Looking at Age Assurance Practicalities

Martin Thomson, 2025-08

A great many governments worldwide are enacting laws that seek to limit the ability of children to access content that is designated for adults only. At the most abstract of levels, this is a good thing. Some material requires maturity and experience to handle. It is worth ensuring that children are not exposed to things that they are insufficiently prepared for.

The challenge is in determining the right implementation; a question that has policy and technical aspects.

Any system that involves restrictions will unavoidably have some consequences for adults. This is appropriate, as protection of children is a social obligation shared by all. The difficult political question is then what costs — to factors like privacy and choice — are acceptable.

This brief looks at a much narrower question: given the technical solutions being developed in response to legislative efforts — or even those mandated in law — are there better alternatives?

This is inspired by comments from US Supreme Court Justice Kagan, who — in a dissent on a case regarding an age verification law in Texas — asks:

> *But what if Texas could do better—what if Texas could achieve its interest without so interfering with adults' constitutionally protected rights in viewing the speech H. B. 1181 covers?*

In a thicket of thorny policy questions, this is a question that submits to technical analysis.

This document starts by briefly summarizing popular approaches to age assurance and the consequences of each. It then examines two approaches that avoid many of the shortcomings identified in those systems. It concludes that Texas could indeed do better by adopting either alternative approach, though one in particular has distinct advantages over all alternatives.

Age Assurance Systems

A common approach taken in legislation is to levy the responsibility for protection of children on the distributors of content. Distributors are made responsible for ensuring that they do not provide adult content unless they are sure that the recipient is an adult. This generally results in distributors seeking proof of age before distributing content. Or exiting the affected market.

Legislation often does not stipulate exactly what level of proof is acceptable. Drafters of laws prefer to be technology neutral, instead providing guidelines or deferring to industry codes. This might give distributors some flexibility in terms of how they implement the requirement.

Age assurance can be technically complex, which means that distributors are often not in a position to manage compliance on their own. As a result a rich ecosystem of independent service providers have emerged.

Assurance Approaches

There are quite a number of different age assurance systems that are considered for deployment. Overwhelmingly, the technology tends to trace to government-issued identity documents, which is often labeled age *verification* as it provides strong evidence of age. The notable exception is a small class of age *estimation* techniques, notably including those that estimate age using AI models and face shape. Age estimation might entail a higher error rate than verification.

Age assertion, where people unilaterally assert their age without any supporting evidence, is increasingly considered unacceptable.

Ofcom in the UK lists 6 age assurance methods that they consider acceptable. Of these, facial age estimation is the only approach that does not ultimately trace to government-issued identity documents. Most methods rely on institutions that have existing Know-Your-Customer (KYC) obligations in law, such as banks, internet providers, and credit card companies.

In some jurisdictions, like Spain, a government agency provides an age assurance service directly. However, most systems involve a third party that takes responsibility for

performing the age assurance, reporting the outcome of the process to the content distributor.

This allows for a degree of centralisation. The government can look to the small set of age assurance services and make determinations about their practices. Some laws could mandate an accreditation scheme for these services as a precondition of use. Content distributors indemnify their risk of being found non-compliant by choosing reputable or accredited age assurance services. Using an external age assurance service also insulates distributors from any sensitive personal information.

What this does not address is the question of whether people — those subject to these requirements — can trust the system.

Privacy and Age Assurance

Knowledge about online activities of individuals, particularly when it is connected to verifiable identities, has great commercial value. It is also potentially very interesting to law enforcement.

A content distributor that performs age verification will receive identity documents or biometric data directly from people. This gives distributors the ability to connect identity to individual content consumption patterns.

A third-party age assurance service gains the same identity documents or biometrics, but does not gain detailed information about content consumption. They only learn which distributors are requesting age assurance.

Laws might require that information be adequately secured and even destroyed, but there are typically no technical safeguards in place that might prevent misuse of data. Many countries have provisions for retention and access to records by law enforcement that might override any privacy measures. Participants in the Australian trial were found to be overcollecting and over-retaining data in anticipation of retention requirements that did not exist. No law can safeguard against attacks or data leaks. No system can guarantee no risk of data leaks, unless it is for data that never enters the system.

