The Gate and the Playground:

Age Attestation, Exclusion, and the Architecture of Verified Peer Space

Travis Gilly
ORCID: 0009-0007-2954-6313
Real Safety AI Foundation
Illinois, United States

Working paper, July 2026 (v0.3)

ABSTRACT

The age assurance debate has settled into a standoff between child safety and privacy in which each side concedes the other has a real interest and neither produces an architecture that serves both. This paper argues that the framing is wrong. The binding constraint on age assurance is neither safety nor privacy. It is exclusion, and exclusion is a distribution problem rather than a cryptographic one. The cryptography required to prove that a person is over a given age without disclosing the person's name, birthday, face, or any other attribute has existed since 1983, and the hardware required to bind a credential to the human presenting it is a deployed W3C standard. What does not exist is a credential that reaches everyone. Every scheme now on the table requires the subject to apply, to pay, to travel, to produce documents, or to own a device manufactured by one of the two companies the scheme is meant to regulate, and each of those requirements falls on the same populations: the poor, the undocumented, and the disabled. This paper proposes the distributive inversion. A boolean age predicate, blind signed by the issuing state, delivered on free standalone hardware, issued at birth alongside the identifiers a government already assigns, and entitled rather than applied for. It further argues that age assurance requires two credentials and not one, because a binary attestation that a holder is over eighteen forces responsible guardians to make platforms blind to the children they are trying to protect, and that the correct deliverable of the architecture is composition rather than exclusion: a space in which every occupant is a verified peer, every occupant's guardian has consented, and intensive supervision therefore raises no adult privacy objection because there is no adult inside to raise one. It argues that supervision is not an addition to the gate but the only mechanism that detects a compromised credential, which converts the rule against treating a gate as a discharge from a policy preference into an architectural necessity. It rejects the zero retention instinct on the ground that a child groomed at thirteen and disclosing at sixteen needs the record to exist, and proposes instead that the objection to retention was never duration but custody and purpose, resolved by escrow. The paper closes by identifying what the architecture does not solve, including the guardian who is himself the threat, the impossibility of binding an infant, the severing of relationships at band boundaries, and the bystander, who never approaches a gate at all and for whom no credential architecture can substitute for a rule keyed to the data subject.

Keywords: age assurance, age verification, zero-knowledge proof, blind signature, hardware credential, digital identity, COPPA, child safety, disability, exclusion, content moderation, verified peer space

CONTENTS

I. Introduction

Picture the playground you already know. Grass, some equipment, a fence around the perimeter, and a gate with a latch a small child cannot work. The fence is real. It does something. A toddler does not wander into the street and a dog does not wander in.

Now watch what no parent has ever done. Nobody walks a four year old through that gate, latches it behind them, and drives away. Not once, anywhere, in the history of playgrounds. The fence is good and the fence is not the program, and every parent on earth knows this without being told, in their body, before they could explain why.

There is a place parents do leave. The indoor play zone at the mall, where you sign a child in, get a wristband, and go have a coffee for an hour. Parents leave there. And the reason they leave is not that the fence is better. The reason is staff. Trained adults, in the room, watching, whose whole job is that nothing happens. The wristband tells you who belongs in the room. The staff are why you are willing to walk out of it.

That is the entire argument of this paper in two buildings. The age assurance debate has spent a decade arguing about the fence. This paper is about what the fence makes possible, which is the room you can leave a child in, and about the fact that a fence with nobody watching behind it is worse than no fence at all, because it persuades the parent to leave.

Every argument about age verification runs on the same axis. One side says children are being harmed and something must check who they are. The other side says that checking who they are means collecting their faces, their documents, and their browsing, and that the collection is itself the harm. Both sides are correct about the other. Neither has produced a system that resolves the objection, and the standoff has now lasted long enough that legislatures have begun to break it by force, mandating verification schemes whose privacy costs the sponsors do not dispute so much as accept.

This paper argues that the axis is wrong. Privacy and safety are not in tension in age assurance, and the appearance of tension is an artifact of a design assumption so widely shared that almost no one states it: that the subject must present something to obtain the credential.

