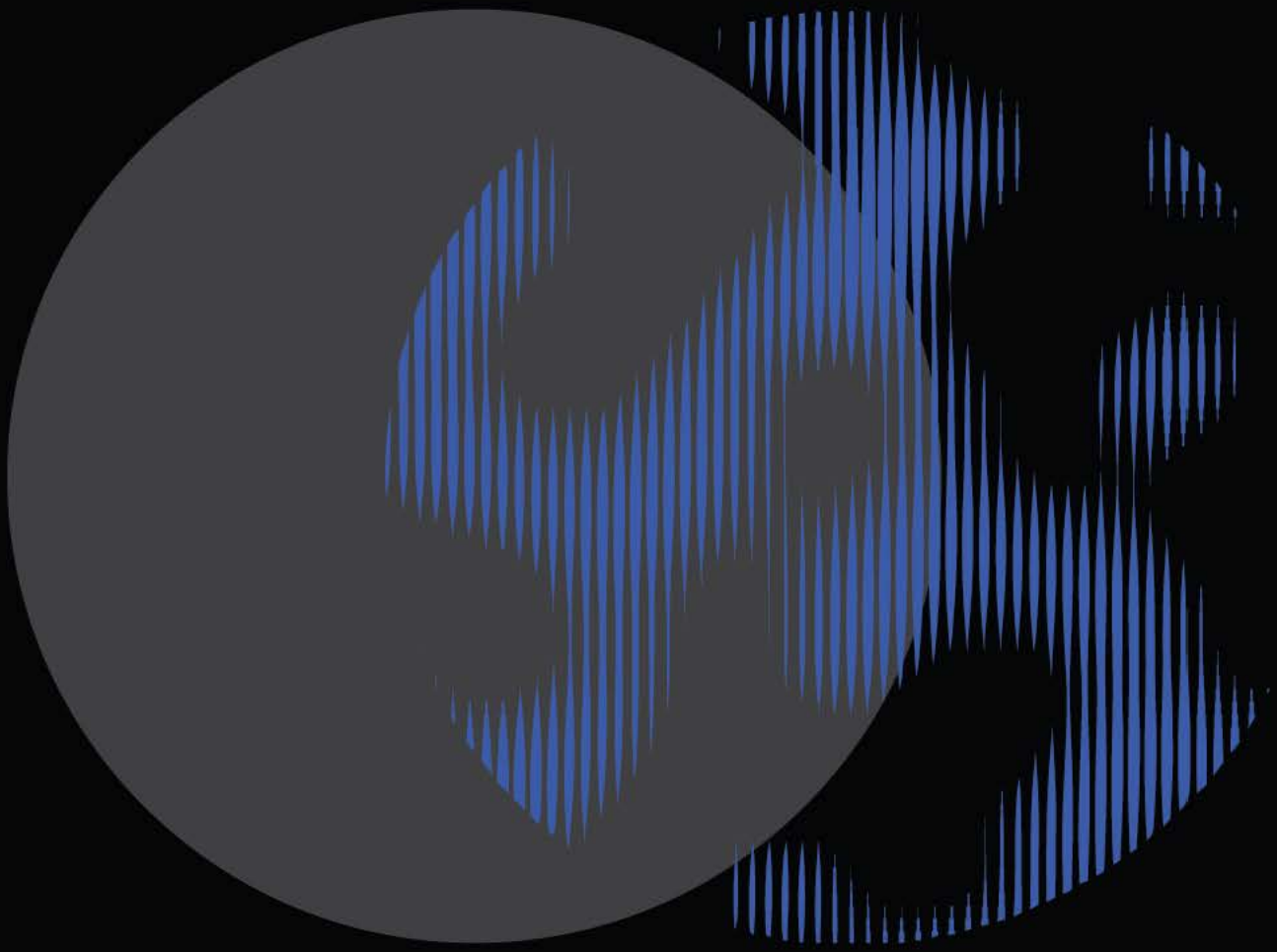
Human Technology
Institute



Inquiry into social media & Australian society

Submission to the Joint Select Committee on Social Media and Australian Society

July 2024

About the Human Technology Institute

The Human Technology Institute (HTI) is building a future that applies human values to new technology. HTI embodies the strategic vision of the University of Technology Sydney (UTS) to be a leading public university of technology, recognised for its global impact specifically in the responsible development, use and regulation of technology. HTI is an authoritative voice in Australia and internationally on human-centred technology. HTI works with communities and organisations to develop skills, tools and policy that ensure new and emerging technologies are safe, fair and inclusive and do not replicate and entrench existing inequalities.

The work of HTI is informed by a multi-disciplinary approach with expertise in data science, law and governance, policy and human rights.

For more information, contact us at hti@uts.edu.au

Authors: Sophie Farthing, Lauren Perry, Sarah Sacher and Prof Edward Santow

Acknowledgement of Country
UTS acknowledges the Gadigal people of the Eora Nation, the Boorooberongal people of the Dharug Nation, the Bidiagal people and the Gamaygal people upon whose ancestral lands our university stands. We would also like to pay respect to the Elders both past and present, acknowledging them as the traditional custodians of knowledge for these lands.

Executive summary

The Human Technology Institute (HTI) welcomes the opportunity to make a submission to the Australian Parliament Joint Select Committee on Social Media and Australian Society inquiry into the influence and impacts of social media.

