models:** The open innovation challenge could help explore innovative technical approaches beyond FAE (e.g. capacity- testing) and reduce the risk of technological lock-in with AA provision.
- **Easier and faster to pilot and scale:** Challenge winners can be supported through grants, pilots, or regulatory sandboxes, which are typically easier and faster to operationalise than a database solution.
- **Smaller capital commitments:** Given the sizeable data collection costs involved in the database solution; a grant-funding intervention is likely to be less costly.

4.3 What are the costs?

The exact amount of funding to allocate for an open innovation challenge will depend on a multitude of factors including the scope and technical complexity of the challenge, the resources likely required for participants to address the challenge, the outcomes expected from the innovation (e.g. concepts, prototypes, or pilots) as well as the number of awards. Benchmarking against similar initiatives, such as UK and EU tech challenges, is a good starting point. For example, the UK government awarded initial grants of £85,000 each to five projects through the first Safety Tech

Challenge Fund in 2021, totalling £425,000, to develop innovative technologies that combat online child abuse while preserving user privacy[footnote 34]. Total funding should also account for operational and support costs. Lastly, the level of investment should reflect the expected benefits in terms of improved AA provision that can be expected from successful innovations.

4.4 What are the limitations?

While open innovation challenges could contribute to promoting innovation in the field of AA, they have certain limitations:

- **Additionality:** There is currently no strong evidence that a lack of funding is the primary barrier to progress across all areas of AA research. Moreover, funding innovation in the AA sector might come at odds with industry concerns that the UK government's digital ID plans could limit the role of private providers and stifle innovation in the future.[footnote 35]
- **Complementarity with high-quality and representative datasets:** Depending on the challenge focus, solutions may enhance but not replace the need for high-quality, representative datasets, e.g., in cases such as improving FAE accuracy for individuals with darker skin tones in poor lighting conditions. Innovations generated through a challenge may still require validation and refinement using robust datasets.

A key next step would be for the government to define a clear research focus where targeted innovation funding could offer the greatest value. This could be supported by deeper engagement with the research community, e.g., by participating in academic and technical forums related to age assurance, biometric and behavioural age estimation, and online safety technologies.

5. Conclusion

This study examined the limitations in data access impacting the development and performance of AA tools, with a focus on FAE tools. It finds that while the accuracy and fairness of FAE tools have improved in recent years, critical performance gaps remain, particularly for children and other underrepresented demographic groups. These gaps are partly driven by the lack of access to high-quality, diverse training

and testing datasets. Addressing this issue is crucial for advancing online safety for children and ensuring inclusive, effective AA provision.

A central proposal explored in this study is the development of a facial image database solution to support the external validation, and potentially also the training and internal validation of FAE tools. The rationale for this stems from the significant commercial, ethical, and technical barriers that prevent the private sector from developing such datasets independently. A public database could help overcome these barriers and enable more accurate and less biased FAE.

However, the development of such a solution presents substantial challenges. These include high data collection costs and challenges (particularly for ethically collecting children's facial images), complex legal and data protection requirements, the need for a clear governance and consent framework, and significant uncertainties around the extent of real-world benefits. Indicative estimates on the FAE vendor market size and willingness-to-pay suggest that the database is unlikely to be commercially viable without public funding, and even with government support, proportionality of costs and benefits remains uncertain. Further work is needed to assess the magnitude of achievable improvements in FAE performance, and whether these translate into measurable gains in online safety and digital access. Key next steps involve:

- Coordinating with existing benchmarking efforts (e.g., NIST) to avoid duplication
- Identifying specific use cases, sectors and demographic groups for which gaps with FAE accuracy has real implications and quantifying those impacts
- Undertaking a larger-scale and multi-round consultation process with FAE industry, technical experts, relevant public sector and third sector organisations (incl. charities/NGOs representing children) to identify/refine and validate the following[footnote 36]:
  - Purpose and applications of the FAE database solution
  - Expected improvements in FAE accuracy and corresponding impact on specific use cases and demographic groups
  - Data requirements (including number of data subjects per relevant demographic sub-group)
  - Data collection and other database costs
  - WTP for the database solution
  - Governance structure
  - Any other potential uses for the database beyond FAE and corresponding

cost/benefit implications

- Undertaking a cost-benefit analysis of the solution with the intention to proceed only if the solution is found to be viable and proportionate
- Undertaking a comprehensive legal assessment and DPIA

In sum, the facial image database solution offers meaningful potential but requires careful design, robust justification, and phased development. In parallel, the study looked at an alternative, lower-cost intervention: an open innovation challenge focused on advancing AA tools. This approach could stimulate innovation across a broader range of technical methods (e.g., behavioural and privacy-preserving techniques) and is easier to implement and scale in the short term. However, they may not address all research gaps, as there is no evidence to suggest that funding is the primary barrier in every area and solutions often still depend on access to high-quality, representative datasets for validation and refinement.

Annexes

Annex 1 Legal assessment of external validation facial imagery database solution

1. Introduction

1.1. London Economics (**LE**) has been engaged by the Department for Science, Innovation and Technology (**DSIT**) to conduct a study into age assurance to better understand the challenges faced by age assurance tools and any possible solution to assist with these challenges.

1.2. As part of the engagement, LE has proposed a facial imagery database testing solution (the **Proposal**).

1.3. This memo relates only to data protection laws in the UK. References to Articles are to the UK GDPR (as defined in the Data Protection Act 2018).

1.4. The Proposal is still in its infancy with many elements yet to be finalised. Therefore, any recommendations highlighted in this memo are an initial view only and a more thorough data protection analysis should be conducted once more

information is known.

1. The Proposal

2.1 As a summary,[footnote 37] the Proposal will operate as follows:

2.1.1. an organisation / organisations (not yet known) will collect information relevant to facial age estimate from individuals, including children;
2.1.2. the information will be collected with the individual's consent;
2.1.3. this information will be held in a database (the Database);
2.1.4. third-party facial age estimation tool developers (Developers) will be able to validate the effectiveness of their tools in estimating a person's age against the Database;
2.1.5. the outcomes of this validation will be recorded and fed back to the Developers; and
2.1.6. the Proposal may also offer a certification following the testing outcomes.

2.2. The Database will contain the following information in respect of each individual:

2.2.1. age;
2.2.2. facial images;
2.2.3. gender;
2.2.4. ethnicity; and
2.2.5. heath data (i.e. disability information which may be pertinent to any facial imagery age estimation technology).

2.3. Identity documents will also be collected from these data subjects in order to verify the age of the data subject, but these documents and any additional personal data contained within these documents (such as name or place of birth) will not be included as part of the Database.

2.4. The Proposal does not intend that Developers will be granted access to the Database directly, but instead will test their solutions via a 'privacy-first' measure, for example: (i) an anonymised API blackbox; or (ii) via an independent third-party testing entity.

2.5. We understand that the 'validation' will operate as follows:

2.5.1. Developers who have an age assurance solution will engage with the Database;
2.5.2. testing of these solutions is undertaken by, or on behalf of, Developers

against the Database;
2.5.3. the Database (or more specifically, those operating the Database) will assess the accuracy and effectiveness of the age assurance solution; and
2.5.4. the Database (or those operating the Database) will feedback the results of the assessment to the Developers (i.e. 90% accuracy overall; 98% accurate on females aged 13-15; 82% accurate on males aged 13-15).
2.6. This memo will use the term **validation** to reflect the process outlined at paragraph 2.5.

3.0. Information not yet known

3.1. The Proposal is still very much in its nascent stage and there are many unknown elements as to how precisely it will operate. These unknowns include:

3.1.1. the ages of those that are intended to be included in the Database (i.e. is there a minimum age; a maximum age; children only; children and young persons);
3.1.2. whether the information for the Database will be collected directly from individuals or via a third party or both – and in each case, how this information will be collected;
3.1.3. the health data that will be collected save that it will be confined to health data that will impact the performance of facial age estimation;
3.1.4. the ethnicity data that will be collected;
3.1.5. the intended size of the database (both in terms of volume of personal data and volume of data subjects);
3.1.6. the number of facial images that will be collected;
3.1.7. the precise purpose of the Proposal, specifically whether the Proposal is intended to:

a. test and certify facial age estimation technology;
b. improve performance gaps in facial age estimation technology;
c. test against potential spoofing and circumvention of facial age estimation technology; d. operate as a standalone offering; or
e. supplement existing facial age estimate certification schemes and/or other testing solutions;

3.1.8. how long information will be retained in the Database;
3.1.9 who will operate the Database;
3.1.10. how many entities will be involved in the operation and governance of the Proposal and the roles of each;
3.1.11. the role of the Developer; and
3.1.12. the criteria required for a Developer to gain access to the Database.

3.2. This not an exhaustive list of the unknown elements of the Proposal nor is it a criticism of the Proposal. The unknown elements of the Proposal mean that a detailed assessment of whether or not it will (or could) comply with data protection legislation is not possible.

3.3. However, this memo will make certain assumptions in respect of the unknown elements of the Proposal in order to provide as helpful feedback to LE as possible and to help inform any future decisions about the Proposal.

3.4. One such assumption will be that the personal data is collected directly from data subjects.

4.0. UK GDPR

4.1. The information held about individuals described in paragraphs 2.2 and 2.3 will be considered personal data under the UK GDPR.

4.2. The Proposal will involve several different processing purposes, including collection, storage and (when applicable) deletion. The purposes will ultimately be determined by the not-yet-known controller. However, this memo will focus on the key processing purpose: the validation of Developers' solutions against the personal data within the Database (the **Validation Purpose**). This memo will assume that the Validation Purpose operates as a standalone offering to Developers for the purpose of allowing Developers to validate the effectiveness of their facial age estimation solutions. The controller (or controllers) will be the organisation[footnote 38] that determines the purposes and means of processing the personal data. Whether an organisation is a controller (or processor or joint-controller) is a question of fact which will be determined by the actual processing and decision-making that is made by the relevant organisations. The Information Commissioner's Office (**ICO**) sets out the following criteria for determining whether an organisation is a controller, by asking which organisation decides:

4.2.1. to collect personal data in the first place;
4.2.2. the lawful basis for doing so;
4.2.3. what types of personal data to collect;
4.2.4. the purpose or purposes the data are to be used for;
4.2.5. which individuals to collect data about;
4.2.6. whether to disclose the data, and if so, to whom;
4.2.7. what to tell individuals about the processing;
4.2.8. how to respond to requests made in line with individuals' rights; and
4.2.9. how long to retain the data or whether to make non-routine amendments to the data.[footnote 39]

4.3. Whilst the Proposal touches upon some of these elements, nothing has been concluded and it is not yet known if a single entity will make these decisions or if these decisions will be demarcated across numerous entities.

4.4. In order to provide the most meaningful feedback to LE, this memo will assume that a single organisation is the controller in respect of the management of the Database and has responsibility for the collection of personal data (the **Database Controller**). It may be that each Developer will act as a controller or joint-controller with the Database Controller in relation to the Validation Purpose, however this memo will also assume that the Database Controller has responsibility for compliance with the UK GDPR principles, lawful bases and data subject rights in respect of the Validation Purpose. Should there be more than one controller, as is likely going to be the case, then each controller will be responsible for comply with data protection legislation, including the principles set out in UK GDPR and demonstrating compliance with the same.
4.5. Based on these assumptions, this memo will now consider the application of the data protection principles to the Proposal.

Lawfulness, Fairness and Transparency Principle[footnote 40]

4.6. **Lawfulness:** The lawfulness element for the Validation Purpose will only be satisfied via appropriate lawful bases, as discussed in paragraph 11 below. The Proposal envisages that consent will be the lawful basis; and explicit consent will be the appropriate lawful basis for any special category personal data.