The cryptography needed to prove a predicate without disclosing the underlying attribute has existed since 1983 (Chaum, 1983). The cryptography needed to make such a credential unlinkable across presentations has existed since 2001 (Camenisch & Lysyanskaya, 2001). The hardware needed to bind a credential to the specific human presenting it, without transmitting a biometric anywhere, is a W3C Recommendation with a decade of deployment behind it (W3C, 2021). National identity systems built on physical credentials with cryptographic authentication have been running continuously in Estonia since 2002 and in Germany since 2010. Nothing in the technical stack is missing.

What is missing is anyone who can get one.

The binding constraint on age assurance is exclusion, and exclusion is not a cryptographic problem. It is a distribution problem. Every scheme currently proposed requires the subject to do something in order to be credentialed: to apply, to pay a fee, to travel to an office, to produce a birth certificate, to hold a credit card, to submit a face, or to own a smartphone. Each of those requirements is a filter, and the filters are not random. They fall on people without money, people without documents, people who cannot travel, and people whose bodies do not present the attributes the scanner expects. The literature on identification barriers is not ambiguous about who this is. It is the same population every time.

The proposal in this paper is a distributive inversion. Stop asking who can obtain the credential. Issue it. Free, to everyone, at birth, alongside the identifiers the state already assigns without anyone applying for them, on standalone hardware the state provides, revocable and reissuable, with no attribute inside it except a signed date and a boolean that

flips when the date passes.

The paper proceeds in nine further parts. Part II inventories what is already solved and argues that the field has misclassified solved primitives as open questions. Part III specifies the three failures that any real architecture must survive: binding, exclusion, and custody. Part IV develops the distributive turn and turns the analysis back on the proposal itself, identifying the population the author's own design would exclude if built carelessly. Part V argues that age assurance requires two credentials rather than one, and that binary attestation actively degrades child safety by forcing responsible guardians to render platforms blind. Part VI specifies the presentation ceremony. Part VII develops the constructive payoff, verified peer space, and the moderation inversion it produces. Part VIII identifies the failure mode most likely to defeat the architecture in practice, which is the treatment of a gate as a discharge. Part IX states plainly what the architecture does not solve. Part X concludes.

One clarification belongs at the front. This paper is about the child who approaches. A credential architecture governs the moment a person presents themselves to a system and asks to be admitted. It has nothing to say about a person who never presents anything, never asks, and never knows the system is there. That population is real, it is growing as ambient sensing moves into domestic space, and it requires a rule keyed to the data subject rather than to the requester. Part IX returns to this. The two proposals should never be confused, and the reader is asked to hold them apart from the outset.

II. What Is Already Solved

The age assurance debate proceeds as though the technical questions were open. They are not. Three components that the debate treats as aspirational are in fact settled, published, standardized, and in several cases running at national scale. Restating them is not pedantry. The misclassification is doing real work, because as long as the primitives are treated as open questions, the policy conversation gets to defer the distributive question indefinitely on the grounds that the technology is not ready. The technology is ready. The distribution is what is not.

A. The Predicate Proof

The central privacy objection to age verification is that proving your age reveals your age, and revealing your age requires revealing the document that establishes it, and the document contains your name, your address, your photograph, and your full date of birth. A person who wants to read a news site should not have to hand a stranger a driver's license.

This objection was answered before the web existed. Chaum's blind signature construction allows an issuer to sign a message it cannot read (Chaum, 1983). The signed object can then be presented to a verifier who confirms the signature is authentic without ever contacting the issuer and without learning anything the signature does not carry. Camenisch and Lysyanskaya extended this into a full anonymous credential system in which a holder can prove possession of a credential, and prove statements about its contents, without the presentations being linkable to each other or to the issuance (Camenisch & Lysyanskaya, 2001).

Nor is this a whiteboard claim. The construction has been built against real documents: zk-creds converts existing identity documents such as passports into anonymous credentials without modifying the document or coordinating with its issuer, and completes an anonymous age assertion in under 150 milliseconds against exactly the use case of age restricted video (Rosenberg et al., 2023).

