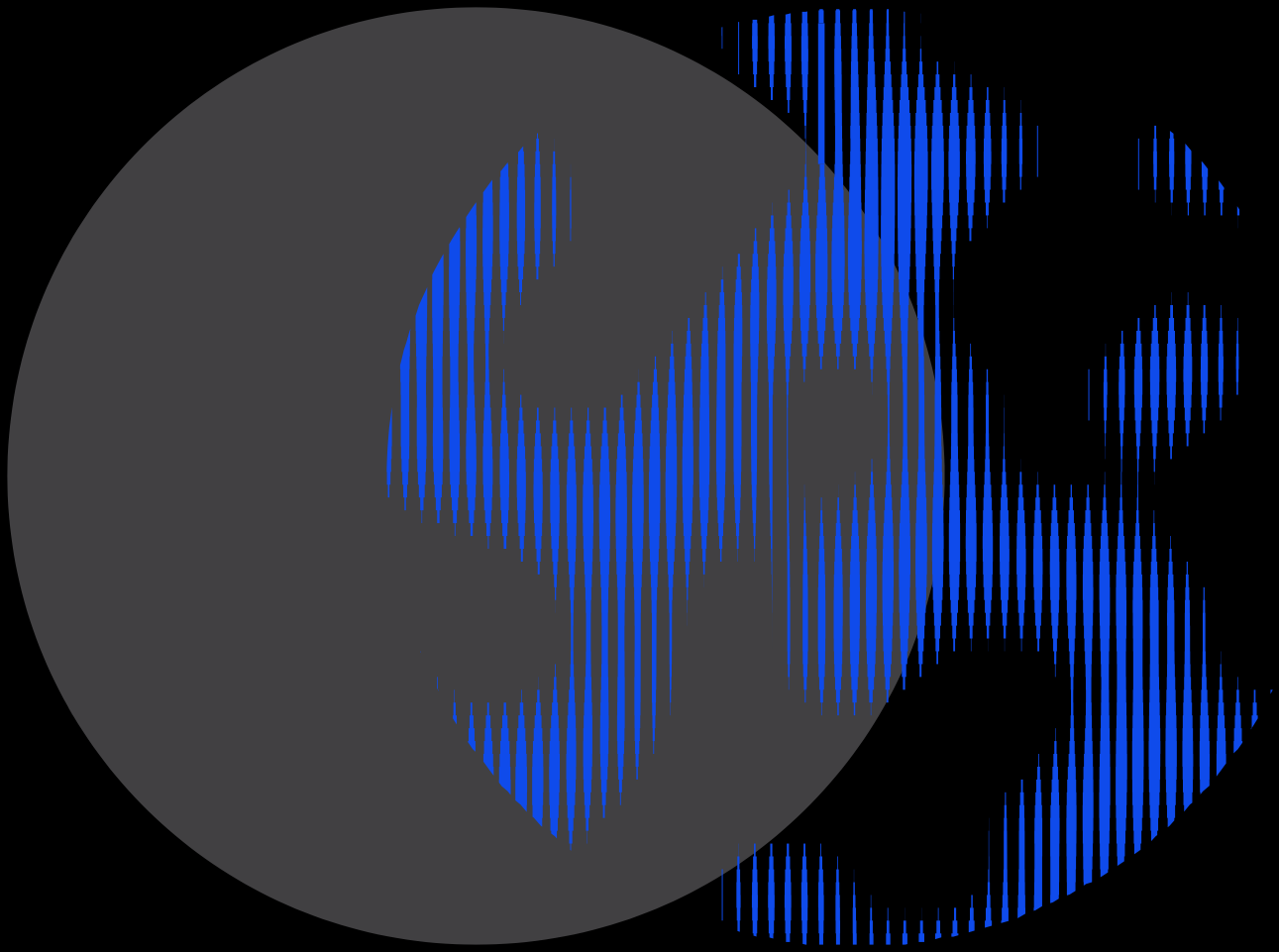
Human Technology Institute



Statutory Review of the Online Safety Act

Submission to the Department of Infrastructure, Transport, Regional Development, Communication and the Arts

5 July 2024

5 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 Professor 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.

5 July 2024

Contents

About the Human Technology Institute 2

Executive summary 1

Safeguarding Australians online 3

A human rights approach to ‘online safety’ 3

The human rights engaged by online activities 3

Balancing human rights in regulating online safety 4

The protection and promotion of human rights should be included as an objective of the *Online Safety Act 2021* 5

Proposed changes to the remit and powers of the eSafety Commissioner 6

Accountability and oversight of the eSafety Commissioner 7

Age assurance and age verification 7

Understanding age assurance technologies 8

Age assurance and age verification to restrict access to online pornography: human rights analysis 10

Executive summary

The Human Technology Institute (HTI) welcomes the opportunity to provide this submission to the Statutory Review of the *Online Safety Act 2021* (Cth).

