Document Sensitivity: Public

D-2.9 Evaluation Proposal

06/02/2025

Dr Asad Ali, Dr Koliya Wedanage, Adrian Ugray, George Billinge, Dr Mark Pedersen, Surya Ramessh

Work Package: **WP2 - Context, Design & Validation**
Task Reference: **T2.9 – Evaluation Proposal**

Funded by

Australian Government
Department of Infrastructure, Transport, Regional Development, Communications and the Arts

Project by

Age Check Certification Scheme

Document Control Information

Settings	Value
Document Title:	D-2.9 Evaluation Proposal
Work Package:	**WP2 - Context, Design & Validation**
Tasking Reference:	AATT/WP2/T-2.9/D-2.9
Document Author:	Dr Asad Ali, Dr Koliya Wedanage, Adrian Ugray, George Billinge, Dr Mark Pedersen, Surya Ramessh
Work Package Lead:	Asad Ali & Koliya Wedanage
Task Leader:	Asad Ali & Koliya Wedanage
Doc. Version:	0.3
Sensitivity:	Public
Date:	06/02/2025

Document Approver(s) and Reviewer(s):

All Approvers are required. Records of each approver must be maintained.
All reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Asad Ali / Koliya Wedenage	Task Leader (always required)	Approved	31-01-2025
Asad Ali / Koliya Wedenage	Work Package Leader (always required)	Approved	03-01-2025
Keith Robinson	Finance Director (only required if allocation of funding or expenditure is required)	N/A	
Danielle Bradbury	Risk Management (only required if new risks identified)	N/A	
Andrew Hammond	Deputy Project Director (optional)	Reviewed	02-01-2025
Tony Allen	Project Director (optional)	Reviewed & Approved for Publication	02-02-2025
Prof. Toby Walsh	External Reviewer	Reviewed & Validation Statement	26-01-2025
eSafety Commissioner	External Reviewer	Reviewed	12-01-2025
OAIC	External Reviewer	Reviewed	17-01-2025

Configuration Management: D2.9 Evaluation Proposal

The latest version of this controlled document is stored in D2.9 Evaluation Proposal

Document history:

The Document Author is authorized to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting and spelling
- Clarification

To request a change to this document, contact the Document Author or Project Owner. Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Revision by	Short Description of Changes
0.1	25-11-2024	Asad Ali	Added Table of Contents for review
0.2	16-12-2024	Asad Ali	First draft for internal review
0.3	29-12-2024	Asad Ali	Second draft addressing all comments and including all sections except statistical design
0.4	02-01-2025	Project Team	Various revisions and additional material incorporated into the document from feedback from the project team
1.0	03-01-2025	Tony Allen	Major Version Created, comments resolved and ready for submission for external validation
1.1	17-01-2025	Asad Ali	Updated to address all comments from all external reviewers
1.2	31-01-2025	Asad Ali	Updated following Stakeholder Advisory Board review
2.0	31-01/2025	Tony Allen	Addition of Final Validation Statement from Prof. Toby Walsh and Approval to send for Print

Table of Contents

Validation Statement

The project plan (D-6.1) included a requirement that the project team submit the evaluation proposal for independent validation by Professor Toby Walsh of the University of New South Wales, this is his independent assessment.

The proposal does a very good job of scoping out a trial to evaluate the effectiveness of age assurance technologies in Australia. The proposal is especially strong with respect to: (1) the comprehensive evaluation criteria; (2) addressing evidence gaps; (3) explicit ethical principles; (4) a standards-based approach; (5) a commitment to open scientific reporting; (6) and recognition of children's rights.

I identified a few minor issues in the initial draft where I recommended some attention such as addressing combinations of age assurance methods, sample sizes for minority groups, and child friendly project outputs (given this group will be directly impacted by age assurance).
All these issues have been adequately addressed in the final evaluation proposal.

In summary, the trial has been scoped out well and looks set to deliver high quality results on the capabilities of age assurance technologies. I commend the work that the team has put in so far.

Professor Toby Walsh
Scientia Professor of Artificial Intelligence at the University of New South Wales
FAA FAAAI FAAAS FACM FEurAI FRSN

Toby Walsh is Scientia Professor of Artificial Intelligence at the University of New South Wales in Sydney and CSIRO's Data61. He is the winner of the prestigious Celestino Eureka Prize for Promoting Understanding of Science and was named on the international "Who's Who in AI" list of influencers. He has given over 100 talks on the subject of AI to a wide range of audiences, from industry associations, conferences for the general public, corporate events, board meetings, policy meetings and many others.

As well as speaking at leading companies and stages globally like CeBIT, the World Knowledge Forum, World Summit AI and TEDx, Toby also appears regularly on TV and radio, has been profiled by the New York Times and has authored four books on AI for a general audience, the most recent ones entitled "Machines Behaving Badly" and "Faking It: Artificial Intelligence in a Human World" (Fall 2023). He is a Fellow of the Australia Academy of Science and was named by the newspaper The Australian as one of the "rock stars" of Australia's digital revolution. He has won both the Humboldt Prize and the NSW Premier's Prize for Excellence in Engineering and ICT. His Twitter account was voted in the top ten to follow to keep abreast of developments in AI.

1. Introduction

The purpose of this Age Assurance Technology Trial ("the trial"; "AATT") is to assess the effectiveness of a wide range of age assurance (AA), parental consent and parental control technologies available in the market, against a broad range of criteria including accuracy, interoperability, reliability, ease of use, minimisation of bias, protection of privacy and data security. The AATT will not design, implement or endorse any particular technology or policy position.

Digital age assurance technologies are increasingly being mandated by governments[1] and deployed in both online and offline contexts to keep people, particularly children, safe. Some of their uses include preventing users below a certain age from accessing harmful online content, providing children with age-appropriate online experiences or restricted children's access to goods like knives or alcohol.

Age assurance refers to various methods which are used to determine a person's age online, encompassing age verification, age estimation and age inference technologies[2]. These are defined as follows:

- Age verification relies on calculating the difference between a verified year or date of birth of an individual, typically from an identity document and a subsequent date (e.g. the date when the age check takes place).
- Age estimation analyses the biological or behavioural features of humans that vary with age, such as their face or voice.
- Age inference is based on verified information which indirectly implies that an individual is over or under a certain age or within an age range, for example holding a credit card implies that the holder is 18 years old or above in Australia.

Other child protections measures include parental consent and parental controls[3]:

- Parental Consent: Consent from a person holding parental authority over a child. Critical to this is the procedure to validate the authenticity of the parental consent. It also usually, but not always, contains the procedure to attest to the age of the child.
- Parental Control tools: These allow an adult responsible for a person under the age of 18 a degree of control over what content the child can access. Parental control systems can be applied at the network or device level or through linking accounts between the child and parent or caregiver.

[1] AV around the world – AVPA

[2] International Organisation for Standardization (ISO), "ISO Draft International Standard (DIS) 27566-1:2025 Age Assurance Systems- Part 1: Framework," 2025. Available https://www.iso.org/standard/88143.html

[3] S. Smirnova, S. Livingstone and M. Stoilova, "Understanding of user needs and problems: a rapid evidence review of age assurance and parental controls," 11 2021. Available: https://eprints.lse.ac.uk/112559/

The design, evaluation and implementation of age assurance and other child protection measures raise several ethical issues and so strong ethical principles must underpin any research in this area. The AATT will be delivered in line with a set of guiding ethical principles: respect, transparency, accountability, fairness, privacy and safeguarding children.[4] These principles will be operationalised through the delivery of several activities contained within Work Package 1: Data, Ethics and Impartiality. In addition, activities across all work packages will be subject to scrutiny from the project Ethics Committee and, where necessary, an independent impartiality panel.

This document is the culmination of Work Package 2 for the Trial. This covers understanding the specific context in use of age assurance technology in the Australian context, including online safety, privacy and digital identity legislation, the Australian Signals Directorate's Information Security Manual[5] and the specific programme requirements. It includes a literature review of research relevant to the evaluation of age assurance technologies from domestic and international sources.

This document covers the design and development of a standardised and replicable evaluation process using ISO/IEC 25040 – Systems and software quality requirements and evaluation[6]; applying the five core characteristics identified in ISO/IEC DIS 25766 – Age assurance systems – Part 1: Framework and the specific indicators identified in IEEE 2089.1 – Online age checking systems[7] and the Software Engineering test design methods in ISO/IEC 29119[8]. All these practices and standards are captured within the Age Check Certification Scheme's existing ISO 17065[9] accreditation.

1.1. Previous Studies

The trial is considering a potential evidence gap regarding how well age assurance, parental consent and parental controls systems work on their own and in combination. Various recent research studies have evaluated these systems theoretically[10] [11] or focused on gathering user perceptions of them[12] [13]. The National Institute of Standards and Technology (NIST) Face Age Technology Evaluation (FATE) Age Estimation testing[14] is a prominent example of independent and

[4] See A-1.1.1 Ethics Handbook.
[5] Information Security Manual (ISM) | Cyber.gov.au
[6] ISO/IEC 25040:2024 - Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality evaluation framework
[7] IEEE SA - IEEE 2089.1-2024
[8] ISO/IEC/IEEE 29119-1:2022 - Software and systems engineering — Software testing — Part 1: General concepts
[9] ISO/IEC 17065:2012 - Conformity assessment — Requirements for bodies certifying products, processes and services
[10] M. Sas and J. T. Mühlberg, "Trustworthy Age Assurance?," The Greens Cluster: Social & Economy. In The European Parliament., 2024.
[11] M. R. Shaffique and S. van der Hof, "Research report: Mapping age assurance typologies and requirements.," Directorate-General for Communications Networks, Content and Technology, Better Internet for Kids (BIK), European Commission, 2024.
[12] "Questions, doubts and hopes. Young people's attitudes towards age assurance and the age-based restriction of access to online pornography.," The eSafety Commissioner, 2023.
[13] "Public perceptions of age verification for limiting access to pornography.," The eSafety Commissioner, 2021.
[14] National Institute of Standards and Technology, "Face Analysis Technology Evaluation (FATE) Age Estimation & Verification", Available: https://pages.nist.gov/frvt/html/frvt_age_estimation.html

transparent technical testing in this space. It assesses the accuracy and potential bias of age estimation by analysing an image of a face (face age estimation). It also uses a fully automated testing approach in a laboratory environment, rather than real-world conditions. Furthermore, the test dataset used consists primarily of images from USA immigration visas, arrest mugshots, border crossings and immigration offices. These differ to images captured when users are browsing online services or physical locations like retail stores.

Another research study is Enex TestLab's assessment[15], which evaluates a range of age assurance systems both theoretically and technically. In this research, the technical evaluation focused on two age assurance systems, using their publicly available demos hosted online, with 14 participants.

Driven in part by the current evidence base and pilots conducted elsewhere, the Australian eSafety Commissioner's Roadmap for Age Verification[16] includes a key recommendation that age assurance technologies should be trialled in Australia. In the Australian Government's response[17], the roadmap's recommendation of conducting a pilot was noted. Following this, the government announced[18] that it would 'provide resourcing to conduct a pilot of age assurance technology to protect children from harmful content, like pornography and other age-restricted online services. This pilot is the AATT.

The UK's online safety and data protection regulators, Ofcom and the ICO, have conducted and published research to explore age assurance technologies and practices. Their work investigates how these methods verify user ages, assess their accuracy and effectiveness and ensure compliance with privacy standards. This research informs guidelines for responsible deployment, balancing safeguarding and user privacy. It aims to support organisations in implementing age assurance mechanisms that align with legal and ethical standards, particularly in digital and online environments. Their work includes:

- **Families' attitudes towards age assurance**: This research explores the perspectives of children and parents on age assurance methods, balancing considerations such as privacy, online safety and ease of use[19].
- **Measurement of Age Assurance Technologies**: A technical study providing insights into the accuracy levels achievable by different age assurance systems, prompting further reflection

[15] Enex TestLab, "Age Verification Technology Evaluation," Appendix 8, Appendices to the Background Report - AV Roadmap. The eSafety Commissioner, 2023.
[16] The eSafety Commissioner, "Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography," 2023.
[17] Australian Government, "Australian Government response to the Roadmap for Age Verification," August 2023. Available: https://www.infrastructure.gov.au/department/media/publications/australian-government-response-roadmap-age-verification
[18] Prime Minister of Australia, "Tackling Online Harms." Available: https://www.pm.gov.au/media/tackling-online-harms.
[19] https://www.gov.uk/government/publications/families-attitudes-towards-age-assurance-research-commissioned-by-the-ico-and-ofcom

on measuring their overall effectiveness. This work was carried out by the Age Check Certification Scheme team[20].

- **How online businesses are using age assurance**: This study examines how organizations implement age assurance measures, the drivers and challenges they face and their future plans in this area[21].

The European Commission is advancing age assurance through the **Better Internet for Kids+ (BIK+) strategy**[22], which aims to create a safer digital environment for children, promoting age-appropriate content and experiences. Complementing this, standardization efforts, including work by the European Telecommunications Standards Institute (ETSI), focus on developing technical guidelines to harmonize age verification practices across Europe[23]. These initiatives seek to balance robust protection for minors with user privacy and accessibility standards in online services.

1.2. Problem statements

The AATT sets out to address three key problems:

1. Overreliance on theoretical evaluation: most age assurance assessments rely on theoretical evaluation, lacking insight into their quantitative performance in real-world and test laboratory environments.
2. Lack of comprehensive technical evaluation: existing technical assessments have focused on a limited set of age assurance systems. Therefore, there is a lack of evidence for the effectiveness of the broader range of systems available in the market.
3. Underrepresentation of Australian subpopulations in studies: existing studies performing technical assessments often fail to incorporate statistically significant sample sizes that represent the diverse subpopulations within Australia, including Aboriginal and Torres Strait Islander peoples as well as multi-ethnic diverse communities. This limits applicability and relevance to the Australian context.

A structured approach to the research and analysis of existing age assurance and other child protection technologies, taking the above problem statements into account, will support policy makers in developing evidence-based approaches to reducing online harms.

1.3. Aim

The Department of Infrastructure, Transport, Regional Development, Communications and the Arts has commissioned this study. In the original tender, the Department set requirements to:

[20] https://www.drcf.org.uk/publications/papers/measurement-of-age-assurance-technologies
[21] https://ico.org.uk/media/about-the-ico/documents/4030926/20240704-ico-age-assurance-report.pdf
[22] https://digital-strategy.ec.europa.eu/en/policies/strategy-better-internet-kids
[23] https://www.etsi.org/committee/1400-hf

a) evaluate the maturity, effectiveness and readiness for use of available age assurance technologies that determine whether a user is 18 years of age or over, to permit the user to access age-restricted online content; and
b) evaluate the maturity, effectiveness and readiness for use of available age assurance technologies that determine the age of a user in the 13-16 years age band, to permit the user to create an account on a social media website or application.

We recognise that since the award of this tender, the Australian Parliament[24] has passed the Online Safety Amendment (Social Media Minimum Age) Act 2024. As a result of this change, which sets an age limit of 16 for social media platforms, the trial will include consideration of age assurance technologies that determine whether a user is 16 years of age or over.

The evaluation will focus on criteria including, but not limited to:

- Accuracy (how well the technology can detect a user's age)
- Interoperability (how well the technology can be used across multiple online platforms)
- Reliability (how consistently the technology can produce the same result)
- Ease of use (how simple the technology is to operate)
- Minimisation of bias (how well the technology avoids racial or other bias)
- Protection of privacy (how well the technology protects users' personal information, including data minimisation techniques)
- Data security (how well the technology safeguards users' personal information from unauthorised access, breaches or theft through, for example, the use of security by design principles and resistance to presentation attacks)
- Human rights protections (i.e. accessibility for all users, including people with disability, as well as applicable rights under the UN Convention on the Rights of the Child)

The evaluation will consider each age assurance technology in the context of broader requirements and guidance applied in Australia, including:

- eSafety: Safety by Design principles[25]
- Department of Industry, Science and Resources: Voluntary AI Safety Standard[26]
- Information Security Registered Assessors Program (IRAP) assessment[27]

The evaluation is not an assessment of conformity against these, but will note if any of the proposed technologies are likely to be generally incompatible with the underlying principles of such guidance.

[24] Online Safety Amendment (Social Media Minimum Age) Bill 2024 – Parliament of Australia
[25] https://www.esafety.gov.au/industry/safety-by-design
[26] https://www.industry.gov.au/publications/voluntary-ai-safety-standard
[27] https://www.cyber.gov.au/irap

1.4. Out-of-Scope

The Age Assurance Technology Trial is **not** designed to develop new technologies or tools for age assurance. Instead, it focuses on assessing existing systems and their practical applications. It is **not** creating a comparative league table of technologies to rank their performance, nor is it focused on approving or certifying specific technologies for compliance or endorsement. The trial does not aim to prescribe specific methods or mandate their use in regulatory frameworks.

Furthermore, the trial is **not** setting industry standards or defining universal benchmarks for age assurance performance. It does not evaluate the commercial viability of these systems, assess their business models or provide a comprehensive policy recommendation for deployment, although the outcome of the trial may assist policy makers, such as the eSafety Commissioner for Australia, to make evidence-based, informed policy decisions.

The trial's goal is to gather structured, scientific and impartial evidence on the effectiveness, usability and limitations of current technologies. This is to inform future research, development, policy, guidance and guidelines as to how the technologies could effectively work in practice to help Australia achieve its policy and online safety objectives.

2. Literature Review

2.1. Age restrictions in Australia

To define the age restrictions in-scope of the study, it is critical to understand the major age restrictions already in place across Australia, in both online and offline contexts. Age restrictions for various products and services are set out below.

2.1.1 Knives and Controlled Items

In the retail sector, knives and other controlled items are restricted from sale to individuals below a certain age[28]. This varies across states as follows:

State/Territory	Minimum Age for the purchase of Knives and Controlled Items
New South Wales (NSW)	16
Western Australia (WA)	18
South Australia (SA)	16
Victoria (VIC)	18
Australia Capital Territory (ACT)	16
Queensland	18
Northern Territory	No specific legislation applies
Tasmania	No specific legislation applies

2.1.2 Smoking

Across all of Australia, it is illegal to sell or supply tobacco products to people under the age of 18[29]. In some states, the police can confiscate cigarettes or other tobacco products found in the possession of those under 18 years old.

2.1.3 Vaping

Vapes and vaping products, which may or may not contain nicotine, can only be sold by pharmacies to support people quit smoking or managing nicotine dependence[30]. The Public Health (Tobacco and Other Products) Act 2023 was passed to facilitate this on 1st July 2024[31].

[28] National Retail Association, "National Sale of Knife Laws | Comparison Table", Available: https://www.nationalretail.org.au/app/uploads/2024/06/National-Knife-Laws-Comparison-Table-v2.pdf
[29] Department of Health and Aged Care, Australian Government., "Smoking and tobacco laws in Australia", Available: https://www.health.gov.au/topics/smoking-vaping-and-tobacco/about-smoking/laws-in-australia
[30] Department of Health and Aged Care, Australian Government, "New laws for vapes", Available: https://www.health.gov.au/vaping/new-laws
[31] Australian Government, "Public Health (Tobacco and Other Products) Act 2023", Available: https://www.legislation.gov.au/C2023A00118/latest/text

Since 1st October 2024, the minimum age to purchase vapes from participating pharmacies with a nicotine concentration of 20 mg/mL or less without a prescription is 18, where state and territory laws allow.

People under 18 years of age can access vapes but only with a prescription, where state and territory laws allow, to ensure they get appropriate medical advice and supervision.

2.1.4 Alcohol

The minimum legal age for the purchase or consumption of alcohol in a licensed venue or from a retail shop is 18[32].

In some states and territories, it is legal to supply alcohol if you have approval from a child's parent or guardian. In others, it is only legal if you are the parent or guardian.

In all states and territories, it is illegal to supply people under 18 with alcohol if responsible supervision is not provided[33].

Responsible supervision refers to:

- if the adult supplying the alcohol is intoxicated
- if the young person is intoxicated
- the age of the young person
- the type and amount of alcohol supplied and over what period of time
- if the young person has eaten food with the alcohol
- how the young person is supervised by the adult supplying the alcohol[34] [35] [36] [37] [38] [39].

The laws across different states and territories are as follows:

- In the ACT, NSW, SA, TAS, VIC and WA alcohol can be provided to minors in a private home **if**:
    - provided by the parent/guardian or with permission of the parent/guardian
    - provided with responsible supervision.[40]
- In the NT and QLD alcohol can be provided to minors in a private home if:
    - provided by the parent/guardian, step-parent or adult who has parental rights and responsibilities

[32] Department of Health and Aged Care, Australian Government, "Alcohol Laws in Australia", Available: https://www.health.gov.au/topics/alcohol/about-alcohol/alcohol-laws-in-australia
[33] Alcohol and Drug Foundation, "Providing alcohol to under 18s", https://adf.org.au/insights/understanding-secondary-supply/
[34] State of Queensland. Supplying alcohol to under 18s 2022
[35] Government of New South Wales. Alcohol and Young People
[36] Attorney-General's Department. Information for parents of minors 2023
[37] Department of Police Fire and Emergency Management. Youth and Alcohol 2022
[38] State Government of Victoria. Minors and alcohol 2023
[39] Alcohol Think Again. Alcohol laws for under 18s 2022
[40] ACT Policing, "Alcohol – Supply of alcohol to minors", https://police.act.gov.au/community-safety/alcohol-and-drugs/alcohol

- provided with responsible supervision[41].