Applied to age, the construction is trivial. The credential contains a signed date. The presentation proves a single boolean: the signed date is more than N years in the past. True or false. No name, no birthday, no photograph, no document, no linkage. The verifier learns exactly one bit, which is the bit it has a legitimate interest in and no other.

It is worth being precise about the vocabulary, because imprecision here costs credibility. This is not a probabilistic estimate and it is not an inference. Age estimation systems, which guess an age band from a face, are probabilistic, and their error rates are the source of much of the objection to them. What is described here is deterministic. The predicate is false until a date and true afterward, computed from a signature the state applied once. There is no model, no inference, and no confidence interval.

B. Binding

A credential proves something about its holder. It does not prove who is holding it. This is the failure mode that has defeated every age verification scheme deployed to date, and it is the reason critics are right to call most of them theater. A twelve year old with an adult's phone passes a wallet based age check perfectly, every time, permanently.

The answer is not new either. FIDO2 and WebAuthn specify a hardware authenticator holding a private key in a secure element, unlocked locally by a biometric or a PIN, where the biometric is matched on the device and never transmitted to anyone (W3C, 2021). The remote party learns only that an authorized human unlocked a specific authenticator that is physically present. It never receives the fingerprint. It never stores the fingerprint. It cannot leak the fingerprint, because it never had it.

The same design point has now reached the age verification literature directly: biometric bound credentials cryptographically tie an age credential to the holder's biometric features without storing a template anywhere, so only the physically present enrollee can present it and the credential cannot usefully be shared (Poh et al., 2025).

Applied to age, this closes the binding gap that wallet based schemes cannot. A credential in a phone can be handed over. A credential in a token that will not unlock without an enrolled fingerprint cannot be handed over in any useful sense, because the token without the finger is inert.

C. National Deployment

Estonia's Identity Documents Act was enacted on 15 February 1999 and establishes both a document obligation and the framework for issuing identity documents carrying certificates, whose suspension, restoration, and revocation the Act governs directly (Isikut tõendavate dokumentide seadus, 1999). Card issuance began in 2002 and the credential now underwrites banking, health records, and voting.

Germany's Identity Card Act was enacted on 18 June 2009, and its implementing regulations, which govern the card and the electronic proof of identity function, took effect on 1 November 2010 (Personalausweisgesetz, 2009; Personalausweisverordnung, 2010). Parliament returned to the subject in 2017 with a statute promoting use of the electronic identity function, and in 2019 extended a card with the same function to Union citizens and European Economic Area nationals resident in Germany (eID-Karte-Gesetz, 2019).

These are not pilots. They are load bearing civic infrastructure with more than two decades and more than one decade of operational history respectively, which disposes of the objection that a state issuing cryptographic identity credentials at population scale is utopian. It has been done, twice, by countries smaller and poorer than the United States.

Two features of the German scheme are worth extracting, because they are the two things this paper is about and both are visible in the statute book rather than in commentary.

The first is that the German card is issued on application and carries a fee, set by its own dedicated fee regulation promulgated under the Act (Personalausweisgebührenverordnung, 2010). A state with a mature, cryptographically sound, decade old national eID nevertheless built an application and a price into the front of it. That is not an oversight.

It is the default assumption of the entire field, which is that a credential is a thing the subject goes and gets, and it is the assumption Part IV proposes to abandon. The most advanced deployed system in this space contains, in its own subordinate legislation, exactly the barrier this paper identifies as the binding constraint.

The second is the direction of travel. In February 2021 the German federal government introduced a bill to establish electronic proof of identity on a mobile device (Bundesrat Drucksache 139/21, 2021). This is not a hypothetical. The migration from a standalone state issued credential to a credential resident on consumer hardware is a live legislative programme in the country with the second longest running national eID in Europe, and Part III.C explains why that migration hands the identity layer to the parties the identity layer is supposed to govern.

One severance must be stated before the failures, because the Estonian statute invites a conflation this paper refuses. Section 1 of the Estonian Act makes the identity document an obligation, and a citizen can face consequences for not holding one (Isikut tõendavate dokumentide seadus, 1999, §1). This paper takes Estonia's distribution and declines its compulsion. The proposal is an entitlement: issued to everyone, applied for by no one, mandatory for no one. A credential every citizen holds because the state mailed it is a different object, legally and politically, from a credential every citizen must hold on pain of sanction, and nothing in the architecture requires the second. The distribution rail is the lesson. The obligation is not.