Transparency: In respect of transparency, the Database Controller must implement a privacy notice which sets out the information required under Article 13 of the UK GDPR.[footnote 41]

4.7. This privacy notice may be hosted online and accessible via a hyperlink on an appropriate website. However, it is also advisable that the Database Controller takes a 'layered' approach to transparency. Rather than including all information within a website privacy notice, the Database Controller should provide specific information to data subjects at the point of collection via consent forms. These consent forms should clearly set out the Validation Purpose and inform data subjects how they can withdraw their consent. They may also include information (or choice) about the retention of personal data.

4.8. Where personal data is collected directly from children, the Database Controller should also have in place a child-friendly privacy notice or ensure that its privacy

notice can be understood by data subjects of all applicable ages.

4.9. **Fairness:** Fairness under data protection is linked to the expectations of the data subject and any possible harms. The Database Controller will need to consider more than simply whether its processing for the Validation Purpose is lawful, but also whether it should be processing personal data for this purpose and how any potential harms can be mitigated.

4.10. Fairness is admittedly an ambiguous obligation. However, an example of unfair processing would be to allow Developers access to the names of data subjects where this is neither needed to notified to the data subject.

Purpose Limitation Principle[footnote 42]

4.11. A controller will only be permitted to use personal data for specified, explicit and legitimate purposes. A controller will not be able to use personal data in a manner incompatible with those purposes.

4.12. At present, there are too many unknowns about the Proposal for the Validation Purpose to be considered specific, explicit or legitimate.

4.13. However, the Proposal clearly intends for the Validation Purpose to be narrow. Once the unknown elements of the Validation Purpose have been resolved, the Database Controller should be in a position to ensure that personal data will only be processed for the specific Validation Purpose (and any other legitimate purposes). As discussed above at paragraph 5.3, to ensure that the Validation Purpose is explicit, it should be notified to data subjects clearly via the consent forms as well as restated with the Database Controller's longer-form privacy notice.

4.14. It is possible that the Database Controller (or a third party) may wish to further process personal data within the Database for purposes of scientific or statistical research – for example, in order to understand better the validation process and effectiveness of the validation. This is conjecture and beyond the scope of this memo, but if such further process is intended, the Database Controller would need to ensure proper safeguards are in place to limit the processing and ensure that the minimum amount of personal data is used for this secondary purpose.

Data Minimisation[footnote 43]

4.15. The data minimisation principles states that controllers must ensure that the personal data collected is adequate, relevant and necessary in relation to the purpose.

4.16. The Proposal clearly focusses on data minimisation. Reasons have been considered and given in respect of each type of personal data collected, and the suggestions regarding limiting the use of identity documents are commendable.

4.17. However, the Database Controller will be held to a 'high-bar' regarding the personal data it collects, particularly as the Proposal includes the collection of health data and ethnicity data which are special categories of personal data. The Database Controller will need to be able to satisfy itself that each personal data point is adequate, relevant and necessary.

4.18. The ICO has raised queries regarding the necessity for ethnicity data and the Database Controller will need to consider these comments alongside its own analysis to determine whether or not gender, health data and ethnicity data is genuinely required for the Validation Purpose – or, for example, if environmental and non-ethnicity features are more relevant metrics.

The Accuracy Principle[footnote 44]

4.19. Controllers must ensure that personal data is accurate and, where necessary, kept up to date. The Database Controller will need to ensure that the personal data included within the Database and used for the Validation Purpose is accurate. The Proposal, of course, has accuracy at its core as it seeks to aid the improvement of the accuracy of facial age estimation, so on the face of it this principle should be fairly straightforward. Furthermore, the Proposal envisages that identity documents are provided in order to verify the age of a data subject who submits their personal data to the Database.

4.20. However, improving accuracy of age assurance is distinct from ensuring that the personal data held in the Database is accurate. One key issue is that it is not understood how the Proposal will ensure the accuracy when it comes to ethnicity, gender and health data (if these are collected). If data subjects are simply offered the opportunity to provide their ethnicity, gender and any health data via a drop-down selection mechanic then there is clearly the possibility that individuals will, deliberately or accidentally, input inaccurate information. If the Database were to contain inaccurate personal data, the Validation Purpose could be frustrated entirely.

The Storage Limitation Principle[footnote 45]

4.21. Personal data must be kept for no longer than is necessary. This will be somewhat of a challenge for the Database Controller as the Validation Purpose is ongoing and it is not known for how long Developers may wish to undertake validation nor for how long the Validation Purpose will be necessary / useful to the

industry.

4.22. Furthermore, the ongoing development of photograph technology means that older facial images within the Database may no longer be as useful as newer facial images.

4.33. As the Proposal envisages a consent-based approach to processing personal data, it may be that the retention periods are put into the hands of the data subjects provided that any retention periods are assessed as necessary by the Database Controller (i.e. please select how long you would like us to retain your personal data for: one year; two years; three years etc.). The Database Controller may also consider contacting data subjects annually or at the end of the requested retention period to ask them if they would like their personal data to be retained or deleted.

4.24. Should the Validation Purpose no longer be necessary for any reason, then the personal data held within the Database must be anonymised or deleted.

The Integrity and Confidentiality Principle[footnote 46]

4.25. The Database Controller must ensure that it has appropriate security measures in place to protect the personal data within the Database (and during collection, deletion, anonymisation and validation).

4.26. The appropriate technical and organisation measures that must be put in place will depend entirely on factors that are not yet known and therefore are not considered within this memo. Moreover, advice should be sought from Information Security experts as well as legal support when considering the Integrity and Confidentiality Principle.

4.27. If the Database Controller is unable to implement appropriate technical and organisation measures to protect the personal data from unauthorised or unlawful processing and against accidental loss, destruction, damage or unavailability, then the Proposal will be frustrated.

4.28. When considering what are ‘appropriate’ measures to keep personal data secure, consideration must be given to the likelihood and severity for the data subjects. As the Proposal intends for the storage of special categories of personal data (and the collection of identity documents – even if briefly), there will be a significant risk to data subjects if their personal data is not kept secure and therefore the ‘appropriate’ measures must reflect this in their robustness.

Lawfulness

Each processing purpose, including the Validation Purpose, requires a lawful basis. [footnote 47] Where special categories of personal data are processed, an additional condition is required.[footnote 48]

1. As mentioned, the Proposal intends to operate on a consent-based model.

5.1. Under UK GDPR consent must be freely given, specific, informed and unambiguous.[footnote 49] This means that clear information must be provided to the data subject about how their personal data will be processed. Any consent given must relate to a single purpose and cannot be given in respect of multiple purposes. The data subject must have a free choice to either give or not give their consent. Furthermore, consent must be active meaning that any form of 'opt-out' mechanic will not be lawful consent.[footnote 50] The English courts have determined that there is a "relatively high bar to be met for valid consent."[footnote 51] Additionally, controllers must be able to demonstrate that the data subject has consented to the processing of their personal data[footnote 52] – meaning that the Database Controller must retain a record of the consent given by each data subject (including how and when it was given) and the precise wording against which the consent was given.

5.2. Whilst it is not known precisely how personal data will be collected from data subjects, the principle of a consent form whereby data subjects can voluntarily submit their images and other personal data to the Database Controller could, in theory, meet the requirements of consent under the UK GDPR provided that the elements set out in paragraph 11.3 are met.

5.3. If the Proposal does proceed with ethnicity data and health data, these will be special categories of personal data and explicit consent will be required to process this personal data. The UK GDPR does not define "explicit consent", but to ensure compliance it is likely that the Database Controller should include an additional, specific opt-in mechanic alongside any provided ethnicity and heath data.

5.4. By way of example, if the Database Controller asked the data subject to select an ethnicity from a drop-down selection it would not be enough to rely on the fact that the data subject had selected a specific ethnicity rather than the 'prefer not to say' option. Instead, in addition to the ethnicity selection, the Database Controller should include a separate opt-in mechanic (such as a tick-box) with a clear statement that by ticking the box the data subject consents to the processing of their ethnicity data for the Validation Purpose.

There will also be the issue regarding children. The guiding principle for the Database Controller is that children should be afforded additional protection and specific appropriate measures should be implemented to ensure that children

understand how their personal data will be processed.[footnote 53] These measures would include that an consent form (and any related privacy notice) should be "in such a clear and plain language that a child can easily understand,"[footnote 54]

Regarding children and consent, there is no specific age for when a child can or cannot give their consent in England, Wales and Northern Ireland,[footnote 55] but consent will not be 'informed' (and therefore not valid) if the child is not able to understand the processing. The minimum age for a child giving consent in respect of an Information Society Service[footnote 56] is 13 years old.

5.5. The Validation Purpose will not be a concept that will be understood by children of all ages. Therefore, the Database Controller will need to consider an appropriate age in which a child would generally be able to understand the Validation Purpose. If a child is too young to understand the Validation Purpose, parental consent will be required. The Database Controller will need to consider how it can obtain (and reasonably rely) on parental consent, including verifying that the adult is the legal parent / guardian of the child.

5.6. Furthermore, where health data (if collected) indicates that a child may have a lower competency than other children of the same age, this should be factored into account by the Database Controller.

5.7. The UK GPDR and ICO guidance is silent on whether or not a child can give 'explicit consent' in respect of processing special categories of personal data. Given that special categories of personal data are likely harder to understand for a child than, say, their name, additional safeguards should be put in place which will likely raise the age at which parental consent is required.

5.8. Finally, given that the Proposal is concerned, at least in part, with ethnicity and heath data that could affect the accuracy of facial age estimation, it is possible that even where a data subject does not consent to processing of special category personal data that the Database Controller (or Developer's tools) will draw inferences from the photographs. If this is the case, the images will be treated as special category personal data.

Annex 2

A2.1 Methodology

A literature review was undertaken as a first stage of this study exploring the data access issues relevant to the development of effective age assurance systems and

viable solutions to the identified data access issues.

The literature review was selected and analysed with the aim of establishing an initial understanding in the following research areas:

- the type and magnitude of training and testing data access issues hindering the development of age assurance tools;
- the features and requirements of the solution datasets, i.e. the datasets that – if made accessible for training and testing - could help improve the performance of age assurance tools;
- best practices and lessons learned from similar data access and testing environment initiatives – extending beyond age assurance provision to initiatives that have addressed similar challenges around enabling access to highly sensitive data to promote research and innovation; and,
- related to the above, the technical, legal and practical challenges and considerations relevant to building this type of testing environment solution.

Over 60 different sources were reviewed in total, including policy and research papers from public bodies in the UK and abroad (notably the EU and Australia), age assurance industry and consulting reports, publications from charities and non-profit/non-governmental organisations, technical roadmap and guidance documents as well as academic literature in the fields of information technology (IT), computer science, biometric science and human-computer interaction (HCI) science.

The initial literature search was conducted online using (a combination of) key words and phrases, including age assurance/verification/estimation; tools/methods/techniques, systems/technologies; training/testing; state-of-the-art/new/emerging; data/dataset(s)/database; access/availability; accuracy/effectiveness, issues/risks/limits/limitations; child(ren); sensitive; online/digital identity/identification; innovation; research (&) development, trust(ed)/ safe, privacy, protect(ion); environment/data(base)/initiative/sandbox. The search was extended to grey literature, to ensure comprehensiveness of the review.

The identified literature was skimmed through to efficiently assess their relevance to any of the research areas (listed above) and overall credibility/quality. Selected sources were then reviewed in more depth, alongside sources shared by DSIT and those identified in the proposal response. Findings were structured by research area, with more weight given to more recent findings (i.e., last two to three years).