This submission is informed by HTI's expertise in the human rights implications of new and emerging technologies. This includes 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 Australia, there is growing political impetus to address a broad range of harms that arise online, particularly considering new challenges raised by the rapid advent of generative AI technologies. These online harms and questions of 'safety' intersect with complex social policy questions, contested democratic values and live public debates.

The statutory review of the Act is taking place against a background of a range of law and policy reforms that are relevant to the remit of the eSafety Commissioner. Long overdue reform of Australia's privacy law, for example,[1] is likely to be relevant to the powers and focus under the Act. Also relevant is the Australian Government's commitment to protect Australians from the harms posed by high-risk AI, and the ongoing work of the Department of Industry, Science and Resources to develop mandatory guardrails (i.e., legislative reform) in respect of the development, deployment and use of AI.[2]

HTI does not seek to answer all the questions in this significant and in-depth review; rather, it focuses on two key issues.

First, HTI recommends that the objectives of the Online Safety Act be expanded to include 'the protection and promotion of human rights'. The Australian Government is required to protect human rights in online spaces, including by creating safe and accountable digital platforms and online environments. Including human rights as a third objective would ensure the eSafety Commissioner considers and balances human rights in the exercise of her powers under the Act.

Secondly, HTI applies a human rights approach to the consultation proposal to use an age assurance process to restrict access to online pornography. 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 methods 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. In this part of the submission, HTI undertakes a human rights assessment of the proposal to use age verification to restrict children under 18 to age-appropriate content through the use of age assurance technology.

Recommendation 1

In order to promote online safety under the *Online Safety Act 2021* (Cth):

- **the objects clause of the Act should be amended to incorporate the purpose of upholding human rights, and balancing rights appropriately in the exercise of powers under the Act**

- **the Act should be assessed using a human rights approach. Any adjustments to existing powers, or proposals for new powers, should align with Australia's international human rights law obligations.**

Recommendation 2

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

Recommendation 3

- **Alternative means of addressing the harms of online pornography beyond age verification should be explored and invested in.**

Recommendation 4

- **Any form of age verification, facial analysis or any other technology that would unjustifiably restrict human rights should not be adopted.**

Safeguarding Australians online

Consultation questions

Question 1: Are the current objects of the Act to improve and promote online safety for Australians sufficient or should they be expanded?

Question 26: Are additional safeguards needed to ensure the Act upholds fundamental rights and supporting principles?

Question 28: What considerations are important in balancing innovation, privacy, security, and safety?

A human rights approach to 'online safety'

The *Online Safety Act 2021* (Cth) (**the Act**) has two objectives: to *improve* and *promote* online safety for Australians. The Act gives the eSafety Commissioner the function of promoting online safety, including by establishing a complaints systems for online bullying and abuse, a complaints and objections system for the non-consensual sharing of intimate images, and an online content moderation system.

The term 'online safety' is not defined in the Act, and there is no universally-agreed definition of what 'safety' means in the online context. 'Safety' is an exceedingly broad term, which can encompass many different, and potentially conflicting, policy and legal objectives.

At a high level, 'safety' may be understood as the prevention and mitigation of harms that arise online, but even that definition raises further questions about *which* harms are being considered and *how* they are being prioritised and balanced.

One common way of approaching this issue is to consider the definition of 'online safety' by reference to international human rights law. International human rights law exists to uphold individual dignity, and among other things to prevent many forms of harm to humans. As a body of law applied globally, albeit imperfectly, the harms to which international human rights law is addressed are clearly defined with a corpus of jurisprudence and expert analysis that assists in applying the law to address relevant harms. It also contains a mechanism, especially via the proportionality test, to allow non-absolute human rights and other legitimate interests to be balanced, without one concern swamping all others.

HTI acknowledges that there are other lenses through which one might legitimately view the issue of online safety. Nevertheless, adopting a human rights approach is useful in an online safety context to define potential harms, and to balance different stakeholder interests. It also reflects Australia's international law obligations to protect human rights in online spaces, including by creating safe and accountable digital platforms and online environments.

The human rights engaged by online activities

The online environment covers a broad range of activities and interactions, from social media to news websites, messaging apps, gaming platforms and dating apps. A range of human rights are engaged in this context, covering both the realisation of some rights, while at the same time limiting the enjoyment of others. In many ways, the

protection and promotion of human rights is central to establishing safe online environments.

The right to privacy, for example, is *essential* to online safety. The right to privacy is a multifaceted human right, enshrined in Article 17 of the *International Covenant on Civil and Political Rights* (ICCPR). 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 shield against unnecessary or unwanted interference, including via 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. Strong privacy protections ensure that individuals can safely engage online—without being subjected to doxxing, abuse or stalking by a violent partner, harmful targeted advertising, identity fraud, data breaches or personalised scams—among many other examples.