III. The Three Failures

Any proposed age assurance architecture should be tested against three failures. The first is technical and well known. The second is the one this paper argues is dispositive and is almost never modeled. The third is structural and, to the author's knowledge, has not been named in the policy literature at all.

A. Binding Failure

Stated above and dispatched. A credential that lives in software on a general purpose device cannot bind to a human. The device can be handed to anyone. Every wallet based proposal inherits this and none of them survive it. The only known answer is a hardware authenticator with local biometric or PIN unlock, and any proposal that does not include one is proposing a system whose defeat requires a child to ask a parent for a phone.

B. Exclusion Failure

This is the failure the debate does not model, and it is the one that decides whether the architecture is civil rights infrastructure or a new barrier wearing a safety label.

Enumerate what a person must have in order to obtain a credential under each proposal currently on the table. A government issued photo identification, which requires a fee in most jurisdictions, an office visit, a birth certificate or equivalent breeder document, and in many jurisdictions a permanent address. A credit card, which requires a bank account and a credit history. A smartphone of recent enough vintage to hold a wallet credential, which requires several hundred dollars. A face that an estimation model reads within its trained distribution. A fingerprint that enrolls.

Each of those is a filter, and the filters compose. A person with no money, no permanent address, no documents, and no smartphone fails all of them simultaneously. A person with a disability that affects their hands may pass every filter except the last one and be excluded anyway, silently, by a sensor. A person without immigration status will not approach an office regardless of what the office would do.

The exclusion is not incidental to these designs. It is constitutive of them, because every one of them begins from the assumption that the credential is something the subject goes and gets. The moment the credential is something you

obtain, obtaining becomes a test, and tests have failure populations, and the failure population is always the same one.

This is not speculative. India's Aadhaar programme is the largest biometric identification system ever built, enrolling over 90 percent of Indian adults, roughly 1.2 billion people, and it has been studied ethnographically and empirically for over a decade. The finding that recurs across that literature is that a system designed to reduce barriers to access produced new ones. Mukherjee and Sahay document the mid day meal scheme, in which children already enrolled at government schools were required to furnish an Aadhaar number to receive food, and conclude that Aadhaar, contrary to its stated promise of reducing barriers to access, is potentially creating barriers (Mukherjee & Sahay, 2019). Rao and Nair's collection traces the same pattern through enrolment, authentication, and welfare delivery practice (Rao & Nair, 2019), and Chaudhuri's ethnography of the Aadhaar enabled public distribution system shows the algorithmic sorting enacted through a shifting assemblage of machines, documents, and human actors that behaves nothing like its specification (Chaudhuri, 2022).

The exclusion is measurable. A 2017 to 2018 survey across three Indian states attributed between 0.8 and 2.2 percent of public distribution system exclusion to Aadhaar related failures, a category comprising seeding failures, authentication failures, connectivity and electricity failures, and the beneficiary's inability to be physically present to authenticate. In those three states alone that is more than two million people, and the entitlement at issue was subsidised food.

Two features of that finding matter here. The first is that the percentage sounds small and the population does not, which is the arithmetic that lets an exclusion of millions be reported as a rounding error. The second is that the failure modes are heterogeneous: a person can be excluded because their fingerprint will not read, because the shop's connection is down, because the power is out, or because they are too ill to travel to the machine. Any architecture that makes access contingent on a successful authentication event inherits all four, and only the first is a biometric problem.

What is excluded matters as much as who. The debate has narrowed to pornography access, where exclusion sounds like a joke to people who do not need the thing being gated. Widen the aperture. If age assurance becomes the price of admission to platforms generally, the excluded person loses the educational content, the medical information, the disability communities, the support networks, and the ordinary participation that everyone else retains. The public health literature has begun saying the same thing about the current wave of age gating: verification requirements fall hardest on young people with limited digital access, literacy, or identity documents, excluding them from mainstream platforms and pushing them toward less regulated spaces (Carah et al., 2024). This is the pattern documented for population level bans in the analysis of Australia's under-16 prohibition (Gilly & Mattheus, 2026): the aggregate statistics register a ban as working because the children it harmed, there the LGBTQ+ youth and the children in abusive homes who lost their communities and their lifelines, were not counted before it and are not counted after it.