2.1.5 Gambling

The minimum legal age for gambling is 18 years. There is no single law that sets this universally across Australia, instead this is governed by a combination of federal, state and territory laws. For example, in Victoria, Section 10.7.3 of the Gambling Regulation Act 2003 states that "A gambling provider must not allow a minor to gamble"[42]. This fixed legal age applies to lottery ticket purchasing, sports betting and race betting, bingo, casino games, poker and all real money betting. Sports betting is the only form of online gambling legally permitted to be offered to Australian residents from domestically based sites[43].

2.1.6 Films (including Adult Films) and Computer Games

The Guidelines for the Classification of Film 2012 legislation[44] and the Australian Classification Board (ACB)[45] define the follows advisory ratings and associated age restrictions:

Rating Type	Rating	Description	Age Restriction
Advisory	General (G)	The content is very mild in impact	None
Advisory	Parental Guidance (PG)	The content is mild in impact	Not recommended for viewing by children under the age of 15 without guidance of a parent or guardian.
Advisory	Mature (M)	The content is moderate in impact	Not recommended for children under the age of 15. But they may legally access this content
Legally Restricted	Mature Accompanied (15+)	The content is strong in impact	Minimum age of 15. But a parent or adult guardian must purchase a ticket and accompany a person under 15 for the duration of the film at a cinema or be with them to purchase a MA 15+ film or game
Legally Restricted	Restricted (R 18+)	The content is high in impact	Minimum age of 18 years old
Legally Restricted	Restricted (X 18+)	Adult films containing sexually explicit activity	Minimum age of 18 years old

[41] Northern Territory Government. Young people, alcohol and drugs 2016
[42] Victoria State Government, "Gambling Regulation Act 2003", Available: https://www.legislation.vic.gov.au/in-force/acts/gambling-regulation-act-2003/107
[43] Gamblinglaws.org, "Gambling Laws in Australia 2024", Available: https://www.gamblinglaws.org/au/
[44] Australian Government, "Guidelines for the Classification of Films 2012", Available: https://www.legislation.gov.au/F2012L02541/asmade/text
[45] Australian Classification Board, "What are the ratings?", Available: https://www.classification.gov.au/classification-ratings/what-are-ratings

Legally Restricted	Refused Classification	Contains content that is outside generally accepted community standards and exceeds what can be included in the R 18+ and X 18+ ratings.	Cannot be sold, hired, advertised or legally imported in Australia

2.1.7 Online Safety

Services in scope of the Online Safety Act 2021[46] include search, social media, app distribution, hosting, internet carriage and providers of equipment and operating systems.

The Act defines two classes of material (in the form of text, data, speech, music, other sounds, visual images moving or otherwise, any other form or a combination of forms):

- Class 1 material that is classified or is likely to be classified, as 'RC' under the National Classification Scheme
- Class 2 material that is classified or is likely to be classified, as either X 18+ or R 18+ under the National Classification Scheme.

The Act requires the development of separate industry codes targeting class 1 and class 2 material. Industry codes for specific sectors, addressing class 1 material, include measures related to age restrictions. For example, Compliance Measure 6(c) in Schedule 1 – Social Media Services Online Safety Codes (Class 1A and Class 1B material)[47] requires Tier 1 and Tier 2 social media services to, at a minimum, "take reasonable steps to prevent an Australian child that is known to be under the minimum age permitted on the service from holding an account on the service…". Further, it suggests the implementation of age estimation technology to determine a user's age as an example of a reasonable step.

The draft Phase 2 codes[48] published in October 2024 set out a non-exhaustive list of examples of age assurance measures that will be considered appropriate, which include:

- matching of photo identification;
- facial age estimation;
- credit card checks;
- digital identity wallets or systems;
- attestation by a parent or guardian of age or whether an Australian end-user is a child;

[46] Australian Government, "Online Safety Act 2021", Available: https://www.legislation.gov.au/C2021A00076/latest/text

[47] The eSafety Commissioner, "Schedule 1 – Social Media Services Online Safety Code (Class 1A and Class 1B Material)", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-06/Schedule-1%E2%80%93Social-Media-Services-Online-Safety-Code-%28Class-1A-and-Class-1B-Material%29.pdf?v=1733862184332.

[48] The eSafety Commissioner, "Draft Phase Two Codes - Consolidated Industry Codes of Practice for the Online Industry (Class 1C and 2 Material)", October 2024, Available: https://onlinesafety.org.au/phase-two-codes/

- other measures meeting the requirements of section 8 (Confirmation of age) of the Online Safety (Restricted Access Systems) Declaration 2022; and
- relying upon appropriate age assurance measures implemented in respect of the relevant end-user by: (1) another party (whether another industry participant, a third party vendor or another third party) and

A Restricted Access System is an access-control system that meets the requirements under the Online Safety (Restricted Access Systems) Declaration 2022 (Cth) (RAS Declaration). This sets out the minimum requirements for access-control systems used by social media services, relevant electronic services and designated internet services provided from Australia[49]. Its primary aim is to restrict children's access to R18+ content (i.e. a subset of class 2 material) online, upon receiving a notice from eSafety.

Rather than mandating specific technologies or processes, the RAS Declaration states that an access-control system must:

- require an application be made by a person to access the relevant material, declaring they are at least 18
- incorporate reasonable steps to confirm an applicant is at least 18
- give warnings about the nature of the material and safety information about how a parent or guardian may control access to the material
- limit access to the material unless certain steps are followed

The development of industry codes addressing class 2 material began on 1 July 2024[50] and consultation on draft codes finished in late November 2024. The eSafety Commissioner has not endorsed the draft codes and will undertake an assessment of whether submitted codes meet the statutory requirements after industry submits the codes for registration by February 28, 2025.

The eSafety Commissioner's Roadmap for Age Verification includes a recommendation that if a service allows pornography, "it should apply settings to prevent it from being accessed by and recommended to children. Among other things, this requires robust age assurance measures"[51]. The phase 2 codes position paper[52] provides further, more robust guidance on where and how the eSafety Commissioner expects to see age assurance deployed to reduce the risk of children's access to pornography. These include providing protections across every level of the technology stack, such as age assurance measures, filters, parental controls, safety settings and others. Another

[49] The eSafety Commissioner, "Development of Phase 2 Industry Codes under the Online Safety Act – eSafety position paper", July 2024, Available: https://www.esafety.gov.au/sites/default/files/2024-07/Development-of-Phase-2-Industry-Codes-under-the-Online-Safety-Act-eSafety-position-paper_0.pdf

[50] The eSafety Commissioner, "Industry codes and standards," Available: https://www.esafety.gov.au/industry/codes

[51] The eSafety Commissioner, "Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-08/Roadmap-for-age-verification_2.pdf

[52] The eSafety Commissioner, "Development of Phase 2 Industry Codes under the Online Safety Act – eSafety position paper", July 2024, Available: https://www.esafety.gov.au/sites/default/files/2024-07/Development-of-Phase-2-Industry-Codes-under-the-Online-Safety-Act-eSafety-position-paper_0.pdf

element of this guidance is leveraging digital ecosystems for privacy-protecting, dataminimising age assurance and complementary safety measures. eSafety believes that existing internet ecosystems can effectively leverage existing enduser information gathering processes for privacy-protecting, data-minimising age assurance and complementary safety measures. The last element is building on pre-existing regulatory schemes which aim to protect and prevent children from accessing class 2 material. These schemes include the Restricted Access Systems (RAS) declaration, the Basic Online Safety Expectations (BOSE) determination, the Phase 1 Codes, and the Phase 1 Standards.

A minimum age of 13 years old is common for accessing several social media services, including BeReal[53], BlueSky[54], Discord[55], GoodReads[56] and Pinterest[57]. This age limit is specified in their terms of use, but it is currently rarely enforced by any means other than self-declaration. Until recently, this also applied to the most popular social media services in Australia. However, the parliament has passed new legislation in November 2024, known as the Online Safety Amendment (Social Media Minimum Age) Act 2024[58]. This defines 'age-restricted social media platforms' which will have an age limit of 16 and are to including SnapChat, TikTok, Instagram and X, among others. The new laws will come into effect no later than 12 months from passage of the bill.

2.2. Consumer Attitudes

Considering that user journeys will be impacted by the deployment of age assurance technologies, a key focus in the literature is on consumer attitudes of the technology. Prominent studies of consumers in Australia and the UK are described below.

Boichak, Humphry and Hutchinson[59] studied Australian teenagers aged 12-17 and their parents, using focus groups, co-design workshops and a national survey of 1200 participants conducted between 2022 and 2023. Their findings suggest that age verification is generally supported, with 72% of young people and 86% of parents believing that more effective age limits would improve online safety for young people. However, participants think it likely would not work. They considered other approaches as better options to keeping people safe online, including more safety education, face-to-face dialogue and accountability from social media companies. Concerns about data protection and privacy were raised by both children and adults, specifically around sharing identity documents with online services to prove age and the risk of data breaches and leaks of sensitive information.

The Family Online Safety Institute (FOSI) conducted research which includes includes a cross-country comparison of views on age assurance held by parents and children in the US, UK and

[53] https://bereal.com/terms/
[54] https://bsky.social/about/support/tos#who-can-use
[55] https://discord.com/terms/#2
[56] https://www.goodreads.com/about/terms
[57] https://policy.pinterest.com/en/terms-of-service
[58] Australian Government, "Online Safety Amendment (Social Media Minimum Age) Bill 2024", Available: https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r7284
[59] J. Humphry, O. Boichak and J. Hutchinson, "Emerging Online Safety Issues – Co-creating Social Media with Young People – Research Report," The University of Sydney, Sydney, 2023.

France[60]. This consisted of two parts, firstly a qualitative part with seventy-one parents and children across the US, UK, and France. The second part was a quantitative survey of 3000 parents and children evenly distributed across the three countries. The report has 10 key findings:

1) Parents are highly engaged with their children's digital lives and are invested in facilitating a safe, positive online experience.
2) Children, like parents, want safe and positive online experiences, and children understand that parents monitor online activity with good intentions.
3) Parents see themselves as having the most responsibility for managing their children's access to age-appropriate content, more so than technology companies or the government.
4) Even as parents feel this strong sense of responsibility, they also want more involvement from relevant partners to help safeguard their children.
5) Children also desire an active role in the processes that will shape their digital lives, even if they are not always comfortable discussing their online activities with parents.
6) Age assurance is seen by parents and children as being more about restricting access to content, rather than ensuring safe and beneficial online experiences.
7) There is no clear 'winner' or standout approach when respondents are asked about their preference for current age assurance methods.
8) This ambivalence appears to come down to a question of balancing invasiveness vs. effectiveness.
9) The applied use of biometrics appears to be a promising method of age assurance, as parents and children view it as effectively assessing age.
10) Parents seek age assurance solutions that are effective yet convenient, and they gravitate toward settings that achieve both.

The eSafety Commissioner researched the attitudes of young people towards age assurance and the age-based restriction of access to online pornography[61]. This mixed methods research comprised of an online survey of 1004 participants followed by online focus groups, both with young people aged 16-18. One key finding was that participants were generally in favour of the regulation of online pornography for people under the age of 16. However, they thought that age assurance would be of limited efficacy and expressed concerns about its implementation. Despite this, the young people surveyed thought that pornography sites, dating sites and social media services should use age assurance tools to restrict underage access to online porn.

[60] Family Online Safety Institute (FOSI), "Making Sense of Age Assurance: Enabling Safer Online Experiences", 2022, Available: https://cdn.prod.website-files.com/5f47b99bcd1b0e76b7a78b88/636d13257232675672619f45_MAKING%20SENSE%20OF%20AGE%20ASSURANCE%20FULL%20REPORT%20-%20FOSI%202022_compressed.pdf

[61] The eSafety Commissioner, "Questions, doubts and hopes. Young people's attitudes towards age assurance and the age-based restriction of access to online pornography", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-08/Questions-Doubts-and-Hopes.pdf

Another study by the eSafety commissioner focused on adults' perceptions of age verification for limiting access to pornography[62]. Using a survey of 1200 adults, the study identified broad support for age verification as a safeguard for children, even though the general community was unfamiliar with it conceptually and in practice. There was ambivalence and scepticism on how the technology would work in practice. Participants were concerned about the security and processing of personal data as part of any age verification system. They considered the government was best placed to process their data to ensure data security and to ensure that the system worked in practice. The study identified the need for further communication and education to build knowledge and awareness in several areas. Two of these areas include, firstly, the effectiveness of age verification tools and, secondly, what makes a secure and privacy-preserving tool for ensuring online pornography remains accessible to adults only.

Recently polls have been conducted by Resolve[63] and YouGov[64] to garner public sentiment on the Online Safety Amendment (Social Media Minimum Age) Bill 2024. These polls showed high levels of support for the age restriction, such as 77% of respondents to the YouGov survey. However, the findings also showed doubts whether the age restriction will be efficacious.

In the UK, digital regulators have commissioned several studies into public perceptions of age assurance. One of these is Ofcom's 2022 research into adult user's attitudes to age verification on adult sites[65]. Through an online survey of 2158 adults following by focus groups, it was found that there is broad support for age verification measures to prevent under-18s from accessing online pornography. Also, participants were more likely to accept age verification measures where they expected those measure to be in place. Furthermore, using a credit card was the preferred means for proving age for paid access to pornography. The study also found serious concerns among participants about how user data may be processed and/or stored for the purposes of age verification, with a very low level of trust in the data privacy practices of adult sites. These privacy concerns could be addressed by increased transparency in data management practices, having a choice of methods to verify age and using an independent third-party for the age check rather than the porn sites themselves.

Ofcom commissioned a follow-up survey focused on trust in age assurance measures[66]. They sought to understand the experiences of people aged 16 years old and over accessing pornographic content online and passing any age checks, if encountered, through a survey of 5242 participants. An additional focus was their attitudes towards proving their age on adult sites and the importance of different factors in encouraging them to comply. One key finding was that, of those participants

[62] The eSafety Commissioner, "Public perceptions of age verification for limiting access to pornography", 2021, Available: https://www.esafety.gov.au/research/public-perceptions-age-verification-for-limiting-access-pornography
[63] The Sydney Morning Herald, "Australians like banning teens from social media", 2024, Available: https://www.smh.com.au/politics/federal/australians-like-banning-teens-from-social-media-they-just-don-t-think-it-ll-work-20241210-p5kx9l.html
[64] YouGov, "Support for under-16 social media ban soars to 77% among Australians", 2024, Available: https://au.yougov.com/politics/articles/51000-support-for-under-16-social-media-ban-soars-to-77-among-australians
[65] Ofcom, "Adult Users' Attitudes to Age Verification on Adult Sites", 2022, Available: https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/vsp/attitudes-to-age-verification/2022-adult-attitudes-to-age-verification-adult-sites.pdf
[66] Ofcom, "Barriers to Proving Age on Adult Sites", 2023, Available: https://www.ofcom.org.uk/online-safety/protecting-children/barriers-to-proving-age-on-adult-sites/

that had never been asked to prove their age to access pornographic content online before, 29% said they would comply with the age check and 55% said they would leave the site at the age check. Of those who choose not to prove their age to access pornographic content online:

- 86% had personal data concerns,
- 24% did not think the age check would be accurate and/or reliable.

Ofcom and the UK's Information Commissioner's Office (ICO) commissioned a joint study to explore parents' and children's attitudes towards potential age assurance methods[67]. The research used in-depth interviews followed by deliberative focus groups. There are several relevant key findings from this work. Firstly, most parents felt that services should have age assurance measures, but it can sit in tension with their desire for control and flexibility over what their children do online. Secondly, most children had circumvented current age assurance methods themselves (typically self-declaration on social media platforms) or knew someone who had. Thirdly, many families felt the type of platform the age assurance method was being used on was critical context for which method felt the most appropriate. Overall, however, parents and children felt that hard identifiers such as a passport or driving licence were the most effective age assurance method. Fourthly, both parents and children had concerns about the amount of effort required to use methods. Finally, some parents and children raised concerns about the amount of data sharing required.

2.3. Parental Consent and Parental Control tools

Smirnova et. al. conducted a rapid evidence review of parental controls in everyday life[68]. The evidence reviewed as part of this research consisted of studying user experiences and perceptions of these technologies. They found that parents use controls to limit the time children spend online, filter content or restrict access to it, limit the people who can contact their child, switch off device functions and limit access to specific applications and/or websites. They also found that, to function effectively, such technical measures must address the needs of both children and parents. The evidence suggests that parental control measures that were developed based on identifying family values were perceived more positively by their users.

Stoilova et. al. conducted a follow up study using the same evidence base, focusing on analysing the context and outcomes of use of parental control tools[69]. As before, the evidence included consists of research studying user experience and perceptions of the tools. This review of the effectiveness of parental controls reveals mixed results: some uses of parental controls bring benefits, for example to children's safety, but others have no effect or limit children's opportunities and some have adverse results, for example to family communication.

[67] Ofcom and The ICO, "Families' attitudes towards age assurance", 2022, Available: https://www.gov.uk/government/publications/families-attitudes-towards-age-assurance-research-commissioned-by-the-ico-and-ofcom

[68] S. Smirnova, S. Livingstone and M. Stoilova, "Understanding of user needs and problems: a rapid evidence review of age assurance and parental controls", 2021, Available: https://eprints.lse.ac.uk/112559/

[69] M. Stoilova, M. Bulger and S. Livingstone, "Do parental control tools fulfil family expectations for child protection? A rapid evidence review of the contexts and outcomes of use", 2023, Available: https://doi.org/10.1080/17482798.2023.2265512

In addition to the findings, these studies reveal the need for further functional and non-functional testing of the technical effectiveness of parental consent and parental control tools.

2.4. Children's Rights and Best Interests of the Child

While age assurance technologies are designed to protect children online, they may have a negative impact on children's rights if not implemented carefully. Over recent years, there has been growing awareness of the need to consider children's rights in the digital world.

The United Nations Convention on the Rights of the Child (CRC) was adopted in 1989.[70] It formally codifies children's rights, setting out the freedoms and protections all countries must offer children and young people under 18 years old. It is the basis upon which much domestic legislation in relation to children rests around the world. In 2021, the Council on the Rights of the Child adopted General Comment 25, which makes explicit that children's rights apply in the digital world and provides guidance on how to apply children's rights to the digital world.[71] General Comment 25 (GC25) sets out four principles to guide the implementation of all other rights under the Convention. These are summarised below:

A. **Non-discrimination.** All children should have equal and effective access to the digital environment in ways that are meaningful to them and States parties should take all measures necessary to overcome digital exclusion. States parties should take proactive measures to prevent discrimination, including on the basis of sex, disability, socioeconomic background, ethnic or national origin, language and discrimination against minority or Indigenous children. GC25 also explicitly calls for the prevention of discrimination against children deprived of liberty or in other vulnerable situations.

B. **Best interests of the child.** The best interests of the child is a dynamic concept that requires an assessment appropriate to the specific context. The digital environment was not originally designed for children, yet it plays a significant role in children's lives. In considering the best interests of the child, states parties should have regard for all children's rights, including the rights to seek, receive and impart information, to be protected from harm and to have their views given due weight and ensure transparency in the assessment of the best interests of the child.

C. **Right to life, survival and development.** Opportunities provided by the digital environment play an increasingly crucial role in children's development. Risks relating to content, contact, conduct and contract encompass, among many other things, violent and sexual content, cyberaggression and harassment, gambling, exploitation and abuse, including sexual exploitation and abuse and the promotion of or incitement to suicide and other life-threatening activities. States parties should

[70] UN General Assembly, "Convention on the Rights of the Child," 1989, Available: https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child.

[71] UN Committee on the Rights of the Child, "General comment No. 25 (2021) on children's rights in relation to the digital environment," CRC/C/GC/25, 2021. Available: https://digitallibrary.un.org/record/3906061/files/CRC_C_GC_25-EN.pdf.

identify and address the emerging risks that children face in diverse contexts, including by listening to their views on the nature of the risks that they face.

D. **Respect for the views of the child.** Children reported that the digital environment afforded them crucial opportunities for their voices to be heard in matters that affected them. The use of digital technologies can help to realize children's participation at the local, national and international levels. States parties should involve all children during the development of legislation or during the undertaking of other activities, which affect them. They should ensure that digital service providers actively engage with children, applying appropriate safeguards and give their views due consideration when developing products and services.