This submission is informed by HTI's expertise on facial recognition technology and digital identity, and a detailed understanding of the regulatory environment for new and emerging technologies, including artificial intelligence (AI). In 2022, for example, HTI outlined a model law for facial recognition technology in Australia.[1] HTI has also provided independent expert advice to government on various areas including digital identity, and provided input to government and parliamentary consultation processes on the *Identity Verification Services Act 2023* (Cth) and Rules, as well as the *Digital ID Act 2024* (Cth).

This submission applies a human rights approach to analysing these complex policy areas. The Australian Government is required to protect human rights in online spaces, including by creating safe and accountable digital platforms and online environments. Proposals to exclude children from social media sites using age verification technology engage a range of human rights—both for children and for all Australian adults.

In this submission, HTI responds to two of the Terms of Reference (ToR) for this inquiry.

First, HTI considers **the use of age verification to protect children from social media (ToR (a))**. There are many ways to undertake both age verification and age estimation (referred to collectively as 'age assurance'), with differing human rights impact depending on the method and technology adopted.

All age assurance technologies engage human rights, particularly the right to privacy. Whether any limitation on human rights can be justified will depend, in large part, on whether it is reasonable, necessary and proportionate. HTI undertakes a human rights assessment of the proposal to restrict social media access to people aged 16 and over through the use of age assurance technology. We note that the Government and the Opposition are each separately considering this idea. The Government has committed to a $6.5 million pilot of age assurance technology to prevent children from accessing harmful online content, the details of which have yet to be publicly released.[2]

Second, HTI comments on **ToR (c), 'the important role of Australian journalism, news and public interest media in countering mis and disinformation on digital platforms'**. HTI notes the crucial role that Australian journalism, news and public interest media play in the digital ecosystem in upholding the rights of individuals to access information from diverse and informed voices, and to exercise their freedoms of opinion and expression. In turn, these processes help to counter the harmful impacts of mis and disinformation which is so prevalent across social media and other digital platforms.

HTI makes six recommendations based on these two Terms of Reference:

1. Any age-based restriction on social media access, and the associated use of age verification procedures, must comply with international human rights law. The Government should publicly explain how any proposed reform to this end would restrict human rights no more than is necessary and proportionate to protect children.

2. The Committee should recommend against using, for age verification, facial analysis or any other technology that would unjustifiably restrict human rights.
3. The Committee should recommend that the Government explore other means, beyond a blunt age-restriction law, to protect children from harm associated with access to social media.
4. The Committee should recommend that the Government introduce its proposed reform to Australia's Privacy Act as soon as possible. Reform should include safeguards for all Australians' personal information online, and specific protections for the personal information of children online.
5. The Committee should recommend that the Government develop legislation to regulate the development and use of facial recognition technologies in Australia, either through the Privacy Act or through stand-alone legislation.
6. The Committee should recommend that the Government introduce a Bill to address mis- and disinformation on digital platforms, with drafting that upholds human rights including freedom of expression.

ToR (a): The use of age verification to protect Australian children from social media

Understanding age assurance technologies

There are several existing laws that restrict access to certain goods and services by reference to an individual's age. For example, only people over the age of 18 are permitted to buy alcohol and tabaco, or to enter pubs, clubs and casinos. The efficacy of these legal rules relies in large part on a suite of age-verification procedures. These are the procedures by which people responsible for selling age-restricted goods and services must determine whether an individual meets the minimum age requirement.

Some age-verification procedures involve the handling of personal information, or even sensitive personal information. It is common for these procedures to involve manual document-checking, such as checking the date of birth on an individual's driver's licence or passport, or the use of human judgment where the individual appears to be clearly over or under the requisite age.

In other words, age restrictions are not a new phenomenon, nor is age verification a new phenomenon. However, new and emerging technology—including sophisticated record digitisation, artificial intelligence, and digital identity—offers novel procedures for carrying out age verification. Some of these new procedures enable age verification to take place online and at a population-wide scale. This can have far-reaching consequences (positive and negative) for a range of human rights, including the right to privacy.

It is critical to ground any discussion about age verification in a sound understanding of the technology and procedures being proposed. Different human rights implications will attach to different technologies. This inquiry's terms of reference include 'age verification', but this term does not connote a specific procedure or even a specific technology for

determining whether an individual meets a legal or other age restriction. That is significant because the efficacy (that is to say, the accuracy and reliability), as well as the impact on an individual's human rights, can vary dramatically depending on what specific procedure is used to conduct age verification. Without knowing which specific procedure, or technology, the Government may choose to mandate for age verification, it is very difficult to provide comprehensive advice to this Committee's inquiry.

This submission adopts the following definitions of three foundational concepts:

- **'Age assurance'** is the process of establishing an individual's age or age range. It is an umbrella term which refers to both age verification and age estimation methods.[3]
- **'Age verification'** implies a process of *accurately* determining a person's age, such as by checking a copy of someone's birth certificate before permitting them to obtain a learner's driving permit.
- **'Age estimation'** refers to less precise processes of inferring someone's age or the age range they fall into. For example, in NSW, anyone who *appears to look under 25 years old* may be asked by a security guard to provide proof of age when entering a licensed venue.[4]

Some technologically-enabled forms of age assurance, particularly those relying on facial recognition or facial analysis technology, can be particularly intrusive on the right to privacy, and a range of associated human rights such as the right to equality and non-discrimination. The *Privacy Act 1988* (Cth) provides that biometric data is 'sensitive information', and therefore subject to stronger protections than many other forms of personal information. However, the Privacy Act was drafted before the rise of many forms of biometric technology, such as facial recognition, became widely available and so it does not contain adequate safeguards for the full range of privacy violations that can arise following the misuse of such technologies.[5]

Two types of age assurance that rely on new technology—facial analysis and AI profiling—are particularly problematic. Each is dealt with in turn below.