Other human rights engaged in the context of online safety include the foundational principles of dignity, equality and mutual respect; the right to freedom of association and non-discrimination; and the right to be free from exploitation, violence and abuse. Several UN reporting mechanisms have incorporated the digital dimension of human rights abuse in their interpretation of international human rights law; the United Nations Committee on the Elimination of Discrimination Against Women, for example, has identified the need for legal reform to protect women from technology-facilitated violence.[3]

Children are particularly vulnerable in online environments. But while children must be protected from harms online, from mass surveillance, to exposure to harmful violent material, their rights to privacy, participation, freedom of expression and association in online environments must also be realised. To put this bluntly, simply denying children access to vast online spaces is no real solution to the genuine problem that children face serious threats to their safety online. On the contrary, adopting the approach laid out in the *Convention on the Rights of the Child* (CRC),[4] governments must find a more nuanced approach that facilitates children's online engagement while simultaneously addressing the things that can cause them harm.

Balancing human rights in regulating online safety

The eSafety Commissioner has an important role in upholding the rights of children and adults, by protecting them from harms associated with cyberbullying, cyber-abuse, non-consensual image-sharing, and exposure to illegal or age-inappropriate material. It is important, when exercising those rights, that these are reconciled with a broader range of human rights, including the human rights to privacy, freedom of expression, and children's rights.

A good example of the complexity of interests that arise in the online safety context is in relation to end-to-end encryption, considered in detail in **Box 1**, below.

In this complex regulatory environment, the proportionality test, recognised in international human rights law, is a practical tool that can support effective regulation. The proportionality test allows for relevant non-absolute human rights to be weighed and balanced, on the basis that certain human rights may be subject to limitations where those limitations are lawful, and can be demonstrably justified in a free and democratic society.

In determining whether a limitation of a human right is reasonable and justified, several factors should be considered. These include whether the limitation is in pursuit of a legitimate purpose, such as online safety, whether the limitation has a rational

connection to the purpose to be achieved, and whether the limitation is necessary to achieve that purpose.

In the second part of this submission, HTI outlines a proportionality approach to the proposal to use age verification to restrict the access of children to age-inappropriate content.

Box 1: End-to-end encryption

The Issues Paper states that measures such as end-to-end encryption, which can serve to uphold the right to privacy, also can result in potential 'online safety harms'. It observes that end-to-end encryption is an 'important defence against security breaches that would otherwise have serious consequences for online users', but concludes that encrypted communications can 'conceal harmful conduct or hinder investigation of the distribution of harmful and illegal [material]'.[5]

HTI submits that it would be simplistic to characterise end-to-end encryption as itself a threat to online safety. While law enforcement may view end-to-end encryption as a challenge, many powers and investigative techniques are available to combat crime. Law enforcement powers are not absolute. The limitation of those powers is important to prevent arbitrary interference with individual rights, including the right to privacy. Notably, the right to privacy protects against excessive state surveillance, which chills rights to freedom of expression and association, and undermines democratic principles and the rule of law.

End-to-end encryption can bring a number of benefits related to online safety: it can increase digital security and the integrity of communications; it can reduce the risk of cybercrime; and it can safeguard journalistic sources, activists and those fleeing violence. Encrypted communications also protect against interference by foreign actors, who can exploit encryption 'back doors' to gain access to information that undermines Australia's national security.[6]

This is quintessentially an area where a balance is needed. It would be disproportionate to undermine end-to-end encryption—compromising cybersecurity, privacy, safety and freedom of expression for Australians—to enhance administrative convenience for law enforcement and regulators. Where law enforcement and other regulatory bodies have other tools to fulfil their functions, including those that aim to protect the community from harm, without compromising the integrity of communications, they should be expected to access those other tools.

The protection and promotion of human rights should be included as an objective of the *Online Safety Act 2021*

As explained above, in fulfilling the Office's functions under the Online Safety Act, the eSafety Commissioner should weigh up a range of considerations, some of which need to be held in tension.

One way to ensure the Act better equips the eSafety Commissioner to exercise these functions and powers to address the full range of online harms would be to include the protection and promotion of human rights as a regulatory object in the Act itself. This would help to ensure the interests of all stakeholders are taken into account and balanced when the eSafety Commissioner exercises her powers in law and policy.

The inclusion of human rights as a regulatory objective is consistent with several current provisions in the Act, as well as proposed changes outlined in the Issues Paper. The eSafety Commissioner is already required, for example, to have regard, where appropriate, to the *Convention on the Rights of the Child* in the performance of her functions.[7]

Similarly, a human rights approach would support the incorporation of a new expectation being considered as part of the review of the Basic Online Safety Expectations, which would require regulated services to consider the best interests of the child principle.[8]

In addition, a human rights approach would support interoperability of the work of the eSafety Commissioner with other online regulators in overseas jurisdictions. A human rights approach to the regulation of online safety has been expressly adopted by the Global Online Safety Regulators Network, of which Australia's eSafety Commissioner is the current Chair.[9] The Network has recognised the importance of coordination and coherence in domestic regulatory approaches to online services, particularly given the specific challenges for individual countries regulating 'a global industry that has accrued significant scale, power and resources'.[10] To support effective, internationally-interoperable regulatory approaches, the Network has adopted a human rights approach. In its 2023 'Position Statement on Human Rights & Online Safety Regulation', the Network articulated support for