C. Custody Failure

The third failure is structural and it is the one that should be alarming to anyone who has watched regulatory capture operate. It is best introduced through the way the demand for age assurance is ordinarily phrased.

The demand is real and it is not hypothetical. In April 2025, Common Sense Media published a risk assessment of social AI companions that rated the category an unacceptable risk for anyone under eighteen, declined on stated safety grounds to publish links to the products it had tested, and issued recommendations. The first recommendation was that no young person under eighteen should use a social AI companion. The second was that developers need robust age assurance beyond self-attestation (Common Sense Media, 2025).

Read the addressee. Developers. The most widely circulated child safety assessment of the period identifies self-attestation as the failure, and assigns the remedy to the firms whose products it has just declared unsafe. It is not an

unusual formulation; it is the standard one. Age assurance is almost always specified as something platforms must implement rather than something a public issuer must distribute, and the question of who holds the credential is treated as an implementation detail beneath the notice of the recommendation.

That formulation is the custody failure in miniature, and the rest of this section describes what it produces at scale.

Nor is the failure hypothetical or American. The European Union has now written it into law. The revised eIDAS framework requires every member state to issue a European Digital Identity Wallet, free of charge to natural persons and voluntary in use, with age attestation among its intended functions (Regulation (EU) 2024/1183). Free and voluntary are concessions to exactly the distribution argument this paper makes, and they are welcome. What the framework does not concede is custody. The wallet is an application, an application runs on a phone, and the phone market's operating layer belongs to the same two companies. The framework's own technical work concedes the predicate half of this paper's argument, since the Commission's age verification profile and the wallet's architecture reference framework explicitly explore zero knowledge approaches (Zafeiropoulou et al., 2025); the cryptography is accepted, and only the custody is surrendered. Europe has legislated free distribution of a credential whose custody it has delegated to the regulated. The proposal here accepts the two concessions and rejects the delegation, which is the entire difference a standalone token makes.

Every major age assurance proposal now moving through legislatures and standards bodies contemplates the credential living in a mobile wallet. The wave is not small; in the United States alone, more than fifty state bills regulating adolescent platform access had been adopted or enacted by late 2024, with age verification mandates among the most common instruments (Thimm-Kaiser & Keyes, 2025). The mobile wallet market is a duopoly. That means that under these proposals, two companies become the custody layer for the identity credential that governs access to the internet, and both of those companies are among the parties the regulation is aimed at.

Consider what that means operationally. Every citizen must obtain a device manufactured or operating system controlled by a regulated party in order to be governed. The regulated party becomes a mandatory intermediary in its own regulation. It gains a hard dependency of every citizen upon its product, a new lever over every downstream platform that must consume the credential, and a permanent seat at every table where the credential's future is decided. It also gains, as a matter of engineering fact, the ability to know which credentials were presented, unless the wallet is built to prevent that, and the entity building it to prevent that is the same entity that would benefit from not preventing it.

There is a reason this has not been argued loudly. The companies in question are the primary funders of the standards work, the primary employers of the engineers who would build it, and the primary participants in the multistakeholder processes where age assurance is being designed. The custody question is not being suppressed. It simply has no natural constituency, because the people who would notice are the people being paid.

A standalone government issued token has no such problem. It is not a phone. It does not need a phone. It works on a library computer, a school Chromebook, a borrowed laptop, and a desktop that is eleven years old. It costs the citizen nothing and it costs the regulated companies their intermediary position, which is the reason it will be opposed on grounds that will be described as technical.

IV. The Distributive Turn

A. Entitlement, Not Application

The proposal is one sentence. Do not make people get the credential. Give it to them.