The dangers of facial analysis

Facial analysis is a form of facial recognition technology which draws inferences about the characteristics of a person based on the physical features of their face. These techniques rely on *biometric information* to do this. Biometric information demands a higher level of privacy protection under the Privacy Act compared with 'ordinary' personal information. HTI is deeply concerned by some current reported uses of facial analysis for age assurance on social media platforms, including by Meta.[6]

Facial analysis differs from other forms of facial recognition which can be used in *identity* verification processes, like facial verification (one-to-one matching of a face to a single, stored image of that same face – as is used in many digital identity systems) and facial identification (one-to-many matching of a face within a broader database of face images).

Where facial analysis is used to assess characteristics about an individual, especially subjective characteristics such as an individual's mood or emotions, the technology can be subject to high rates of error.[7] While an individual's age is not subjective, in the sense that one's age is a question a fact, one's age is not immediately or readily apparent from one's face. This might be contrasted with a facial analysis tool that sought to identify people with blue or brown eyes.

While providers of age estimation technology claim high overall rates of accuracy, error rates can vary across demographic groups. There can be higher error rates in using facial analysis to estimate a child's age. For example, Yoti (the company engaged by Meta Australia in June 2024 to commence verifying the ages of some users) notes it has a Mean Absolute Errors rate of 1.4 years for 13–17-year-olds.[8] This raises concern for how effective this solution would be if rolled out at scale as a method for distinguishing between users under or over the age of 16.

The use of facial analysis technologies on children also raises elevated privacy concerns given the particular sensitivities around collecting biometric information of children, and their legal capacity to provide free, informed and otherwise genuine consent to this process. Some parents have indicated concern for this approach; a recent survey conducted with parents by the UK Children's Commissioner examined different methods of age assurance to restrict access to social media, with only 8% preferring the option of having their child's face scanned.[9] Even if a facial analysis tool purports to operate on an anonymous basis (in that the tool does not link its age estimation of a face with the identity of the individual whose face is being used), there remains a reasonable risk that any biometric information collected via this method could become linked to the individual's social media profile information, or it could be saved in a database for AI training or other purposes.

Facial recognition and analysis technologies are also historically less accurate for people of colour and people with disability.[10] While technical, lab-tested accuracy is improving year on year, the precision of these sorts of tools will decline once deployed in real-world settings.[11] This can be due to low light levels, unstable internet connections, or camera quality in users' personal devices – the exact conditions which many social media users would likely experience in their homes when faced with an age estimation app.

Finally, a number of case studies highlight just how easy these facial analysis tools are to circumvent. In June 2024, an Australian journalist applied an aging filter to an image of a child on their smart phone and successfully duped Yoti's age estimation app.[12]

The dangers of AI profiling

AI profiling refers to the automated analysis of personal data to make decisions or predictions about an individual. Personal information used in AI profiling can be collected across a wide range of sources and can include internet search data, online spending habits, social media engagement and surveillance data.[13]

AI profiling has been used to assess the age of a user based on their online behaviour. For example, a username, hashtag usage or IP address can all be used to estimate an individual's likely age range.[14] However, AI profiling cannot determine an exact age and has a wider margin for error as compared with other age estimation methods.[15] While some studies have applied machine learning analysis to social media profiles to ascertain demographic data for research, the results highlight that age prediction from online behaviour can be highly variable in accuracy.[16] There are also concerns that the behavioural indicators relied on for age estimation are subjective and based on unscientific assumptions of 'mature' online interactions, "conflating numeric age with life stage."[17]

The use of AI profiling for age estimation raises significant privacy concerns. AI profiling relies on the collection and analysis of personal information. However, individuals are often not meaningfully informed about how and when their data is being used. This impacts their ability to provide consent for the use of their information for age estimation purposes. Further, AI profiling can reveal highly personal information about a user beyond estimating their age. These processes rely on the 'mosaic effect'[18] of collating a trove of

behavioural and activity-based data which essentially can make an individual reasonably identifiable, irrespective of whether the AI-profiling tool claims to formally identify an individual or just estimate their age. The tool would then use this linked-up profiling data and its own AI-generated analysis to make a significant decision about that person's eligibility to access a social media service.

Using a human rights approach to analyse the proposal to use age verification to limit social media access for people under 16

As previously noted, there are many procedures or processes by which age assurance can take place. All rely on at least some personal information—be it, say, the date of birth on one's birth certificate, or biometric data in a facial analysis system. The extent to which the process collects, stores and uses that personal information depends on the technology and methodology adopted. This in turn determines the extent to which an individual's human rights will be affected by an age-assurance process.

Any decision to employ age verification as a means to enforce social media restrictions requires careful consideration. It is reasonable—indeed desirable—for the Government and Parliament to take steps to make children safe online, however, there is real complexity in determining what steps they should take, especially given some measures can have unintended consequences on the children who are intended to be protected.

In this submission, HTI has applied Australia's international human rights law obligations to some of the idea of applying age verification to restrict children under the age of 16 from access social media platforms. The Australian Government is bound to follow international human rights law. In addition, in the context of age verification to protect children from social media harm, human rights law is particularly helpful to analyse how to characterise competing interests and to balance those interests in crafting a solution.

In this submission, HTI applies a human rights analysis to assess the proposal to use age verification to protect children from social media harm. This involves assessing:

1. what, if any, human rights are affected by the proposal
2. whether the proposal pursues a legitimate aim
3. whether any limitation on human rights is lawful, necessary and proportionate to achieve the legitimate aim.