In addition to the above principles, GC25 highlights the importance of respecting the **evolving capacities of the child** as an enabling principle that addresses the process of their gradual acquisition of competencies, understanding and agency. This process has particular significance in the digital environment, where children are mor likely to be unsupervised.

On the adoption of GC25, the 5Rights Foundation, who acted as the Chair of the Steering Committee on General Comment 25, published a child-friendly version of the document.[72] This provides a summary of GC25 in accessible language, highlighting that the digital world must take children's ages into account when providing for their needs.

Livingstone, Nair, Stoilova, van der Hof & Caglar combined legal and social research methods to comprehensively examine the legal, technical and practical challenges of age assurance from the perspective of children's rights.[73] The authors find that current approaches vary in effectiveness and are often ineffective, with different approaches being taken in different sectors. They call for clear guidelines on the security, transparency and inclusiveness of age assurance methods with specific attention to those methods which utilise artificial intelligence. The authors advocate for a child rights-respecting approach to age assurance, emphasising the need for privacy-preserving approaches and consideration of children's evolving capacities.

In their analysis of age assurance in the lives of children and families, they find that little research has asked how age assurance is managed in the domestic context. Given the diversity of families, it is unlikely that age assurance will have uniform impacts in all domestic contexts. Where policymakers rely on parental management of children's digital activities, outcomes are likely to be iniquitous given parents' different resources, competencies, etc. However, there is little evidence exploring the use of age assurance across diverse family groups at present. Age assurance could ease the task parents face of raising children in a rapidly changing digital world. However, the authors argue that it is ultimately not clear whether age assurance can be designed in ways that respect children's rights holistically without stimulating new and creative workarounds.

Having said this, the authors go on to highlight that to safeguard children's rights in the digital environment, it is necessary to know when users are likely to be children, unless digital services are

[72] 5Rights Foundation, "In our own words: Young people's version of General Comment No. 25 on children's rights in relation to the digital environment," 2024. Available: https://5rightsfoundation.com/wp-content/uploads/2024/09/In-our-own-words-young-peoples-version-of-GC25.pdf.

[73] S. Livingstone, A. Nair, M. Stoilova, S. van der Hof and C. Caglar, "Children's rights and online age assurance systems: The way forward," *The International Journal of Children's Rights*, vol. 32, no. 3, pp. 721–747, 2024. Available: https://doi.org/10.1163/15718182-32030001.

made appropriate for all ages by design. Age assurance may be necessary to safeguard children's rights, but age assurance tools themselves should be designed in a way that protects children's rights. They find that although age assurance remains controversial, there are strong grounds for age assurance as a norm, together with privacy-by-design and that it is plausible that age assurance could be designed in ways that respects children's rights.

The UK Council for Internet Safety (UKCIS) Digital resilience Working Group, chaired by Vicki Shotbolt and Dr Richard Graham, was established to develop and coordinate a digital resilience strategy aiming to enable the development of digital skills, emotional understanding and effective responses to online problems. The Working Group produced the Digital Resilience Framework and an online hub to support the dissemination, application and development of a resilience-based approach to online safety.[74] The framework suggests that supporting the development of digital resilience is an effective way to ensure children are safer online. Resilience adapts and grows through activation, in response to context, experience and learning. The authors stress that this does not mean children should simply be expected to cope with bad situations, which would lead to a toxic environment. Instead, children should be supported to understand when they are at risk online, know how to seek appropriate help, learn from their experience, adapt their future choices, and recover when things go wrong by receiving the appropriate level of support.

2.5. Effectiveness of Age Assurance

The US National Institute of Standards and Technology (NIST) Face Age Technology Evaluation (FATE) Age Estimation is the most prominent example of technical testing of age assurance systems[75]. It provides a comparison of the accuracy of a range of face age estimation systems, submitted by providers on a voluntary basis up to four times a year. It also describes the variance of accuracy of each system across different demographics / features, including gender and skin tone, to assess bias in performance. However, their evaluation focuses solely on testing the criteria of accuracy and bias, of one kind of age assurance: face age estimation. It also uses a fully automated testing approach which does not provide insight into real-world performance and the user experience. Another limitation is that the dataset used for testing consists of images from USA immigration visas, arrest mugshots, border crossings and immigration offices. These differ to images taken when users are browsing online services or physical retail stores.

In 2024, two significant theoretical evaluations – without any technical testing – were conducted on a range of theoretical approaches to age assurance. These evaluations did not assess the actual systems that are available on the market.

[74] UK Council for Internet Safety, "Digital Resilience Framework," 2019. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/831217/UKCIS_Digital_Resilience_Framework.pdf.

[75] National Institute of Standards and Technology, "Face Analysis Technology Evaluation (FATE) Age Estimation & Verification", Available: https://pages.nist.gov/frvt/html/frvt_age_estimation.html

Sas and Mühlberg[76] reviewed 9 age assurance methods, including self-declaration, age declaration coupled with email confirmation, vouching, AI profiling, biometric analysis, capacity testing, hard identifiers, digital identities and proxies for official documentation. They also considered age proof transmission methods, including direct collection by online service providers, connection with a third-party account and variations of using age tokens. Age-token approaches included age tokens directly transmitted to service providers, the "double-blind" method, age tokens on a centralised digital wallet, age tokens on decentralised wallets, age tokens on the user's terminal and age tokens at browser-level. These were all assessed for the likelihood of occurrence of several risks of age assurance:

- User identification
- Loss of online anonymity
- Privacy intrusion
- Commercial profiling
- Victim targeting
- Identity theft
- Data fraud
- Restriction of user's autonomy
- Restriction of user's fundamental rights
- Exclusion and marginalisation
- Biases and inaccuracy
- Feasibility challenges
- Circumvention

The assessment was performed using interviews with researchers, civil society organisations, age assurance system providers and regulatory authorities. This was followed by desk research and analysis of relevant legal frameworks and age assurance literature.

The report highlights the potential privacy and inclusivity concerns of age assurance systems and the impacts of their deployment on user's fundamental rights. However, promising avenues are also identified, such as privacy-preserving techniques using double-blind transmission methods. The scope of this study did not include a technical evaluation of age assurance technologies available in the market today. Risks such as biases and inaccuracy, exclusion and marginalisation and feasibility challenges require a technical assessment of system in the market.

Shaffique and van der Hof also performed a theoretical evaluation of age assurance methods[77]. Using desk research and analysis, ten main methods of age assurance were assessed: (1) Self-declaration; (2) Hard identifiers; (3) Credit cards; (4) Self-sovereign identity; (5) Account holder

[76] M. Sas and J. T. Mühlberg, "Trustworthy Age Assurance?" The Greens Cluster: Social & Economy. In The European Parliament., 2024.

[77] M. R. Shaffique and S. van der Hof, "Research report: Mapping age assurance typologies and requirements," Directorate-General for Communications Networks, Content and Technology, Better Internet for Kids (BIK), European Commission, 2024.

confirmation; (6) Cross-platform authentication; (7) Facial age estimation; (8) Behavioural profiling; (9) Capacity-testing; and (10) Third-party age assurance service. Ten key requirements of age assurance systems were also defined and explored: proportionality, privacy, security, accuracy, functionality, inclusivity, participation, transparency, notification mechanisms and considering the child's perspective. The study found that age assurance is a complex matter, including that different age assurance methods have different strengths and weaknesses. For example, there is sometimes a tension between the level of assurance in the user's age and the level of privacy, where methods that may offer a higher level of assurance may require further personal data from the user. Another factor is that some methods may be easy to use, but that may also make them easy to circumvent. It is, therefore, difficult to determine the right approach to age assurance for any given situation. As with the previous report, this study also does not include any technical testing. Therefore, some requirements considered in this study require further investigation through technical testing of systems in the market to provide robust research evidence. Some of the relevant requirements include accuracy, inclusivity, functionality and participation.

Another prominent assessment of age assurance technologies is Enex TestLab's 2022 evaluation commissioned by the eSafety Commissioner[78]. It is notable for including technical testing of two age assurance systems, alongside theoretical evaluation of these and several other systems. The sample size used for the technical testing was 15 participants, using publicly available demo versions of age assurance technologies, which may not be as effective as production versions. These limitations may impact confidence in some findings from the study.

In the UK, Ofcom's draft guidance (due to be updated in January 2025) for services providers publishing pornographic content[79] defines four key technical criteria and two principles for effective age assurance. Alongside the criteria, they provide a theoretical evaluation of age assurance systems that justify them.

The technical criteria are:

- Technical Accuracy: how an age assurance method can correctly determine the age of a user under test lab conditions.
- Robustness: the degree to which an age assurance method can correctly determine the age of a user in unexpected or real-world conditions.
- Reliability: the degree to which the age output from an age assurance method is reproducible and derived from trustworthy evidence.
- Fairness: the extent to which an age assurance method avoids or minimises bias and discriminatory outcomes.

[78] The eSafety Commissioner, "Appendix 8: Independent assessment of age assurance and safety technologies, Roadmap for age verification - background report", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-08/Appendices-to-background-report.pdf

[79] Ofcom, "Guidance for service providers publishing pornographic content - Consultation on draft guidance on age assurance and other Part 5 duties", 2023, Available: https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/272586-consultation-guidance-for-service-providers-publishing-pornographic-content/associated-documents/consultation-guidance-for-service-providers-publishing-pornographic-content-online/?v=368673

The principles are:

- Accessibility: this covers the principles that the age assurance should (a) be easy to use and (b) work effectively for all
- Interoperability: re-using the result of an age check across multiple services allowing different providers of age assurance methods to share this information in line with data privacy laws.

These criteria and principles align with characteristics, criteria and requirements discussed elsewhere in the literature. The draft guidance specifically raises the lack of evidence of the effectiveness of age assurance systems in practice (whether deployed in lab environments or production environments).

Two important studies address the measurements of age assurance systems. The UK Information Commissioner's Office (ICO) commissioned part 1 of a study into the criteria and metrics required to assess the effectiveness of age assurance techniques and to understand the potential for consistency, comparability and standardisation of measurement[80]. Key metrics were defined and classified by the type of age assurance output produced; either continuous where the output is the age computed by the system or binary where the output is either 'yes' or 'no' indicating that the age meets a threshold or not. For continuous approaches, metrics include:

- Mean Absolute Error (MAE): The central value of the absolute errors (i.e. difference between output age and real age, ignoring whether it is higher or lower) of the sample.
- Standard Deviation (SD): The amount of variation or spread over the distribution of absolute errors in the sample.

For binary approaches, the metrics include

- True Positive Rate (TPR): the sensitivity of the technology's ability to correctly detect people who are over the age threshold.
- False Positive Rate (FPR): the technology's probability of false alarm (i.e., incorrectly identifying someone as being over the age threshold).
- Positive Predictive Value (PPV): the proportion of the sample correctly identified as being over the age threshold given that they have been predicted as being over the age threshold.

The ICO study did not cover measurement of the overall effectiveness of age assurance systems. It instead focused primarily on the criterion of accuracy, while also describing some considerations for inclusion, security and privacy.

Follow-up research was commissioned by Ofcom and the ICO to further explore the measurement of accuracy levels achievable by different age assurance systems[81]. The research provided self-

[80] The UK Information Commissioner's Office (ICO), "Measurement of Age Assurance Technologies", 2022, Available: https://ico.org.uk/media/about-the-ico/documents/4021822/measurement-of-age-assurance-technologies.pdf

[81] ICO and Ofcom, "Measurement of Age Assurance Technologies Part 2 – Current and short-term capability of a range of Age Assurance measures", 2023, Available: https://www.drcf.org.uk/publications/papers/measurement-of-age-assurance-technologies

reported accuracy levels for a range of age assurance technologies from providers that participated in the study. The study also proposed the expression of a single 'headline' accuracy metric for all age assurance systems, which should be accompanied by further specific metrics including MAE, TPR, FPR, etc. Future work was identified to establish ranges of accuracy by performing technical assessment of age assurance systems. This research also has certain limitations, including:

- As companies' accuracy figures were self-declared, the study's hypothetical indicators of confidence should be considered illustrative.
- Technical accuracy is only one dimension of age assurance systems and the report does not cover other aspects of the overall effectiveness of these technologies.
- Difficulties in acquiring adequate, independent data sets for testing age assurance technologies pose an additional challenge for measuring technical accuracy.

2.6. The eSafety Commissioner's Call for Evidence Responses, Consultation and Cross-Sector Workshops

As part of its work on age assurance, The Australian eSafety Commissioner conducted a call for evidence[82], two rounds of stakeholder consultations[83]and several cross-sector workshops[84].

According to these pieces of eSafety research, more than three in four Australian adults support government implementation of age assurance for online pornography. However, there are concerns about effectiveness, privacy, and security. These themes – and concerns about accessibility, fairness and bias – were echoed by young people and multi-sector stakeholders.

These considerations informed the development of assessment criteria which an independent test lab applied to review age assurance technologies available on the market, including biometric (age and voice) estimation and identity document (ID)-based tools. They also reviewed a recent European age assurance pilot and international standards for age assurance.

The independent assessment found the age assurance market is immature but developing. Each technology has benefits and trade-offs. For example, ID-based solutions can provide a high level of certainty but risk excluding those without access to ID, whereas facial estimation technology is promising but may offer a lower level of certainty, and may vary in accuracy based on skin tone,

[82] The eSafety Commissioner, "Appendix 2: Call for evidence request and thematic summary of submissions, Roadmap for age verification - background report", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-08/Appendices-to-background-report.pdf?v=1733502339463

[83] The eSafety Commissioner, "Appendix 5: Consultation Summaries, Roadmap for age verification - background report", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-08/Appendices-to-background-report.pdf?v=1733502339463

[84] The eSafety Commissioner, "Appendix 7: Cross-sector workshop summary, Roadmap for age verification - background report", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-08/Appendices-to-background-report.pdf?v=1733502339463

gender, and physical differences. Consumer choice to select the option users are comfortable with, and which works for them, is a key lesson from the European pilot.

For these and other reasons explored in these pieces of research, age assurance technologies should be trialled in Australia, based on lessons from pilots conducted elsewhere, before being mandated. While eSafety should be involved in the development, implementation, and evaluation of any such pilot, they did not at the the time have the resources, capabilities, or expertise to lead its delivery.

2.7. Australia Signals Directorate (ASD): The Information Security Manual (ISM)

The Australian Signals Directorate (ASD) produces the Information Security Manual (ISM)[85]. The ISM is a cyber security framework that an organisation can apply, using their risk management framework, to protect their information technology and operational technology systems, applications and data from cyber threats. It is a set of guidelines published by the Australian Cyber Security Centre (ACSC) to help organizations protect their information and systems from cyber threats. It is primarily targeted at Australian government agencies and the software systems they operate and/or procure from suppliers.

The ISM draws from National Institute of Standards and Technology (NIST) Special Publication (SP) 800-37 Rev. 2, Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy[86]. Broadly, the risk management framework used by the ISM has six steps: define the system, select controls, implement controls, assess controls, authorise the system and monitor the system.

The manual also defines several Cyber Security Principles, followed by several Cyber Security Guidelines, with the latter subdivided to target specific organisational roles, processes or infrastructure elements. These subdivisions contain guidance and recommended controls to implement the cyber security principles.

2.8. Australia's Digital ID System

Australia's Digital ID System[87] is relevant to age assurance systems, as a digital identity may be able to prove an individual's age. Although an individual's age is an attribute of their identity, it is not necessarily the case that establishing the full identity of an individual in a global context is needed to gain age assurance[88]. In many cases, establishing an individual's age should not establish their

[85] Australia Signals Directorate (ISM), "The Information Security Manual (ISM)", Available: https://www.cyber.gov.au/resources-business-and-government/essential-cyber-security/ism

[86] National Institute of Standards and Technology (NIST), "NIST SP 800-37 Rev. 2 Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy", 2018, Available: https://csrc.nist.gov/pubs/sp/800/37/r2/final

[87] Australian Government, "Australia's Digital ID System," Available: https://www.digitalidsystem.gov.au/

[88] ISO/IEC DIS 27566 Introductory Text

full identity. As such, while the process of age assurance may in some instances be connected to identity verification, it can also be performed in ways other than via identity verification. The approach to assessment of age assurance systems can, therefore, be informed by existing approaches to the assessment of digital identity systems, such as that specified in Australia's digital ID system.

This system is made up of two parts: (a) the voluntary accreditation scheme for digital ID service providers and (b) the Australian Government Digital ID System.

The voluntary accreditation scheme is open to all government and private sector digital ID service providers across the economy. Accreditation demonstrates that a provider meets strict rules and standards for:

- privacy protection
- security
- usability
- accessibility
- risk management
- fraud control and more.

These services are also subject to additional privacy safeguards. These are set out in law, with civil penalties for non-compliance. Accreditation is only mandatory if a provider wants to join the Australian Government Digital ID System.

Accredited providers will be able to display a Trustmark. This shows they provide accurate, trustworthy digital ID services that protect people's personal information and that they are regulated.

Since 1 December 2024, the Digital ID Regulator is responsible for the Accreditation Scheme and accrediting digital ID providers.

The Australian Government Digital ID System is designed to provide a secure, convenient and voluntary way for people to verify who they are online. It is delivered and supported by six agencies:

- Department of Finance
- The System Administrator
- Services Australia
- The Australia Competition and Consumer Commission
- Office of the Australian Information Commissioner
- The Treasury's Data Standards Body
- Australian Taxation Office

Some accredited providers are participating in the Australian Government Digital ID System, thus allowing their use to access government online services.

There are also many government services who use the system to verify their users. By December 2026 private businesses will be able to apply to join the Australian Government Digital ID System.

The eSafety Commissioner stated[89] that before the use of specific age assurance technologies is prescribed, stakeholders reported to them that measures need to be in place to alleviate concerns about privacy and security, and also satisfy the implementation factors raised above. These include the need for independent oversight, strong governance, transparency, trustworthiness, fairness, and respect for human rights. To promote international harmonisation, this work should be aligned with relevant international standards which are in place or under development.

There is substantial work already well underway to develop such a framework for Australia's Digital Identity System. The Australian Government should build on this work to establish a similar regulatory accreditation regime to the Trusted Digital Identity Framework for age assurance

Establishment of such a regulatory scheme should include consideration of a strong, independent regulator or accreditation body with functions including:

- accreditation
- compliance and enforcement related to accreditation-enabling capabilities, such as:
  - register of accredited providers
  - application portals for prospective providers
  - any enabling IT infrastructure for the regulatory regime
- general regulatory functions – reporting, publication of guidance etc.

Based on eSafety Commissioner's consultation across government, at this stage, there is likely no existing regulator or accreditation body that has the full breadth of experience and capability to provide all the necessary functions, particularly in relation to this type of digital accreditation. However, building on the work of equivalent accreditation regimes in government such as the Trusted Digital Identity Framework could provide a good basis for starting discovery work on how an accreditation scheme could operate.

In addition to this, further criteria may be required specifically for the accreditation of age assurance systems. This is because the wide array of age assurance systems works in various ways, relying on either a verified date of birth, the analysis of an individual's biological or behavioural features that vary with age or verified information which indirectly implies than an individual is over or under a certain age. This is distinct from digital identity, where only the former option is relevant (i.e. reliance on a verified date of birth). The following additional criteria, as per the literature review and DITRDCA's guidance, are also needed:

[89] The eSafety Commissioner, "Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography", 2023, Available: https://www.esafety.gov.au/sites/default/files/2023-08/Roadmap-for-age-verification_2.pdf

- Accuracy (how well the technology can detect a user's age)
- Interoperability (how well the technology can be used across multiple online platforms)
- Reliability (how consistently the technology can produce the same result)
- Minimisation of bias (how well the technology avoids racial or other bias)
- Human rights protections (i.e. accessibility for all users, including people with disability, as well as applicable rights under the *UN Convention on the Rights of the Child*)

Therefore, digital identity systems accredited under voluntary accreditation scheme may need to be further assessed against the additional criteria. This is an area for further exploration with the agencies responsible for the Digital ID system:

1) Department of Finance
2) The Office of the System Administrator
3) Services Australia
4) The Australia Competition and Consumer Commission
5) The Office of the Australian Information Commissioner
6) The Data Standards Body
7) Australian Taxation Office

3. Materials and Methods

This trial will use primarily quantitative methods, along with supplementary qualitative methods, to assess the systems in scope against the specified criteria. There are four pillars to the project methodology.

The first pillar is the development of an ethical framework, including activities relating to data protection, child safeguarding, impartiality and inclusion of Aboriginal Australians and Torres Strait Islander peoples. This will inform all trial activities and deliverables.

The second is the assessment methodology, including the evaluation criteria and test strategy, which will be derived from and informed by all relevant ISO standards.

The third pillar is stakeholder engagement, including the recruitment of user participants, the recruitment of technology providers, vendor interviews and strategies to inform all relevant stakeholders.

The fourth and final pillar is project management and risk assessment.