> an approach to regulation that is human rights-respecting and proportionate and recognises the shared ethical duty for governments, regulators, businesses and service providers to preserve the human rights and dignity of users online, to mitigate and prevent online harms, and promote user safety, empowerment and autonomy.[11]

Recommendation 1

In order to promote online safety under the *Online Safety Act 2021* (Cth):

- **the objects clause of the Act should be amended to incorporate the purpose of upholding human rights, and balancing rights appropriately in the exercise of powers under the Act**
- **the Act should be assessed using a human rights approach. Any adjustments to existing powers, or proposals for new powers, should align with Australia's international human rights law obligations.**

Proposed changes to the remit and powers of the eSafety Commissioner

The Issues Paper canvasses a range of potential harms that could be addressed by the Online Safety Act and the eSafety Commissioner. These include consideration of powers to address cyberflashing, volumetric attacks, online hate, technology facilitated abuse, online abuse of public figures, body image harm, self-harm promotion, and 'boasting' about crimes.

The precise meaning of some of these harms will require clarification, and should be defined in a way that is consistent with international human rights law. For example, harms such as 'online hate' engage competing democratic values and rights. Some

types of hate speech are already defined in law—most notably in respect of racial hatred in the *Racial Discrimination Act 1975* (Cth). The legal definition for any new categories of hate speech, and any defences such as those based on freedom of expression, should be set out in primary legislation as well.

In this way, the eSafety Commissioner's content moderation powers should be carefully circumscribed in legislation that has been the subject of full parliamentary debate and review. More specifically, any new powers given to the eSafety Commissioner should be clearly defined in primary legislation; address a harm that is not currently covered by existing powers or laws; be specifically targeted to that harm with safeguards to prevent over-reach; and exercised in the context of a human rights assessment that takes into account countervailing rights.

Accountability and oversight of the eSafety Commissioner

To ensure that human rights considerations are properly taken into account, and human rights interferences curtailed, the eSafety Commissioner's powers must be subject to appropriate accountability measures and safeguards.

Limits on the exercise of discretionary powers by the eSafety Commissioner, and procedural fairness requirements, should be clearly outlined in the primary legislation. This is an important rule of law measure, that would also prevent inadvertent interferences with human rights by both current and future office holders.

Suggestions for improvements along these lines were raised when the Online Safety Bill was first introduced to Parliament. The Parliamentary Joint Committee on Human Rights (PJCHR) report on the Bill made a number of recommendations for amendments. For example, the PJCHR noted that Part 9 of the Bill (the Online Content Scheme), included overly-broad discretion.[12] The PJCHR observed, for example, that there is no requirement in law for the eSafety Commissioner to 'consider the context or purpose for which [online material] was published in determining whether to issue a remedial notice', including to take into account 'public interest' considerations. The PJCHR concluded that Part 9 is not 'sufficiently circumscribed such that it constitutes a permissible limitation on the right to freedom of expression'.[13]

HTI suggests that the PJCHR's recommended safeguards to address this concern, along with its other recommendations, be implemented in future amendments to the Online Safety Act.

Individual exercises of powers under the Online Safety Act should also be procedurally fair, transparent and reviewable. There are opportunities to improve existing processes in this regard. For example, through thorough reporting requirements that illustrate the *reasoning* behind content moderation decisions and informal requests (in relation to removal notices, blocking requests and so on); and reporting on the nature and outcomes of requests for internal reviews of decisions. The eSafety Commissioner's annual report does not include this level of detail.[14]

Age assurance and age verification

Consultation question

Question 12: What role should the Act play in helping to restrict children's access to age-inappropriate content (including through the application of age assurance)?

HTI recently made a submission to the Joint Select Committee on Social Media and Australian Society's *Inquiry into social media and Australian society.* In that submission, HTI assessed the Government's proposal to restrict access to social media to those over 16 years, through the adoption of age verification tools. We refer to that submission for a detailed human rights analysis of age verification in relation to social media.[15]

Each proposed use case for age verification should be assessed independently, in light of the particular human rights issues that are raised. The Issues Paper explores the possibility of using age verification technology to restrict access to pornography. The eSafety Commissioner previously released a *Roadmap for age verification,* which also focused on pornography restrictions.[16] HTI therefore focuses on age verification in respect of pornography access in this submission, noting that no specific model or proposal has been put forward in the Issues Paper.

This part of the submission:

1. explains age-assurance technologies, noting that there are a range of options available, with differing privacy implications
2. outlines a detailed human rights analysis of the use of age verification technologies in respect of access to pornography.

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 tobacco, 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. 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.

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.[17]
- **'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.[18]

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 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.[19]

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 draw 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.[20]

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.[21] 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.[22]

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.[23] 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 other data collected about the individual, 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.[24] While technical, lab-tested accuracy is improving year on year, the precision of these tools can substantially decline once deployed in real-world settings.[25] 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.[26]

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.[27]

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.[28] However, AI profiling cannot determine an exact age and has a wider margin for error as compared with other age estimation methods.[29] 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.[30] 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."[31]

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'[32] 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 decisions regarding access to restricted materials.