As an outcome of this review, the literature search was further extended in two ways. First, it was extended to new keywords and phrases identified in the reviewed

literature as highly relevant to the research areas in scope. These new words largely included more specific and targeted terms than those included in the initial search, helping refine the scope of the literature search. Examples of searched terms (mostly used in different combinations) include statistical/algorithmic/AI/ML tools, biometric/facial/voice/behavioural/ capacity testing age estimation/verification/assurance,

biometric open-source datasets, bias, spoofing, liveness detection, trusted research environment. The literature search was also extended through citation searching, i.e. additional papers addressing similar or related research topics were identified through references in the initially reviewed literature. This also helped triangulate key findings from the initially reviewed literature to ensure the quality of the review. The quality of the literature review was further ensured by documenting the search and iterating the review over multiple rounds undertaken by different members of the project team.

Lastly, this review suggests that there is limited available literature measuring (and cross-comparing) the gaps in performance of the different age assurance technologies, and more specifically those "caused" by training and testing data access issues, as well as literature investigating the practical implications of these performance gaps for the use cases that these technologies can safely and robustly enable.

A2.2 Findings

A2.2.1 Data access issues relevant to age assurance

There are still evident limitations with the current technology in estimating the age or age band of a child to provide age-appropriate services online. Lack of access to high-quality and representative training and testing datasets has been widely raised as a key challenge and area of concern in the development of age assurance solutions (ACCS, 2023[footnote 57]; European Commission (EC), 2024[footnote 58]; Sas & Muelberg, 2024[footnote 59]; Unicef, 2021[footnote 60]). In a workshop held by the Age Check Certification Scheme (ACCS) in 2023, the issue of appropriate testing datasets received the most traction among participants, which included representatives across the UK age assurance industry, statistical and measurement analysts, biometric scientists, and conformity assessment specialists. They noted the legal, ethical and practical challenges in collecting, curating and maintaining suitably age-labelled, reliable and ethically obtained data sets of sufficient quality, particularly for under-18 age groups (ACCS, 2023).

Type and magnitude of data access issues

- **What age assurance techniques are impacted by limited data access for training and testing and how?**

Statistical age assurance techniques are impacted by limited access to high-quality and representative testing and training datasets (ACCS, 2023). These include biometrics analysis, behavioural profiling and inference, environmental analysis and capacity testing. A description of these techniques in the context of age assurance is provided in the table below.

Table 3 - Statistical age assurance techniques

Age assurance technique	Description
Biometric analysis	Biometric analysis is an algorithm-based method which involves assessing a user's biometric characteristics such as their voice, facial features or fingerprints to estimate their age[footnote 61]. For example, GoBubble (children safe social media platform) uses facial estimation technology for age assurance. This involves a face scan in combination with self-declaration, where the child can only access the website where the age the child entered is within the range estimated by the face estimating technology. These scans are instantly deleted following the process (5Rights foundation, 2021[footnote 62].
Behavioural profiling and inference	Behavioural profiling and inference involve analysing a user's pattern of internet activity and/or internet interactions, often using an algorithm-based approach to estimate the user's age. For example, Meta (previously Facebook) developed technology which looks at a user's birthday messages, such as 'Happy 18th birthday' in combination with the age self-declared by the user as a method of age assurance (Meta, 2021[footnote 63]). A study by the University of Cambridge found that by analysing a person's Facebook likes, they can predict a range of factors including a person's age with high levels of accuracy (University of Cambridge, 2013[footnote 64]; Kosinski et al, 2013[footnote 65]). Ying et al (2012)[footnote 66] analysed data on behavioural features such as types of application usage per day, calendar event creation per day and text usage per day to predict the age group of users.

Environmental analysis	Environmental analysis is an algorithm-based approach that involves using data collected from the user's physical surroundings or the digital infrastructure they interact with to estimate their age. The Verification of Children Online (VoCO) listed technology environment and audio environment as examples of environmental analysis, where confidence in a user's age can be increased where locational factors align with the self-declared age (VoCO, 2020[footnote 67]. Ying et al (2012) used data on environmental features captured such as types of wireless devices connected, average time stayed per location and similarities between Bluetooth devices used at home and outside of the home to predict the age group of users with a Multi-Level Classification Model.
Capacity testing	Capacity testing involves estimating a user's age (often as a range) based on their cognitive or problem-solving abilities by getting them to complete an assessment such as solving a puzzle or completing an aptitude test. For example, the Chinese app 'Baby Bus' can be set by parents to shut down after it has been used for a chosen period. After it shuts down, the user must perform a cognitive test by verifying they understand traditional Chinese characters. This prevents young children, who will most often find this challenging, from passing the test but is simple for most adults (5Rights foundation, 2021).

These techniques are still maturing, but biometric analysis and more specifically facial analysis, seems to be the most popular technology choice where hard identifiers-based authentication is not available or limited and user-reported age assurance is insufficient. They are typically used in combination with other age assurance techniques to enhance the level of confidence in identifying users' age or age range (ACCS, 2023). Research informed by interviews with children, teenagers, their parents and carers as well as representatives from technology companies, age assurance providers, regulators, policy officials, and child safety organisation, indicates that there is no one-size-fits-all approach to age assurance (Hilton & King, 2023[footnote 68]). No single age assurance solution was found to work for all user groups interviewed[footnote 69]. As a result, it is important for the market to foster innovation and continue to develop and improve the effectiveness of a wide range of age assurance techniques.

The level of assurance of statistical techniques highly depends on the quality of the datasets used to train them (EC, 2024). For example, facial age estimation tools still exhibit limitations in their ability to accurately differentiate children between narrow age bands (ACCS, 2023) due to being trained on datasets that lack sufficient scale and representativeness (Unicef, 2021). There are open-source datasets available for training biometric age estimation tools, and in particular facial analysis (1000+ to 100,000+ subject data points, i.e. images, videos or recordings). These include Morph II, FACES, FG-Net, MEDS and Adience, etc. (ACCS, 2023). Still, there remain limitations in the available datasets, whether in scale, representativeness (in particular of children of younger age and ethnic minorities), availability and accuracy of age-labelling (in particular with the lack of hard identifier-based "verified" age data being made available alongside biometric data for children) or environmental variations (e.g., should include higher variation in pose, lighting condition and image quality and not just "perfect" images) (Dahlan, 2021[footnote 70]). Academic reviews of age assurance tools seem to commonly test these open datasets individually (i.e. each tool is tested against one dataset at a time) and compare performance across different datasets for the same tool or across different tools using the same dataset.

The lack of representative data and unequal access to hard identifiers associated with the data sets that are available increases the bias of AE tools (ACCS, 2023; Sas & Muelberg, 2024). Bias is measured in the literature as the difference in the average error between subjects with a specific characteristic and those without (ACCS, 2023). Bias can be caused by different factors and at different stages of development and implementation of AI-based models. Poor data quality and representativeness is cause for algorithmic bias: if some groups are under-, overrepresented in the training data, the algorithm may become less accurate for the underrepresented group (Panic & Marjanovic, 2024[footnote 71]; Zheyan et al., 2021[footnote 72]). Additionally, bias present in humans often carry over into trained models, including age estimation models. A recent study found that AI models tend to overestimate the age of smiling faces (even more than humans do) and that the magnitude of the decrease in accuracy caused by facial expressions correlates with the participant's age. This highlights the importance of including a diverse range of facial expressions in the training and testing datasets (Ganel et al., 2022[footnote 73]). Other examples of prevalent bias impacting AI facial analysis-based age estimation tools include ethnicity and gender (Voelke et at., 2012[footnote 74]; Clapes, 2018[footnote 75]; Dahlan, 2021). Facial estimation-based technology led to lower accuracy rates for females than males or darker skinned people compared to those with lighter skin (eSafety Commissioner[footnote 76], 2023).

- The lack of access to appropriate independent testing datasets with live data (as opposed to static data) and high variation (not just perfect data points, images and

recordings) also hinders the development of age assurance tools that are effective against spoofing and other circumvention risks (ACCS, 2023). For example, a recent study finds that state-of-the-art face age verification algorithms struggle to distinguish replayed images from genuine images, leaving them vulnerable to replay attacks[footnote 77](#_ftn1) (Korshunov et al., 2024[footnote 78]).

- **What is the impact of these data access issues on end-users, i.e. consumers and children in particular?**

There is a lack of publicly available quantitative evidence on the impact of age assurance tools on online harm prevention and user safety. This could be driven in part by challenges in quantifying the counterfactual, i.e. what would happen without age assurance. In the same vein, there is limited evidence that evaluates the practical implications of the data access issues and related limitations in age assurance technologies on the use cases these technologies can support.

The enforcement of regulatory requirements for the provision of age assurance (in particular with the enactment of the Online Safety Bill in October 2023) will likely contribute to the wider business adoption of age assurance techniques beyond self-declaration[footnote 79].

The Australian government has recently awarded a tender to the ACCS to evaluate the effectiveness of different age assurance technologies in the market.[footnote 80] This initiative could help gain a better understanding of the gaps in performance affecting different age assurance technologies and the practical implications for the use cases these technologies can safely and robustly enable. This initiative comes in to support the implementation of the Online Safety Amendment (Social Media Minimum Age) Bill[footnote 81] passed by the Australian federal government on 29 November 2024. This is the world first legislation banning social media platforms from allowing users under 16 to access their services, threatening companies with fines of up to AU$50 million (US$32m) if they fail to comply.

Still, the inaccuracy and bias affecting AI-based age assurance tools have raised significant concerns as they have the propensity to foster discrimination, while insufficiently reducing the risk of online harms to children (Sas & Muelberg, 2024; Unicef, 2021). Facial estimations tools often have a margin error of several years, which can lead to prohibiting adults or older children from accessing content or services they are entitled to, while potentially enabling adults to access to online spaces reserved for children (Sas & Muelberg, 2024).

The lack of diversity in the training datasets used by statistical models and algorithmic approaches for age estimation leads to higher inaccuracies in estimating the age of users with features that are underrepresented in the training dataset,

which typically belong to minorities and marginalised communities. This is not only problematic from a performance perspective, but also because it could perpetuate structural and societal biases and further hinder the access of underrepresented groups to age-appropriate and age-restricted services online (Sas & Muelberg, 2024; 5Rights Foundation, 2021; EDRi, 2023[footnote 82]).

- **Is there a market failure and is there scope for government intervention?**

The lack of access to appropriate training and testing datasets are driven by resource, practical and commercial constraints as well as ethical and legal concerns. Age assurance providers are mostly composed of startups and small-sized companies, which individually lack the significant resources and investment needed to obtain datasets of the required scale and quality. There are also a few larger players, which involve firms that both develop and use their own age assurance solution (e.g. online platforms like Meta, YouTube, TikTok). However, their in-house solutions seem to involve leveraging on their own user data, which is seemingly not easily accessible to the rest of the market. The first part of this observation is based on looking at the age assurance policies and measures reported to be taken by some of these online platforms. For example, in addition to having an initial industry-standard age-gate that requires people to fill in their self-declared birthdate when signing up for their services, TikToK reports having an online safety moderation team monitoring signs that an account may be used by a child under the age of 13. They also report using other information provided by their users to identify potential underage accounts[footnote 83].

Similarly, Meta reports developing AI technology to find and remove accounts belonging to people under the age of 13, for example by looking at the age reported in birthday messages posted on a user account[footnote 84], [footnote 85]. Meta has partnered with third-party age assurance provider Yoti, but this seems to mainly involve Yoti providing an additional or complementary layer of age verification using their own facial estimation technology (trained and tested on their own data)[footnote 86].