Given that this Committee is not considering a fully fleshed-out proposal for age verification, let alone a bill, any analysis about human rights compliance must make a number of assumptions about the technology adopted for age assurance or verification, and what safeguards are set out in law. We have set out what we consider to be the necessary assumptions in the remainder of this section of the submission, which follows a conventional three-step human rights analysis. On this basis, we consider that there are serious human rights risks in adopting a blunt minimum age requirement, backed up by current age verification or assurance procedures.

Step 1: the human rights affected by this proposal

As is increasingly being recognised in comparable democracies overseas, including most recently by the United States Surgeon General, social media platforms 'have not been proven safe' for children and young people.[19] While exclusion from social media platforms may seem a straightforward solution to protect children from online harms, in practice this proposal has significant implications that are both negative and positive for the human rights of children. More specifically, the proposal is likely to engage a range of human rights, including privacy and freedom of expression.

The right to privacy

The adoption of an age assurance procedure is likely to require all users to provide some personal information in order to access a social media service, not only children. This directly engages the right to privacy.

Any age assurance procedure for social media would also take place within a context of existing community mistrust about how personal data is being captured and used (or misused) by social media companies, which are engaging in unchecked harvesting of personal information.[20]

The right to privacy is a multifaceted human right, enshrined in international law by Article 17 of the *International Covenant on Civil and Political Rights* (ICCPR), to which Australia is a party. The right to privacy underpins many other fundamental rights—such as freedom from discrimination and freedom of association, religion, thought and expression—because it provides an important brake against the misuse and overuse of individuals' personal data. While the right to privacy is not an absolute right, this right cannot be limited or restricted arbitrarily. International law sets the default position that an individual's right to privacy must be respected.

Freedom of expression

In the digital world, social media platforms can be an important way for people to participate in cultural, social and political processes which support the flourishing of human life. Online communication, including via social media, can support the right to freedom of expression, encompassing the right to access information, communicate freely, and hold and express beliefs. Social media can also support related rights such as the right to assemble and freely associate. Some researchers have referred to the Arab Spring as an example where social media was used to support peaceful assembly and protest.

Social media can be an especially important gateway for young people under the age of 18 who possess fewer avenues to access information. Online platforms have become a forum for young people to develop their thoughts, opinions and beliefs, and contribute to democratic functioning. There also can be benefits for children and young people accessing education, health information and communities online. For young people with marginalised identities, such as those in the LGBTQIA+ community; and those who are physically or socially isolated, due to disability, living in a remote location, or abuse in the home, social media can provide access to life-saving support lines and communities. Studies have found, for example, that some LGBTQIA+ young people, including those from 'hostile home environments', find safe spaces via social media where they can meet supportive peers, and access online resources about LGBTQIA+ topics—including information about physical, mental and sexual health.[21] Other research highlights similar benefits for young people with disability; an Australian study found that learning to use social media led to increases in social participation among young people in rural areas with communication disabilities.[22]

Protection of children

There are also human rights imperatives associated with *restricting* children's access to social media. There is increasing evidence of significant harm associated with, and perhaps inherent in, the interaction of children and young people with social media platforms. There is a growing body of research, for example, supporting a connection between social media engagement and poor mental health outcomes for children and

young people.[23] There is also evidence of children suffering significant physical and psychological harm from online predatory behaviour on social media, such as sextortion scams; cyberbullying; and the use of generative AI to create and circulate offensive images of children.[24]

The United Nations Committee on the Rights of the Child has recognised the significant implications for children's privacy presented by 'routine' digital practices including 'automated data processing, profiling, behavioural targeting, mandatory identity verification, information filtering and mass surveillance'.[25]

Step 2: would excluding children from social media be a legitimate aim?

Any proposal regarding age verification in this area is premised on the idea that this would protect children under the age of 16 from harms associated with social media. The analysis above recognises that there are human rights benefits and harms associated with children accessing social media. The threshold question whether a blanket restriction on children under the age of 16 accessing social media constitutes a legitimate aim is, therefore, not straightforward to answer.

Given the margin of appreciation that international law affords to nation states in seeking to protect their communities from harm, it is likely that it would not be *inherently* contrary to Australia's international human rights law obligations to seek to restrict children under the age of 16 from accessing social media. While acknowledging that this is not certain, this submission proceeds from that assumption.

In any event, it seems clear that any such proposal necessarily would limit some human rights (especially the rights to privacy and to freedom of expression). Hence, whether this proposal is justified in limiting these human rights will turn on the third step in the human rights analysis—considering whether the specific law and age assurance or verification procedures adopted are reasonable, necessary and proportionate to achieve the protective aim that is sought.

Step 3: is age verification a reasonable, necessary and proportionate approach to protecting children from social media harm?

As noted above, if one assumes that restricting social media access to people over the age of 16 is a legitimate aim for the purposes of international human rights law, the next question is whether any restriction on human rights—especially on freedom of expression and the right to privacy for affected children and adults, as described above—is no greater than is reasonable, necessary and proportionate to achieve that aim.

Reasonableness and necessity

If we accept, for the purposes of this inquiry, there is a significant risk of harms for children, associated with access to social media, in principle a legal restriction on that access is not unreasonable.