Age assurance and age verification to restrict access to online pornography: human rights analysis

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.

While it is reasonable—indeed desirable—for the Government and the eSafety Commissioner to take steps to make children safe online, there is real complexity in determining what steps they should take, especially given some measures can have unintended consequences for the human rights of both adults and children.

HTI has applied Australia's international human rights law obligations to the idea of applying age verification to restrict children under the age of 18 from accessing pornography. The Australian Government is bound to follow international human rights law. In addition, in the context of age verification to protect children from 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 the eSafety Commissioner has not outlined a specific proposal for age verification in the Issues Paper, our analysis about human rights compliance must make a number of assumptions about the technology that may be adopted for age assurance or verification, and what safeguards may be 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.

Step 1: What human rights would be affected?

Age verification and age assurance measures engage the right to privacy. The adoption of an age assurance process is likely to require all users to provide some personal information in order to access pornography, not only children. This directly engages the right to privacy (Article 17, ICCPR). 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.

Any restrictions on online content and communication engages the right to freedom of expression (Article 19, ICCPR). Freedom of expression is also not an absolute right and can be limited – including by law for the purpose of protecting children from harm.

Age verification measures also engage children's rights, which are enshrined in the CRC, including children's right to privacy (Article 16). Age verification measures seek to realise children's rights to protection from harm (Articles 19, 3, 34). HTI is not expert in the impact of pornography on children and young people. However, we note that while studies are mixed, and often qualified by difficulties showing causation, there is evidence indicating that viewing pornography, particularly at an early age, can be harmful to children, and may adversely affect their healthy development.[33]

Step 2: Would restricting access to pornography to people under 18 be a legitimate aim?

As noted above, there is evidence of harms linked to children being exposed to pornography online. Protecting children from harm is a legitimate aim under human

rights law and so it may be assumed that preventing children from being exposed to pornography is a legitimate aim.

Whether pursuing age verification in respect of pornography is justified in limiting human rights will turn on the third step in the human rights analysis—considering whether potential law and age assurance or verification procedures that may be adopted are reasonable, necessary and proportionate to achieve the protective aim that is sought.

Step 3: is age verification reasonable, necessary and proportionate approach to protecting children from the harms of pornography?

Reasonableness

If we accept that there are risk of harms for children associated with access to pornography, the introduction of a tool to restrict access to pornography based on age is not unreasonable in principle.

Existing law recognises these harms and community expectations to address them. Current state and territory laws prohibit adults from selling or showing pornography to young people—although it is not illegal for someone under the age of 18 to view pornography.[34]

It is reasonable that legal and related measures aimed at inhibiting access, by children, to pornography should apply in both online and offline environments. If a person wants to purchase pornography in a physical store, they will be asked to present an identity document. It is reasonable also for government to consider what would be an appropriate corresponding measure in the online environment.

The eSafety Commissioner has existing powers relating to online pornography. The Online Content Scheme relies on the National Classification Scheme to assess explicit material. Depending on the classification level, and whether the material in question is provided from Australia, the Commissioner can issue enforceable removal notices, or enforceable remedial notices requiring content to be placed behind a 'restricted access system'.[35] The Online Safety (Restricted Access System) Declaration 2022 sets out criteria for access control systems—such as requiring a person to declare they are over 18, and incorporating reasonable steps to confirm the person is at least 18.[36]

However, there are a particular considerations that arise when pornography restrictions are enforced through age verification tools. This is considered below.

Proportionality

Least privacy restrictive approach

Where an age verification 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 or verification, 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 verification or 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?

There are also specific privacy considerations associated with pornography, which must be taken into account. Depending on how age verification tools handle data, there is a risk that users will have data related to their pornography habits stored, hacked, sold or shared.[37] Such data has the potential to reveal sensitive personal information, such as a person's sexuality, which is a protected attribute under discrimination law. Data may also be used by bad actors to extort or blackmail others with embarrassing or reputation-damaging information.[38] Adults over the age of eighteen should be able to engage in legal activity, such as consensual viewing of pornography, with the expectation of privacy. These particular sensitivities further underline the importance of ensuring that the most privacy protective model for age verification is chosen.

Some age assurance technologies, such as the age estimation methods outlined earlier, are likely to have a disproportionate negative impact on the privacy of *all* users given the amount of sensitive personal information that will need to be collected by a private company. Relying on these types of technologies 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,[39] Bunnings and Kmart.[40]

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 pornography. 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 considering restricting access to adult online content using the European Digital Identity Wallet.[41]

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, not everyone will possess the required documentation to do so. Therefore, there would need to be alternative, privacy-protecting mechanisms for adults to verify their age and avoid being arbitrarily denied access to age-restricted material. 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.[42]

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.