The pillars are visualised below and each one is described in detail in the following four sections.

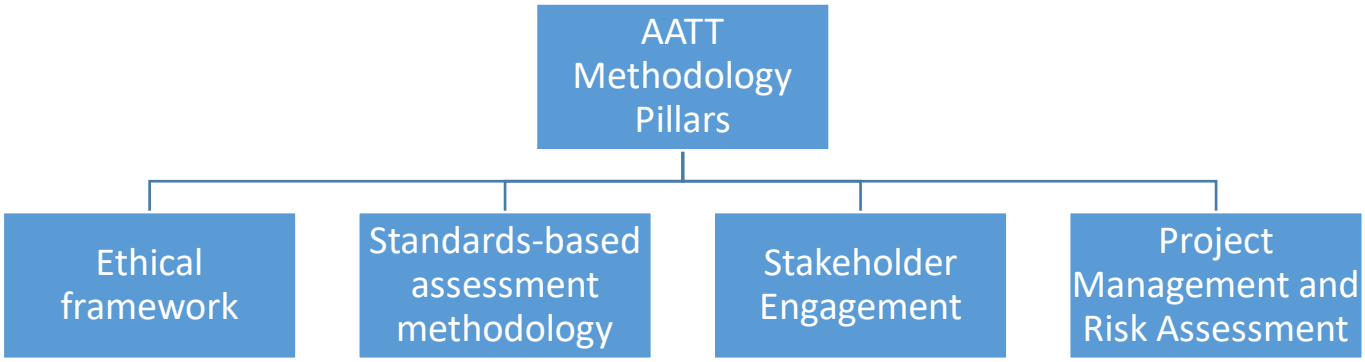

Figure 1: Four pillars of the AATT methodology

4. Test Strategy

4.1. Introduction

4.1.1 Overall Scope

The scope of the test effort includes technologies which provide:

- Age Assurance (AA), including:
  - Age verification (AV)
  - Age inference
  - Age estimation (AE)
- Parental controls
- Parental consent

for the following age gates:

- 18+ age gate
- 16+ age gate
- 13+ age gate

Out of scope:

1. Exact age
2. 13-16 age range (as this is covered by the gates above)

4.1.2 Audience

This test strategy provides guidance to the testing team executing the technology trial and stakeholders wishing to understand the process used in the evaluation.

4.1.3 References

The test strategy for this project is aligned to ISO/IEC 29119 and uses the following standards as reference:

- ISO/IEC 25010:2023 Systems and software Quality Requirements and Evaluation (SQuaRE) - Product Quality Model
- ISO/IEC 25040:2011 Systems and software Quality Requirements and Evaluation (SQuaRE) – Evaluation process
- The ISO 29119 Software and systems engineering – Software testing series
- The ISO/IEC 30107 Biometric Presentation Attack Detection series

- ISO/IEC DIS 27566-1:2025 Age assurance systems – Part 1: Framework
- IEEE 2089.1-2024 IEE Standard for Online Age Verification
- ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts

4.2 Vendor Interviews

Each participant vendor or platform will be invited to participate in a face-to-face or virtual interview with the Age Assurance Technology Trial team. The interview will be structured to elicit the vendor's approach to securing age assurance. Interviews will be on-the-record and recorded, but it is not expected that vendors would be asked to reveal anything that is secret or not openly available on the information pages about their product (even if it might be challenging to find). We may quote extracts from the transcripts in the final report. We intend to share a summary of the transcript with the vendor and, subject to fact checking, intend to make this publicly available.

Vendors that are unwilling or unable to complete a face-to-face interview will be provided with a questionnaire. We intend to publish a list of any vendors that are participants in the trial but declined to participate in the interview process.

Each interview will be bespoke to the vendor, but the types of questions that may be asked include:

Introduction

Thank you for participating in this interview as part of the Australian Age Assurance Technology Trial. The purpose of this session is to gain a deep understanding of your system, how it operates and the safeguards it has in place. We are looking to learn about your approach, capabilities and any limitations to assess your system's suitability and effectiveness in this field.

Section 1: Background and Overview

1. Can you provide an overview of your system and its primary functionality?
2. What is it that you do to ensure age assurance?
3. What inspired the development of your technology and what key problem(s) does it aim to solve?
4. What industries or use cases are currently utilizing your system?
5. Have you carried out any assessments of the impact your technology may have on fundamental rights, including the rights of children?

Section 2: Privacy and Data Protection

6. What types of data does your system collect, process and store?
7. How do you decide what data is collected?
8. For what specific / discrete purposes do you collect data?
9. Have you implemented any specific measures to comply with Australian privacy laws, including the Privacy Act and the Children's Privacy Guidelines?

10. Do you have measures in place to handle data deletion requests or cases where a user withdraws consent for their data to be held and / or processed?
11. Do you employ any anonymization or pseudonymization techniques?
12. Do you have measures in place to limit the impact of your system / technology on the user's digital footprint?

Section 3: Security Measures

13. What security protocols are in place to protect user data, both in transit and at rest?
14. Have you conducted independent security audits or penetration tests? If so, could you share the outcomes or certifications obtained?
15. Do you have measures in place to protect your system from misuse or attacks, such as spoofing or fraudulent attempts to bypass age verification?
16. How quickly can you respond to and address vulnerabilities or breaches?
17. How do you ensure that your system is fail safe?

Section 4: Accuracy and Effectiveness

18. Can you describe the accuracy rates of your system and how they are measured (e.g., false positives/negatives)?
19. How is your system testable?
20. What external validation or third-party testing has been conducted to verify your claims of accuracy?
21. How does your system handle edge cases, such as where an individual's age is estimated as on / just above the particular age threshold?
22. Do you have mechanisms in place to ensure that your system continually improves in accuracy over time?

Section 5: Accessibility and Inclusivity

23. How do you ensure your system is accessible to users with disabilities?
24. How do you ensure your system is accessible to those with limited technical literacy?
25. How do you measure any discrepancies in performance across different demographic groups, and what measures do you take to mitigate these discrepancies?
26. Have you taken any measures to account for the particular cultural context of Australia, such as consulting with First Nations peoples?
27. How does your system address children from looked after care (i.e. who either do not live with or do not have parents)?
28. Do you have mechanisms in place for users to report issues, provide suggestions for improvement, or raise complaints? Could you describe these?

Section 6: Transparency and User Trust

29. What information, if any, do you offer users to help them understand why or how age decisions are made?
30. Do you publish white papers, explainer videos, or other resources in accessible languages?
31. Do you provide users with access to a record of their data and how it was processed?
32. Have you taken steps to communicate the workings of your system to end-users, especially minors and parents, in an understandable way?

33. Are there any user-facing tools or dashboards that provide insights into your system's decisions?

Section 7: Configurability and Scalability

34. How configurable is your system for different use cases or compliance requirements in Australia?
35. Can your system integrate seamlessly with existing platforms and systems, such as e-commerce, gaming or social media sites?
36. How scalable is your system for organizations with varying sizes and levels of user activity?

Section 8: Safeguards and Ethical Considerations

37. What safeguards have you implemented to ensure your system is not overly intrusive or restrictive for users?
38. Can your system operate in a way that respects the autonomy of older teens while still ensuring the safety of younger children?
39. How do you address the evolving capacity of children to make their own decisions?
40. In considering the role of parents, how do you manage the boundaries of individuality of the parents and of their children?
41. Does your system support children to access age appropriate content to develop their own individuality, resilience and communities of interest?
42. How do you manage the ethical expectations of your stakeholders?

Section 9: Research and Continuous Improvement

43. What research has been conducted to develop and validate your system? Are there any peer-reviewed studies or white papers available?
44. How do you stay up to date with advancements in age assurance technologies and ensure your system remains state-of-the-art?
45. Are there mechanisms in place for continuous feedback and iteration based on real-world deployment?

Section 10: Closing and Additional Insights

46. In your opinion, what is the biggest strength of your system compared to competitors?
47. What are the most significant challenges or limitations your system faces?
48. Is there anything we have not covered that you believe is crucial to understanding your system or its role in the Australian Age Assurance Technology Trial?

Conclusion

Thank you for providing detailed insights into your system. This information will contribute to assessing and understanding age assurance technologies in the context of the trial. If we have follow-up questions, would you be open to providing further details?

4.3 Testing Objectives

The technology trial seeks to evaluate whether the technology provided by trial participants, whether individually or in combination, meets the Australian government's requirements for a range of online services to implement age assurance and parental control and consent.

The quality criteria for the technology trial are outlined in the tender and described in Section 4.3.1. The risk assessment and test design processes summarised in this document use ISO/IEC 25010 quality attributes and ISO/IEC 30107 biometric security criteria as points of reference.

The goal of the testing conducted within the technology trial is not intended to replace any detailed software testing process which would normally be undertaken by a company during the acquisition and deployment of any specific age assurance technology which may be selected to promote user safety and/or comply with relevant regulatory frameworks. Rather, the goal is to assess that the capability of technologies available within market at the present time.

4.4 Evaluation Scope and Approach

4.4.1 Evaluation/Quality criteria in scope

1) **Accuracy**: how well the technology can detect a user's age. Assessing the variance of accuracy across different environmental conditions and contexts is in scope to a certain degree, including but not limited to:
   - Some, but not all, lighting conditions
   - Some, but not all, audio conditions
   - Some, but not all, end user devices
   - Some, but not all, operating systems
2) **Interoperability**: how well the technology can be used across multiple online platforms
3) **Reliability**: how consistently the technology can produce the same result
4) **Ease of use**: how simple the technology is to operate, including how the system offers functionality appropriate to the capacity and age of a child or adult, up to and including those of retirement age, who may use the service
5) **Minimisation of bias**: how well the technology avoids racial or other bias, recognising that the complete elimination of bias is unattainable.
6) **Protection of privacy**: how well the technology protects users' personal information
7) **Human rights protections** i.e. accessibility for all users, including people with disability, as well as applicable rights under the *UN Convention on the Rights of the Child*
8) **Data security**: how well the technology safeguards users' personal information from unauthorised access, breaches or theft through, for example, the use of security by design principles and resistance to presentation attacks
9) **Circumvention**: Resistance to certain kinds of attacks – Clause 9 of ISO 27566-1, including:

- Biometric presentation attacks: where individual presents pictures, replayed videos or 3D masks to the system.
- Spoofing attack: when an individual is attempting to try to fool the age estimation method, e.g., by trying to look older than they really are while wearing a hat, glasses, a fake beard or a fake moustache.

10) **Technology Readiness Level (TRL):** According to the Australian Government Department of Defence, the use of TRLs enables consistent, uniform discussions of technical maturity across different types of technology[90]. TRLs are based on a scale from 1 to 9 with 9 being the most mature technology. The New South Wales (NSW) Government's Invest NSW initiative provides a tool to calculate the TRL level for a technology system[91]. The table of TRL levels is set out below.

TRL	Definition
TRL 1	**Basic Research**: Initial scientific research has been conducted. Principles are qualitatively postulated and observed. Focus is on new discovery rather than applications.
TRL 2	**Applied Research**: Initial practical applications are identified. Potential of material or process to solve a problem, satisfy a need or find application is confirmed.
TRL 3	**Critical Function or Proof of Concept Established**: Applied research advances and early stage development begins. Studies and laboratory measurements validate an
TRL 4	**Lab Testing/Validation of Alpha Prototype Component/Process**: Design, development and lab testing of components/processes. Results provide evidence that performance targets may be attainable based on projected or modelled systems.
TRL 5	**Laboratory Testing of Integrated/Semi-Integrated System**: System Component and/or process validation is achieved in a relevant environment.
TRL 6	**Prototype System Verified**: System/process prototype demonstration in an operational environment (beta prototype system level).
TRL 7	**Integrated Pilot System Demonstrated**: System/process prototype demonstration in an operational environment (integrated pilot system level).
TRL 8	**System Incorporated in Commercial Design**: Actual system/process completed and qualified through test and demonstration (pre-commercial demonstration).
TRL 9	**System Proven and Ready for Full Commercial Deployment**: Actual system proven through successful operations in operating environment and ready for full commercial deployment.

[90] https://www.dst.defence.gov.au/sites/default/files/basic_pages/documents/TRL%20Explanations_1.pdf
[91] https://www.investment.nsw.gov.au/assets/Grants-and-rebates/MVP-Technology-Readiness-Level-Tool.xlsx

4.4.2 Criteria Out-of-Scope

1. Volume and stress testing of each system. We will assess responsiveness from a user experience perspective. But it is not feasible to stress test the system to identify, for example, the number of age checks that can be successfully performed per hour.
2. The variance of an age assurance system's accuracy in every possible environmental condition and context is not feasible in the timeframe of the AATT. In-scope environmental conditions and contexts are set out in the previous subsection. As a result:
    a. Some lighting conditions will be out-of-scope. We will test common lighting conditions but not all.
    b. Some audio conditions will be out-of-scope. We will test common audio conditions but not all.
    c. Some end user devices will be out-of-scope. We will test common end user devices but not all.
    d. Some operating systems will be out-of-scope. We will test some common operating systems but not all.
3. Technical testing for protection of privacy, such as the correct implementation of data encryption algorithms and protocols.
4. Technical testing for security mechanisms, such as the correct implementation of de-identification protocols or access controls.
5. Technical testing of security vulnerabilities in the age assurance system, such as penetration testing. This is distinct from the circumvention criterion in which some biometric presentation attack detection tests are in-scope.
6. Technical testing of fraud attacks – distinct from the circumvention criterion listed above – where users make deliberate, concerted efforts to evade the age assurance check which are beyond reasonable expectations for providers to mitigate. Examples include adults assisting a child to provide a fake age by completing the age check on their behalf.

4.4 Evaluation Matrix

The ISO/IEC 25010[92] standard, part of the Systems and Software Quality Requirements and Evaluation (SQuaRE) series, provides a robust framework for evaluating software and system quality. This analysis aligns the standard's product quality model with the specific requirements of the Age Assurance Technology Trial (AATT). By integrating ISO/IEC 25010 into the evaluation design, the trial ensures consistency, precision, and adherence to international best practices.

[92] ISO 25010:2023 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Product quality model, 2023, https://www.iso.org/standard/78176.html

This table below details how ISO/IEC 25010's nine quality characteristics and sub-characteristics are mapped to the trial's core contexts: accuracy, reliability, interoperability, minimisation of bias, ease of use, privacy protection, data security, and human rights protections.

Final Evaluation Criteria	Sub-criteria	Mapping to ISO 25010	Mapping to ISO 25040	Mapping to IEEE 2089.1	Mapping to ISO 27566-1
Accuracy	FPR, FNR, TPR, TNR	Functional correctness	Specification of quality metrics	Accuracy, classification errors	Functional effectiveness
Interoperability	API success rate, data exchange capability	Interoperability	Specification of the evaluation plan	Interoperability	Functional interaction
Reliability	Fault tolerance, MTBF	Reliability	Execution of evaluation tests	Fault recovery	Risk identification
Ease of use	Usability score, accessibility testing	Interaction capability	Quality requirements validation	Accessibility considerations	Interaction inclusivity
Minimisation of bias	Parity analysis, demographic fairness	Interaction capability	Specification of the evaluation plan	Bias mitigation measures	Inclusivity and fairness
Human rights protections	Data minimisation, ethical compliance	Security, flexibility	Evaluation report conclusions	Human rights impact	Ethical protections
Protection of privacy	Encryption, purpose limitation	Security	Specification of quality metrics	Privacy by design	Privacy and confidentiality

Data security	Audit trails, data breach prevention	Security	Execution of evaluation tests	Security effectiveness	Security resilience
Circumvention	Resistance to bypass attempts	Reliability, security	Execution of evaluation tests	System robustness	Functional robustness

Table 1 Mapping ISO 25010 to the trial's evaluation criteria

The **evaluation matrix** for the trial is described in Table 2 below. It is designed to systematically assess systems and software quality within the AATT based on the ISO/IEC 25040[93] standard. The matrix incorporates the product quality model (ISO/IEC 25010) and aligns it with project-specific requirements (e.g., accuracy, privacy, security, and usability). This structure ensures that all critical quality aspects are evaluated comprehensively and objectively.

ISO/IEC 25010 Characteristic	**Sub-characteristic**	**Evaluation Metric**	**Weight (%)**	**Threshold**	**Measurement Method**
Functional Suitability	Completeness	% of age scenarios covered	15	≥ 95%	Test coverage analysis
	Correctness	% of accurate verifications	20	≥ 98%	Accuracy testing
	Appropriateness	% of tasks successfully performed	10	≥ 90%	User testing
Performance Efficiency	Time behaviour	Response time for verification	10	≤ 2 seconds	Load testing
	Resource utilisation	CPU/memory usage during peak operation	5	≤ 75%	System profiling

[93] ISO 25040:2024 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality evaluation framework, 2024, https://www.iso.org/standard/83467.html

Compatibility	Co-existence	Interference with other systems	5	No observable impact	Integration testing
	Interoperability	% of APIs successfully integrated	10	≥ 95%	API testing
Interaction Capability	Ease of use	Average usability score	10	≥ 85%	User surveys, usability testing
	Inclusivity	% of users achieving tasks	5	≥ 90%	Accessibility testing
	Engagement	User satisfaction score	5	≥ 80%	User feedback surveys
Reliability	Maturity	Mean Time Between Failures (MTBF)	10	≥ 1,000 hours	System monitoring
	Fault tolerance	% of fault recovery scenarios passed	10	≥ 95%	Fault injection testing
Security	Confidentiality	% of encryption protocols passing tests	15	≥ 99.9%	Security testing
	Integrity	% of transactions with no data corruption	10	≥ 99%	Data integrity checks
	Accountability	% of audit trails complete	10	≥ 98%	Audit trail analysis
Flexibility	Scalability	Performance with 10x traffic increase	10	No more than 20% degradation	Stress testing

	Adaptability	% of environments supported	10	≥ 90%	Compatibility testing
	Configurability	Time to update configurations	5	≤ 30 minutes	Configuration testing

Table 2 The AATT Evaluation Criteria

4.5 Test Approach

Step 1: Practice Statements

The test approach begins with the 'functional claim' provided by the participating age assuring provider, relying party (that deploys age assurance on their online service) or intermediary (which provides functions to support the age assurance process). This claim will be set out in their **practice statement**[94]**,** which is the documentation of the practices, procedures and controls employed by an organization to fulfil a service. It will include information about how their system works and how it fulfils the evaluation criteria. Assessing the practice statement will lead to the identification of the key areas to test as described in Step 2 below. Information expected to be provided in the practice statement includes (but is not limited to):

1) Age eligibility requirements: which age-related eligibility requirements the system supports.
2) Age assurance components: which age assurance components (e.g. kinds of age verification, age estimation and/or age inference) are used by the system.
3) Binding process: how the system undertakes binding of the age assurance result to the correct individual.
4) Privacy and data protection: how the age assurance provider approaches protecting the privacy of users, including the data protection laws and obligations
5) Ease of use: how the system offers functionality appropriate to the capacity and age of a child or adult, up to and including those of retirement age, who may use the service.
6) Security: how the system addresses security requirements, including secure-by-design principles
7) Human rights protections: the extent to which the system is accessible and inclusive to users, does not unduly restrict access of users who should have access and provides sufficient and meaningful information for a user to understand its operation.

[94] Clause 3.1.13, ISO DIS 27566-1:2025 Age assurance systems – Part 1: Framework

Step 2: Evaluation Process

Assessing the practice statement will lead to the identification of the areas that require evaluation. Using the ISO 29119 Software Testing series, the evaluation process will focus on the following two test levels[95] to structure the test activities required:

- system testing
- acceptance testing

to structure the test activities required. The high-level test types are set out as follows. Detailed, low-level test designs will be created as part of the evaluation activities in Work Package 4.

Automated[96] functional testing[97] will be used to evaluate the accuracy of each participating technology. Where feasible, functional tests will include coverage of presentation attack detection features based on ISO/IEC 30107.

Automated non-functional testing[98] will focus on core quality attributes such as minimisation of bias and reliability.

Automated testing will consist of the following steps, described in further in the next section of this report:

1. Gather inputs e.g., the test dataset,
2. Set up communications between the test lab and the system being tested i.e. target of evaluation,
3. Conduct the tests i.e. construct and perform API calls to the system being tested,
4. Record the results of testing.

Manual[99] usability[100] and acceptance testing[101] will be used to cover features which are not suitable for automation.

Manual functional testing will be used to test the **interoperability[102]** aspects of each technology by a combination of manual tests, such as to confirm that a given technology works on various device platforms. This will also cover the **interoperability of parental consent and parental control systems.** In addition to typical manual functional testing, this will involve the creation of artificial accounts known as 'avatars' where required, for example on mobile device operating systems.