Turning then to the question of necessity, excluding a child from a social media service is a blunt measure that would be difficult to implement in practice. One reason for this is that the proposal is unnecessarily broad. 'Social media' encompasses a wide range of online interactions. The *Online Safety Act 2021* (Cth) defines a 'social media service' to include those services 'where the sole or primary purpose is to enable online social interaction and where the service allows end-users to interact with other end-users and post material'.[26]

As noted in the eSafety Commissioner's submission to this inquiry, the aim of removing children from 'social media' to protect them from harm is difficult to achieve as social media 'forms part of a converged and integrated contemporary media environment' with a 'fluid interplay between social media, websites, messaging apps, gaming platforms, dating apps, as well as ephemeral media'.[27] This means that a child excluded from a social media platform such as Instagram, WhatsApp or TikTok, could easily be exposed to the same harmful content on a Google search engine, for example. In short, there is a real risk that this proposed legal restriction may not be sufficiently efficacious in protecting children from harm to meet the necessity requirement.

Parliament also should consider, therefore, other law and policy reform to achieve this aim. To this end, the National Children's Commissioner, Anne Hollonds, stated:

> Alternative and additional approaches to protecting children from online harm and securing their privacy should be considered alongside age verification and parental consent. These include stronger privacy protections for *all* users such as default privacy settings that are opt-in; requiring websites to be easily filterable by parental control software to better protect younger children; and providing education on human rights, online safety and privacy for parents and children.[28]

This is similar to the approach taken in some jurisdictions overseas. The Californian Age Appropriate Design Code, for example, doesn't impose restrictions on social media use, but rather requires companies to embed specific design features, privacy protection and harm minimisation settings to all products and services likely to be accessed by children under the age of 18.[29]

Other jurisdictions have taken a more targeted approach to protecting children online, such as requiring age verification for online content that is lawfully restricted to adults. The Canadian *Protecting Young Persons from Exposure to Pornography Bill S-210,* for example, proposes to restrict young persons' online access to sexually explicit material. Regulations will require a "prescribed age-verification method" to be used; for a method to be prescribed, it must be reliable, maintain user privacy, use personal information solely for age verification purposes, destroy any personal information once verification is completed, and comply with best practices for age verification and privacy protection.[30]

HTI also urges the Committee to take into account concurrent reform processes and consultations that have a bearing on the age verification proposal. Given the pace and breadth of law and policy reform related to new and emerging technologies across parliament, regulators and government, it is crucial that there is co-ordination and coherence. The outcomes of the current consultation reviewing the Online Safety Act review, for example, are likely to be relevant to the deliberations of the Committee, including its consideration of the question of whether reform is needed to require industry to act in the best interests of the child, and if Australia should legislate a statutory duty of care similar to the UK's *Online Safety Act 2023*.[31]

Proportionality

Where an age assurance process identifies an individual, particularly where sensitive biometric data is used to verify or estimate that person's age, there is considerable intrusion on the right to privacy. A measure, such as age assurance, will likely be considered a proportionate limitation on the right to privacy where it is the *least restrictive means possible* to achieve the harm-prevention aim. In this context, any procedure for age assurance will need to be scrutinised by reference to the following sorts of questions:

- To what extent is personal data being collected, stored and used beyond that which is absolutely necessary to fulfil the age verification task?
- Is sensitive biometric data involved, or other sensitive information about the user and the nature of their online activities?
- In what circumstances would personal data be shared with others beyond the organisation running the relevant social media platform?
- Is the age verification procedure designed in a way that preserves the anonymity of relevant individuals, to the maximum extent possible?
- Is personal data being retained, and by whom? Is any age-related or other data being deleted immediately? Is data being retained by the social media platform, and potentially used for other purposes, such as targeted advertising, or brokered to a third party?

Some age assurance technologies, such as the age estimation methods (facial analysis and AI profiling) outlined earlier, are likely to have a disproportionate negative impact on the privacy of *all* social media users given the amount of sensitive personal information that will need to be collected by a private company. Relying on these types of technologies to exclude children from social media also risks normalising the collection of sensitive biometric data by private companies, and can facilitate 'function creep'. In Australia, we have already seen an increase in businesses' adoption of non-consensual face scanning and scraping of face data, as evidenced by the practices of Clearview AI,[32] Bunnings and Kmart.[33]

Conversely, some age verification technologies may offer greater privacy protections by ensuring that *less* personal information is collected, used and disclosed. For example, the use of a well-constructed and tightly-regulated digital identity system could prevent the identification of individuals seeking to verify their age for the purpose of accessing a social media platform. This may be the least intrusive age verification process currently available, and is being considered in comparable jurisdictions overseas—the European Union Taskforce on Age Verification, for example, is currently looking into restricting access to adult online content using the European Digital Identity Wallet.[34]

In order to comply with Australia's international human rights law obligations, an age verification system relying on a digital identity must be clearly established in law, with robust safeguards that:

- impose rigorous technical and cybersecurity standards to protect the privacy of users' personal information and their online activities
- ensure system usability and equal access to services for all entitled users
- guarantee strict use limitations on collected data, to ensure that data can only be used for the immediate purpose of verifying age in that exact use context
- provide access to remedy should age assurance processes fail, leading to harms such as identity fraud or being arbitrarily blocked from accessing goods and services.

With the above principles in mind, and subject to the two caveats described below, HTI considers that undertaking age verification through the Australian Government's legislated Digital ID scheme could present an option that minimises the negative human rights impact, as compared with other age verification and assurance procedures.