Limitations and difficulties in accessing appropriate datasets (sufficient quality and scale) can create significant market entry barriers and carries the risk of hampering competition and innovation in the age assurance market in the future. Hypothetically, this could pave the way for the larger market players (including very large social media platforms, search engines, online marketplaces) with the resources and/or access to vast amounts of user data to almost exclusively innovate and operate in this market. This would ultimately harm competition, innovation and consumer welfare in the long run. Similar competition concerns have been raised with regard to cross-platform authentication as an age assurance method. Cross-platform

authentication involves leveraging on existing user accounts on large online platforms to authenticate the age of a user in order to access new products and services. This could further entrench the market dominance of these large tech firms (5Rights Foundation, 2021).

The industry is calling for greater support and intervention from the government and regulators to address this gap in the market, including initiatives to improve access to datasets, conduct landscape review research, and publish relevant findings. Several previous initiatives, such as the Home Office's Age Verification Regulatory Sandbox for alcohol sales, the ICO's Regulatory Sandbox, and the Safety Tech Challenge Fund, are highlighted as positive advancements. The industry also notes large uncertainty around the ethics and lawfulness of collecting biometric and personal data about children which adds to the challenge of gathering, curating, maintaining, and providing suitable data sets (ACCS, 2023).

A.2.2 Datasets that can help improve training and testing of age assurance tools

- **What are the requirements and considerations relevant to the datasets that can help mitigate the data access issues impacting the development of effective age assurance tools?**

The literature provides valuable insights on the type and features of the datasets that would benefit the development AI-based age assurance tools (ACCS, 2023), namely:

- **Representative and fairly distributed**: the datasets need to be representative of age, gender, skin tone, different socio-economic groups, etc. However, it is essential to carefully select subjects so that the dataset remains unbiased and accurately represents the population. Key weaknesses in existing datasets lie in the lack of data of children as well as that of ethnic minorities and people with disabilities and atypical facial traits.
- **Age-labelled and accurate:** datasets should be accurately age-labelled and verified, e.g. individual biometric data should be available alongside accurate data on the subject's age. Current datasets typically involve either 'data scraping' the internet, with automatic age labelling, or using self-reported age. Moreover, children typically lack hard identifiers (e.g. passports), adding to the challenge of obtaining ground-truth data that can be used to train age estimation algorithms.
- **Very large scale/sample size:** Sample size formulae are readily available to determine the number of samples needed to accurately estimate overall accuracy, taking into account the technology's estimated accuracy, the desired margin of error, and a specified confidence level. Preliminary calculations indicate that

achieving strict or enhanced accuracy thresholds requires a very large sample size.

- **Ethical, lawful and commercially usable:** datasets should be ethically collected and in accordance with GDPR and other applicable laws and regulations
- **High variation:** datasets should not just contain "perfect" quality visual or voice data, but should introduce more environmental and quality variations to allow age assurance tools to improve their performance in the case environmental perturbations, e.g. dataset of facial images should be representative of orientation and position, single or multiple faces in an image, image size, resolution, the direction of light, luminosity, colour, etc. Academic findings highlight the need for a diversity of facial expressions to be present within training and validation datasets (Ganel et al., 2022; [footnote 87]).
- **Live data:** Static data is proving to be insufficient to train and test AI-based age assurance tools against risks of spoofing and circumvention (e.g., usage of pre-prepared photos, modulating voices, deepfakes) and enable them to conduct robust liveness and authentication checks.
- **Different data types needed to help and enable different technologies, user groups and use cases:** Different use cases and different user groups need different age assurance technologies which need different datasets, e.g. facial vs. voice age estimation. To optimise inclusion as well as data minimisation, it is important to adopt a layered approach to age assurance and provide users with a range of age assurance options (Hilton & King, 2023).
- **Independence of testing and training datasets:** the training and testing datasets need to be completely independent from one another to avoid overfitting and data leakage and ensure robust performance. When the same data is used for both training and testing, AI-based age estimation can "memorize" the patterns in the dataset rather than learning to generalise. This can lead to overfitting, where the tool performs well on the training and testing data but fails to accurately predict outcomes on new data. Age assurance tools rely on subtle patterns, such as facial features, behavioural traits, or voice characteristics. Any overlap in datasets could artificially inflate accuracy without genuinely improving the model's ability to estimate user age or age band. Ensuring independence also avoids data leakage, which occurs when information from the testing set inadvertently influences the training process, leading to misleading performance metrics. Obtaining independent testing datasets is a key challenge highlighted by industry representatives.
- **Synthetic data:** One consideration in the curation of the potential solution datasets is to introduce synthetic data (Schroff, 2021[footnote 88]). There are limitations with the use of synthetic data to train age assurance models, partly due

to the limitations with the original datasets themselves used to create the synthetic data. However, if the original real-world datasets collected as part of the solution are of sufficient quality and scale to create high-quality synthetic data, then that could help mitigate some of the privacy concerns related processing real data on people.

An important consideration in the exploration of the solution datasets is the trade-off between the potential for accuracy and effectiveness achievable by age assurance tools (from being able to access training and testing datasets of a certain level of quality and quantity) and the cost involved in collecting, curating, maintaining and providing access to these datasets. This trade-off would also have to consider the use cases that can or cannot be enabled depending on the different levels of accuracy and effectiveness achievable by different age assurance tools, and the impact on end-users, and in particular children, in terms of online safety, harm prevention and inclusion. There seems to be limited research in the literature undertaking this type of analysis. For illustrative purposes, the graph below shows the type of trade-off analysis that would be undertaken to assess the “sweet” spot whereby the accuracy threshold required for a hypothetical use case (requiring high accuracy of 90% in this example) is met while the cost of the datasets is below a maximum viable cost threshold (hypothetically “5” cost units in this example).

Figure 5 - Hypothetical trade-off between accuracy and cost of appropriate training and testing datasets

A2.2.3 Solutions to the data access issues

Trusted research environment (TRE)

Technical interventions in the form of sector-specific sandboxes (e.g. gaming), government-funded competitions, better open-sourced datasets and trusted testing environments have been raised as potential solutions to the data access issues impacting innovation in age assurance. A trusted testing environment, also commonly referred as a trusted research environment (TRE), could prove to be the most viable solution, by striking the right balance between enabling valuable research and innovation through access to large amounts of highly sensitive data and protecting that data with the highest of security standards and processes. The literature provides valuable insights from relevant TREs on best practices, key challenges, and lessons learned in the development of a TRE.

Building a successful Trusted Research Environment

The requirements for building a TRE are outlined in the Trusted Research Environments Green Paper (UKHDRA, 2020[footnote 89]. These requirements are based on the 'five safes' of data privacy defined by the ONS, including safe people, safe projects, safe settings, safe outputs and safe data (ONS, 2017[footnote 90]). They ensure that data can be accessed for research purposes, but data privacy is maintained (ONS, 2017). By their very design, the data of individuals held within a TRE is privacy protected (UKHDRA, 2020). A TRE acts as one of the five safes,

providing a ‘Safe Setting’ or ‘Data Safe Haven’ approach, which allows access to sensitive data whilst minimising the risk of re-identification of individuals from de-identified data and limiting unauthorised use of the data.[^91]

A paper by the UK Health Data Research Alliance outlines their principles and best practises for building a TRE. A TRE should securely hold data to prevent individual data export, allow secure remote access for researchers to perform analysis, and include tools for analysis within the environment while maintaining strict barriers between the TRE and the external world (UKHDRA, 2021[^92]).

Key challenges when building a Trusted Research Environment and viable solutions

There may be **legal challenges** with obtaining the appropriate permissions to process sensitive/verified data. There may be issues/time delays in getting permission from the data controller or being able to assume data controllership.

There may be challenges with **user experience**, for example you cannot copy and paste within a TRE or access full software capabilities within the TRE due to privacy concerns. This may hinder the productivity of researchers accessing a TRE.

One potential limitation of TRE’s outlined by the UK Health Data Research Alliance is that they require detailed applications for each project, including methodology, funding, ethical approval, and timeframes, which could be burdensome for startups or even act as a deterrent for new entries into the age assurance market (UKHDRA, 2021).

Furthermore, while TREs typically contain standard statistical tools, there is a risk that researchers may need additional software or custom algorithms, and restrictions on these requests could limit the effectiveness of training age assurance technologies such as complex facial recognition using algorithms and AI. There may be concerns over researcher productivity when working within the TRE, or it may be time consuming to adapt the TRE to support the type of analysis needed e.g. to accommodate for various coding languages and statistical analysis software (UKHDRA, 2020; HDR UK, 2020[footnote 93]. Formal safeguards and processes, while vital for privacy, may slow technological progress (UKHDRA, 2021).

The Data Analytics and Research Environments UK (DARE UK) programme was created to explore the potential for TRE’s to enable research and innovation involving sensitive data (DARE, 2022[footnote 94]). Phase 1, which gathered information about the needs of research communities, confirmed the ‘consensus’ that TRE’s are the way forward for being able to use sensitive datasets for research and innovation (DARE, 2024[footnote 95]). Phase 2 is looking to ‘Build, test and

Establish' TRE's and is currently underway (DARE, 2024[footnote 96]). The investigation in phase 1 delved into the challenges faced in designing TRE's and potential solutions, resulting in 31 recommendations for the design and delivery of trustworthy research infrastructure involving national data (DARE, 2022). Some key recommendations from the phase 1 summary report include providing a straightforward researcher accreditation framework to allow trustworthy researchers to more easily access sensitive data in the name of public benefit. It also recommends centralising and standardising the processes enabling access to sensitive data across the UK, as well as developing 'architecture' for TRE's to use as a reference (DARE, 2022).

DARE UK also recommended developing privacy enhancing technologies (PET's) to be used by TRE's as well as modifying data lifecycle to allow cross-domain research (DARE 2022). This could help solve the issue that 45% of respondents to the Alan Turing Institute survey said they often needed to combine data from multiple data sources for research and this was a barrier when they needed to access sensitive data from multiple environments (Alan Turing Institute, 2022[footnote 97]).

Despite these challenges with TRE's, alternatives such as the 'Data Release' method, where processed or de-identified datasets are released to researchers, are more vulnerable to data re-identification than the TRE method (UKHDRA, 2020). In a TRE, there should be Airlocks preventing unauthorised import and export to and from the TRE, and only summary data can be exported from the TRE. Researchers can only access the sensitive data via a Virtual Desktop Interface (VDI), and they cannot connect to other software repositories whilst they are in the safe setting/TRE (UKHDRA).

Another key consideration is that there is a risk that algorithmic models can fully memorise the data points they are trained on (and even more so in the case of less represented groups and outliers within the datasets) (Carlini et al., 2019[footnote 98]; Feldman, 2021[footnote 99]. In the context of age assurance tools, this means that, for example, facial analysis tools trained within the testing environment, can memorise the faces of the data subjects it encountered within the testing environment and ultimately "extract" this sensitive data out of the testing environment. There are ways to mitigate this memorisation risk within the design of the algorithm by including privacy-preserving features. Such features should be considered when reviewing and approving the tools that can access the testing environment.

Alternative solutions to Trusted Research Environments

Regulatory sandboxes are an alternative solution to a TRE which can encourage innovation in various sectors. Examples of past regulatory sandbox projects relevant

to the development of age assurance technologies include the ICO's Regulatory Sandbox and the Home Office's Age Verification Regulatory Sandbox for alcohol sales. The ICO's Regulatory Sandbox more specifically addresses the data access issues impacting age assurance providers.