[95] Clause 4.2.4.2, ISO 29119-1:2022 Software testing - Part 1: General concepts
[96] Clause 4.4.7, ISO 29119-1:2022 Software testing – Part 1: General concepts
[97] Annex A.2.5, ISO 29119-4:2021 Software testing – Part 4: Test Techniques
[98] A list of non-functional testing types is set out in Figure 2, Clause 4.2.5, ISO 29119-1:2022 Software testing – Part 1: General concepts
[99] Clause 4.4.7, ISO 29119-1:2022 Software testing – Part 1: General concepts
[100] Annex A.2.15, ISO 29119-4:2021 Software testing – Part 4: Test Techniques
[101] Clause 4.2.4.2, ISO 29119-1:2022 Software testing – Part 1: General concepts
[102] Interoperability testing is described further in Annex A.2.7, ISO 29119-4:2021 Software testing – Part 4: Test Techniques

Static reviews[103] will focus on evaluation of features relating to privacy, data security, compliance with human rights requirements and technology readiness assessment. It will also cover these requirements for **parental consent and parental control systems.** Dynamic testing of these features for each participating technology is beyond the scope of the current trial, however dynamic testing of these features is recommended for any technology being deployed.

Deployments of AA, Parental Consent and Parental Control systems on online services will be tested using a blend of the appropriate test approaches listed above, depending on the criteria to test and type of system deployed.

Figure 2 provides an overview of the overall test approach.

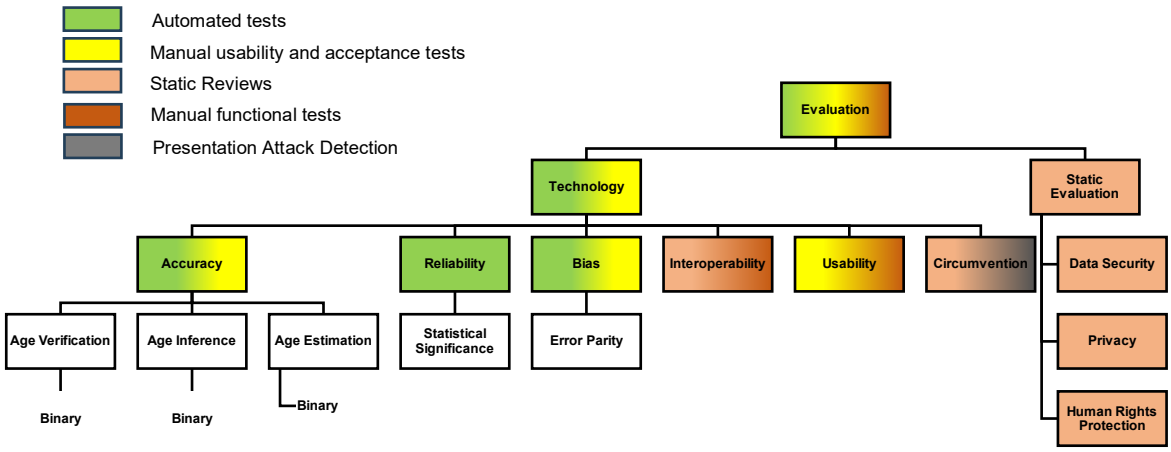

Figure 2: Test approach overview

Figure 3 summarises the application of test activities to relevant effectiveness/quality criteria. For many criteria, two test approaches are applicable and will be used, but one will be considered primary and the other secondary. A test approach is considered primary if it is better suited to fully test the criterion. Choices for primary and secondary approaches are described as follows:

- Accuracy: manual usability & acceptance tests can be used to assess accuracy; however manual testing limits the breadth and depth of coverage in terms of the total number of tests and the number of subgroup divisions of participants that can be tested. Automated testing significantly increases the test coverage which can be achieved; therefore, this will be used to test accuracy.
- Interoperability: static reviews can describe theoretical analysis gained from reviewing technical specifications, but manual functional tests will provide more meaningful results from live testing.

[103] Clause 4.1.5, ISO 29119-1:2022 Software testing – Part 1: General concepts

- Ease of use: static reviews can describe theoretical analysis, but manual usability & acceptance tests will provide more meaningful results from live testing with users.
- Minimisation of bias: manual usability & acceptance tests can be used to assess accuracy, which can then be investigated for any variance amongst subpopulations. However manual testing limits the breadth and depth of coverage in terms of the total number of tests and the number of subgroup divisions of participants that can be tested. Automated testing significantly increases the test coverage which can be achieved; therefore, this will be used to test minimisation of bias.
- Circumvention: static reviews can describe theoretical analysis, but presentation attack detection tests will provide more meaningful results from live testing of age assurance systems.

Each high-level test type is described in the following sub-sections.

Test Approach / **Quality Criteria**	Automated Tests	Manual usability & acceptance tests	Manual functional tests	Static reviews	Presentation Attack Detection
Accuracy	✓	O			
Interoperability			✓	O	
Reliability	✓				
Ease of Use		✓		O	
Minimisation of bias	✓	O			
Protection of privacy				✓	
Human rights protections				✓	
Data Security				✓	
Circumvention				O	✓
Technology Readiness				✓	

✓ = Primary O = Secondary

Figure 3: Effectiveness/Quality criteria test coverage

4.5.1 Automated Testing

Functional & Non-functional

The focus for this aspect of testing will be to confirm that key features of each age assurance technology in scope functions accurately, reliably, minimises bias and is acceptably resilient to presentation attacks when used with the expected Australian usage conditions.

Test automation allows for many tests to be conducted consistently across a range of input data and age assurance systems. This allows for assessment of the accuracy of each system when used in a variety of expected usage conditions by a range of different users from different demographics

groups. In this way, reliability and minimisation of bias can also be assessed. The key evaluation metrics for these tests will be:

- False Positive Rate (FPR),
- False Negative Rate (FNR),
- True Positive Rate (TPR),
- True Negative Rate (TNR),
- Classification Accuracy,
- Outcome Error Parity (OEP).

Important considerations:

- It is only possible to perform automated testing of AA services that can be communicated with via an application programming interface (API) or through a software development kit (SDK).
- Ideally, the AA system would be deployed in the test lab, rather than hosted by the provider. This allows observing the AA system in the controlled environment of the lab. However, installation of the AA system in a self-contained environment may not always be feasible.
    - In the case where the AA system is not deployed in the test lab i.e. it is hosted on external infrastructure such as a cloud service, the test data will leave the perimeter of control of the test lab during the tests. This presents a risk to the project as providers may then train their system on the test data. We must, therefore, enter into a legal agreement with providers that ensures that they will not record and/or store the test data during or after the testing process.
- It is also possible to perform automated testing with the AA system hosted by the provider in their environment. The provider will need provide access from our test lab via an API.
- If AA system is unable to provide suitable interface for automated testing, evaluation may be limited to manual testing and static reviews. This may restrict the ability to properly assess key aspects such as reliability and minimisation of bias, which require many repeated tests.
- Suitable test data is a key requirement for automated testing[104]. This is discussed in more detail in section 4.5.10.
- Techniques such as the use of the Fitzpatrick Scale[105] of skin tones will be used for designing the dataset to sufficiently perform bias testing.
- Generally, each kind of age assurance system will require different test data as follows:

[104] A-1.1.2 Ethical Data Collection Protocol sets out further detail on data minimisation and the categories of data that the AATT will collect.

[105] Fitzpatrick TB. The validity and practicality of sun-reactive skin types I through VI. Arch Dermatol. 1988 Jun;124(6):869-71. [PubMed]

(1) Age Verification: test data will include samples of the verified date of birth (e.g. images of identity documents held by the participant and images of the user that bind them to the identity document) that enables calculating the user's age from that date to the current date.
(2) Age Estimation: test data will include samples of the biological information (e.g. images of the face, audio recording of the voice, video recording of hand gestures or gait, etc.) or behavioural information (e.g. typing patterns, etc.) that vary with age.
(3) Age Inference: test data will include samples of the verified information (e.g. credit card details, etc.) that indirectly implies the individual is over or under a certain age.

4.5.2 Automated testing process

The overall process that will be run for the automated testing of each age assurance system:

1) Gather inputs – create and prepare the test dataset for testing in line with the trial's ethical data collection protocol[106]. Test datasets could include photos, videos, audio files. They will also include demographic information including but not limited to ground truth age, gender, ethnicity, skin tone. Specific data variables for the datasets will be required for each age assurance method – these will be defined once participating providers are confirmed. Example of such data variables include images of photo IDs, sensor quality, image lighting, etc. Test data preparation will include modification of input data to simulate expected usage conditions (e.g. low light, poor quality image capture).
2) Set-up & establish communications between test lab and AA system. This may involve deploying the AA system in the test lab and/or implementing API communications functionality on the AA system.
3) Conduct the tests by constructing & performing API calls to the AA system.
4) Record results of testing, including:
   a) Age output from the AA system.
   b) Time taken to output result.
   c) Data and metadata captured by the AA system.

4.5.3 Presentation Attack Detection

Functional testing will include specific testing of each system's ability to resist circumvention or perform "Presentation Attack Detection" (PAD), including:

1) Presentation attacks: where, for voice age estimation systems as an example, the individual plays a pre-recorded audio clip performed by an adult.

[106] See A-1.1.2 Ethical Data Collection Protocol.

2) Spoofing attack: when an individual is attempting to try to fool the age estimation method, e.g., by trying to look older than they really are while wearing a hat, glasses, a fake beard or a fake moustache.

The key difference is that spoofing attacks involve the actual presence of the individual, but some aspect of their presence is manipulated to deceive the system (e.g., wearing a mask or using altered facial features). In contrast, presentation attacks involve using an external object or "payload," such as a photo, video or fake fingerprint, that does not represent the individual directly.

It is not feasible to physically test (using PAD) all circumvention methods for all AA systems, due to the project's timeline and available resources. Therefore, we will take a risk-based approach, selecting the highest impact circumvention methods per AA method to test**.** As outlined below, we have defined "highest impact" as those methods which are most readily available, according to factors such as the cost and technical literacy required to perform the attack. PAD will be supplemented by on-paper evaluation of circumvention of each AA method that goes beyond the current state-of-the-art as per recent studies[107] [108].

The ISO/IEC 30107 Biometric Presentation Attack Detection series does not categorise different kinds of attacks. It allows testing bodies to define custom categorisation of attacks. This recognises the fact that different kinds of attacks may require different test protocols and resources. Furthermore, the kinds of attacks will develop over time, so providing a categorisation of attacks aids in developing uniform test strategies to cover each category.

There is no consensus or agreed standard for categories of presentation attacks. Two examples of categories include the iBeta Quality Assurance[109] presentation attack detection levels and the FIDO Alliance's spoof presentation attack levels[110]. They differ in the number of levels they define; iBeta define 2 levels whereas FIDO Alliance define 3 levels. But they use the same criteria to categorise, which include the time, expertise and resources required to perform a presentation attack. We will therefore use the same criteria.

The evaluation will include circumvention testing leveraging the extensive experience of the ACCS from its auditing of national identity schemes. The report will note where a solution has had independent testing on its circumvention detection (eg under ISO 30107-3).

[107] M. Sas and J. T. Mühlberg, "Trustworthy Age Assurance?" The Greens Cluster: Social & Economy. In The European Parliament., 2024.
[108] M. R. Shaffique and S. van der Hof, "Research report: Mapping age assurance typologies and requirements.," Directorate-General for Communications Networks, Content and Technology, Better Internet for Kids (BIK), European Commission, 2024.
[109] https://www.ibeta.com/iso-30107-3-presentation-attack-detection-confirmation-letters/
[110] Appendix A, https://fidoalliance.org/specs/biometric/requirements/Biometrics-Requirements-v4.0.1-fd-20240522.html#TriagePAD

The circumvention evaluation will aim to assess the extent to which each age assurance technology is vulnerable to Presentation and Injection Attacks that children in Australia are considered likely to be able to use today. The team will use tools and techniques that are publicly available via the internet and on smartphone devices, including paid-for tools, and it shall research fraud methods publicised on social media sites. The testing team will necessarily employ a risk-based approach to its circumvention testing but is confident that this approach will ensure that the testing team can report on the effectiveness of each tested solution.

Within this framework, circumvention criteria will explicitly assess resistance to AI-driven threats such as deepfakes and synthetic media. Liveness detection techniques will be tested for effectiveness against AI-generated manipulations. We will also design and implement adversarial scenarios to simulate real-world attacks. This will position the evaluation to address emerging risks effectively. We recognise the speed of development of AI and deepfake technologies and note that all approaches will need to monitor for new attacks and continuously develop defences against them.

The specific presentation attacks to test are dependent on the age assurance method being tested. Therefore, they will be defined along with the associated test protocols when the participating providers have been confirmed.

4.5.4 Usability and Acceptance testing

Overall Approach

Manual usability and acceptance testing will involve constructing a digital survey that requires mystery shoppers/users[111] , i.e. usability and acceptance (UA) testers, to undertake age checks by a subset of AA systems. UA testers will answer usability questions about each system after completing its age check. To avoid overburdening UA testers, only a subset of AA systems will be tested by each person, up to a maximum of 6. This is because each person will only be able to complete a small number of AA checks before experiencing fatigue. This could lead to abandonment of the process or negatively biased evaluations of systems tested later in the sequence.

Figure 4 on the following page shows the user journey for one participant completing the maximum of 6 AA checks.

[111] See Section 6.4 Call for Participation for further details.

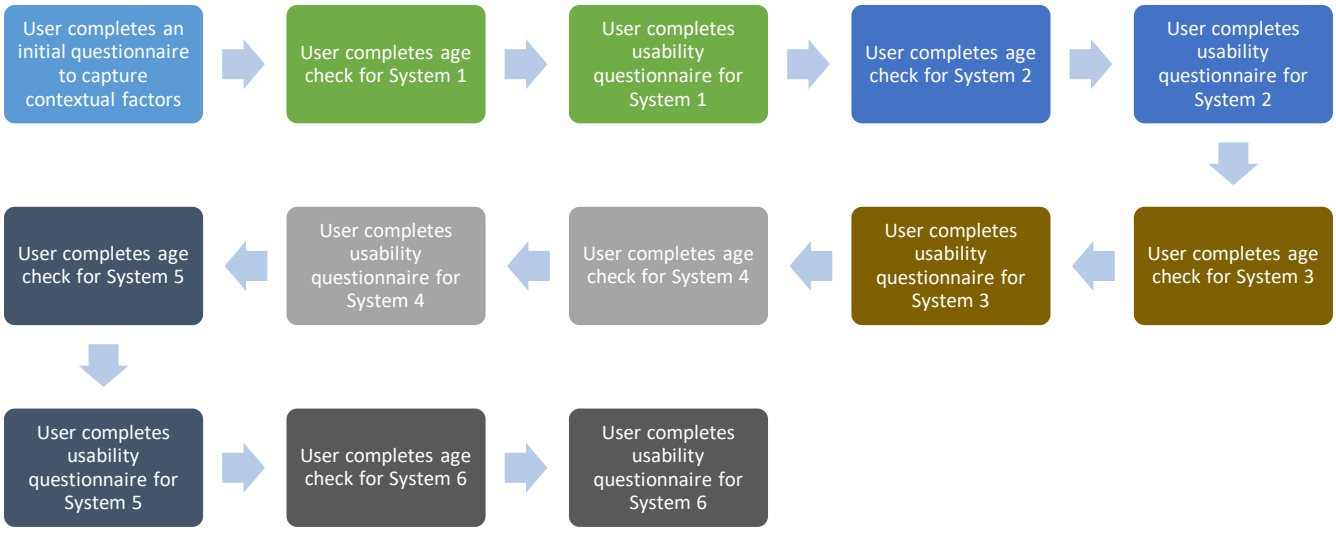

Figure 4: User journey for digital survey used for usability and acceptance testing

There are three approaches proposed:

A. In-person focus groups.
This method is where participants travel into the test lab to undertake the usability and acceptance tests whilst being observed by the host. As it is observed, it is a more reliable method. It is challenging for UA testers to travel into the test lab, particularly those from minority subpopulations, therefore the total number of participants in this case would be smaller than other approaches.

B. Remote.
This method is where participants are provided an online survey to complete remotely. It allows for a larger number of participants to be reached more easily than with in-person testing. It is possible to obtain a statistically significant sample size that is representative of the Australian population, as defined in Section 4.5.10. However, this approach is less reliable as it is not observed so participants could, for example, rush it through. Mitigation techniques for this can be employed, such as keeping track of time and discouraging rushed responses. Another challenge is that producing the remote survey that integrates with several age assurance systems and directs each user through the subset of systems assigned to them is a challenging and lengthy task, which we estimate will take approximately 2 months.

C. Hybrid observation.
This method involves gathering participants into focus groups that are run via online conferencing software. Each participant is provided an online survey, but the observer guides the focus group participants through the user journey, asking usability questions and recording the responses. This option allows for more participants than the in-person approach and it may be possible to obtain a statistically significant sample size that is

representative, as defined in Section 4.5.10. This approach also leads to greater reliability in the results when compared with the remote option, as it will include human observation.

The detailed design of UA testing will be specified during the early stages of Work Package 4 "Evaluation Activity", guided by the approaches set out above. However, we provide below the evaluation criteria to assess usability and provide an example of the questionnaire to be used during testing.

4.5.5 Usability and Acceptance criteria and questionnaire design

ISO 9241-11:2018 Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts defines several criteria that can be used the trial to assess ease of use. These can be used to inform the questionnaire that users will complete after each age check.

Key criteria for ease of use include:

- Satisfaction: the extent to which the user's physical, cognitive and emotional responses that result from the use of a system, product or service meet the user's needs and expectations
- Effectiveness: the accuracy and completeness with which users achieve specified goals.
    - Note that this is how 'effectiveness' is defined under ISO 9241 and its context of human-system interaction. This is distinct from the 'overall' effectiveness of age assurance systems, which is considered to mean how well the technology performs against several relevant criteria as defined in Section 4.3.3 Evaluation Criteria.
    - Note also that we will capture the accuracy of the age check as part of the UA testing, but this is the secondary test approach for assessing accuracy rather than primary. The primary test approach for assessing accuracy is Automated Functional Testing as described in Section 4.5.1.
- Efficiency: resources used in relation to the results achieved

To assess these criteria, a draft questionnaire with five questions along with their range of possible answers is included below. For each question, the factor being assessed is also stated.

1. [Effectiveness] Were you able to complete the age check?
    a. Yes, without any issues
    b. Yes, but with some issues
    c. No
2. [Efficiency] How would you rate the amount of effort needed to complete the age check?
    a. Likert scale: 1 (Too Much Effort) to 5 (Minimal Effort)
3. [Efficiency] How would you describe the time it took to complete the age check?
    a. Quicker than expected
    b. Roughly as expected
    c. Longer than expected
4. [Satisfaction] How confident did you feel in your ability to use the age assurance system to complete the age check?
    a. Likert scale: 1 (Not Confident) to 5 (Very Confident)
5. [Satisfaction] How would you rate your overall experience using the age assurance system?
    a. Likert scale: 1 (Very Dissatisfied) to 5 (Very Satisfied)

4.5.6 Manual Functional Testing: Interoperability testing

Using manual functional tests, the interoperability criteria we plan to evaluate for an AA system include:

- Which devices are supported?
- Which operating systems are supported?
- Which browsers are supported?
- Which interoperability standards are supported?
- Which interoperability and reusability approaches does the system implement (e.g. age tokens)?
- With which online services does the AA system's interoperability mechanism work currently?

This information can be provided directly from the age assurance provider via issuing them a survey, which can then be assessed and validated with manual functional tests.

Comprehensive testing (of all devices/OS/software) is not feasible for this trial due to the timeline and available resources.

4.5.7 Manual Functional Testing: Avatar tests for Parental Consent and Parental Controls

Parental consent and parental control systems are unlikely to support automated testing, as they are primarily driven by user interfaces. These systems will be tested using a type of manual functional test, 'avatar tests.' An avatar is an artificial identity, which in this case will be created as an artificial account on an online service or a device or operating system, to test specific functions.

The manual functional testing of these systems will cover the **interoperability** criterion. The testing of example systems, such as Android or iOS family accounts, will include for example:

- Which devices are supported?
- Which operating systems are supported?
- Which functions of the digital service can be restricted?
- Does the system integrate with other digital services (e.g. app stores or online services like social media apps)? If yes, which ones?
- What kind of restrictions can be put in place, both on the digital service itself and any other services it integrates with?
    - Which are default and which are opt-in?
    - How granular are they?
    - Can they be calibrated for children's evolving capacities as they mature
- How can restrictions be removed?
- What monitoring information is provided to the parent/guardian and the child?

During the trial, we will create two avatars on each parental consent and parental control system in-scope. One avatar will represent a child user and one will represent its parent/guardian. After creation, they will be linked together as parent and child on the digital service. Then, the restrictions available for application by the parent account, both on the digital service and any

integrated online services, will be explored and applied. Each restriction will then be tested by using the child account.

4.5.8 Static reviews

Static reviews will be used to evaluate the suitability of AA systems in terms of how well their features support the following qualities in the context of usage in Australia.