However, there are two immediate caveats to this approach in the context of considering the current age verification proposal. First, while the Digital ID Accreditation Rules allow for people to set up a Digital ID from the age of 15, not everyone will possess the required

documentation to do so—and this may be especially problematic for teenagers aged 16-18 years. Therefore, there would need to be alternative, privacy-protecting mechanisms for individuals to verify their age and avoid being arbitrarily denied access to social media platforms. Second, while there is now federal legislation in place to govern the use of digital identity, there are flaws in that legislative scheme, including the ability for law enforcement to access sensitive personal data at a low threshold.[35]

In addition, as noted above, the Privacy Act, which regulates the handling of biometric data and other personal information, is in need of major reform. Australia's privacy legislation reform is long overdue, and it is as yet unclear how proposed privacy reform, anticipated in the second half of 2024, will tackle some of the more difficult questions raised by the use of age verification.

There is also no dedicated law for facial recognition technology in any Australian jurisdiction, despite the significant privacy implications of the increasing use of such technology both domestically and overseas. Following the publication of HTI's world-leading report outlining a model law for facial recognition, HTI has called for the introduction of specific laws governing the use of facial recognition technologies (including facial analysis tools) to adequately protect Australians from the very real risks of surveillance and discrimination.[36]

HTI's recommendations based on ToR (a):

1. Any age-based restriction on social media access, and the associated use of age verification procedures, must comply with international human rights law. The Government should publicly explain how any proposed reform to this end would restrict human rights no more than is necessary and proportionate to protect children.
2. The Committee should recommend against using, for age verification, facial analysis or any other technology that would unjustifiably restrict human rights.
3. The Committee should recommend that the Government explore other means, beyond a blunt age-restriction law, to protect children from harm associated with access to social media.
4. The Committee should recommend that the Government introduce its proposed reform to Australia's Privacy Act as soon as possible. Reform should include safeguards for all Australians' personal information online, and specific protections for the personal information of children online.
5. The Committee should recommend that the Government develop legislation to regulate the development and use of facial recognition technologies in Australia, either through the Privacy Act or through stand-alone legislation.

ToR (c): The important role of Australian journalism, news and public interest media in countering mis and disinformation on digital platforms

Independent media underpins trust in Australian democracy

Independent journalism, news and public interest media underpin a well-functioning democracy. The media has a vital role to inform the public, fact-check falsehoods especially in the context of political discourse, and provide a platform for a diversity of voices and public debates. The importance of a healthy information environment has been recognised as foundational to Australian democracy, including by implication in the Australian Constitution.

In recent years, Australians have increasingly received information and news from digital platforms. Social medial platforms like Facebook, Instagram, X and TikTok are said by some to resemble a 'digital town square', where people share news, information, and engage in political debates and advocacy. (That claim is also disputed by others, especially on the basis that the companies operating these platforms generally exist for profit-making purposes, rather than to provide a social good.) However, unlike the physical public square, digital platforms, in their current form, are not 'neutral' platforms. Their AI-based business model shapes our access to information, and the debates we have in new and sometimes insidious ways. Fundamentally, digital platforms have undermined public access to verifiable information about civil and political issues.

It is estimated that approximately 4 billion people globally use social media platforms.[37] The extraordinary rise of social media has been accompanied by a rise in mis- and disinformation, fuelled particularly during times of uncertainty and heightened social anxiety (such as during the COVID-19 pandemic or USA elections).[38] The growing availability of generative AI applications, especially in the last 18 months, has also provided new avenues to create and spread mis- and disinformation.

Recent studies indicate declining public trust in authoritative statements of fact and events, even when they are supported by clear evidence, expertise or traditionally reliable sources, and a downward trend of confidence in government amongst OECD countries, including Australia.[39] Unless this trend is reversed, we are likely to see further erosion of the quality of our public discourse, and a loss of legitimacy for our democratic institutions.

In this context, the importance of Australian journalism, news and public interest media cannot be overstated. The challenge, considered by this Committee and elsewhere, is how to ensure access to reliable, truthful and diverse information and media in the context of the online environment.

How Australians receive news and information

There is a clear distinction between how Australians receive news and information on digital platforms compared with traditional print and broadcast media.

First, while both traditional for-profit media and digital platforms rely on advertising revenue, there is huge scope for digital platforms to secure considerably higher revenues by intensive data collection about users to personalise their news feed. Personal data

feeds the engines of algorithms, which are designed to increase the level of engagement users have with content presented on their feeds (this is sometimes referred to as the 'attention economy'). Content that is highly emotional, divisive, shocking and/or controversial is more likely to receive higher audience attention and interaction. The algorithms that sort content on social media platforms tend to prioritise this kind of content and are capable of sharing it rapidly and at great scale with audiences, often at the expense of truthful content. As a result, factual, balanced reporting can be deprioritised or drowned out, while false or misleading news likely to illicit emotion is frequently bumped up and reshared.

Additionally, social media content is personalised and dynamic, users are presented vastly different information on their feeds, which over time affects perceptions of truth and widens the gap between individual experiences of reality. This can create echo chambers and feed polarisation. It can be difficult for users to understand the nature of content curation on their feeds, due to this personalisation and the opacity of the platforms. Australians today are less likely to benefit from a shared understanding of relevant public information which was more characteristic of the age of legacy media.

It is also becoming increasingly difficult for people to source or fact check information on social media. For example, in the past, the content which Facebook users saw on their home pages was presented chronologically, and it was possible to scroll back to posts and engagements from days prior in order to review them. This is no longer the case, which makes it challenging to assess when and where information is coming from and hold authors to account. The advent of generative AI, with its capacity for synthetic content creation and rapid dissemination has supercharged the challenge of discerning fact from fiction online. As deep fakes, bots, and other forms of AI tools become more realistic and prevalent on social media, people are less able to trust the evidence they see and read.