The ICO introduced their Regulatory Sandbox to encourage the development and innovation of technologies that safely utilise sensitive data and have the potential to deliver public benefit. One phase of the Sandbox focused on innovations involving either data sharing or children's privacy. Yoti entered this phase of the Sandbox in 2021 to extend its age estimation technology to younger users (aged 6-12) and provide age assurance to online platforms accessible only to children (e.g. gaming sites). Yoti uses a privacy preserving approach to age assurance where the user's age is estimated by capturing an image which is immediately deleted after the age estimation is completed (ICO, 2022)[footnote 100]. However, the accuracy of Yoti's facial estimation algorithm technology depends on having access to a large and diverse set of facial images of individuals, including children of the relevant age brackets, alongside corresponding ages. The ICO helped Yoti navigate regulatory requirements and parental consent forms applicable to collecting this data. Following their participation in the Sandbox, Yoti reported that their age estimation model was accurate to within 1.28 years on average. Moreover, the 6–12-year-old bracket achieved the highest accuracy across all trialled age brackets. A key outcome of Yoti's participation in the Sandbox was the decision by the ICO that Yoti's facial estimation tool was not processing special category data, despite using biometric data, as it did not intend to confirm the unique identification of a person. This resulted in the ICO updating their guidance accordingly to clarify that if biometric data is processed but not for the purpose of uniquely identifying someone, it does not qualify as special category data processing (Yoti, 2024)[footnote 101].

The Home Office's sandbox was conducted in 2022 to explore digital alternatives to employees in the retail and hospitality sectors manually checking IDs, that could help prevent under 18's buying age restricted products while enhancing consumer experience (Home Office, 2022)[footnote 102]. As part of this, various age assurance technologies from providers such as Yoti, 1account, MBJ technology and Innovative Technology Ltd were paired with retailers such as supermarkets (Asda, Co-op, Morissons and Tesco), bars and clubs across nine different trials. For example, 1account tested the use of its digital identity (ID) mobile app in a nightclub. A user could set-up a digital ID within the app by scanning an identity document, which would then be authenticated using a live photo of the user. The in-app digital ID could then be used to verify users' age to enter the nightclub (Home Office, 2022) [footnote 103]. Another trial involved integrating facial age estimation technology developed by Yoti to self-scan checkouts in Asda stores to estimate consumers'

ages, with anyone deemed to likely be under 25 (Challenge 25) to require a manual check by an employee before approving the sale of alcohol and other age-prohibited items. Two retailers who participated in the trial adjusted the Yoti facial age estimation technology to require a manual check for individuals estimated to be younger than either 28 or 30. In both cases, this resulted in a 100% rejection rate of 18 to 19-year-olds and proved to be more accurate than humans assessing whether a customer needs a 'Challenge 25' check (RASG, 2023[footnote 104]).

Open innovation challenges are another way that the government and regulators can encourage innovation of new and emerging technologies. For example, the first round of the Safety Tech Challenge Fund supported 5 projects developing innovative automated tools to help keep children safe online, specifically using end-to-end encryption (UK Gov, 2021)[footnote 105]. Round two focused on the issue of sharing links containing child sexual abuse material (CSAM) and funded three companies developing solutions to detect and/or disrupt links routing offenders to CSAM. All three companies (C4FF, Camera Forensics, Vistalworks) reported making technological progress that would not have been possible without the resources from the open innovation challenge (DSIT, 2024[footnote 106]. Still, technical, legal and process challenges in accessing representative and real data for training and testing were highlighted as limitations to the development and technical performance of the solutions (DSIT, 2024).

A2.2.4 Other key considerations relevant to improving age assurance provision

Our review of the literature relevant to the state of play of age assurance provision in the UK highlights the need for key developments and interventions - other than those looking to solve the data access issues in scope of our study – that likely need to take place in order to promote a healthy market and innovation in the age assurance space.

The lack of interoperability across age assurance systems, third-party providers, online services, and other technologies operating the age assurance market has been highlighted as an area requiring further intervention. There needs to more coordination and efforts between the different age assurance ecosystem players to enable interoperability between age assurance products and services. There is currently a large variability across providers in the age check response formats (e.g., discrete or continuous) they provide as well as remediation and integration processes[footnote 107] they involve for their users (e.g., online platforms). This limits cross-recognition of age assurance checks by other providers and adds to the perceived uncertainties on the legal liability applicable to age assurance providers and the platforms and websites using their solutions. Interoperability is also key in enhancing the end-user experience and helping somewhat decrease the reluctance

to age assurance adoption currently prevalent among consumers and businesses (FOSI, 2023[footnote 108]). With interoperable age assurance systems, end-users could re-use the verification from one provider across different websites requiring age verification, only needing to re-verify in the case of higher risk content or services. This would prevent users from being repeatedly prompted to verify their age on each new site (FOSI, 2023).

The development and maintenance of minimum standards is key to promote interoperability in the market. This could involve minimum standards around performance benchmarks, verification methods, privacy and data protection, data formats and protocols, user experience, compliance and certification, risk management and security. In the UK, the Age Verification Providers Association (AVPA) and ACCS are leading the efforts in developing these standards. However, there are divergent views on the working standards as well as industry concerns that the current market is too premature for detailed and sophisticated standards. The ICO and ACCS are specifically leading the research on technical measures and accuracy metrics for age assurance tools. Additionally, there is some scope for the age assurance market to leverage on the standardisation efforts and lessons learned so far from the Digital Identity and Attributes Trust Framework (DIATF) [footnote 109]. A key objective of the DIATF is to drive interoperability in the digital identify verification market, nationally and internationally, while maintaining the upmost highest minimum standards of quality and security applicable to the digital identification products and services “participating” in the framework. This involves maintaining close cooperation with industry, technical experts, relevant regulatory bodies and third sector representatives, in the UK and abroad, throughout the lifetime of the framework (beyond the development phase, even when fully operational). The DIATF also tackles the issues of trust and reluctance in adopting digital identity verification among consumers and businesses and investigates government as well as cross government-industry interventions that could help drive consumer trust and wider adoption of digital ID verification in the UK. Again, there is potential for the age verification market to leverage on these efforts. The euConsent initiative, funded by the European Union, aims to develop a standardised, interoperable, and privacy-compliant framework for age verification and parental consent management across digital services in the EU[footnote 110]. As part of this initiative, the EU is also funding the development of a hybrid solution that revolves around the use of digital “age” tokens[footnote 111]. The tokens will be created through interoperable age assurance methods (part of the euConsent framework) and re-usable across digital services, such as apps and websites. These tokens could help resolve concerns around data minimisation looming over existing age assurance provision systems. In the latest news, an initial proof of concept is expected to be built by the end of 2024[footnote 112]. In this pilot phase, age tokens seem to have only

been made available to adults and children above the digital age of consent [footnote 113] (euConsent, 2022)[footnote 114]). Children below the digital age of consent would instead require parental consent every time they want to access age-restricted (or personal data collecting) online services. However, parents in the pilot were able to use age tokens to verify their own age as part of the process of providing parental consent, which helped create some efficiencies.

The industry has also raised a need for further legal guidance from the government and regulators on the requirements and liabilities with regard to sharing data relevant to age assurance, as well as clarity on which techniques cannot be applied to what use cases (ACCS, 2023). The industry also calls for further development and resources in independent testing. There are still very few parties offering that service exist[footnote 115], especially for evaluating accuracy for under-13 age thresholds.

To conclude, there is the potential for a testing environment-type of solution to promote innovation and positive developments in the age assurance market, beyond solving the data access issues and performance of age assurance tools. This type of solution could help deepen our understanding and confidence of age assurance techniques (through being tested on high-quality and real-world data). This could help develop clear and robust guidance on the applicability and trade-off of different age assurance tools vis-à-vis different use cases (e.g., of varying risk profiles and accuracy needs). It could also support comparability of performance outcomes between different age assurance tools (ACCS, 2023) and help define appropriate performance metrics and thresholds as well as minimum standards more broadly

Annex 3 Stakeholder consultations

A3.1 Engagement statistics

Individual interviews

Table 2 shows the number and type of stakeholders interviewed as part of the stakeholder consultations phase.

Stakeholder type	Number of interviews	Proportion of total
Age assurance providers (including testing organisations)	6	20%
Private sector user side	7	23%

Public sector	6	20%
Third sector (charities and academics)	11	37%
Total	30	100%

Workshop
The study also involved a workshop with 9 participants representing different facial age estimation solution providers and testing organisations.

A3.2 1. Key findings
Individual interviews
Views on the age assurance market
There was consensus among interviewed stakeholders around the limitations of existing age assurance tools. Some of the key areas are as follows:

- Accuracy challenges and vulnerabilities to circumvention and spoofing risks
  - No age assurance method is 100% accurate
  - Children and minorities, e.g. darker skin tones, women, individuals with disabilities likely to experience higher error rates
- Lack of accountability
  - Insufficient mechanisms for recourse/appeal/opt-out when age estimation tools produce errors
- Privacy and ethical concerns
  - Risk of data breaches or misuse
  - Risk of public distrust
- Costs
  - Age assurance solutions (beyond self-declaration) are prohibitively expensive for small businesses (e.g., subscription to age estimation and multi-layered ID checks systems can cost over £250/month while custom API integration can cost up to £4000)
  - High costs of age assurance also contribute to inconsistent implementation across online platforms

There was also a consensus about the limitations of the age assurance market more broadly. Stakeholders noted a lack of coordination, consistency, standardisation and interoperability in the age estimation market. Providers often work in isolation and use proprietary data which leads to fragmented solutions and inconsistent user

experiences. On top of this, there is no clear government guidance on minimum requirements for age assurance methods, and although some international standards (e.g., ISO 27566) are emerging, there is still lots of variability in the age assurance market.

Stakeholders emphasised the need for a simple, universally accepted system, potentially aligned with DIATF, and voiced concerns about dominance by a few large age assurance providers. Additionally, there is insufficient research on the age assurance market, specifically a lack of research on the real-world risks of children accessing inappropriate content, making it hard to consider the true value of improved data access in the market.

Views on potential solutions to data access issues

There was limited consensus on prioritising a database solution for age assurance, with greater scepticism from industry and users compared to the public sector. However, there is broad agreement that such a solution could potentially support fair and consistent testing of age assurance tools and contribute to improved standardisation in the market. Even stakeholders hesitant about a database solution acknowledge the benefits of a high-quality dataset for independent testing and performance benchmarking.

Some stakeholders have pointed to international efforts addressing data and standardisation gaps, such as NIST's datasets, Australia's use of mystery shoppers, and Google Canada's development of a facial image dataset. However, there is no evidence of existing initiatives have been identified that provide comprehensive, representative datasets which include images of children. Additionally, a secure database could be scalable beyond age assurance to mitigate other online harms, such as grooming and illegal content detection, while fostering industry collaboration and ethical oversight.

However, there are drawbacks to a database solution. A key challenge is the high upfront costs and limited commercial viability. On top of this, charities and academics have raised concerns about public trust, consent, and the risks of data privacy and potential misuse, particularly regarding the collection children's data.

Some stakeholders have mentioned the potential to use synthetic data. However, some age assurance providers and experts have warned that synthetic data is useful for spoofing tests but limited in training real-world algorithms. Synthetic data could potentially help scale up a dataset but in order for this to be effective, the original dataset has to representative.

Industry stakeholders also note that a database solution may have limited

applicability in sectors requiring strict legal age verification, such as gambling, where precise personal data verification is necessary. If the applicability of the solution is narrow or limited, it may affect the value for money of a database solution. Given these challenges, some stakeholders suggest alternative government interventions, such as clearer regulatory guidance or incentives to drive innovation, rather than using a database approach.