The quality criteria and the specific factors to evaluate under each of them, are described as follows:

1) **Protection of privacy:** how well the technology protects users' personal information, including through following data minimisation practices both in the capture of information from the user and the sharing of information with the relying party. We recognise that privacy is a key concern for users and a variety of stakeholders and that the assessment of privacy protection is a considerably complex task. It is not feasible to perform a comprehensive privacy assessment in this trial, therefore we will take a cautious approach to the assessment of privacy which shall be informed by the following considerations:
   a) Privacy factors can be assessed from manually gathering information, such as the privacy policy and data required from the user, from the age assurance provider.
   b) Some key privacy factors associated with age assurance systems are defined in the ISO 27566-1 Age Assurance Standards – Framework, Clause 8:
      i) Data minimization
         (1) Collection limitation
         (2) Non-disclosure of age-related data:
         (3) Compliance with legal obligations
         (4) Purpose limitation
         (5) Access control
      ii) Data disposal
      iii) Avoidance of adding to digital footprint
      iv) User awareness
      v) Audit Logs
   c) Importantly, we will also use the Australian Privacy Principles[112], as far as is feasible, for the assessment in-scope systems. We will focus on the following principles:
      i) Open and transparent management of personal information (APP 1)
      ii) Anonymity and pseudonymity (APP 2)
      iii) Collection practices, including data minimisation (APP 3)
      iv) Notification of the collection of personal information (APP 5)
      v) Use or disclosure of personal information, including the use of personal information for purposes other than age assurance (APP 6)
      vi) Consent, including the higher standards that apply to any collection, use or disclosure of sensitive information (APP 3 and APP 6)

[112] Office of the Australian Information Commissioner, "Australian Privacy Principles", 2022, Available: https://www.oaic.gov.au/privacy/australian-privacy-principles/read-the-australian-privacy-principles

vii) Limitations around direct marketing (APP 7)
viii) Cross-border disclosure of personal information (APP 8)
ix) Use of government related identifiers (APP 9)
x) Quality of personal information (APP 10)
xi) Security of personal information (APP 11)

2) **Human rights protections**: In addition to privacy, the extent to which the age assurance is accessible for all users to evaluate potential impact on freedom of expression. In addition, an assessment of efforts that have been made to respect other relevant fundamental rights, including children's rights as set out in the *UN Convention on the Rights of the Child.* Human rights protections can be assessed from gathering information, such as operating procedures and system requirements, from the age assurance provider. Key factors include:
   a) The percentage of the Australian population that hold the necessary documentation required by the age assurance system e.g. holders of valid passport or driving license,
   b) Efforts that have been taken to make the age assurance system to internet users with particular needs,
   c) Which languages the age assurance system supports,
   d) The complexity of the age assurance process and the availability of transparent information for users.
3) **Data security**: if user data is stored and, if so, if it is secure and or de-identified. Data security factors can be assessed from gathering information, such as security policies and certifications, from the age assurance provider. Key factors are defined in the ISO 27566-1 Age Assurance Standards – Framework, Clause 9:
   a) Security by design and default – which may be demonstrated by security certifications such as ISO 270001 or IRAP assessments aligned to the ASD Information Security Manual.
   b) Freshness, Reuse and Forwarding of age assurance result
      i) Freshness of an age assurance result: An age assurance result shall be protected from unplanned reuse.
      ii) Forwarding of an age assurance result: An age assurance result shall be protected from unplanned forwarding to a third party.
      iii) Planned memorization or reuse of an age assurance result: An age assurance provider or a relying party may provide for the planned memorization or reuse of an age assurance result.
4) **Technology Readiness Level (TRL):** The AATT will use the TRL calculator provided by The New South Wales Government[113], which divides the TRL into 3 factors: technology, product development and product definition/design.

Static reviews: Step-by-step process

- Step 1: Desk research to gather necessary inputs for the evaluation including: academic literature, industry literature and documents from providers (e.g. system requirements, privacy policy, etc.).
- Step 2: Analyse the method against the specific characteristics/criteria using the inputs.
- Step 3: Record the results of analysis against the characteristics/criteria.

[113] https://www.investment.nsw.gov.au/assets/Grants-and-rebates/MVP-Technology-Readiness-Level-Tool.xlsx

4.5.9 Deployments of in-scope systems across the Tech Stack

Age assurance, parental consent and parental control systems may be deployed at various points on the tech stack i.e. key points on the user journey from first accessing the device to finally accessing a piece of content.

The key points on the tech stack are:

1) Device and Operating System
2) Optionally, App Store
3) Network Access (ISP or Mobile Network Operator)
4) Optionally, Search service
5) Online Service:
    a) App or
    b) Website or
6) Optionally, User Account
7) Content

The effectiveness of an age assurance, parental consent and parental control system may vary in accordance with the point at which it is deployed on the tech stack.

The testing of systems on different points on the tech stack will be conducted either:

- Using static evaluation, if at a particular point on the tech stack there no live deployments in practice or none participate in the trial,
- Using a blend of test approaches described above if a live deployment exists and participates in the trial. The blend of test approaches will depend on the type of system to test and the criteria to test, as described earlier in this report.

4.5.10 General data requirements

To ensure an ethical approach to data gathering, incorporating privacy and human rights protections, the protocols set out in various Work Package 1 deliverables associated with the following activities will be followed:

- A-1.1.2 Approach to privacy impact assessment
- A-1.1.3 Data collection ethical protocol

These are described further in Section 7 below.

In designing an evaluation proposal for sampling age groups and sub-population demographics, our approach prioritises achieving statistical significance while managing the project's cost-effectiveness. The desired statistical validity directly influences the number of trial participants required to ensure confidence in key evaluation measures, such as classification accuracy (e.g., false accept rate, false reject rate, failure to acquire rate), binding accuracy (linking the age assurance output to the correct individual), and outcome error parity (minimizing bias across demographic groups).

Using Australia's population of 26 million as a baseline, we propose an initial sample size calculation based on a confidence interval of 95% (Z-score = 1.96) and a margin of error of 0.03. Using the Sample Size Calculator provided by Australian Bureau of Statistics[114], this results in a minimum sample size of approximately 1067 participants for population-wide analysis. For specific sub-population categories, such as age or demographic groups, we propose a slightly wider margin of error of 0.05, which reduces the required sample size to 384 participants per subgroup.

In our proposed approach, we start by calculating the overall sample size required for population-wide analysis using standard statistical methods. With a 95% confidence interval (corresponding to a Z-score of 1.96), we ensure that the evaluation results have a high degree of reliability, meaning there is only a 5% chance that the true results fall outside the calculated range. The margin of error represents the acceptable level of uncertainty in the results. By setting this margin at 0.03 (or 3%), we limit the range of potential deviation in the population-wide evaluation metrics. These parameters lead to a required sample size of 1067 participants for the entire population. This ensures that the analysis can support robust conclusions about age assurance system performance across Australia.

For sub-population categories, such as specific age brackets (e.g., 10–12, 13–15) or demographic groups (e.g., Aboriginal and Torres Strait Islander peoples), we adjust the sampling approach to account for the practical challenges of recruiting participants. By applying a wider margin of error of 0.05 (or 5%), the required sample size decreases to approximately 384 participants per subgroup. While this slightly increases uncertainty, it still provides a statistically meaningful representation of these sub-populations, allowing for balanced insights while optimizing resource use. This flexible, stratified approach ensures representational fairness and cost-efficiency, aligning with project goals.

Recruiting sufficiently large sample sizes, such as 1067 participants for population-wide analysis or 384 participants for each sub-population, requires a strategic and multi-faceted approach to ensure representational fairness, feasibility, and cost-effectiveness. Our approach to addressing this challenge includes:

1. **Partnering with Community Organizations and Stakeholders**
   Collaborations with community organizations, schools, universities, and advocacy groups can provide access to diverse populations, including hard-to-reach subgroups such as Aboriginal and Torres Strait Islander peoples or specific age brackets (e.g., 10–12, 13–15). These organizations often have established relationships and trust within their communities, facilitating recruitment.

2. **Leveraging Digital and Social Media Platforms**
   Digital platforms allow for targeted outreach campaigns to recruit participants. Tools such as social media advertising, email outreach, and recruitment websites can be tailored to

[114] Australian Bureau of Statistics, "Sample Size Calculator", 2024, Available: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Sample+Size+Calculator

demographics based on age, location, and other factors. Incentives, such as vouchers or compensation, can encourage participation.

3. **Using Existing Databases and Networks**
   Where ethically permissible, we would leverage existing participant databases from prior research or studies. Partnering with institutions such as research councils, market research firms, or governmental organizations can provide access to pre-screened participants who meet the study's demographic requirements.
4. **Implementing a Tiered Recruitment Strategy**
   To ensure proportional representation while managing recruitment efforts, we would use a tiered approach:
   a. **General Population Sampling**: Focus recruitment efforts broadly across urban and rural areas using random sampling techniques.
   b. **Sub-Population Sampling**: Employ stratified sampling to ensure specific subgroups are represented adequately, focusing additional efforts on minority or underrepresented populations.
5. **Offering Flexible Participation Options**
   Providing participants with flexibility in how they engage with the study can increase recruitment rates. For instance, participants could be involved through in-person assessments, online platforms, or mobile app interactions, depending on their preferences and the study's logistical constraints.
6. **Collaborating with Educational Institutions**
   Schools and universities are excellent sources for recruiting participants in specific age groups, particularly younger demographics. Engaging with parent-teacher associations, student organizations, and academic networks can streamline recruitment for youth participants.
7. **Conducting Localised Outreach Campaigns**
   Localised campaigns in areas with high population densities or target demographics can help meet recruitment goals. Advertising in local newspapers, radio stations, or community boards can generate interest and reach specific groups effectively.
8. **Employing Incentives and Rewards**
   To motivate participation, we would offer incentives such as gift cards, charitable donations on behalf of participants, or small cash payments. Clear communication about the purpose of the study and its benefits to the community or society can further encourage engagement.

9. **Monitoring Recruitment Progress**

Recruitment will be closely monitored to ensure quotas for both overall and sub-population samples are met. Adjustments, such as increasing outreach efforts in underrepresented areas, will be made as needed to balance demographic representation

As a guiding principle for the test dataset design, we are aiming for a margin of error of 3% across the entire Australian population and wider 5% margin of error on specific subpopulations where necessary. In practice, we may encounter cases where the margin of error increases beyond 5%. In these cases, the margin of error will be reported together with the associated test results. If the margin of error exceeds 20%, we will not report the results of testing.

4.5.11 Combination of systems & successive validation

Interoperability between systems is already included in the evaluation criteria. To explore the potential advantages of combining multiple age assurance methods, the evaluation framework will also incorporate a dedicated methodology for testing the combinations of age assurance systems. Experimental scenarios will be designed to assess how methods complement one another, either in sequence or in parallel, under laboratory conditions and real-world conditions such as social media registration or accessing restricted content. The analysis will include technical compatibility, operational challenges, and combined system performance metrics, such as accuracy improvement and enhanced user satisfaction.

An important consideration for evaluation of the effectiveness of combined systems is that such systems do not produce a cumulative output i.e. one system's output is not used as an input to the next one in the sequence. One system may trigger another one, for example if an age estimation check assesses that the user is under-25, this may trigger an age verification check. But the under-25 age result is not used as an input to the age verification system.

4.6 Test Environment

4.6.1 Overview

The test environment and all associated testing practices will be set up to follow guidelines, principles and frameworks as specified in the Evaluation Proposal. This includes following the framework in the Information Security Manual (ISM) as produced by the Australian Signals Directorate.

The test environment is described here at a high-level, which will guide the low-level design to be produced as part of Work Package 4 evaluation activities, once it is clear which systems will participate in the AATT.

The trial will ensure that all appropriate measures are taken to ensure privacy and data security, including through conducting Privacy Impact Assessments (PIAs)and following other relevant Work Package 1 protocols.

4.6.2 Hardware/Software Requirements

Infrastructure
The testing environment will be hosted on a secure cloud platform (for example, AWS, Microsoft Azure or Google Cloud). Environments will include isolated and secure network configurations with the use of "Private Clouds". For any on-premises systems, dedicated machines will be housed in restricted access locations with appropriate physical security.

Hardware
Servers or virtual machines with adequate processing power and memory will be used to simulate high-volume tasks, including encryption/decryption operations and data analysis. With the use of cloud-based services, this can be scaled up as necessary at low cost.

Software Platforms
Modern programming languages such as Python and JavaScript will be used to build test infrastructure, as both provide a multitude of reusable testing libraries and are used extensively in the commercial setting. It is expected that a custom-developed test infrastructure may be built on top of these base libraries, to model the concepts specific to AA testing. Environments such as Linux or Docker containers will be used for running the task software. Where required compatibility with vendor-provided APIs or SDKs for testing will be developed.

Security Features
All hardware and software will support end-to-end encryption (TLS), secure authentication protocols (for example, OAuth 2.0) and role-based access control to ensure data integrity.

4.6.3 Test Data

Input Data
Where possible, only anonymized and aggregated data will be used to protect user identities and any Personal Identifiable Information (PII) will be removed or hashed before testing. Where this may not be possible, to determine accurate results, appropriate measures, as set out in A-1.1.3 Data Collection Protocol, will be in place to protect the data.

Data Security
Input and result data will be encrypted both in transit and at rest using modern encryption algorithms. Access to the data will be restricted to authorized personnel and systems only.

Data Retention
A formal data destruction policy will be implemented to ensure all collected input and result data are securely deleted at the conclusion of the trial, including overwriting residual data where applicable.

4.6.4 Test Tools

Test Case Management
Tools such as Atlassian's Jira or OpenText ValueEdge will be used to document, track and manage test cases. Tickets will categorise appropriately and priority levels assigned to case tickets.

Defect Tracking

All defects will be logged and monitored using a dedicated tool such as Jira/ValueEdge. All defects will be categorised by provider and/or by test infrastructure. Defect tracking of participant systems may be important in establishing usability levels and whether the participant system is considered ready. Defects of the testing infrastructure will be managed in a similar way.

Automation
Automated testing tools such as Selenium or Postman will be used for API testing and functional verification. Custom Python scripts will handle repetitive test setups and data preparation tasks.

4.7 Test Design

Once technology trial participants have been confirmed, detailed test plans and test designs will be produced for each of the age assurance mechanisms and parental control and parental consent mechanisms in scope. These plans and test designs will provide specific guidance on how to evaluate each system, following the approach outlined in this test strategy.

Providers will by necessity have some input into the test designs, but this will be kept to a minimum to ensure independence and impartiality of the AATT. The key input to the test designs from providers will be the practice statements described in Section 4.3.5. Test designs can be produced once the functional claim in the practice statement is assessed, to understand how the system works and which areas need testing. There will also be engagement with the provider to set up communications between the test lab and the system to test, however this will not impact evaluation results under any of the evaluation criteria.

4.8 Mapping of Relevant Aspects of ISO/IEC 25010

The ISO/IEC 25010 standard, part of the Systems and Software Quality Requirements and Evaluation (SQuaRE) series, provides a robust framework for evaluating software and system quality. This analysis aligns the standard's product quality model with the specific requirements of the Age Assurance Technology Trial (AATT). By integrating ISO/IEC 25010 into the evaluation design, the trial ensures consistency, precision, and adherence to international best practices. This report details how ISO/IEC 25010's nine quality characteristics and sub-characteristics are mapped to the trial's core contexts: accuracy, reliability, interoperability, minimisation of bias, ease of use, privacy protection, data security, and human rights protections.

4.8.1 Quality Characteristics

The ISO/IEC 25010 model categorises product quality into nine high-level characteristics:

1. **Functional Suitability**: Measures how well the software meets stated and implied needs.
2. **Performance Efficiency**: Evaluates system performance regarding responsiveness, resource utilisation, and throughput.
3. **Compatibility**: Focuses on interoperability and co-existence with other systems.
4. **Interaction Capability**: Includes inclusivity, usability, and user engagement.
5. **Reliability**: Ensures stability and fault tolerance during operation.

6. **Security**: Encompasses confidentiality, integrity, and accountability of information.
7. **Maintainability**: Evaluates ease of modification and support over time.
8. **Flexibility**: Covers adaptability, scalability, and configurability.
9. **Safety**: Addresses operational safety, risk management, and fail-safe mechanisms.

4.8.2 Sub-characteristics

Each characteristic is broken into sub-characteristics that provide granular quality dimensions. For example:

- **Functional Suitability**: Functional completeness, correctness, and appropriateness.
- **Security**: Confidentiality, integrity, and non-repudiation.
- **Project Contexts in Use**: The AATT project evaluates systems with unique operational and ethical requirements. These include:
  - **Accuracy**: Precision in verifying user age.
  - **Interoperability**: Seamless integration with diverse platforms and technologies.
  - **Reliability**: Stable operation in varying scenarios.
  - **Ease of Use**: Accessible design for various users, including children or adults up to and including those of retirement age.
  - **Minimisation of Bias**: Avoidance of systemic biases in age verification algorithms.
  - **Privacy Protection**: Ensuring compliance with legal frameworks (e.g., GDPR) and maintaining user trust.
  - **Data Security**: Protecting sensitive information from breaches or misuse.
  - **Human Rights Protections**: Aligning with ethical standards, including child safety online and data minimisation.

4.8.3 Mapping Table

The table aligns ISO/IEC 25010 characteristics and sub-characteristics with project requirements while also mapping to ISO/IEC 25040, IEEE 2089.1, and ISO/IEC 27566-1:

Final Evaluation Criteria	Sub-criteria	Mapping to ISO 25010	Mapping to ISO 25040	Mapping to IEEE 2089.1	Mapping to ISO 27566-1
Accuracy	FPR, FNR, TPR, TNR	Functional correctness	Specification of quality metrics	Accuracy, classification errors	Functional effectiveness
Interoperability	API success rate, data exchange capability	Interoperability	Specification of the evaluation plan	Interoperability	Functional interaction

Reliability	Fault tolerance, MTBF	Reliability	Execution of evaluation tests	Fault recovery	Risk identification
Ease of use	Usability score, accessibility testing	Interaction capability	Quality requirements validation	Accessibility considerations	Interaction inclusivity
Minimisation of bias	Parity analysis, demographic fairness	Interaction capability	Specification of the evaluation plan	Bias mitigation measures	Inclusivity and fairness
Human rights protections	Data minimisation, ethical compliance	Security, flexibility	Evaluation report conclusions	Human rights impact	Ethical protections
Protection of privacy	Encryption, purpose limitation	Security	Specification of quality metrics	Privacy by design	Privacy and confidentiality
Data security	Audit trails, data breach prevention	Security	Execution of evaluation tests	Security effectiveness	Security resilience
Circumvention	Resistance to bypass attempts	Reliability, security	Execution of evaluation tests	System robustness	Functional robustness

4.8.4 Addressing Identified Gaps

- **Minimisation of Bias:** ISO/IEC 25010 lacks explicit provisions. IEEE P7000 standards and custom metrics for fairness analysis can be applied to mitigate this.
- **Ethical Considerations:** Augment with frameworks addressing human rights and ethical design, as described above in Section 4.5.8.

4.9 Test Execution

The initial general approach to test execution, followed by the kinds of reports that will be produced to ensure sufficiency of an audit trail, is defined below. The approach is divided into the different kinds of tests and the different systems under test. These will be refined further during the beginning of the testing process.

4.9.1 Automated Tests

1. Entry and Exit Criteria: Define conditions for starting and stopping tests.
    a. Entry criteria:
        i. Dataset gathered.
        ii. Test environment ready – system to test is either deployed in the test lab or accessible from it.
        iii. Test scripts created and tested.
    b. Exit criteria:
        i. The age check completes successfully.
        ii. A classification of the user's age being equal to and above or below, each threshold is produced. The thresholds are 13+, 16+ and 18+
        iii. Metadata of the test is also recorded.
2. Execution Plan: Schedule testing activities.
    a. Timeline (this is subject to change):
        i. Age Verification systems to be tested first in M3 (January) and M4 (February).
        ii. Then Age Estimation systems to be tested in M4 (February) and M5 (March).
        iii. Then Age Inference systems to be tested in M5 (March) and M6 (April).

4.8.2 Manual usability and acceptance testing

1. Entry and Exit Criteria:
    a. Entry Criteria:
        i. Participants recruited and assigned a maximum of 6 AA systems per participant.
        ii. Test environment ready – the survey, including integration with each AA system and the build of all questionnaires, is deployed and tested. This will require the team members leading evaluation activity and the user participants to have the necessary access to the AA systems to test.
    b. Exit Criteria:
        i. The survey, including all age checks, completes successfully.
        ii. A classification of the user's age being equal to and above or below, each threshold is produced. The thresholds are 13+, 16+ and 18+
        iii. The user input to all questionnaires has been recorded.