International example: France

In France, the Digital Space Regulation Law (*SREN Law*) mandates that websites and video-sharing platforms which broadcast pornography content implement age verification systems to ensure it is not accessible to minors.[43]

The age verification method used must comply with the technical standards which may be updated as needed based on the opinion of the French Data Protection Authority (CNIL).[44] A public consultation is currently underway for the proposed standards which cites facial biometric analysis with liveness detection and verification of physical identity documents as acceptable examples.[45]

CNIL, in partnership with cryptography researchers, has developed an open-source 'double anonymity' model for age verification. It adds a digital intermediary between a restricted website and an age-verification service. The system prevents the website from accessing information that could identify a user beyond their age, while a third-party age verifier cannot detect which site a user is visiting.[46]

Appropriately targeted to the harm being addressed

If age verification processes in relation to pornography are not carefully targeted, there is a potential for over-reach that arbitrarily interferes with freedom of expression.

Terms such as obscenity and pornography are notoriously difficult to define, and pornography restrictions have a history of being used to police sexuality based on subjective morality standards.[47] More recently, there have been instances where major

technology platforms, such as those provided by Alphabet (the company that owns Google) and Apple, in an attempt to moderate explicit material, disproportionately targeted and censored LGBTI+ content.[48]

In the present case, much depends on how pornography is defined, what content is required to be age-restricted, and how that content is practically restricted across different online platforms. Relevantly, the eSafety Commissioner has noted that the National Classification Scheme, which forms the basis of the current regulatory framework under the Online Safety Act for online pornography, was created in a context that is 'very different to the modern online environment'. The Commissioner has also reported stakeholder feedback that the Scheme is 'outdated and problematic'.[49] This points to a need to re-assess approaches to identifying and classifying pornography, before age verification measures are introduced.

Additionally, safeguards should be included to prevent the use of age verification for the restriction of content and communications beyond pornography – for example, to include legal and non-explicit content deemed offensive or inappropriate – as this would likely tip the balance into unjustifiable policing of freedom of expression.

To ensure a proportionate approach, the process for identifying and restricting pornography should be clarified in primary legislation, so that it is sufficiently circumscribed, with associated accountability and transparency requirements for both online platforms, and the eSafety Commissioner.

Necessity

There are practical difficulties associated with age verification as a means of effectively achieving the legitimate purpose. This may point to the need to consider alternative policy measures.

There are complexities associated with the blurred boundaries of the online environment, where pornography is generally free of charge, and available on a range of platforms beyond dedicated pornography websites, including on social media. In this way pornography is distinct from other use cases where age verification is more straightforward, such as to restrict alcohol purchases—it is comparatively easy to distinguish transaction points and identify restricted products in the latter scenario.

Age verification that is limited to major pornography sites may also result in the unintended consequence of people accessing riskier, unmoderated sites, including on the dark web, to avoid age verification tools. Others may simply adopt a VPN as an easy workaround.[50]

There are also hurdles associated with attempting to regulate major international platforms. France and German regulators are currently in protracted litigation against a number of pornography websites, arguing that they failed to restrict access to pornography.[51] When age verification laws were introduced in Virginia, Mississippi and Utah, Pornhub made the decision to block all residents in those states from access to its website, rather than complying with age verification requirements.[52]

Aside from the practical difficulties, there are deeper questions around whether age verification is the best possible policy response to address harms associated with pornography consumption by young people. The eSafety Commissioner currently

pursues parallel measures, such as education for parents and children, and safety by design initiatives, both of which are crucial to achieving the policy intent. Meanwhile, implementing effective age-verification tools will be a complex and resource intensive endeavour – particularly to the privacy standard that is required—and results are not guaranteed. There is no silver bullet. Our Watch has observed that 'simplistic approaches that seek to simply ban or discourage [children] from watching [pornography] are unlikely to be effective'.[53] Another Australian study found that that 'systems being proposed to automate age verification…divert resourcing that could be spent on strategies that are proven to support healthy sexual development.'[54]

In light of the potential pitfalls of pursuing age-verification, consideration should be given to pursuing alternative measures that would achieve the same policy intent.

Conclusion

The extent to which age verification measures are proportionate will depend on the model chosen, and the construction of any future laws. Proposals for age verification will only be proportionate if the privacy settings are right, and the law appropriately targeted. In the absence of strong Privacy Act reforms, HTI would caution against pursuing age verification at this time, as it will be difficult to guarantee respect for the right to privacy without this building block in place.