In the United States, the state already assigns an identifier to essentially every person at birth, without an application in the ordinary sense, without a fee, and without an office visit, through a hospital enumeration process that most parents

experience as a form they sign while still in the maternity ward. That identifier is already universal, already tied to a birth date the state already holds, and already the basis for interaction with federal systems throughout life.

The proposal attaches one thing to that existing act: the issuance of a credential containing a blind signed date and nothing else, delivered on a physical token, at no cost, at the same moment. For people already born, the same credential is obtainable through the same channel the state already operates for the identifier itself, with the identifier as the basis.

The scale of application failure is documented. An estimated one billion people hold no official proof of identity at all (Martin et al., 2020), and the mechanism is not mysterious: where identity documents must be applied for, administrative burdens systematically exclude the poorest and least connected, and the exclusion then cascades, because every benefit and service keyed to the document falls with it (Chudnovsky et al., 2021). An applied for credential inherits the application's failure demographics. An issued credential has none.

The point of at birth issuance is not elegance. It is that there is no application to fail. There is no fee to afford, no office to travel to, no document to produce, and no line to stand in. The credential arrives before anyone has to want it. That is the entire contribution of this paper and every other component is off the shelf.

B. Why Free Is Not a Detail

The instinct of a system designer reading the previous paragraph is to file the cost question as an implementation detail to be resolved at appropriations. That instinct is the error the whole paper exists to name.

Cost is not a detail of an identification scheme. Cost is the mechanism of exclusion. It is the filter that selects who is inside the system and who is outside it, and it does so with a precision that no explicit criterion could achieve without being challenged. An identification requirement with a fee attached is a means test that never has to admit it is one, and the population it excludes is documented in a large empirical literature on identification barriers that has been running for decades in the voting context and reaches the same finding every time.

Attaching a fee to an age credential would be worse than the voting case, because the excluded person does not lose one act on one day. They lose ordinary participation in the medium through which education, healthcare information, employment, community, and civic life are increasingly conducted. And they lose it invisibly, because a person who cannot get in generates no complaint, no ticket, no data point, and no line item. The system will report that it is working.

C. The Author's Own Design, Audited

The Harm Blindness Framework, a systematic stakeholder analysis methodology that identifies preventable harm at four decision points and defines harm blindness as the systematic failure to identify stakeholder harm (Gilly & Mattheus, 2026), asks before anything else who is affected and who is invisible to the people making the decision. Applied to the proposal above, the framework returns a finding against the proposal, and a paper that has run this diagnostic against a national statute owes its own design the same treatment.

One clarification first, because the term is ambiguous and the ambiguity hides the problem. The distributive turn removes application failure. There is no form to submit and no fee to fail to pay, so nobody is excluded at issuance. That is the point of it and it holds.

What it does not remove is biometric enrollment failure, which is a different event at a different moment. The metric has a standard name, failure to enroll, defined as the proportion of a population for whom a system cannot generate a usable template, and its documented causes are precisely the population this paper claims to protect: worn ridges, common among manual labourers and the elderly, dermatological conditions, injury, amputation, and extreme dryness or moisture.

The author declines to assert a single population wide failure to enroll figure, and the reason is worth stating rather than hiding. Vendor literature quotes rates in the low single digits. Those figures are drawn from cooperative subjects under favourable conditions, they are published by parties with an interest in them, and they are not disaggregated by disability. The one thing every serious treatment agrees on is that failure to enroll is not a fixed property of the technology; it is sensitive to operator training, environmental conditions, and the demographics of the enrolling population, which is another way of saying that the rate for the general population tells you nothing about the rate for the people this paper is about. A number that is not disaggregated is not evidence for a claim about a subpopulation, and citing one would be the error this paper accuses others of making.

The disaggregated evidence that does exist is instructive, and its condition is the finding. Between April and December 2004 the UK Passport Service, with the Home Office Identity Cards Programme and the Driver and Vehicle Licensing Agency, ran a biometrics enrolment trial of roughly 10,000 volunteers, comprising a 2,000 person quota sample selected to reflect the general population, an opportunistic sample of 7,266, and, critically, a cohort of 750 disabled participants. Atos Origin conducted it and the report was published in May 2005 (UK Passport Service, 2005; *UKPS Biometric Enrolment Trial*, 2005). The disabled cohort took longer to enrol than either of the others.