Equitable Access

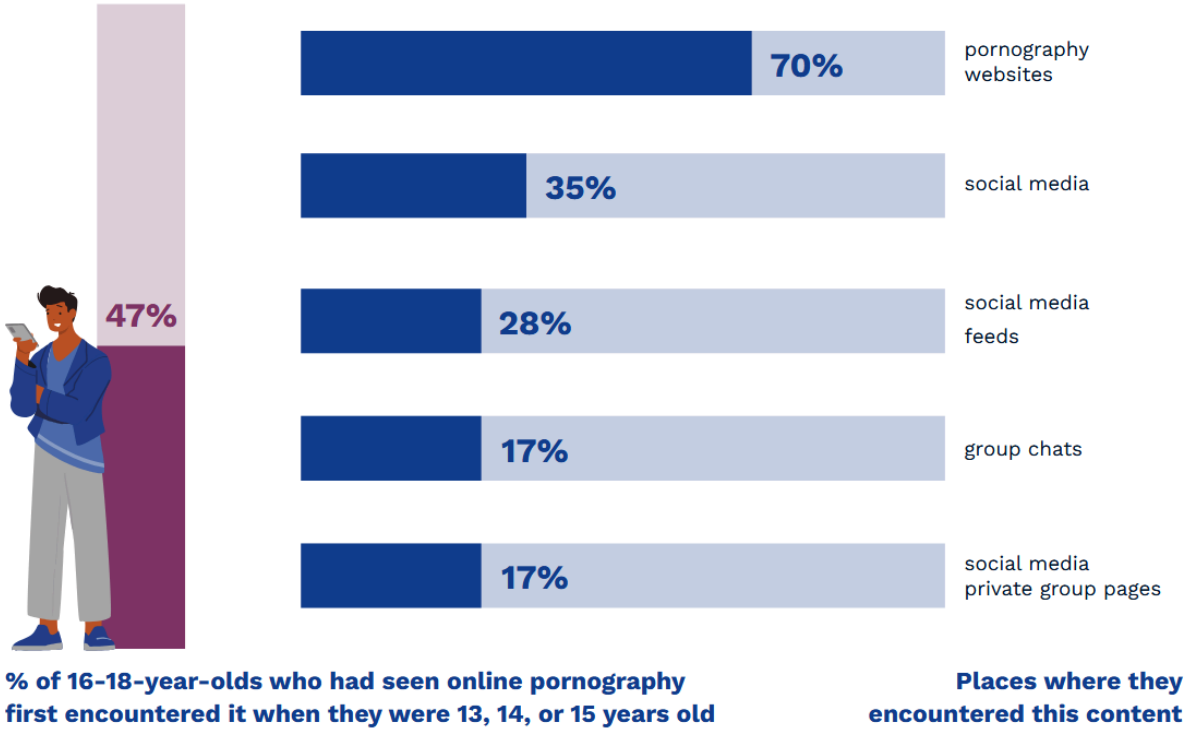
A reliance on government-issued identity documents creates a number of opportunities for unjustified exclusion, particularly for those who are unable to produce satisfactory evidence of age.

For age verification approaches, this can include those who are unable to produce recognized documents for any reason. This includes those who lack documentation, but also those whose documentation is not recognized. Jurisdictions might eventually agree on interoperable documentation, but until that happens, parallel implementation of laws will exclude many adults.

For age estimation, the likelihood of errors is a factor. A proportion of people over any given age threshold will fail to be correctly identified by something like facial age estimation. This error rate can be significantly higher for some demographic groups as these systems are often trained on biased training data. Gender and skin tone are known to have a significant effect on error rates.

Efficacy

It is understandable that governments seek to levy the costs associated with compliance on the industry that directly profits from the distribution of adult content. However, direct distribution of adult content is not the only way that children encounter adult content.



Source: Australian eSafety commission research (pdf) (please note the graphing errors)

This research suggests that a focus on the distribution of content by a small number of providers has some effect in addressing the risk of exposure, but does not offer a comprehensive solution. This research predates widespread use of generative AI systems, which introduce their own challenges. This shortcoming is often recognized, but rarely addressed.