Regulation is required to address the challenges posed by mis- and disinformation

HTI supports policy and law to protect access to independent and fair media and public interest journalism, including to counter mis- and disinformation.

Digital platforms are already subject to Australian laws, and regulator oversight via bodies such as the Australian Communications and Media Authority (ACMA) and the eSafety Commissioner. The problem, however, is that the growth of mis- and disinformation online, and especially via social media, is not well addressed by existing law. While journalists and media organisations are subject to some laws designed to promote truthful content, those laws generally do not apply to social media and other digital platforms. Given the challenges posed by mis- and disinformation on digital platforms, it is incumbent on government to act.

Regulation in this area is difficult to draft in a way that upholds freedom of expression and of the press. There is a risk that overly-tight regulation of news and online content can limit these freedoms unduly. That said, there are numerous examples of Australians laws that appropriately engage the right to freedom of expression. The *Broadcasting Services Act 1992,* for example, contains provisions to maintain a standard of journalism in line with public expectations, and it provides consequences when these standards are not met.[40]

The challenges posed by regulating digital platforms to address mis and disinformation are well known, as was evident in the in-depth consultation on the Exposure Draft Communications Legislation Amendment (Combatting Misinformation and Disinformation)

Bill 2023. In that consultation, concerns were raised regarding the Exposure Draft Bill's potential impact on freedom of expression, including due to its broad definitions of mis- and disinformation, and the low threshold of harm established. These concerns, including those raised by the Australian Human Rights Commission, some civil society organisations and the legal profession,[41] highlight the potential for unintended consequences, and the need to adopt a human-rights centred approach when redrafting and consulting on the next version of the Bill. On the other hand, HTI considers that reform to address mis- and disinformation is necessary and so the Government should move swiftly to introduce a Bill that sets an appropriate balance.

In addition, regulation to address mis- and disinformation should not be considered in a vacuum, but rather alongside current law reform process and exercise of regulatory powers. How businesses use personal information, for example, including in relation to targeted advertising, is being considered in the context of privacy reform.[42] Similarly, the work of the eSafety Commissioner to ensure the accountability of digital platforms, such as requiring transparency around algorithmic recommender systems and takedown notices for harmful online content, as well as the ongoing review of the Online Safety Act, is also relevant to the questions being considered by the Committee.[43]

HTI's recommendations based on ToR (c):

6. The Committee should recommend that the Government introduce a Bill to address mis- and disinformation on digital platforms, with drafting that upholds human rights including freedom of expression.

---

[1] Davis, N, Perry, L & Santow, E (2022) *Facial Recognition Technology: Towards a model law*, Human Technology Institute, The University of Technology Sydney.
[2] See: Karen Middleton, 'Albanese follows Dutton's lead with tougher position on children's social media ban' (13 June 2024) *The Guardian*; Australian Government Department of Prime Minister and Cabinet, 'Tackling online harms' (Media Release, 1 May 2024); Josh Taylor, 'Children could be clocked from social media as well as porn sites in Australian trial of age check technology' (2 July 2024) *The Guardian*.
[3] eSafety Commissioner, *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography* (Report, March 2023) 16.
[4] 'Evidence of Age', *Liquor and Gaming NSW* (Web Page).
[5] Davis, N, Perry, L & Santow, E (2022) *Facial Recognition Technology: Towards a model law*, Human Technology Institute, The University of Technology Sydney; Corge, J and Svantesson, D 'The five generations of facial recognition usage and the Australian privacy law' *International Data Privacy Law 2024*, ipae007.
[6] Chris Burt, 'Facebook introduces Yoti age estimation in Australia ahead of global rollout: Nation grapples with teens' social media use', *Biometric Update* (online, 6 June 2024).
[7] Davis, N, Perry, L & Santow, E (2022) *Facial Recognition Technology: Towards a model law*, Human Technology Institute, The University of Technology Sydney, 16.
[8] 'Facial Age Estimation white paper', *Yoti* (Blog Post, 15 December 2023).
[9] 'What we've learned about methods of age assurance on social media', *Children's Commissioner for England* (Blog Post, 22 November 2023).
[10] See for example, United General Services Administration (GSA) *Executive Order 13985 – Equity Action Plan* (20 January, 2022), 8; Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (Conference Paper, Conference on Fairness, Accountability and Transparency PMLR 81, 2018) 77; K. S. Krishnapriya et al, 'Characterizing the Variability in Face Recognition Accuracy Relative to Race' (Conference

Paper, IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,2019) 2278; Inioluwa Deborah Raji and Joy Buolamwini, 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products,' (Conference Paper, AAAI/ACM Conference on AI, Ethics, and Society, Association for Computing Machinery, 2019) 429.

[11] Patrick Grother, Mei Ngan and Kayee Hanaoka, Face Recognition Vendor Test (FRVT) Part 2: Identification (No NIST IR 8271, National Institute of Standards and Technology, September 2019) 6.

[12] Cam Wilson, 'How to Fool a Selfie AI Age Verification Tool in Seconds with a Simple Filter Trick' (14 June 2024) Crikey.

[13] 'What is automated individual decision-making and profiling?', *Information Commissioner's Office* (Web Page).

[14] eSafety Commissioner, Submission No 191 to Standing Committee on Social Policy and Legal Affairs, House of Representatives, *Inquiry into age verification for online wagering and online pornography* (November 2019) 8.