That trial is more than two decades old. Its sensors are two generations obsolete and no serious person should quote its error rates as current performance, any more than one should quote a 2004 assessment of a 2026 product. The point is what has happened since, which is that the only large scale public trial to deliberately recruit a disabled cohort and report against it was run before the iPhone existed, and it has not been repeated.

The absence is the argument. Twenty-two years into the deployment of biometric identification as civic infrastructure, there is no current, public, disability disaggregated performance data for the population this paper is about, in any jurisdiction known to the author. The rate is not unknown because it is unknowable. It is unknown because measuring it has never been anyone's job, and a harm that is nobody's job to measure does not appear in any evaluation of whether the system works. The gap is documented from inside the field itself: a 2023 review of biometric mobile authentication for elderly users found no comprehensive accounting of how elderly performance compares with other users (Wang et al., 2023), and a 2024 usability study reported both measurable accessibility disparities for disabled users and a perception gap in which nondisabled people rate biometric systems as more inclusive than disabled users experience them to be (Ólafsdóttir et al., 2024). That is the same structure this paper identifies in Part III.B, where an exclusion of two million people registers as a figure between 0.8 and 2.2 percent, and it is the reason a proposal in this space has an obligation to name its own excluded population rather than wait to be told.

What can be asserted is the direction, the mechanism, and the scale of the consequence, and Part III.B supplies the empirical anchor: in a deployed national system, biometric authentication failure was one of several mechanisms that excluded millions of people from subsidised food. Whatever the laboratory rate, the deployed rate multiplied by a national population is a number of human beings, and they are disproportionately poor, disabled, and old.

There is a second enrollment problem, and it is structural rather than statistical. An infant cannot enroll a fingerprint. Neonatal friction ridges are present but small, and they change substantially as the hand grows, so a print captured in a maternity ward will not authenticate the same person at seven. This means the credential can be issued at birth and cannot be bound at birth. There is a window, running from birth to whatever age enrollment becomes stable, in which the token exists and is either unbound or bound to a guardian.

The paper does not treat this as a footnote. A guardian bound token is a delegation, and a delegation held over a child who cannot yet revoke it is precisely the abuse surface Part IX identifies as unsolved. The window is where the worst case lives. Three partial answers are available and none of them is clean. First, defer binding: the token is inert until the child can enroll, which means a young child has no credential and therefore no access to spaces designed for young children, which is an exclusion. Second, bind to the guardian and re enroll to the child at a specified age,

which works and hands an adult an unsupervised child credential for years. Third, bind to the guardian and require the guardian credential to be co present at every presentation during the window, which narrows the abuse surface without closing it and makes the young child's access wholly dependent on an adult being in the room, which is arguably what it should be at that age and is nevertheless a dependency.

The author's present view is that the third is correct for the youngest band and that the honest position is to state the residual, which is that the architecture cannot bind a person who does not yet have a stable body. Part VII is where that residual gets absorbed, because a compromised credential in a watched room is caught by the watching.

A design that fails silently for disabled users and then locks them out of the network is the shape every accommodation failure takes. It cannot be bolted on afterward, because retrofitted accommodation is how the shape of compliance gets built instead of compliance.

Three requirements follow, and they are load bearing rather than gestural. First, multiple biometric modalities, so that a hand that will not enroll is not the end of the process. Second, a non biometric fallback, which is to say a PIN path that is a first class citizen of the design and not a degraded mode, accepting the binding cost that comes with it and stating that cost out loud. Third, an assisted enrollment path for people who cannot complete enrollment independently, which is a human process and not a technical one, and which the state already operates for other credentials.

Naming the third requirement produces an uncomfortable result that the paper should state rather than bury. Assisted enrollment means a person other than the subject participates in binding the subject's credential. That is a delegation, it has an obvious abuse surface, and it is the same abuse surface Part IX identifies as unsolved. The honest position is that assisted enrollment trades a certain exclusion for a possible abuse, that the trade is worth making because the exclusion is certain and the abuse is not, and that the trade should be made with the abuse surface documented rather than denied.