Alternative Content Sources

Where demand for content exists and barriers are raised to access, there is the risk of people seeking alternative sources.

Legislation that is limited to specific jurisdictions and implemented by content distributors is trivially circumvented online by VPN users. Content distributors use crude tools like IP geolocation to determine which laws apply to any given transaction. A VPN user is able to access content as though they were geographically located in a jurisdiction of their choice.

There are sources of content that simply do not comply with local rules. This might be due to a lack of business presence in the jurisdiction in question, lack of awareness, or even

deliberate disobedience. Governments can prevail on ISPs and CDNs to deny such services carriage, but that sort of blocking is trivially avoided.

The use of messaging to exchange adult content offers challenges for a distributor-centric approach. So does the generation of content using AI.

Censorship, Surveillance, and Accountability

Any system that controls access to content carries a non-trivial risk of abuse by authorities, either for censorship or for surveillance.

For censorship, the process of determining whether content qualifies for restrictions is potentially open to abuse. For example, educational content on gender and sexuality can be a great use to teenagers who are still in the process of learning about themselves. Moreover, there is nothing inherent to the systems that ensures that they are only used to restrict access to sexually explicit material or to certain age groups.

Use of systems to censor or restrict access to content is therefore a significant risk. Systems for ensuring that classification is accountable are therefore necessary. Transparency might allow censorship-related abuse to be addressed through existing democratic accountability mechanisms.

Surveillance is a real risk of any system that involves sensitive information or sensitive content. This system includes both. Consider the Spanish "Pajaporte", which uses linkable tokens issued by the government. If the Ministry were to be compromised or if they chose not to delete issuance records, the tokens they provide to citizens could be used to link logs of content access to government identity records. Claims that the system presently is not usable for surveillance of this nature relies solely on claims by the Ministry that they are not *and will not ever* retain records. This is not the standard by which the internet has been developed and this is clearly an unacceptable privacy outcome.

In discussions of these features, analogies are often drawn to presenting a physical credential at a club. These are not always apt analogies, in part because of innate differences between online and offline interactions. Though these differences continue to erode, given that presenting credentials offline is frequently used for commercial and government surveillance purposes.

Acceptable Approaches

This section examines two approaches to achieving the outcome of ensuring that children cannot access adult content. In summary:

1. Zero-knowledge presentations of digital credentials can be effective, but it has privacy and equity of access issues that make it challenging to deploy at internet scale. The necessary technology is immature and any rollout would need to address several novel problems.
2. Device-based enforcement provides excellent privacy and equity characteristics. The main challenge in relying on devices is a shortfall in efficacy during early phases of deployment.

There are significant challenges with both, but there are significant differences in the severity of potential problems. Overall, device-based enforcement is:

- the most privacy-friendly,
- the most transparent and accountable,
- the most equitable,
- the best potential for comprehensive coverage,
- the least expensive to implement, and
- the best able to provide a smooth transition in implementation.

The primary cost of device-based approaches is that a smoother transition comes with a much longer deployment time scale. During this transitionary period, the efficacy of the system in terms of ensuring that content is not accessed by minors is lower than other alternatives. This is less of a drawback than a distinct advantage, as incremental deployment offers more learning opportunities than a system that depends on disruptive actions.

Zero-knowledge systems are the next best approach, but the technology is unproven at internet scale. That alone makes it difficult to recommend when compared to a device-based approach, but it seems technically feasible to construct a system that achieves key privacy and equity goals.

Zero-Knowledge Presentations of Digital Credentials

The use of zero-knowledge systems can allow people to present compelling evidence of their age, traceable to an authoritative source of information, but without revealing that source. The recipient of a proof can be convinced that the evidence exists, without ever receiving that evidence.

A great deal of focus has been given to the use of modern zero knowledge cryptography for this particular problem. In particular, systems that use government-issued credentials, with their high degree of assurance, to back the proofs that are generated. These are multi-show systems, where a single credential can be presented any number of times. This is appealing because it addresses another problem with other approaches by being generic: systems like Pajaporte reveal that someone is seeking restricted content when they access the system.