[15] 'How do you check age online?', *Age Verification Providers Association* (Web Page); Nina Cesare et al., 'How well can machine learning predict demographics of social media users' (Research Paper, 6 February 2017), 8.

[16] Karen O'Connor et al., 'Methods and Annotated Data Sets Used to Predict the Gender and Age of Twitter Users: Scoping Review' (2024) 26 *Journal of Medical Internet Research*; Nina Cesare, Christian Grant and Elaine Nsoesie, 'Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best Practices' (Research Paper, February 2017); James Marquardt et al., 'Age and Gender Identification in Social Media' (Conference Paper, Conference and Labs of the Evaluation Forum, 2014).

[17] Nina Cesare, Christan Grant and Elaine Nsoesie, 'Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best Practices' (Research Paper, February 2017).

[18] The Centre for Humanitarian Data, 'Mosaic effect' (n.d), United Nations Office for the Coordination of Humanitarian Affairs (OCHA) <https://centre.humdata.org/glossary-2/mosaic-effect/>

[19] Chapman, M 'US surgeon General wants cigarette-style warnings placed on social media' *Sydney Morning Herald* (18 June 2024).

[20] Justine Humphry, Catherine Page Jeffery, Jonathon Hutchinson and Olga Boichak 'Age verification for social media would impact all of us. We asked parents and kids if they actually want it' *The Conversation* (22 May, 2024); Office of the Australian Information Commissioner (2023) *Australian Community Attitudes to Privacy Survey 2023,* Office of the Australian Human Rights Commissioner, Australian Government; David Swan, 'Facebook, Instagram Are Using Your Data and You Can't Opt Out' *Sydney Morning Herald* (13 June 2024).

[21] Linda Charmaraman, 'Social media gives support to LGBTQ youth when in-person communities are lacking' The Conversation (28 September 2021); See also Committee on the Rights of the Child, General Comment No. 25 (2021) UN Doc CRC/C/GC/25 [50].

[22] Parimala Raghavendra et al, 'Enhancing social participation in young people with communication disabilities living in rural Australia: outcomes of a home-based intervention for using social media' Disability and Rehabilitation (2015) 37(17), 1576-90.

[23] For example, Jonathan Haidt, *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness* (Penguin Press, 2024).

[24] For an outline of the type of threats facing children online see: Department of Infrastructure, Transport, Regional Development, Communications and the Arts *Statutory Review of the Online Safety Act 2021, Issues Paper* (April 2024).

[25] UN Committee on the Rights of the Child, General Comment No. 25 (2021) on children's rights in relation to the digital environment, para 68.

[26] eSafety Commissioner *Summary of Reasons – Social Media Services Code* (31 May, 2023); s 13 *Online Safety Act 2021* (Cth).

[27] eSafety Commissioner, Submission to the Joint Select Committee on Social Media and Australian Society (21 June, 2024), 11, available here.

[28] Australian Human Rights Commission *Submission to the Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures Bill 2021)* (1 December, 2021).
[29] *California Age-Appropriate Design Code Act,* Cal, AB-2273 (2023).
[30] Clauses 6, 11.
[31] Department of Infrastructure, Transport, Regional Development, Communications and the Arts, 'Statutory Review of the Online Safety Act 2021' (2021) Department of Infrastructure.
[32] Office of the Australian Information Commissioner, Media Release, 'Clearview AI Breached Australians' Privacy' (2 November 2021) OAIC.
[33] Office of the Australian Information Commissioner, Media Release, 'OAIC opens investigations into Bunnings and Kmart; (12 July 2022) OAIC.
[34] The EU Taskforce on Age Verification, with the European Commission, recently launched a proof-of-concept project using the European Digital Identity Wallet. See, European Commission 'Second Meeting of the Task Force on Age Verification' News Article (20 March, 2024).
[35] Human Technology Institute, *Submission to the Senate Economics Legislation Committee on the Digital ID Bill 2023 and the Digital ID (Transitional and Consequential Provisions) Bill 2023* (23 January 2024).
[36] See Davis, N, Perry, L & Santow, E (2022) *Facial Recognition Technology: Towards a model law*, Human Technology Institute, The University of Technology Sydney.
[37] See: Wong, 'Top Social Media Statistics And Trends Of 2024'(18 May 2023), *Forbes Online.*
[38] Sadiq Muhammed T and Saji K. Mathew, 'The Disaster of Misinformation: A Review of Research in Social Media' (2022) 13(4) *International Journal of Data Science and Analytics* 271.
[39] See e.g., Organisation for Economic Co-operation and Development (2023), *Trust in Government*, doi: 10.1787/1de9675e-en; Sora Park et al, 'Global mistrust in news: The impact of social media on trust' (2020) 22(2) *International Journal on Media Management* 83.
[40] Parliamentary Library, 'Media Policy and Regulation: A Quick Guide' (31 May 2022) Parliament of Australia.
[41] Australian Human Rights Commission, *Finding balance: combatting misinformation and disinformation without threatening free expression, Submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts* (18 August 2023); Digital Rights Watch, *Submission: Combatting Misinformation and Disinformation Online* (30 August, 2023); Law Council of Australia, *Submission, Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023 – Exposure Draft* (29 August, 2023).
[42] Attorney-General's Department *Government Response: Privacy Act Review Report* (Commonwealth of Australia, 2023).
[43] See, for example, eSafety Commissioner, 'Tech companies grilled on how they are tackling terror and violent extremism' Media Release (19 March 2024); Department of Infrastructure, Transport, Regional Development, Communications and the Arts, 'Statutory Review of the Online Safety Act 2021' (2021) Department of Infrastructure.