2. Execution Plan (this is subject to change):
    a. Testing of all participating age assurance systems to be conducted between M3 (January) and M6 (April)

4.9.3 Manual functional testing

1. Entry and Exit Criteria:
    a. Entry Criteria:
        i. Any avatars required are created.
        ii. Test environment ready – the age assurance or parental consent or parental control system is either deployed in the test lab or accessible from it via a user-interface.
        iii. Any other online services to test for interoperability are accessible from the test lab.
    b. Exit Criteria:
        i. For age assurance systems:
            1. A classification of the user's age being equal to and above or below, each threshold is produced. The thresholds are 13+, 16+ and 18+.
            2. The result of successfully performing the age check on a service is recorded, along with any metadata.
        ii. For parental consent and parental control systems, the result of the restriction being tested on the host digital service or another integrated online service is recorded, along with any metadata.
2. Execution Plan (this is subject to change):
    a. Testing of all participating age assurance systems to be conducted between M3 (January) and M6 (April)
    b. Testing of all participating parental consent and parental control systems to be conducted between M4 (February) and M6 (April)

4.9.4 Presentation Attack Detection

1. Entry and Exit Criteria: Define conditions for starting and stopping tests.
    a. Entry Criteria:
        i. Dataset gathered and/or avatars created if required for the test.
        ii. Test environment ready – system to test is either deployed in the test lab or accessible from it.
        iii. Test scripts created and tested if required for the test.
    b. Exit Criteria:
        i. System produces an output indicating either
            1. Attack detected or
            2. The user's age (i.e. attack not detected).

2. Execution Plan (this is subject to change):
    a. Testing of all participating age assurance systems to be conducted between M3 (January) and M6 (April)
    b. Testing of all participating parental consent and parental control systems to be conducted between M4 (February) and M6 (April)

4.9.5 Static Evaluation

1. Entry and Exit Criteria: Define conditions for starting and stopping tests.
    a. Entry Criteria:
        i. Literature gathered, including practice statements from participating providers, academic sources and industry sources.
    b. Exit Criteria:
        i. Result of static evaluation recorded per criteria.
2. Execution Plan (this is subject to change):
    a. Testing of all participating age assurance systems to be conducted between M3 (January) and M6 (April)
    b. Testing of all participating parental consent and parental control systems to be conducted between M4 (February) and M6 (April)

4.9.6 Defect Management

For all the above kinds of tests, defect management will be handled by a Defect Management Policy to be defined at the beginning of the evaluation activity in Work Package 4. This will outline the process for reporting, triaging and resolving defects, including the following steps:

1. Defect to be created in management system (such as JIRA).
2. Mandatory fields and attributes filled in by reporter.
3. All new tickets to be triaged and prioritised by selected Defect Triage Lead.
4. Triage Lead re-assigns to appropriate party for fixing.
5. Evaluation Trial Manager to be notified and defect tracked.

4.9.7 Test Reports

Using ISO 29119-3:2021 Software testing – Part 3: test documentation, we intend to produce the following documentation for the testing process:

- Test completion report: this provides a summary of the testing that was performed.
- Test Results: a record of whether a specific test case has passed or failed; i.e. if the actual results correspond to the expected results or if deviations were observed or if planned execution of the test case was not possible.
- Test Incident Report: a test incident is any issue that is noticed during testing that requires investigation, such as when an actual result deviates from the expected result of a test or

when the system behaves differently to how a tester expected. Test incidents are recorded in test incident reports. There will be one incident report for each unique incident (incident reports may also be known as defect reports, bug reports, fault reports, etc.).

- Test execution log: this records details of the execution of one or more test procedures as a series of events.

All these reports collectively provide a transparent, traceable record of testing throughout the project. They ensure:

1. Traceability: Every test result can be linked back to specific test cases, procedures and data.
2. Accountability: Issues related to data, including anomalies or changes, are documented.
3. Reproducibility: Future testers can replicate results using the same data and procedures.
4. Compliance: Documentation for test designs, processes, implementation and results demonstrate adherence to standards.

These reports will provide the final project report with a clear audit trial, ensuring confidence in the results.

4.10 Test Governance

Roles and Responsibilities: The RACI matrix for all team members associated with test activities are defined in Figure 5 below.

Role	Stakeholder	Test Age Verification	Test Age Estimation	Test Age Inference	Test Parental Consent and Parental Controls	Test Tech Stack & TRLs
Test Lead	Adrian Ugray	I	R	I	I	R
Test Lead	Jason Smart	R	I	I	I	I
Test Lead	Jonathon Cleaver	R	R	I	I	I
Test Lead	Ji Yu Jan	I	R	R	I	I
Test Lead	Seung Roh	I		I	I	I
Test Lead	Stan Potums	I	R	R	I	I
Test Lead	Tracey Rawlinson	I	I	I	R	I
Test Lead	Surya Ramessh	C	C	C	C	C
WP4 Lead	Mark Pedersen	A	A	A	A	A
WP1 Lead	George Billinge	C	C	C	C	C
WP2 Co-lead	Asad Ali	C	C	C	C	C
WP2 Co-lead	Koliya Wedanage	C	C	C	C	C

WP3 Lead	Iain Corby[115]	I	I	I	I	I
WP5 Lead	Rhianne Kiddle	I	I	I	I	I
WP6 Lead	Keith Robinson	I	I	I	I	I
Project Co-director	Tony Allen	A	A	A	A	A
Project Co-director	Andrew Hammond	A	A	A	A	A

Figure 5: RACI Matrix. R=Responsible, A=Accountable, C=Consulted, I=Informed

Review and Approval Process: The test strategy, plans and results will be reviewed and approved using the following process:

1. Prepare Draft (1-5 days)
    - Responsible: Test Lead (i.e. author of the document).
    - Action: Create a draft of the test strategy, plan or results document, ensuring all necessary details are included (e.g., scope, test coverage, results).
2. Internal Review (1-2 days)
    - Responsible: Test Leads, WP4 Lead, WP1 Lead WP2 Co-Leads, Project Co-Directors.
    - Action: Team members review the document for accuracy, completeness, incorporation of ethical framework and alignment with AATT goals.
    - Output: Collect feedback via comments or a review form.
3. Feedback Incorporation (1-2 days)
    - Responsible: Test Lead (i.e. author of the document).
    - Action: Revise the document based on feedback. Clearly note any changes made.
4. Final Review and Approval (1 day)
    - Responsible: WP4 Lead, Project Co-Directors.
    - Action: Review the revised document for final approval. Stakeholders ensure:
        i. Requirements are covered.
        ii. Results are accurate.
        iii. Risks are addressed.
    - Output: Document is formally approved and ready for use. The approved document should be stored with version control for reference and audit purposes.

[115] We are including some reclusion methods for Iain Corby due to the conflict of interest arising from his role at the Age Verification Providers Association (AVPA). We will not inform him of activities in a way that could undermine confidence in the trial.

5. Data Protection and Ethical Framework

Age assurance and parental control technologies may impact the fundamental rights of internet users. The design, evaluation and deployment of these technologies is subject to intense public debate and scrutiny. Transparent ethical frameworks will therefore be implemented throughout the lifecycle of the AATT to give all stakeholders confidence that the trial is conducted in a principled, ethically rigorous manner.[116]

The AATT will be conducted in line with several guiding ethical principles: Respect, Transparency, Accountability, Fairness, Privacy and Safeguarding Children. The Project Ethics Committee will meet on an at least monthly basis to review trial activities and principles, ensuring that the guiding ethical principles are being observed.[117] These principles will be operationalised through several activities and deliverables, primarily sitting within Work Package One: Data, Ethics and Impartiality. Summaries of the AATT's approach to these issues are set out below.

5.1 Privacy and data protection

The impact of age assurance technologies on privacy and data protection are commonly cited concerns during research into user attitudes.[118] The trial will evaluate the potential impact of different age assurance technologies on user privacy, but it is also vital that the trial be conducted in a way that respects individuals' right to privacy. This includes ensuring individuals, including children and other individuals of different capacities, can make informed choices about how their personal information is collected and processed. Failing to include sufficient privacy safeguards during the trial risks harms such as identity theft, discrimination or surveillance. Such failures would significantly undermine confidence in the AATT.

The age assurance technology trial will involve the collection and processing of individual's personal information, including potentially sensitive information. This information may be exposed to algorithmic systems during evaluation activities. As such, strict safeguards will be implemented to respect the privacy of participants, with transparency and accountability mechanisms ensuring individuals are able to exercise their fundamental rights. Data collection will be limited to solely the data that are expressly required for the purposes of the trial and all data collected will be destroyed

[116] See A-1.1.1 Ethics Handbook for further detail.
[117] For terms of reference, see A-1.1.1 Ethics Handbook.
[118] https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/vsp/attitudes-to-age-verification/2022-adult-attitudes-to-age-verification-adult-sites.pdf?v=328580#:~:text=o%20From%20a%20personal%20perspective,online%20activities%20(figure%201).; https://www.drcf.org.uk/publications/papers/families-attitudes-towards-age-assurance-research-commissioned-by-the-ico-and-ofcom/.

once they are no longer needed. In addition, personal information will be pseudonymised and strict information management and security protocols will be followed throughout the course of the AATT.[119] Participants will be offered clear information about the data collection activities in comprehensible language, along with a single point of contact to ask questions, raise concerns or withdraw from participating in the trial.[120] Where particular efforts need to be made to ensure that historically marginalised communities, such as Aboriginal Australians and Torres Strait Islander People, are represented in the trial. First Nations communities will be involved in decision-making processes that affect them, ensuring their autonomy and perspective remain central to trial activities. Ongoing efforts to build relationships with First Nations communities will be respected, with full transparency being provided on the trial's scope, methodology and outcomes.[121]

5.2 Safeguarding children

As the AATT is evaluating technologies designed to protect children, some children will be involved in the trial. Respecting the rights and wellbeing of children, whether participating in the trial or more broadly, is a priority of the trial team. Children share all the fundamental rights of adults but may have diminished autonomy or differing capacities to exercise or advocate for their rights. As such, additional measures will be implemented to ensure that children participating in the trial are kept safe and feel supported at all stages, in line with the Australian Government's National Principles for Child Safe Organisations.[122]

Children participating in the trial will be informed of their rights in accessible language, offered opportunities to participate in decisions affecting them and have their views taken seriously. Child-focused processes will be implemented to respond to complaints and concerns regarding children's wellbeing and steps will be taken to ensure that team members working with children are supported to reflect child safety and wellbeing in practice.[123]

5.3 Impartiality

The AATT presents several risks to impartiality, whether real or perceived.[124] Both real or perceive risks to impartiality or conflicts of interest may undermine confidence in the outcomes trial. Transparent governance processes have therefore been implemented to identify and seek to mitigate threats to impartiality and to identify and mitigate real or perceived conflicts of interest. Key risks are identified in the initial impartiality report.[125] They include risks concerning government influence on the outcomes of the trial, the selection of age assurance technology vendors, potential of conflicts of interest and evaluation methodologies or approaches to reporting that favour one technology over another. Mitigation strategies have been identified and

[119] See A-1.1.2 Approach to Privacy Impact Assessment, A-1.1.3 Data Collection Ethical Protocol for further information.
[120] See A-1.3.1 Human Test Subjects Protocol.
[121] See A-1.3.2 Application of AIASTIS Code of Ethics for Aboriginal and Torres Strait Islander Research.
[122] https://www.childsafety.gov.au/resources/national-principles-child-safe-organisations.
[123] See A-1.2.1 Child Safeguarding Policy for further detail.
[124] Impartiality is defined for the purposes of this trial as "the presence of objectivity," in line with ISO/IEC 17065:2012.
[125] See A-1.4.1 Initial impartiality report.

implemented for each identified risk, newly identified risks to impartiality will be monitored, reported and addressed throughout the course of the trial.

Real or perceived conflicts of interest of AATT team members must be reported in the conflict of interest register.[126] Where appropriate, strict recusal procedures will be in place for members with identified conflicts during specific decision-making processes. The Project Ethics Committee will review potential conflicts and approve appropriate recusal measures. Where appropriate, the ACCS Impartiality Committee, who are independent of the AATT trial team, will be consulted to approve measures to mitigate real or perceived conflicts of interest. All conflicts, risks to impartiality and steps taken to mitigate these, will be published on the trial website. These oversight mechanisms and the trial's commitment to transparency will be fundamental to building confidence in the trial.

5.4 Transparency and Open Data

As part of ensuring transparency, academic rigour and repeatability, the AATT will make use of the Open Science Framework (OSF) platform[127]. This provides tool support to manage several aspects of the research project lifecycle. The key features of OSF that will be utilised in the trial are:

- Pre-registration: the AATT's aims and methods will be registered into the OSF. Importantly, these will not be editable once registration is complete. This counters problems like overfitting, p-hacking, cherry picking or hypothesizing after results are known (sometimes called "HARKing").
- Results data sharing: the AATT's test results will be stored on the OSF and made publicly available. This will enable results to be examined in detail by others interested in the study, for reasons such as repeating the tests, conducting further analysis of the results or conducting future research based on the results.
    - Importantly, test results will be made available publicly after (a) obtaining consent from test subjects and (b) sufficient pseudonymisation or anonymisation of test subject data. We will ensure that results data cannot be linked back to any individuals that acted as test subjects. This is integrated into the project's approach to data protection and ethics, described further in Section 7.

Following approval of the Evaluation Proposal Report (this deliverable), the AATT will be entered into OSF, beginning with the pre-registration step.

5.5 Managing Potential Research Bias

Managing and minimising research bias is a necessary part of the approach to ensuring the trial's results are reliable. Research bias refers to systematic errors that may arise during a study's design,

[126] See A-1.4.2 Conflict of Interest Register; A-1.4.2.1 Conflict of Interest Policy.
[127] The Open Science Framework, https://osf.io/

execution, or analysis, potentially resulting in misleading conclusions. It can emerge at any point in the research process and significantly affect the study's reliability and validity.

The UK's Critical Appraisal Skills Program defines the following kinds of research biases[128]:

- Placebo effect: This is a psychological phenomenon where a patient experiences an improvement in symptoms due to the belief that they are receiving treatment. This can inadvertently distort results of clinical trials where a 'placebo group' believes they are receiving the treatment under study.
- Hawthorne effect: This refers to the alteration of people's behaviour when they are aware they are being observed. This awareness can cause individuals to work harder, skewing the results of studies, particularly those involving human performance.
- Measurement bias: Occurs when data or information is not accurately recorded in a research study. This can stem from errors in data collection, inconsistent measurement tools, or subjective interpretation of data, leading to skewed and unreliable results.
- Publication bias: This is the tendency for researchers and editors to handle the reporting of experimental results that are positive (i.e., showing a significant finding) differently from results that are negative (i.e., supporting the null hypothesis) or inconclusive, leading to a misleading bias in the overall published literature.
- Observer/Experimenter bias: Is when the person conducting the research allows their expectations or beliefs to influence the results of the experiment. This can lead to distorted data, as the researcher may subconsciously favour results that confirm their own preconceptions or hypotheses.
- Reporting bias: Is a type of bias where researchers selectively report or omit information based on the outcome of the research or personal beliefs, which can distort the findings and undermine the integrity of the study.
- Sampling bias: Is when the selection of participants for a research study is not representative of the full population. The skewed sample could lead to a misrepresentation of the data and flawed conclusions.
- Recall bias: This occurs when the participants in a research study may not remember previous events or experiences accurately or they may subconsciously alter their memories. This can lead to skewed data and impact the credibility of the research results.
- Selection bias: Occurs when the method of selecting participants or groups for a study produces an outcome that is not representative of the total population. For instance, if the sample group is not randomised or certain groups are excluded, it could produce skewed or incomplete results.
- Confirmation bias: This is the tendency to favour, seek out, interpret and remember information in a way that confirms one's pre-existing beliefs or hypotheses, whilst giving

[128] UK Critical Appraisal Skills Programme (CASP), "Different Types of Bias in Research", https://casp-uk.net/news/different-types-of-research-bias/

disproportionately less consideration to alternative possibilities. This bias can lead to flawed conclusions as it may prevent researchers from accurately assessing all relevant data in a neutral manner.

In the table below, each type of research bias is assessed to identify whether it applies to this trial and, if it does, how it will be mitigated.

Type of Bias	**Applicable to AATT?**	**Mitigation**
Placebo effect	No, as there will not be any use of placebos	N/A
Hawthorne effect	Yes, during manual usability and acceptance testing with user participants	The survey and questionnaires for manual usability and acceptance tests will be designed to minimise psychological impact on user participants
Measurement bias	Yes	All tools and processes to and measure test results will be tested and internally validated prior to commencing evaluation
Publication bias	Yes	Publications from the trial containing results will only be published after being reviewed and approved by the trial's ethics committee. Also, all experiments will be registered in the Open Science Framework (OSF) and all results will be published after being appropriately anonymised.
Observer/Experimenter bias	Yes	Test designs will be reviewed internally before test implementation. The review process will include review and approval from the trial's ethics committee.
Reporting bias	Yes	Test results and findings will only be published after being reviewed and approved by the trial's ethics committee
Sampling bias	Yes	The sample for creating the test dataset and for manual usability and acceptance tests

		will be representative as described in Section 4.
Recall bias	Yes, during manual usability and acceptance tests	The user journey for manual and user acceptance tests is designed to capture information from user participants immediately after testing one of the systems assigned to them.
Selection bias	Yes	The sample for creating the test dataset and for manual usability and acceptance tests will be representative as described in Section 4.
Confirmation bias	Yes, when analysing practice statements from providers and then designing tests	Provider participants will have minimal input into the testing process. Their engagement will be limited to providing a practice statement to understand how their system works and how communications between the test lab and their system can be set up in order to perform tests. The test review and approval process includes ethical review and approval from the trial's ethics committee. This will ensure the minimisation of confirmation bias.

6 Stakeholder Engagement

6.1 Stakeholder Advisory Board

The project has established a Stakeholder Advisory Board to create a forum for representatives of key stakeholder groups to provide input. It is chaired by Professor Jon Rouse APM.

The trial itself is independent, so the Board is only advisory, but will provide the opportunity for a wide range of experts and people with an interest in age assurance technology and its applications to offer advice and challenge to the project team.

The board will publish minutes of its proceedings to ensure transparency and its membership will be listed on the project website.

6.2 Public and Participant Communications

The project has adopted a policy of transparency by default so as the trial progresses, all relevant documentation will be made available on the document repository in the project website, unless there is a compelling reason not to do so. The specific results of the trial will not be published prior to the final report's publication, but other papers relating to the process, such as testing approaches and plans will be available.

There will be an ongoing opportunity to provide feedback on any published documents which will be shared with the relevant members of the trial team, subject to supervision by the project's Ethics Committee to prevent inappropriate interventions.

The project intends to publish regular newsletters to anyone who signs up on the website to receive these.

Participants that are providers may be invited to discuss draft testing plans for their system and will then be shown provisional results, with time allowed for in the project plan to undertake remedial testing if it is agreed that something went awry.

To engage younger audiences and their guardians effectively, child-friendly materials such as infographics and explainer videos will be developed. Interactive sessions, such as focus groups and Q&A, will ensure clear communication of findings. This approach will foster greater understanding and inclusion among those directly impacted by age assurance technologies.

6.3 Recruitment of Age Assurance Providers and Relying Parties

A critical success factor for the trial is to persuade providers of a wide range of age assurance methods to put them forward for evaluation. The project team will look for opportunities to raise awareness of the trial and to build sufficient confidence in its approach that suppliers will be willing to participate.

There will need to be a particular effort to test methods which are not currently operating in a production environment. A good example is App-store based age assurance, where an app-store manages the process of checking a user's age and then makes that age attribute available to apps and websites to facilitate the application of age-restrictions. This is not in existence at present and the leading app stores are not supporting its development. In such examples, the project team will be as creative as possible to deliver sufficient evidence to allow these alternatives to be considered alongside existing systems.

The project will also seek to encourage relying parties which already implement age assurance to any extent to participate in the trial. This may include the use of third-party age assurance providers as part of wider system, for example by first monitoring user behaviour for contra-indicators that suggest self-declared age is inaccurate and then referring the user for a more extensive age assurance check.

As well as laboratory-based testing, the project intends to assess the user-experience of any operational method of age assurance. Working with a specialist test-purchasing service – also known as 'mystery shopping' - it will recruit a diverse set of test users who will be assigned tasks to test and then provide feedback across a range of methods.

6.4 Call for Participation

The Trial process will begin with a formal "Call for Participation" leading to "expressions of interest" from potential participants.

The success of the trial is dependent on being able to consider the full range of methods, to the extent they are sufficiently developed to be susceptible to testing. The aim of the call for participation is to encourage organisations to put forward a wide range of methods for testing as part of the trial. Even where a method does not yet exist above TRL 3, it needs to be submitted as an expression of interest to allow for desktop testing.