Of note here is the BBS+ signature scheme and Google's longfellow-zk, both of which provide the same basic capabilities. The former being significantly simpler than the latter, which adds considerable complexity in order to gain some level of compatibility with deployed consumer hardware.

Zero-knowledge systems are also the focus of investigations in the EU and trials in Australia (maybe: the term "double blind" is used to describe the mechanism, but that is not a concept used in the academic literature).

The core promise of these systems is that while compelling proof of age is conveyed, the recipient of that proof learns very little about the person offering that proof; at the same time, the source of that proof learns nothing about the use of the credential. It is also the case that the same credential can be used with different services, without those services being able to recognize that this has happened. This property is broadly known as "unlinkability".

Challenging Issues

While this technology is promising, it is worth recognizing that cryptography doesn't solve problems, it only transmutes them. A number of technical issues exist with building out a system based on zero-knowledge proofs.

The system creates significant vectors for reuse of credentials. This means that — absent additional protections — a single adult could generate many age proofs that any number of other people could use.

- Rate limiting is the best defense against this style of abuse, but that has limits, especially with respect to how effective it is at restricting credential sharing.
- Hardware attestations from wallets is a more popular approach, that being the backing of mass-market device vendors, like Apple and Google. This leads to exclusion (what about more open platforms or people without access to the necessary hardware) and could involve secondary flows of information with privacy consequences (attestations that can be traced to a particular vendor, device model, or even serial number range).

Unlinkability is broadly incompatible with revocation of credentials in case they are stolen, incorrectly issued, or otherwise need to be retracted. This is not a necessary feature for low stakes cases like age verification, but it is often presented as such. SD-BLS is one system that seeks to provide unlinkability and revocation simultaneously.

These systems also reveal a small amount of information in addition to whether an age threshold has been reached. Most designs leak the identity of the authority that issued the underlying credential. At a minimum, addressing this requires the use of delegatable anonymous credentials, which ensure that a recipient only learns about the set of possible issuers.

Standards are not sufficiently mature for deployment. In effect, though the topic is well-studied, translation into a system that could be deployed is not yet complete.

The high computation cost of generating a proof comes with significant costs for people with older devices. This can also mean that two sites can reidentify a device based on it being unavailable during expensive proof generation.

Existing systems are not secure against falsification once a cryptographically-relevant quantum computer (CRQC) is developed. A CRQC would allow an attacker to recover secrets used to issue credentials. This limits the lifespan of any solution in this area. On the positive side, privacy is maintained for any proofs that are shared.

Deployment Considerations

Successful deployment of these systems at internet scale also depend on governance structures being created to manage the set of issuers so that those that depend on proofs — either to produce them or consume them — have a shared understanding of what proofs might be accepted.

There are strong analogies to a certificate authority system in the arrangements necessary to oversee credential issuance. It is worth recognizing the inherent fragility of the WebPKI governance infrastructure. This is no easier to address in a system that is much larger: there are many more people to certify than there are websites and there are many more authorities that might perform that certification.

Even with the challenges of implementing a global system, a patchwork of local laws is unlikely to produce good outcomes. A global system maximizes both the efficacy of limitations on access and equity of access for those of the requisite age. That is, a single system ensures that content distributors are able to rely on being able to ask for proof and that many more adults are able to produce that proof.

Another key question to address when it comes to any system that depends on providers applying access controls is the decision when to enable those access controls. Any system that depends on having a non-trivial portion of the population use credentials needs to have those credential systems in place before flipping the switch. Rollout of a system of digital credentials is a significant undertaking, which means that the result is extended timelines, unless the true goal is over-blocking.

Even with the challenges, it's worth recognizing that a system built on zero-knowledge — with all its flaws — is vastly superior to asking content distributors to identify every visitor.

Device-Based Enforcement

A device-based approach distributes responsibility for ensuring that content does not reach children. Content classification is a responsibility shared by content creators and distributors, but is also one that can be shared by end devices (see Apple's sensitive content warning).

From a technology and standards perspective, this requires no novel technology at all. Content is classified and annotated; annotations are checked and content is displayed or

not. The standardisation effort required is limited to choosing and agreeing on the annotations. Standardization is not a trivial exercise but nor is it a technical one.