Recommendations

2. **Any age-based restriction on pornography access, and the associated use of age-verification procedures, must comply with international human rights law. The Government and eSafety Commissioner should publicly explain how any proposed reform to this end would restrict human rights no more than is necessary and proportionate to protect children.**
3. **Alternative means of addressing the harms of online pornography beyond age verification should be explored and invested in.**
4. **Any form of age verification, facial analysis or any other technology that would unjustifiably restrict human rights should not be adopted.**

[1] See, e.g., Human Technology Institute, Submission to the Privacy Act Review Report consultation (March 2023); Human Technology Institute, 'Reform to Australia's privacy laws takes a step forward' (Web Story) 28 September 2023; Human Technology Institute, Submission to the Attorney-General's Department, *Consultation on doxxing and privacy reforms* (March 2024)

[2] Department of Industry, Science and Resources, Safe and responsible AI in Australia consultation (Interim Response, 17 January 2024) 5 <https://storage.googleapis.com/converlens-au-

industry/industry/p/prj2452c8e24d7a400c72429/public_assets/safe-andresponsible-ai-in-australia-governments-interim-r>.
[3] Council of Europe *The digital dimension of violence against women as addressed by the seven mechanisms of the EDVAW Platform'* (Thematic Paper, 17 November, 2022).
[4] Committee on the Rights of the Child, General Comment No. 25 (2021) UN Doc CRC/C/GC/25 [13].
[5] Department of Infrastructure, Transport, Regional Development, Communication and the Arts, *Statutory Review of the Online Safety Act 2021: Issues Paper* (Aprile, 2024) 40.
[6] See, e.g., Mieke Eoyang and Michael Garcia, *Weakened encryption: The threat to America's National Security* (Report, Third Way Cyber Enforcement Initiative, September 2020) <https://www.thirdway.org/report/weakened-encryption-the-threat-to-americas-national-security>.
[7] *Online Safety Act* (Cth) s 24.
[8] Department of Infrastructure, Transport, Regional Development, Communication and the Arts, *Amending the Online Safety (Basic Online Safety Expectations) Determination 2022: Consultation Paper* (November 2023) 4 <https://www.infrastructure.gov.au/sites/default/files/documents/amending-the-online-safety-basic-online-safety-expectations-determination-2022-consultation-paper-november2023.pdf>.
[9] eSafety Commissioner (Australian Government) *The Global Online Safety Regulators Network* (Web page, 27 May 2024).
[10] Global Online Safety Regulators Network *Position Statement: Regulatory coherence and coordination: the role of the Global Online Safety Regulators Network* (April 2024).
[11] Global Online Safety Regulators Network *Position Statement: Human Rights & Online Safety Regulation* (September 2023).
[12] Parliamentary Joint Committee on Human Rights, *Human Rights Scrutiny Report* (Report 5 of 2021, April 2021) 67.
[13] Parliamentary Joint Committee on Human Rights, *Human Rights Scrutiny Report* (Report 5 of 2021, April 2021) 67.
[14] See e.g. Australian Communications and Media Authority and eSafety Commissioner, *Annual Report 2022-23* (2023) <https://www.esafety.gov.au/sites/default/files/2023-10/ACMA-and-eSafety-Commissioner-annual-report-2022-23.pdf?v=1719975316484>.
[15] Human Technology Institute, Submission to the Joint Select Committee on Social Media and Australian Society's *Inquiry into social media and Australian society* (July 2024).
[16] eSafety Commissioner, *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography* (Report, March 2023).
[17] eSafety Commissioner, *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography* (Report, March 2023) 16
[18] 'Evidence of Age', *Liquor and Gaming NSW* (Web Page)
[19] 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.
[20] 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)
[21] Davis, N, Perry, L & Santow, E (2022) *Facial Recognition Technology: Towards a model law*, Human Technology Institute, The University of Technology Sydney, 16.
[22] 'Facial Age Estimation white paper', *Yoti* (Blog Post, 15 December 2023)
[23] 'What we've learned about methods of age assurance on social media', *Children's Commissioner for England* (Blog Post, 22 November 2023)
[24] See for example, 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.
[25] 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.
[26] Cam Wilson, 'How to Fool a Selfie AI Age Verification Tool in Seconds with a Simple Filter Trick' (14 June 2024) Crikey
[27] 'What is automated individual decision-making and profiling?', *Information Commissioner's Office* (Web Page)
[28] 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.
[29] '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) <https://arxiv.org/abs/1702.01807> 8.
[30] 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* <https://www.jmir.org/2024/1/e47923/>; 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) <https://oudatalab.com/papers/cesare2017detection.pdf>; James Marquardt et al., 'Age and Gender Identification in Social Media' (Conference Paper, Conference and Labs of the Evaluation Forum, 2014)
[31] 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) <https://oudatalab.com/papers/cesare2017detection.pdf>.
[32] See 'Mosaic effect' *Centre for humdata* (Webpage) <https://centre.humdata.org/glossary-2/mosaic-effect/>.
[33] Australian Institute of Family Studies, *The effects of pornography on children and young people* (Research report, December 2017) <https://aifs.gov.au/research/research-reports/effects-pornography-children-and-young-people>; United Kingdom Children's Commissioner, *Evidence of pornography's influence on harmful sexual behaviour among children* (Report, May 2023) <https://assets.childrenscommissioner.gov.uk/wpuploads/2023/05/Evidence-on-pornographys-influence-on-harmful-sexual-behaviour-among-children.pdf> ; Gemma Mestre-Bach, Alejandro Villena-Moya and Carlos Chiclana-Actis, 'Pornography use and violence: A systemic review of the last 20 years' *Trauma, Violence & Abuse* (2023) 25(2).
[34] See e.g. *Classification (Publications, Films and Computer Games) Enforcement Act 1995* (NSW) s 13; *Crimes Act 1900* (NSW) s 66EB (2B).
[35] Online Safety (Restricted Access Systems) Declaration 2022 (Cth); eSafety Commissioner, *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography* (Report, March 2023) 21.
[36] Online Safety (Restricted Access Systems) Declaration 2022 (Cth); eSafety Commissioner, *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography* (Report, March 2023) 31.
[37] Zahra Stardust et al, 'Mandatory age verification for pornography access: Why it can't and won't "save the children"' *Big Data & Society* (2024) 11 (2) <https://journals.sagepub.com/doi/10.1177/20539517241252129?icid=int.sj-abstract.similar-articles.3>.
[38] Zahra Stardust et al, 'Mandatory age verification for pornography access: Why it can't and won't "save the children"' *Big Data & Society* (2024) 11 (2) <https://journals.sagepub.com/doi/10.1177/20539517241252129?icid=int.sj-abstract.similar-articles.3>.
[39] Office of the Australian Information Commissioner, 'Clearview AI Breached Australians' Privacy' (2 November 2021) OAIC <https://www.oaic.gov.au/newsroom/clearview-ai-breached-australians-privacy>.