The call for participation will describe what the trial is seeking as follows:

- age verification processes that involve finding, validating and binding a date of birth from a record either somewhere or on something and then binding that to the correct individual;

- age estimation systems that involve the analysis of features or behaviours of humans that vary with age (this is not just about face age estimation);
- age inference techniques that involve any other facts or information about individuals from which it is reasonable to infer their age;
- control systems aimed at the parents or legal guardians of children to pre-set their online experience including any processes that deal with the evolving capacity of children to manage their own experiences as they age up;
- processes that seek and manage consent of parents or legal guardians at the point children encounter an age related eligibility requirement, including systems that manage consent revocation;
- platforms or services with end-to-end customer lifecycle experience management including in-app or in-service indicator and signal management when children are identified or suspected;
- deployments of multi-level successive validation approaches or those that are positioned in the tech stack to provide signals through app stores or at search level to enable providers of age restricted goods, services, content, venues or spaces to manage access through use of those signals.

To be selected for participation, it will be made clear that technology needs to:

- be available for deployment in the Australian context (but participants do not have to be based in Australia)
- have a technology readiness level of 4 or above (the higher the better). We will reinforce that we are interested in learning about tech at a readiness level of 1-3, but these systems will not be tested.
- have a practice statement applicable to your technology that is developed in accordance with clause 11 of ISO/IEC DIS 27566-1 - Age Assurance Systems - Part 1: Framework

Participants will also be advised that the trial team will providing tools to assist in the preparation of practice statements.

Expressions of interest will be captured in a Google form which will be sent to any organization that has completed the form on the project website indicating they might be interested in participating or who responds to the call for participation.

When the project website is live mid-January, the form will be available to complete online there as well. The call for participation will be formally issued on January 6th, 2025, with a deadline of January 25th.

EoI Contact details		
Field Label	**Field type**	**Required field**
Text: Please use this form to express an interest in submitting your age assurance system to the Age Assurance Technology Trial. If you wish to submit multiple but separate methods, please submit an additional form for each method. If you wish to include a combination of methods used successively (a waterfall approach) or in combination, please use one form.	Text to display (HTML)	-
Company name	Single line text Field	Yes
Contact (First name, Last name)	Name field	Yes
Lead contact Email	Email field	Yes
Lead contact Phone	Phone number field	Yes
Home country	Country field	Yes
Australian Office contact email if available	Email field	No
Australian Office contact phone if available	Phone number field	No
Select your area(s) of technology software testing: Age Verification Age Estimation Age Inference Parental Consent Parental Controls Alternative stack systems	Radio select button that allows for multi-select? Please confirm if they can choose more than one. [Asad: yes, more than one option should be allowed]	Yes
Technology Readiness Level (see https://www.gao.gov/assets/gao-20-48g.pdf) 1 Basic principles observed and reported 2 Technology concept formulated 3 Experimental proof of concept 4 Technology validated in lab 5 Technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies) 6 Technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies) 7 System prototype demonstration in operational environment 8 System complete and qualified 9 Actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space)	Radio select button for single selection of a number between 1 and 9	Yes
Brief Description of method (s)	Multi-line text field	Yes

7 Project Management and Risk Assessment

7.1 Quality Control Mechanisms

The AATT deploys a rigorous approach to Project Management, encompassing the tools, guidelines and templates for planning, managing and most importantly delivering the project. The structured Quality Management approach uses a PM² Project Management Methodology, making use of the PM² template set for projects, including the approach to Risk Management, Project Change and Deliverable Acceptance.

The approach to Quality Control is part of the overall management of quality in accordance with ISO 17065 Accreditation and deployment of laboratory testing in accordance with ISO 17025 Accreditation.

Document Protocol

There is a protocol for each document produced as part of the AATT, whether that be a Deliverable, Report or Plan. So, as part of the Quality Control Process at the beginning of each document, there will be key data including information on the Work Package Lead, Task Leader, Document Version and Document Sensitivity level. There is also a list of Document 'Approvers' and 'Reviewers' associated with each Project Document, who have an action either to Approve or Review the document (or both), according to their role within the Project. For instance, the 'Task Leader' for any given Project Task, is always required to Review the document, as is the Work Package Lead for which the task belongs.

Document History

In terms of Document History, to request a change to a particular document, the Team Member in question must contact the Document Author or Project Owner, who has the authority to approve that change.

The Quality Control Mechanisms are central to Milestone 3 and Milestone 4 of the Project, the Delivery of a Preliminary Report in [M6]and Completion of the Final Report in [M8]. For the Preliminary Report, there will be an extensive quality control focus, whereby the Report will be proofread, sense checked and its accuracy validated, by a dedicated sub-set of the Project Team.

When it comes to Final Report Production, the quality control associated with that will include the necessary pre-publication checks and clearances that are required to include:

- The Department;
- eSafety Commissioner and Office of the Australian Information Commissioner (OAIC);

- Ethics Review in accordance with AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research;
- Peer Review by Prof. Toby Walsh at the University of South Wales.

7.2 Risk Management Plan

The Project has a Risk Management Plan[129] which defines and documents the Risk Management Process for the Age Assurance Technology Trial. It describes how risks will be identified and assessed, what tools and techniques can be used, what the evaluation scales and tolerances are, the relevant roles and responsibilities, how often risks need to be revisited, etc. The Risk Management Plan also defines the risk monitoring and escalation process as well as the structure of the *Risk Register* which is used to document and communicate the risks and their response actions.

The purpose of that document is:

- To outline the risk approach and process to be used for the project;
- To identify the roles and responsibilities related to risk management;
- To specify the methodology, standards, tools and techniques used to support risk management.

The Risk Register currently identifies 7 Key Risks. A selection of the main risks include:

01. **Data Ethics associated with Working with Human Test Subjects, including Aboriginal and Torres Strait Islander peoples (WP1 - 6)**. There is the possibility that the Technology Trial may not address the critical risks associated with data, ethics and impartiality that will arise throughout the project when it comes to working with human test subjects.
STRATEGY - PREVENT. The creation of key project documentation and processes, including an Ethics Handbook (A-1.1.1), a Data Collection Ethical Protocol (A-1.1.3), a Human Test Subjects Protocol (A-1.3.1)and Application of the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research for Data Collection Phase (A-1.3.2) will address any issues relating to the data ethics of working with human test subjects.
02. **Exclusion of Aboriginal and Torres Straight Islander peoples (WP1 -6).** The Technical Trial may risk excluding certain demographics, including Aboriginal people, from participation.
STRATEGY - PREVENT. The project needs to ensure that multi-ethnic diverse communities are included in the demographic spread of human test subjects. This will be achieved through the application of the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research and initial and ongoing Equality, Diversity and Inclusion Monitoring.

[129] The latest version of this controlled document is stored in AATT SharePoint WP 6 – Programme Management, Risks & Quality Control – T6.3 – Risk Management.

03. **Lack of public trust in the project.** There is the risk that the general public will not have trust in the age assurance technology trial and its goals and objectives, which would reduce project credibility.
    STRATEGY - ACCEPT. This is a risk that cannot be avoided; public trust can be subjective. However, Work Package 3 deals with Communications, including creation of a Project Website and blog and aims to secure transparency and maintain public confidence in the project. The sub-contractor who will build the website has a wealth of experience and impressive portfolio in this field.

7.3 Project Compliance and Monitoring

As part of the compliance with the AATT's Principal Contract Requirements, there is a requirement for all Project Team Members to keep and update a Timesheet, which they submit to the Project Finance Team (WP6 Project Management Leads) at the end of each month. The Finance Team and The Department must approve any expenses.

Each Project Subcontractor has a Contract and is given a Purchase Order (PO) each month, setting out estimated allocation of hours for them for that month. Timesheets for everyone within a particular Subcontractor (for example for each Team Member within the KJR Team) are collated together and an invoice is submitted monthly for appropriate reimbursement of time and expenses.

Project Compliance and Work Package 6: Programme Management, Risks & Quality Control

Project Compliance is also central to Work Package 6 of the AATT, with Task 6.5 focused on assurance for contract compliance, accounting for the spending of public money, accounting and audit. Various activities are associated with this from the onset of the Project, including identification of contract compliance issues relevant to the Project and as described above, completion of sub-contracting arrangements, contracts and banking payments (including ethics, AML and sanctions compliance in M2. There will also be Accounting for Expenditure at the half-way point of the Project [M4].

Task 6.5 also includes adherence to standards, including existing ISO/IEC 17065 accreditation and the Protective Security Policy Framework, Privacy Act 1988, any Legislative Requirements, Chief Executives Instructions, Archives Act 1983, Public Governance, Performance, Accountability Act 2013and any requirements of the Australian National Audit Office. Project Compliance would not be complete without full adherence to these relevant regulations and laws.

Task 6.6 is linked with T6.5 above and is the Performance Review & Project Evaluation. This Task involves internal and external review of contract performance, including assurance from project legal advisors on compliance, any external audit of expenditure required (as outlined in T6.5) and an analysis of the timelines and acceptability of project deliverables.

Contract Management Meetings

Project Compliance in the day-to-day management of the AATT is vital to achieving consistency and ensuring the appropriate level of monitoring required. A variety of meetings are held and are ongoing, including fortnightly Project Team meetings, monthly Contract Management meetings with the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (DITRDCA). There is also a weekly Project Directors' meeting, whereby the two Project Directors convene to update on a birds-eye view of the AATT, including tasks and deliverables in progress, actions and any other important developments that they need to be made aware of.

8 Appendices

8.1 Glossary of Terms

Term	Definition
Acceptance testing	Testing of a software system's functions by end-users to validate the system's functionality and usability
Accuracy	How well the age assurance system can detect a user's age
Age Assurance	The set of processes and methods used to verify, estimate or infer the age or age range of an individual.
Age Assurance Provider	Entity responsible for providing the result of an age assurance to a relying party.
Age assurance result	information produced by an age assurance system indicating that an individual is a certain age, over or under a certain age or within an age range
Age Assurance Solution	Equivalent to Age Assurance System.
Age Assurance System	A system that uses age assurance methods to determine an individual's age and provide a result about that age.
Age Estimation	Age assurance method based on analysis of biological or behavioural features of humans that vary with age
Age gate	A mechanism used to restrict access to content, products or services based on a user's age. It typically involves verifying or attesting that a user meets the minimum age requirement before granting access.
Age Inference	Age assurance method based on verified information which indirectly implies that an individual is over or under a certain age or within an age range
Age restriction	Equivalent to age-related eligibility decision
Age Verification	Age assurance method based on calculating the difference between a verified year or date of birth of an individual and a subsequent date
Age-related eligibility decision	action by a relying party to determine access to goods, content, services, venues or spaces based on an age limit or an age band.
Authoritative party	Entity that is recognized to have the right to create and manage a record that contains a set of attributes that allows an individual to be uniquely identified within a given context
Automated Testing	Automated testing uses tools to perform tests with none or minimal human intervention.
Avatar	A virtual representation or model of a user or system entity used during testing activities. The avatar is often used to simulate real users or interactions within a system, especially in the context of user experience testing or automated testing.

Bias	A systematic error or deviation introduced into a process, measurement or outcome, leading to results that do not accurately reflect the intended objectives or population
Binding	Property that relates a result of age assurance to the correct individual.
Circumvention	Attacks on age assurance systems that manipulate it to produce an incorrect result. This specifically includes biometric presentation attacks and spoofing attacks.
Class 1 material	Under the Online Safety Act 2021, this is material that is classified or is likely to be classified, as 'RC' according to the Australian Classification Board ratings.
Class 2 material	Under the Online Safety Act 2021, this is material that is classified or is likely to be classified, as either X 18+ or R 18+ according to the ACB ratings.
Classification Accuracy	The accuracy is the proportion of the sample that has been correctly classified as being over or under the age threshold.
Data security	how well the age assurance system safeguards users' personal information from unauthorised access, breaches or theft through, for example, the use of security by design principles and resistance to presentation attacks
Ease of use	how simple the age assurance system is to operate, including how the system offers functionality appropriate to the capacity and age of a child or adult, up to and including those of retirement age
Entry criteria	The set of conditions or requirements that must be met before the testing process can begin.
Exit criteria	The conditions or requirements that must be met to conclude the testing process
False Negative Rate (FNR)	The technology's miss rate (i.e., incorrectly identifying someone as being under the age threshold). It is the proportion of the sample who have been predicted as being under the threshold among those who are over the age threshold.
False Positive Rate (FPR)	The technology's probability of false alarm (i.e., incorrectly identifying someone as being over the age threshold). It is the proportion of the sample who have been predicted as being over the threshold among those who are not over the age threshold.
Minimisation of bias	how well the age assurance system avoids racial or other bias
Functional testing	Testing of a software system's functions to validate the implementation of its functional requirements
Human rights protections	how an age assurance system is accessible and inclusive to users, does not unduly restrict access of users who should

	have access and provides sufficient and meaningful information for a user to understand its operation.
Identity document	physical or digital document issued by an authoritative party containing identifying attributes of an individual.
Individual	Human being, i.e., a natural person, who acts as a distinct indivisible entity or is considered as such
Interoperability	how well the age assurance system can be used across multiple online platforms
Manual testing	Humans performing tests by entering information into a test item and verifying the results
Material	Under the Online Safety Act 2021, this is encompasses various forms of content, including films, computer games, publications and other types of audio, visual or audio-visual content.
Non-functional testing	Testing of a software system's functions to validate the implementation of its non-functional requirements, such as performance and security.
Online Platform	a service that allows end-users to access material using a carriage service; or a service that delivers material to persons having equipment appropriate for receiving that material, where the delivery of the service is by means of a carriage service This does not include a broadcasting service; or (a datacasting service (within the meaning of the Broadcasting Services Act 1992).
Outcome Error Parity (OEP)	A measure of the variance of an accuracy metric across subdivisions of the sample, such as gender skin tone.
Parental Consent	Consent from a person holding parental authority over a child
Parental Control	Parental control systems allow an adult responsible for a person under the age of 18 a degree of control over what content the child can see or hear.
Practice statement	Documentation of the practices, procedures and controls employed by an organization to fulfil a service
Presentation attack	Where an individual presents an external physical or digital object, such as a picture, pre-recorded audio or video clip or 3D masks to the age assurance system to circumvent it.
Presentation Attack Detection (PAD)	Automated discrimination between bona-fide presentations and biometric presentation attacks.
Protection of privacy	how well the age assurance system protects users' personal information, including data minimisation techniques
Reliability	how consistently the age assurance system can produce the same result
Relying party	Entity that relies on the results of age assurance to make an age-related eligibility decision.

Spoofing attack	When an individual attempts to fool the age assurance system, by presenting themselves to the system with artificial alterations such as wearing a fake beard.
Static Review	Evaluation of a test item where no execution of the code takes place and can be performed manually.
Successive validation	Type of age assurance process where multiple independent age assurance methods are used sequentially to establish an age assurance result
System testing	Testing of a software system's functional and/or non-functional performance against a set of requirements
Technology Readiness Level (TRL)	The technical maturity of a technology. They enable consistent, uniform discussions of technical maturity across diverse types of technology
Test case	A test case is a set of input values, execution conditions, expected results and procedures used to determine whether a specific feature or functionality of a system is working correctly.
Test data / Test dataset	data created or selected to satisfy the input requirements for executing one or more test cases
Test Item	Work product to be tested
Test procedure	Sequence of test cases in execution order, with any associated actions required to set up preconditions and perform wrap up activities post execution.
Test script	Document specifying one or more test procedures
Testing	Set of activities conducted to facilitate discovery and evaluation of properties of test items
True Negative Rate (TNR) – Also known as Specificity	The technology's ability to correctly detect people who are not over the age threshold. It is the proportion of the sample who have been predicted as being under the threshold among those who are under the age threshold.
True Positive Rate (TPR) - Also known as Sensitivity	The technology's ability to correctly detect people who are over the age threshold. It is the proportion of the sample who have been predicted as being over the age threshold among those who are over the age threshold.
User	Equivalent to Individual
User journey	The sequence of steps a user takes to accomplish a specific goal while interacting with a product, service or system.

8.2 Detailed Gantt Chart

The project plan, including a detailed Gantt Chart and full description of all work packages is available online[130].

8.3 Risk Matrix

The AATT's risk register is covered as part of Deliverable D-6.3 Risk Management Plan. The risk register is copied from that deliverable in Figure 6 below.

Risk Description & Causes			Mitigation & Contingency Actions		
KEY RISKS					
01. Data Ethics associated with Working with Human Test Subjects, including Aboriginal and Torres Strait Islander peoples. (WP1 - 6). There is the possibility that the Technology Trial may not address the critical risks associated with data, ethics and impartiality that will arise throughout the project when it comes to working with human test subjects.			STRATEGY: PREVENT. The creation of key project documentation and processes, including an Ethics Handbook (A-1.1.1), a Data Collection Ethical Protocol (A-1.1.3), a Human Test Subjects Protocol (A-1.3.1), and Application of the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research for Data Collection Phase (A-1.3.2) will address any issues relating to the data ethics of working with human test subjects.		
Likelihood 7/10	***Impact 8/10***	***Assessment 56/100***	***Likelihood 5/10***	***Impact 7/10***	***Assessment 35 /100***
02. Exclusion of Aboriginal and Torres Straight Islander peoples (WP1 -6). The Technical Trial may risk excluding certain demographics, including aboriginal people, from participation.			STRATEGY: PREVENT. The project needs to ensure that multi-ethnic diverse communities are included in the demographic spread of human test subjects. This will be achieved through the application of the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research, and initial and ongoing Equality, Diversity and Inclusion Monitoring.		
Likelihood 7/10	***Impact 9/10***	***Assessment 63/100***	***Likelihood 5/10***	***Impact 3/10***	***Assessment 15/100***
03. Lack of public trust in the project. There is the risk that the general public won't have trust in the age assurance technology trial and its goals and objectives, which would reduce project credibility.			STRATEGY: ACCEPT. This is a risk that can't be avoided; public trust can be subjective. However, Work Package 3 deals with Communications, including creation of a Project Website and blog, and aims to secure transparency and maintain public confidence in the project. The sub-contractor who will build the website has a wealth of experience and impressive portfolio in this field.		
Likelihood 6/10	***Impact 7/10***	***Assessment 42/100***	***Likelihood 3/10***	***Impact 4/10***	***Assessment 12/100***
MANAGERIAL RISKS					
M01. Project Time Line (WP6). As there are a number of interdependencies between the Work Packages, including requirements for completion of certain milestones before further activity can progress, this runs the risk of project over-run if milestones are missed. Failure to meet timelines would reduce alignment with progress in the age assurance space, as well as being unprofessional.			STRATEGY: REDUCE. The Project will be managed through the PM2 Quality Management Suite and in accordance with ACCS's Integrated Management System. The fortnightly Project Team Meetings, as well as the Monthly Contract Meetings, will maintain impetus and reduce the risk of project over-run.		
Likelihood 5/10	***Impact 7/10***	***Assessment 35/100***	***Likelihood 2/10***	***Impact 7/10***	***Assessment 14/100***
M02. Undertaking this project with Australian and UK sub-contractors. As a variety of different sub-contractors are involved in the technology trial, we may see some risk in project management owing to individuals moving at different speeds.			STRATEGY: PREVENT. This risk will be prevented by maintaining a well-planned Project Plan. Everybody involved in the project in their different capacities are well aware of the fact this is an 8 month project and a pro active working attitude and efficiency is key.		
Likelihood 5/10	***Impact 4/10***	***Assessment 20/100***	***Likelihood 2/10***	***Impact 4/10***	***Assessment 8/100***
M03. Engagement and Communications (WP3). Participants need to be recruited for the technology trial, including test subjects and age assurance providers, but there runs the risk of an insufficient number and types of participants being found.			STRATEGY: REDUCE. Maintaining a comprehensive and thorough Project Plan, including the establishing of a Project Advisory Board, hosting of an initial Project Stakeholder Event to set out the project plan, and outreach to at least six schools geographically spread across Australia to secure test subject participation, will massively reduce this risk.		
Likelihood 8/10	***Impact 7/10***	***Assessment 56/100***	***Likelihood 4/10***	***Impact 8/10***	***Assessment 32/100***
ENVIRONMENTAL, SOCIAL, GOVERNANCE (ESG) AND OTHER RISKS					
ESG01. The Call for Participation, seeking test subjects, including children, poses the risk that children may not be appropriately safeguarded in the technology trial.			STRATEGY: PREVENT. Creation of a thorough and well thought out Project Plan, taking into account the ethics of having children as participants, will prevent this risk. The recruitment process includes establishing the consent mechanisms (T1.1), safeguarding children (T1.3) and the ethical considerations surrounding this (T1.3).		
Likelihood 6/10	***Impact 9/10***	***Assessment 54/100***	***Likelihood 3/10***	***Impact 6/10***	***Assessment 18/100***

Figure 6 Risk Register

[130] https://ageassurance.com.au/wp-content/uploads/2024/11/D6.1-Project-Plan.pdf