Stakeholders generally advocated for complementary or alternative government interventions to enhance the age assurance market rather than prioritising a database solution. A key recommendation is the development of government guidance on age assurance methods and best practices, along with standardised certification or endorsement systems. This aligned with the findings from our workshop, with the point that government guidance can vary across departments and his can hinder investment into the age assurance market.

For companies who use age assurance tools, stakeholders highlighted the need for training resources to help them understand different age assurance techniques, and consider privacy, ethical, and legal aspects when selecting the most appropriate age assurance method. Clear government endorsement of recommended technologies could also help build public and industry confidence.

Workshop
We further discussed the potential uses of a dataset (e.g. for external validation, internal validation or training) and their various advantages and disadvantages. Different uses will be of various benefit do different stakeholders. For example, improved data access for training may be more useful for tool developers, whereas external validation may be more useful to users.

We discussed the required size of potential datasets used to perform these roles. The required size will depend on its use. External validation requires significantly less data than a dataset to be used for training. The 70-20-10 rule is usually used for data, with 70% of available data reserved for training, 20% for internal validation and 10% for external validation. Because of this, building a dataset for external validation would be less resource intensive. This said, there was discussions about costs being reverse exponential, with high fixed costs from setting up systems and then decreasing per unit costs for collecting larger datasets. However, this participant also noted that it cost them £50 per participant to collect data.

There was a consensus that it is better to include a representative dataset for external validation, though a variety of lighting conditions is also essential.

An idea which arose from the workshop was data passing through a continuous

cycle. New data is initially reserved for external validation, then after a period it is moved into internal validation and finally moved into training data. It must be done in this order.

Unclear or mixed messages from the government across departments was raised as a key obstacle to buy-in from industry. Clear and consistent backing of age assurance as a technology from government would have an impact on the sector.

Bibliography

5Rights foundation (2021). But how do they know it is a child? Age Assurance in the Digital World.

Age Check Certification Scheme (2022). Measurement of Age Assurance Technologies- A Research Report for the Information Commissioner's Office (ICO).

Age Check Certification Scheme (2023). Measurement of Age Assurance Technologies, Part 2 – Current and short-term capability of a range of Age Assurance measures. A Research Report for the Information Commissioner's Office (ICO) and the Office of Communications (Ofcom).

Arthur Cox LLP (2024). Navigating Age Assurance in the Online World Insights from the ICO. Available at: https://www.arthurcox.com/knowledge/navigating-age-assurance-in-the-online-world-insights-from-the-ico-2/

Australian Government (2023). Government response to the Roadmap for Age Verification. Available at: https://www.infrastructure.gov.au/sites/default/files/documents/government-response-to-the-roadmap-for-age-verification-august2023.pdf

Australian Government (2024). Update on Age Assurance Trial. Available at: https://mddb.apec.org/Documents/2024/TEL/TEL69-STSG/24_tel69_stsg_010.pdf

BiometricUpdate.com (October 4 2024). The EUDI Wallet was not meant for age assurance: AVPA. Available at: https://www.biometricupdate.com/202410/the-eudi-wallet-was-not-meant-for-age-assurance-avpa

Carletti, V., Greco, A., Percannella, G., and Vento, M. (2019). Age from faces in the

deep learning revolution. IEEE transactions on pattern analysis and machine intelligence.

Carlini, N., Liu, C., Erlingsson, U., Kos, J., Song, D. (2019). The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks. arXiv:1802.08232 Clapes, A., Bilici, O., Temirova, D., Avots, E., Anbarjafari, G., and Escalera, S. (2018). From Apparent to Real Age: Gender, Age, Ethnic, Makeup, and Expression Bias Analysis in Real Age Estimation.Computer Vision and Pattern Recognition (CVPR) Workshop Paper. Congressional Research Service (2023). Challenges with Identifying Minors Online. Available at: https://crsreports.congress.gov/product/pdf/IN/IN12055#:%7E:text=Potential%20Challenges%20with%20Identifying%20Minors,as%20those%20younger%20than%2013

Dahlan, H. A. (2021) A Survey on Deep Learning Face Age Estimation Model: Method and Ethnicity. International Journal of Advanced Computer Science and Applications; West Yorkshire Vol. 12, Iss. 11.

DARE UK (2022). Paving the way for a coordinated national infrastructure for sensitive data research A summary of findings to date from Phase 1 of the UK Research and Innovation DARE UK programme. Available at: DARE_UK-Paving_the_way_coordinated_national_infrastructure_sensitive_data_research-Aug2022.pdf

DARE UK (18 September 2024). £18 million for DARE UK to support secure research on sensitive data. Available at: https://dareuk.org.uk/news-and-events/18-million-for-dare-uk-to-support-secure-research-on-sensitive-data/ DARE UK (5 November 2024).

DARE UK Phase 2 is underway, but what does this mean for everyone?Available at: https://dareuk.org.uk/news-and-events/dare-uk-phase-2-is-underway-but-what-does-this-mean-for-everyone/

DCRF (2022). Families' attitudes to age assurance. Research Commissioned by the ICO and Ofcom.

Digital Trust & Safety Partnership (2023). Age Assurance Guiding Principles and Best Practices.

DSIT (2024) Evaluation of Safety Tech Challenge Fund Round 2.

Eidinger, E., Enbar, R., and Hassner, T. (2014). Age and gender estimation of unfiltered faces. IEEE Transactions on information forensics and security.

El Khiyari H, Wechsler H (2016) Face Verification Subject to Varying (Age, Ethnicity, and Gender) Demographics Using Deep Learning.

eSafety Commissioner (2023). Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography. Available at: https://www.esafety.gov.au/about-us/consultation-cooperation/age-verification#roadmap-andbackground-report

eSafety Commissioner (2024). Tech Trends Issues paper. Available at: https://www.esafety.gov.au/sites/default/files/2024-07/Age-Assurance-Issues-Paper-July2024_0.pdf

EuConsent (2022). Pilot Execution Report. Available at: https://euconsent.eu/download/pilot-execution-report-first-large-scale-euconsent-pilot/

EuConsent (2024). Feasibility Study for AVP Interoperability between Native Mobile Applications. Available at: https://euconsent.eu/download/feasibility-study-for-avp-interoperability-between-native-mobile-applications/

European Commission (2024). Research report: Mapping age assurance typologies and requirements.

European Digital Rights (EDRi) (2023). Position Paper: Age verification can't 'childproof' the internet. Available at: https://edri.org/our-work/policy-paper-age-verification-cantchildproof-the-internet/

Family Online Safety Institute. (2023). Coming to Terms with Age Assurance. Feldman, V. (2021). Does Learning Require Memorization? A Short Tale about a Long Tail. Google Research, Brain Team.

Ganel, T., Sofer, C., and Goodale, M.A. (2022). Biases in human perception of facial age are present and more exaggerated in current AI technology. Scientific Reports 12.1, p. 22519.

GOV.UK (2023). Online Safety Act 2023. Available at: https://www.legislation.gov.uk/ukpga/2023/50/section/157

GOV.UK (2023). UK digital identity and attributes trust framework beta version (0.3). Available at: https://www.gov.uk/government/publications/uk-digital-identity-and-attributes-trust-framework-beta-version/uk-digital-identity-and-attributes-trust-framework-beta-version

Hanaoka, K., Ngan, N., Yang, J., Quinn, G.W., Horn, A., Grother, P. (2024). Face Analysis Technology Evaluation: Age Estimation and Verification. NIST Interagency Report NISTIR 8525.

HDR UK (2020). A national health data research capability to support COVID-19 research questions

Health Data Research UK (2020). Digital Innovation Hub Programme Prospectus Appendix- Principles for Participation. Available at: https://www.hdruk.ac.uk/wp-content/uploads/2020/03/200304-Principles-for-Participationv2pdf.pdf

Hiba, S., Keller, Y. (2023). Hierarchical Attention-based Age Estimation and Bias Analysis. arXiv:2103.09882v2

Hilton, Z., and King, H. (2023). Making age assurance work for everyone: inclusion considerations for age assurance and children.

Home Office (2022). Call for proposals. Available at: https://www.gov.uk/government/publications/age-verification-technology-in-alcohol-sales-regulatory-sandbox/call-for-proposals

Home Office (2022). Details of the trials. Available at: https://www.gov.uk/government/publications/age-verification-technology-in-alcohol-sales-regulatory-sandbox/details-of-the-trials

Home Office (2022). Key learnings from the trial. Available at: https://www.gov.uk/government/publications/age-verification-technology-in-alcohol-sales-regulatory-sandbox/key-learning-from-the-trial

ICO (2022). Age Assurance: Estimating or Verifying the Age of Service Users. Available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/how-to-use-our-guidance-for-standard-one-best-interests-of-the-child/best-interests-framework/age-assurance/

IFF (2024a). How online businesses are using age assurance: Research findings. Available at: 20240704-ico-age-assurance-report.pdf

IFF Research (2024b). Organisational approaches to age assurance in the UK: Technical annex. Report for the ICO.

Information Commissioner's Office (2022). Regulatory Sandbox Final Report: Yoti - A summary of Yoti's participation in the ICO's Regulatory Sandbox. Available

at: https://ico.org.uk/media/for-organisations/documents/4020427/yoti-sandbox-exit_report_20220522.pdf

Japan Digital Agency (2023). Individual Number Cards.

Jarvie, C., and Renaud, K. (2021). Are you over 18? A snapshot of current age verification mechanisms.

Korshunov, P., George, A., Özbulak, G., and Marcel S. (2024). Vulnerability of Face age Verification to Replay Attacks. Published in 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Available at: https://www.pnas.org/doi/pdf/10.1073/pnas.1218772110

Lindmark. E. (2021). The feasibility of face scanning as an age verification tool from a technical- and UX-perspective.

Meta (2021). How Do We Know Someone Is Old Enough to Use Our Apps? Available at: https://about.fb.com/news/2021/07/age-verification/

Ngan, M., and Grother, P. (2014). Face Recognition Vendor Test (FRVT) Performance of Automated Age Estimation Algorithms. National Institute of Standards and Technology

OECD (2019). The role of sandboxes in promoting flexibility and innovation in the digital age. Available at: https://goingdigital.oecd.org/data/notes/No2_ToolkitNote_Sandboxes.pdf

Office for National Statistics (2017). The 'Five Safes'- Data Privacy at ONS. Available at: https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/

Ofgem (2018). What is a regulatory sandbox? Available at: https://www.ofgem.gov.uk/sites/default/files/docs/2018/09/what_is_a_regulatory_sandbox.pdf

Okokpujie, K., Noma-Osaghae, E., John, S. N., Ndujiuba, C., and Okokpujie, I. P. (2021). Comparative analysis of augmented datasets performances of age invariant face recognition models. Bulletin of Electrical Engineering and Informatics. DOI: 10.11591/eei.v10i3.3020

Online Safety Data Initiative. (2022). Building a Trusted Research Environment for

sensitive online harms data

Othmani, A., Taleb, A. R., Abdelkawya, H., and Hadid A. (2020). Age estimation from faces using deep learning: a comparative analysis. Computer Vision and Image Understanding.

Panic, N., Marjanovic, M., Bezdan, T. (2024). Addressing Demographic Bias in Age Estimation Models through Optimized Dataset Composition. Mathematics 2024, 12, 2358. https://doi.org/10.3390/ math12152358

Pontes, Britto Jr, Fookes, Koerich (2016). A flexible hierarchical approach for facial age estimation based on multiple features. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0031320315004458

Praesidio Safeguarding (2023). Making age assurance work for everyone: inclusion considerations for age assurance and children.