[40] Office of the Australian Information Commissioner, 'OAIC opens investigations into Bunnings and Kmart; (12 July 2022) OAIC <https://www.oaic.gov.au/newsroom/oaic-opens-investigations-into-bunnings-and-kmart>.
[41] The EU Taskforce on Age Verification, with the European Commission, recently launched a proof of concept project using the European Digital Identity Wallet. https://digital-strategy.ec.europa.eu/en/news/second-meeting-task-force-age-verification
[42] 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)
[43] Osborne Clarke, 'New French rules on pornographic content to create a safer internet for minors' *Lexology* 16 May 2024 <https://www.lexology.com/library/detail.aspx?g=bc84de02-a88d-46dc-a33a-7bf1d4236a07>.
[44] Osborne Clarke, 'New French rules on pornographic content to create a safer internet for minors' *Lexology* 16 May 2024 <https://www.lexology.com/library/detail.aspx?g=bc84de02-a88d-46dc-a33a-7bf1d4236a07>.
[45] Adan, *Adan response to Arcom consultation on age verification standards* (June 2024) <https://www.adan.eu/en/publication/consultation-arcom-verification-age/>;
[46] Lauren Leffer, 'Online age verification laws could do more harm than good' *Scientific American* 16 April 2024 <https://www.scientificamerican.com/article/online-age-verification-laws-privacy/> ; Frank Hersey, 'Double anonymity to bring age verification to porn and social media in France' *Biometric Update* (Blog) 20 February 2023 <https://www.biometricupdate.com/202302/double-anonymity-to-bring-age-verification-to-porn-and-social-media-in-france>.
[47] Zahra Stardust et al, 'Mandatory age verification for pornography access: Why it can't and won't "save the children"' *Big Data & Society* (2024) 11 (2) <https://journals.sagepub.com/doi/10.1177/20539517241252129?icid=int.sj-abstract.similar-articles.3>.
[48] Kendra Albert and Afsaneh Rigot, 'Apple and Google still have an LGBTQ problem' *Wired* 16 August 2021 <https://www.wired.com/story/apple-google-lgbtq-apps/>.
[49] eSafety Commissioner, *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography* (Report, March 2023) 31.
[50] See e.g., Mack DeGeurin, 'Online porn restrictions are leading to a VPN boom' *Popular Science* 3 April 2024 <https://www.popsci.com/technology/vpn-boom/>.
[51] See, e.g., Arcom, 'Access of minors to pornographic sites: Referral to the President of the Paris Judicial Court' (Media Release) 8 March 2022 < https://www.arcom.fr/presse/acces-des-mineurs-aux-sites-pornographiques-saisine-du-president-du-tribunal-judiciaire-de-paris>; 'Administrative court confirms state media authority ban on Cypriot pornography websites' *IRIS Legal Observations of the European Audiovisual Observatory* (Web Page, 2023) <https://merlin.obs.coe.int/article/9799>.
[52] See e.g., Amanda Silberling, 'Pornhub blocks access in Mississippi, Virginia and Utah amid changing laws' *Tech Crunch* 3 July 2023 <https://techcrunch.com/2023/07/03/pornhub-blocks-access-in-mississippi-virginia-and-utah-amid-changing-laws/>.
[53] Our Watch, *Pornography, young people and preventing violence against women* (2020), 14 <https://media-cdn.ourwatch.org.au/wp-content/uploads/sites/2/2020/11/20022415/Pornography-young-people-preventing-violence.pdf>.
[54] Zahra Stardust et al, 'Mandatory age verification for pornography access: Why it can't and won't "save the children"' *Big Data & Society* (2024) 11 (2) <https://journals.sagepub.com/doi/10.1177/20539517241252129?icid=int.sj-abstract.similar-articles.3>.