Enforcement is applied, based on any classification, by endpoints. The decision is informed by information that a device has about its user, information that does not leave the device.

This approach has a nearly optimal privacy story: no age-related information leaves the device. A distributor learns that content was not displayed, but cannot definitively attribute that to the viewer being underage.

Coverage

The approach also potentially has a nearly optimal coverage story: any medium that can support the carriage of classification information can benefit from the safeguards in the device. That means that protections apply to messaging, social media, generative AI, in addition to the more direct means of distributing content.

This approach also deals with cross-border interactions far better than alternatives. People only need to convince their device of their age, according to the rules of the jurisdiction where the device was purchased.

Services operating outside a jurisdiction are more likely to offer content classification, as this is far less onerous than putting age assurance gates in place. Many services already do.

Despite this potential, coverage is immediately the most obvious drawback of a device-based approach.

Presently, device-based options for content filtering are limited in their efficacy. The choice to use content filtering is almost entirely discretionary and usually put in place by parents and guardians. Compliance with a particular filtering regime is often piecemeal, so filtering systems are often coupled with limitations on what apps can be used, limiting usage to those that are known to properly integrate filtering controls.

Most devices already provide content filtering support to one degree or other. Government mandates might help strengthen and standardize both classification and filtering systems, ensuring better coverage as well as formalizing the process for enabling and disabling filters.

Censorship, Surveillance, and Accountability

Of particular concern in this space is the potential for content restrictions to be used as a tool for censorship or surveillance. This approach largely avoids those questions, because content is delivered to endpoints without any knowledge of the age of the user.

With no reporting to an authority, surveillance risks are almost entirely avoided. Content distributors might learn when content is not displayed, but that is indistinguishable from a bounce (someone leaving a site) or a host of other reasons.

Classification is transparent and therefore more accountable, mitigating censorship risk. The latent risk is inherent in any classification system: misclassification can ensure the suppression of material that might be valued by those who cannot access it.

Where censorship concerns arise is when control over personal devices is taken from people. In the case that devices can only be sold with filtering enabled, it is necessary to closely examine the process of disabling filters. That process can be flexible and accessible, offering a range of options, perhaps drawing from the full suite of age assurance technology on offer. Problems arise when the choices of ways to disable filters do not address the needs of people.

End User Compliance and Rollout

One of the biggest problems with age assurance in general is the effect it has on the market for adult content. The introduction of any mandate results in some portion of the market seeking to circumvent protections.

This is no different with a device-based approach. There are several concerns to be cognizant of and work to mitigate. The primary one being that rollout of a system through software updates might motivate people to disable updates, which could have serious consequences for security and not just for those that react to mandates in this way.

To the extent that any system denies people control over their personal devices, special care is needed. Much of the digital economy depends on the assumption that people can and do trust that their devices work in their interests. A heavy-handed approach that attempts to wrest control of devices from citizens is unlikely to lead to good outcomes. Particular care is needed in this case.

Architectural Perspective

Though a secondary consideration, the provision of key capabilities at endpoints is a long-held architectural principle of the internet known as the end-to-end principle. This perspective is also consistent with the conclusions of RFC 7754.

Conclusion

Policy makers are no longer satisfied with simple age gates where people are able to assert their age without any proof. It seems that there is a widespread assumption that the obvious next step is to build a system with the highest level of assurance available: one backed by government issued credentials.

There is a significant risk of overcorrection leading to the introduction of disruptive measures that include immature technology and privacy violations.

Learned experience with internet-scale deployment of any change shows that iterative and incremental change is far more likely to be successful than large, disruptive changes. The question of how to best restrict access to adult material is not special in that regard.

This paper looked at the prevailing trends toward age assurance technology. It concludes that device-based approaches have more realistic prospects of success. That these are incremental in nature — and so do not deliver on their goals immediately — might be offered as criticism, but this is an important feature for ensuring that the system can be deployed. Device-based approaches are uniformly superior to alternatives when considering all other factors.

This is not a politically popular position. It requires accepting imperfect outcomes in the short term. It depends on greater cooperation from actors not involved in the adult content industry. But these weaknesses are also what makes a device-based approach the strongest choice.