Retail of Alcohol Standards Group (RASG) (2023). RASG Sandbox Evaluation. Available at: https://rasg.org.uk/wp-content/uploads/2023/11/RASG-Sandbox-Evaluation.pdf

Sas, M., and Muehlberg, J. T. (2024). Trustworthy Age Assurance? A risk-based evaluation of available and upcoming age assurance technologies from a fundamental rights perspective. Report for the Greens/EFA in the European Parliament.

Schroff, F. (2021). Synthetic Faces to Improve Privacy and Fairness. European Data Protection Supervisor.

The Alan Turing Institute. (2022). Review of Digital Research Infrastructure Requirements for AI. Available at: https://www.turing.ac.uk/sites/default/files/2022-09/ukri-requirements-report_final_edits.pdf

UK GOV (2021). Government funds new tech in the fight against online child abuse. Available at: https://www.gov.uk/government/news/government-funds-new-tech-in-the-fight-against-online-child-abuse

UK Health Data Research Alliance (2020). Trusted Research Environments (TRE)- A strategy to build public trust and meet changing health data science needs. The Trusted Research Environments Green Paper. Available at: https://zenodo.org/records/4594704

UK Health Data Research Alliance (2021). Building Trusted Research environments.

Available at: [https://www.ed.ac.uk/sites/default/files/atoms/files/5safe_principles-building_trusted_research_environments.pdf] (https://www.ed.ac.uk/sites/default/files/atoms/files/5_safe_principles-_building_trusted_research_environments.pdf)

Unicef (2021). Digital Age Assurance Tools and Children's Rights Online across the Globe: A Discussion Paper.

University of Cambridge (2013). Digital records could expose intimate details and personality traits of millions. Available at: https://www.cam.ac.uk/research/news/digital-records-could-expose-intimate-details-and-personality-traits-of-millions

Van der Hof, S. and Ouburg, S. (2022). 'We Take Your Word For It' - A Review of Methods of Age Verification and Parental Consent in Digital Service. VoCO (2020).

VoCO (Verification of Children Online) Phase 2 Report. Available at: https://assets.publishing.service.gov.uk/media/5faa9cffd3bf7f03a841cfc2/November_VoCO_report_V4__pdf.pdf

Voelkle, M. C., Ebner, N. C., Lindenberger, U. & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. Psychol. Aging 27, 265–277.

World Bank ID4D and Govtech Singapore (2022). National Digital Identity and Government Data Sharing in Singapore. A Case Study of Singpass and APEX.

Ying et al (2012). Demographic Prediction Based on User's Mobile Behaviors. Available at: https://www.idiap.ch/project/mdc/publications/files/mdc-final241-ying.pdf

Yoti (2024). Facial estimation whitepaper. Available at: https://www.yoti.com/wp-content/uploads/2024/11/Yoti-Age-Estimation-White-Paper-September-2024-PUBLIC.pdf

Zheyan, S., Liu, J., He, Y., Zhang, X., Xu, R., Yu, H., and Cui, P. (2021). Towards Out-Of-Distribution Generalization: A Survey.

Footnotes

1. Ofcom. Quick guide to children's access assessments. [16 January 2025 updated version]. Available at: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/quick-guide-to-childrens-access-assessments/ ↩

2. A more comprehensive description of these techniques is provided in Annex A2.2.1. ↩
3. Biometric Update.com (September 2024). Facial age verification, age estimation coming of age across sectors. Available at: https://www.biometricupdate.com/202409/facial-age-verification-age-estimation-coming-of-age-across-sectors [Last accessed 03/04/2025]. ↩
4. Age Check Certification Scheme (2023). Measurement of Age Assurance Technologies, Part 2 – Current and short-term capability of a range of Age Assurance measures. A Research Report for the Information Commissioner's Office (ICO) and the Office of Communications (Ofcom). ↩
5. European Commission (2024). Research report: Mapping age assurance typologies and requirements, ↩
6. Sas, M., and Muehlberg, J. T. (2024). Trustworthy Age Assurance? A risk-based evaluation of available and upcoming age assurance technologies from a fundamental rights perspective. Report for the Greens/EFA in the European Parliament. ↩
7. Unicef (2021). Digital Age Assurance Tools and Children's Rights Online across the Globe: A Discussion Paper. ↩
8. Dahlan, H. A. (2021) A Survey on Deep Learning Face Age Estimation Model: Method and Ethnicity. International Journal of Advanced Computer Science and Applications; West Yorkshire Vol. 12, Iss. 11. ↩
9. Anda, F., Le-Khac, N., and Scanlon, M. (2020). DeepUAge: Improving Underage Age Estimation Accuracy to Aid CSEM Investigation. Forensic Science International: Digital Investigation. ↩
10. Replay attacks in the context of facial age estimation involve an attacker attempting to deceive the system by presenting previously captured or pre-recorded images or videos of a person's face, rather than a live and real-time facial input. This type of attack aims to bypass the system's live detection mechanisms and falsely verify the user's age or identity. ↩
11. Korshunov, P., George, A., Özbulak, G., and Marcel S. (2024). Vulnerability of Face age Verification to Replay Attacks. Published in 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). ↩
12. Panic, N., Marjanovic, M., Bezdan, T. (2024). Addressing Demographic Bias in Age Estimation Models through Optimized Dataset Composition. Mathematics 2024, 12, 2358. https://doi.org/10.3390/ math12152358 ↩
13. NIST (2024 - includes updates as of 11/04/2025). Face Analysis Technology Evaluation: Age Estimation and Verification. NIST Interagency Report NISTIR

8525. Face Analysis Technology Evaluation: Age Estimation and Verification. ↩

14. The MAE measures the average difference between the predicted age and the actual age of an individual. It is a common metric used to assess the accuracy of AE models. ↩
15. Yoti (2024). Facial Age Estimation Whitepaper. ↩
16. Retail of Alcohol Standards Group (RASG) (2023). RASG Sandbox Evaluation. Available at: https://rasg.org.uk/wp-content/uploads/2023/11/RASG-Sandbox-Evaluation.pdf ↩
17. The demographic analysis of ethnicity in the NIST evaluation was done using country-of-birth as a proxy for ethnicity. Only countries for which birth is a more or less reliable indication of ethnicity were included. To increase sample sizes and make it easier to report validation results, countries were further grouped into regions. ↩
18. European Digital Rights (EDRi) (2023). Position Paper: Age verification can't 'childproof' the internet. Available at: https://edri.org/our-work/policy-paper-age-verification-cantchildproof-the-internet/ ↩
19. According to the latest 2021 Census for England and Wales, 86.5% of usual residents held at least one passport. See: Dual citizens living in England and Wales - Office for National Statistics ↩
20. CitizenCard - UK Photo ID card & Proof Of Age ↩
21. Retail of Alcohol Standards Group (RASG) (2023). RASG Sandbox Evaluation. Available at: https://rasg.org.uk/wp-content/uploads/2023/11/RASG-Sandbox-Evaluation.pdf ↩
22. Quick guide to children's access assessments - Ofcom ↩
23. Biometric Update.com (September 2024). Facial age verification, age estimation coming of age across sectors. Available at: https://www.biometricupdate.com/202409/facial-age-verification-age-estimation-coming-of-age-across-sectors [Last accessed 03/04/2025]. ↩
24. The UK Digital identity and attributes trust framework ↩
25. Synthetic data is artificially generated data that mimics real-world data (in this case facial images of real people) but does not directly come from actual data collection. Interviews with academics and technical experts suggest that at the moment the use of synthetic data to create larger representative database of facial images of data subjects that are underrepresented in the existing data is challenging. It could be used to introduce different type of variations to facial images, e.g. angle changes or potentially changes to facial expressions. ↩

26. Best practice typically involves splitting a dataset into 70% for training, 20% for internal validation, and 10% for external validation. Under this structure, a representative internal validation set would generally require twice as many individual data subjects as an external validation set, and the training set would require seven times as many. If a dataset originally intended for external validation is expanded to independently support training and internal validation as well, the total number of required data subjects would typically increase tenfold to maintain representative and independent subsets. ↩
27. Face Analysis Technology Evaluation (FATE) Age Estimation & Verification ↩
28. The legal assessment was undertaken for option 1 – which was the solution initially explored and discussed with key stakeholders. However, the key considerations and challenges outlined by the legal assessment are all applicable to option 2. ↩
29. The legal assessment was undertaken for option 1 (also outlined in Figure 3) in section 3.1.4. – which was the solution initially explored and discussed with key stakeholders. However, the key considerations and challenges outlined by the legal assessment are all applicable to option 2. ↩
30. In this context, an API black box is a secure testing environment where FAE models are submitted for evaluation via an API, allowing them to be tested against the database without granting developers access to the underlying data. Only performance results are returned, ensuring data confidentiality and standardised assessment. ↩
31. Ly S, Reyes-Hadsall S, Drake L, Zhou G, Nelson C, Barbieri JS, Mostaghimi A. (2023). Public Perceptions, Factors, and Incentives Influencing Patient Willingness to Share Clinical Images for Artificial Intelligence-Based Healthcare Tools. ↩
32. Laura Sexton, L., Moreton, R., Noyes, E., Castro Martinez, S., and Laurence, S. (2023). The effect of facial ageing on forensic facial image comparison. ↩
33. 5Rights Foundation (2021). But how do they know it is a child?. ↩
34. Government funds new tech in the fight against online child abuse - GOV.UK ↩
35. UK groups call for government digital ID plan overhaul, limits for Gov.uk ↩
36. Following the roadmap/decision tree provided in section 3.2 outlining key steps required to undertake an assessment of the viability and proportionality of a potential facial database solution ↩
37. More information on the Proposal can be found in LE's 'Milestone 3' document. ↩
38. Article 4 of the UK GDPR uses the term "natural or legal person" but for the

purposes of this memo the term organisation will be used instead. ↩

39. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/controllers-and-processors/controllers-and-processors/how-do-you-determine-whether-you-are-a-controller-or-processor/ ↩
40. Article 5(1)(a) of the UK GDPR. ↩
41. Or Article 14 of the UK GDPR where personal data is not collected directly from the data subject. ↩
42. Article 5(1)(b) of the UK GDPR. ↩
43. Article 5(1)(c) of the UK GDPR. ↩
44. Article 5(1)(d) of the UK GDPR. ↩
45. Article 5(1)(e) of the UK GDPR. ↩
46. Article 5(1)(f) of the UK GDPR. ↩
47. Article 6 of the UK GDPR. ↩
48. Article 9 of the UK GDPR (and additionally) ↩
49. Article 4(11) of the UK GDPR. ↩
50. Planet 49 [2020] 1 CMLR 25. ↩
51. Leave.EU v Information Commissioner [2021] UKUT 26 (AAC). ↩
52. Article 7(1) of the UK GDPR. ↩
53. Recital 38 of the UK GDPR. ↩
54. Recital 58 of the UK GDPR. ↩
55. In Scotland the children aged 12 or over are generally able to give consent for data protection purposes. ↩
56. “any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services” – Article 1(b) of Directive (EU) 2015/1535. ↩
57. Age Check Certification Scheme (2023). Measurement of Age Assurance Technologies, Part 2 – Current and short-term capability of a range of Age Assurance measures. A Research Report for the Information Commissioner’s Office (ICO) and the Office of Communications (Ofcom). ↩
58. European Commission (2024). Research report: Mapping age assurance typologies and requirements, ↩
59. Sas, M., and Muehlberg, J. T. (2024). Trustworthy Age Assurance? A risk-based evaluation of available and upcoming age assurance technologies from a

fundamental rights perspective. Report for the Greens/EFA in the European Parliament. ↩

60. Unicef (2021). Digital Age Assurance Tools and Children's Rights Online across the Globe: A Discussion Pape. ↩
61. Voice prints can be categorised as behavioural elements in some definitions of biometric techniques. However, these are to be distinguished from the behavioural patterns and traits that behavioural profiling and inference techniques look at to estimate user age, which focus on user internet activity and interactions instead. ↩
62. 5Rights foundation (2021). But how do they know it is a child? Age Assurance in the Digital World. ↩
63. Meta (2021). How Do We Know Someone Is Old Enough to Use Our Apps? Available at: https://about.fb.com/news/2021/07/age-verification/ ↩
64. University of Cambridge (2013). Digital records could expose intimate details and personality traits of millions. Available at: https://www.cam.ac.uk/research/news/digital-records-could-expose-intimate-details-and-personality-traits-of-millions ↩
65. Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Available at: https://www.pnas.org/doi/pdf/10.1073/pnas.1218772110 ↩
66. Ying et al (2012). Demographic Prediction Based on User's Mobile Behaviors. Available at: https://www.idiap.ch/project/mdc/publications/files/mdc-final241-ying.pdf ↩
67. VoCO (2020). VoCO (Verification of Children Online) Phase 2 Report. Available at: https://assets.publishing.service.gov.uk/media/5faa9cffd3bf7f03a841cfc2/November_VoCO_report_V4__pdf.pdf ↩
68. Hilton, Z., and King, H. (2023). Making age assurance work for everyone: inclusion considerations for age assurance and children. ↩
69. There seems to be a lack of literature comparing the degree of dataset bias affecting the different statistical age estimation methods. Facial age estimation is sensitive to dataset bias because it relies on physical characteristics and the appearance of age, which vary largely across individuals of different groups. However, voice, capacity testing, and behavioural profiling methods are also vulnerable to dataset bias. Voice analysis can be susceptible to cultural biases in speech for example. Capacity testing and behavioural profiling could be (unintentionally) designed to favour certain cultural, educational, or socioeconomic

backgrounds. The degree of bias of each technique is dependent on the quality and representativeness of the data it is trained and tested on. ↩

70. Dahlan, H. A. (2021) A Survey on Deep Learning Face Age Estimation Model: Method and Ethnicity. International Journal of Advanced Computer Science and Applications; West Yorkshire Vol. 12, Iss. 11. ↩
71. Panic, N., Marjanovic, M., Bezdan, T. (2024). Addressing Demographic Bias in Age Estimation Models through Optimized Dataset Composition. Mathematics 2024, 12, 2358. https://doi.org/10.3390/ math12152358 ↩
72. Zheyan, S., Liu, J., He, Y., Zhang, X., Xu, R., Yu, H., and Cui, P. (2021). Towards Out-Of-Distribution Generalization: A Survey. ↩
73. Ganel, T., Sofer, C., and Goodale, M.A. (2022). Biases in human perception of facial age are present and more exaggerated in current AI technology. Scientific Reports 12.1, p. 22519. ↩
74. Voelkle, M. C., Ebner, N. C., Lindenberger, U. & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. Psychol. Aging 27, 265–277. ↩
75. Clapes, A., Bilici, O., Temirova, D., Avots, E., Anbarjafari, G., and Escalera, S. (2018). From Apparent to Real Age: Gender, Age, Ethnic, Makeup, and Expression Bias Analysis in Real Age Estimation. Computer Vision and Pattern Recognition (CVPR) Workshop Paper. ↩
76. eSafety Commissioner (2023). Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography. Available at: https://www.esafety.gov.au/about-us/consultation-cooperation/age-verification#roadmap-andbackground-report ↩
77. Replay attacks in the context of facial age estimation involve an attacker attempting to deceive the system by presenting previously captured or pre-recorded images or videos of a person's face, rather than a live and real-time facial input. This type of attack aims to bypass the system's live detection mechanisms and falsely verify the user's age or identity. ↩
78. Korshunov, P., George, A., Özbulak, G., and Marcel S. (2024). Vulnerability of Face age Verification to Replay Attacks. Published in 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). ↩
79. Fieldfisher (12 January 2024). Age assurance: a modern coming of age approach to ensure the safety of children online and an age appropriate experience. Available at: https://www.fieldfisher.com/en/insights/age-assurance-a-modern-coming-of-age-approach-to-ensure-the-saftey-of-children-online ↩
80. Gov.au (15 November 2024). Tender awarded for age assurance trial. Available

at: https://www.infrastructure.gov.au/department/media/publications/tender-awarded-age-assurance-trial ↩

81. Online Safety Amendment (Social Media Minimum Age) Bill 2024. Available at: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r7284 ↩
82. European Digital Rights (EDRi) (2023). Position Paper: Age verification can't 'childproof' the internet. Available at: https://edri.org/our-work/policy-paper-age-verification-cantchildproof-the-internet/ ↩
83. See https://www.tiktok.com/legal/page/global/age-appropriate-experiences/en ↩
84. One stakeholder mentioned that the age inference profiling models used by big tech are proprietal, with little or no data shared about efficacy, accuracy or time it takes to complete the checks. This is a concern as the services are likely to be processing a significant amount of children's data while their age inference models are running. ↩
85. See https://about.fb.com/news/2021/07/age-verification/ ↩
86. See https://about.fb.com/news/2022/06/new-ways-to-verify-age-on-instagram/ ↩
87. One stakeholder suggested that data sets would also need to be continually refreshed to match the capabilities of camera's used by the market. Higher quality camera's in smart phones should enable better image resolution to support the FAE. ↩
88. Schroff, F. (2021). Synthetic Faces to Improve Privacy and Fairness. European Data Protection Supervisor. ↩
89. UK Health Data Research Alliance (2020). Trusted Research Environments (TRE)- A strategy to build public trust and meet changing health data science needs. The Trusted Research Environments Green Paper. Available at: https://zenodo.org/records/4594704 ↩
90. Office for National Statistics (2017). The 'Five Safes'- Data Privacy at ONS. Available at: https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/ ↩
91. HDR UK (2020). A national health data research capability to support COVID-19 research questions ↩
92. DARE UK (2022). Paving the way for a coordinated national infrastructure for sensitive data research A summary of findings to date from Phase 1 of the UK Research and Innovation DARE UK programme. Available at: DARE_UK-Paving_the_way_coordinated_national_infrastructure_sensitive_data_research-Aug2022.pdf ↩

93. DARE UK (2024). £18 million for DARE UK to support secure research on sensitive data. Available at: https://dareuk.org.uk/news-and-events/18-million-for-dare-uk-to-support-secure-research-on-sensitive-data/ ↩

94. DARE UK (2024). DARE UK Phase 2 is underway, but what does this mean for everyone? Available at: https://dareuk.org.uk/news-and-events/dare-uk-phase-2-is-underway-but-what-does-this-mean-for-everyone/ ↩

95. The Alan Turing Institute. (2022). Review of Digital Research Infrastructure Requirements for AI. Available at: https://www.turing.ac.uk/sites/default/files/2022-09/ukri-requirements-report_final_edits.pdf ↩

96. Carlini, N., Liu, C., Erlingsson, U., Kos, J., Song, D. (2019). The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks. arXiv:1802.08232 ↩

97. Feldman, V. (2021). Does Learning Require Memorization? A Short Tale about a Long Tail. Google Research, Brain Team. ↩

98. Information Comissioner's Office (2022). Regulatory Sandbox Final Report: Yoti. Available at: https://ico.org.uk/media/for-organisations/documents/4020427/yoti-sandbox-exit_report_20220522.pdf ↩

99. Yoti (2024). Yoti Facial Age Estimation White Paper. Available at: https://www.yoti.com/wp-content/uploads/2024/11/Yoti-Age-Estimation-White-Paper-September-2024-PUBLIC.pdf ↩

100. Home Office (2022). Call for proposals. Available at: https://www.gov.uk/government/publications/age-verification-technology-in-alcohol-sales-regulatory-sandbox/call-for-proposals ↩

101. Home Office (2022). Details of the trials. Available at: https://www.gov.uk/government/publications/age-verification-technology-in-alcohol-sales-regulatory-sandbox/details-of-the-trials ↩

102. Retail of Alcohol Standards Group (RASG) (2023). RASG Sandbox Evaluation. Available at: https://rasg.org.uk/wp-content/uploads/2023/11/RASG-Sandbox-Evaluation.pdf ↩

103. UK Gov (2021) Government funds new tech in the fight against online child abuse. Available at: https://www.gov.uk/government/news/government-funds-new-tech-in-the-fight-against-online-child-abuse ↩

104. DSIT (2024) Evaluation of Safety Tech Challenge Fund Round 2. Available at: https://assets.publishing.service.gov.uk/media/6707a7a3e84ae1fd8592f125/safety_tech_challenge_fund_2_evaluation.pdf ↩

105. Remediation processes involve processes required to resolve issues when an age assurance service fails to function as expected or delivers incorrect results. Large variation in remediation processes across age assurance providers can occur due to differences in verification standards, whereby one service might use facial recognition, while another relies on document scanning, requiring unique fixes for failures; error handling, whereby different providers can have different approaches to handling disputes and errors, e.g. some systems might offer self-service remediation, while others require manual intervention, causing inconsistent workflows. Integration processes refer to processes relevant to incorporating an age assurance system into an organization's existing infrastructure (e.g., websites, apps, platforms). Again, large variation in integration processes can happen because of different APIs and Protocols, whereby each service might use unique APIs, requiring custom development for each integration, diverse data requirements with varying legal and compliance needs. ↩

106. Family Online Safety Institute. (2023). Coming to Terms with Age Assurance. ↩

107. Gov.UK (2023). UK digital identity and attributes trust framework beta version (0.3). Available at : https://www.gov.uk/government/publications/uk-digital-identity-and-attributes-trust-framework-beta-version/uk-digital-identity-and-attributes-trust-framework-beta-version ↩

108. See: https://euconsent.eu/ ↩

109. The age token is essentially a cryptographic data token, i.e. a string of characters, that confirms a user's age status and consent without exposing any personally identifiable information. ↩

110. BiometricUpdate.com (October 4 2024). The EUDI Wallet was not meant for age assurance: AVPA . Available at: https://www.biometricupdate.com/202410/the-eudi-wallet-was-not-meant-for-age-assurance-avpa ↩

111. The digital age of consent is 16 under GDPR laws, with flexibility to decrease this to a minimum of 13 by member states. See: https://euconsent.eu/digital-age-of-consent-under-the-gdpr/ ↩

112. euConsent (2022). Pilot Execution Report. Available at: https://euconsent.eu/download/pilot-execution-report-first-large-scale-euconsent-pilot/ ↩

113. The ACCS is potentially the only one offering this. ↩

↑ Back to top

Help us improve GOV.UK

To help us improve GOV.UK, we'd like to know more about your visit today. Please fill in this survey (opens in a new tab).

Services and information

Benefits

Births, death, marriages and care

Business and self-employed

Childcare and parenting

Citizenship and living in the UK

Crime, justice and the law

Disabled people

Driving and transport

Education and learning

Employing people

Environment and countryside

Housing and local services

Government activity

Departments

News

Guidance and regulation

Research and statistics

Policy papers and consultations

Transparency

How government works

Get involved

Money and tax

Passports, travel and living abroad

Visas and immigration

Working, jobs and pensions