V. Two Credentials

A. Attestation

The first credential says something about the holder. The holder's signed date is more than N years past. True or false. Nothing else is inside it and nothing else can be extracted from it. This is the credential the entire age assurance debate is about and it is the simpler of the two.

B. Delegation

The second credential says something about a decision. A guardian, holding a guardian credential, authorized this holder's access to this service, and this holder's age is fourteen.

The distinction is not academic and the difference between the two is the difference between a system that protects children and a system that erases them. An attestation is a fact about a person that travels everywhere the person goes. A delegation is a decision about a service that travels nowhere else.

C. The Lie the Binary Forces

Consider a fourteen year old with autism, attention deficit hyperactivity disorder, and an intellectual disability. He wants an account on a conversational AI platform. What he wants to do with it is write fan fiction: take characters he already loves, put them in situations he invents, and practice the back and forth of building a story with something that answers. He has asked for months. He asks plainly, because he is an honest kid, and what he says he wants to do with it

is what he wants to do with it.

The platform's terms require users to be eighteen.

His parent considers it and says yes, and the yes is not casual. It is the most researched decision in the house. It is also the kind of decision the evidence says matters most for exactly this child, since adolescents with special educational needs face elevated online risk and active parental mediation is associated with reduced risk for them in particular (Cavallini et al., 2025).

The parent knows the child. He knows which of his son's interests sustain and which burn out in a week, and this one has sustained. He knows that the practice his son is asking for, the structured low stakes rehearsal of a social exchange with something that will not judge him or get bored of him, is the thing the clinical literature says helps. He has read that literature. He knows the meta analytic finding that agent mediated practice produces moderate gains in social and emotional skills for autistic children, and that the gains are largest for the children most severely affected (Zhong et al., 2025). He also knows the part of that literature the vendors do not quote: that every trial producing those gains used a purpose built agent, in a structured session, on a clock, with a human in the room, and that the researchers themselves warn that the support must fade or the child becomes reliant on it.

So he knows what he is buying and he knows what it is not. He is not buying the intervention. He is buying the closest available thing and supplying the missing parts himself.

He supplies them explicitly. Time is bounded, and the boundary was negotiated with his son rather than imposed on him, because a rule the kid agreed to is a rule the kid keeps. The subject matter is bounded to the fiction and his son knows why. The parent has access to the logs and reads them, not as a spot check but as a practice. The permission is revocable at any moment and both of them know it, which is itself part of the arrangement, because a permission that can be withdrawn is a permission that teaches something.

And he knows what to watch for, which is the part that separates him from a parent who merely permits. He knows about the wrongful death litigation against companion platforms. He knows the specific mechanism of parasocial drift, that the register slides from tool toward confidant without anyone deciding it should, because the products are optimized for a metric that rewards the slide. He knows to watch for the conversation becoming about the machine instead of about the story. He knows what the fade is supposed to look like and he knows he is the one who has to enforce it, because the product will not.

He knows all of this because he went and learned it. Nobody made him. There is no course, no requirement, and no institution that offered. He is exactly the parent that every child safety policy in the world is written on the assumption of, and he is doing the job better than the policy asks.

The architecture offers this parent one option. Assert that the child is over eighteen.

Observe what the assertion costs, because it is not a formality. The platform now believes an adult is in the seat. Every protection it maintains for minors is off. Every escalation path keyed to age is inactive. Every content restriction, every session limit, every crisis routing rule, every design choice the platform might have made differently for a fourteen year old is disabled, because the platform has been told, by the child's own parent, that there is no fourteen year old there.

The good parent has been made into the mechanism that blinds the system to their own child. And the good parent then compensates personally, running the safety layer in their own head, reading the logs, watching the tone, because the product cannot run a safety layer for a user it has been informed does not exist.

Nor is the parent in this story rare. The empirical literature finds that many parents facilitate their children's underage platform use themselves, creating the accounts on their children's behalf despite the stated age limits (Rose

& Middling, 2025). The population of guardians the binary forces into lying is not an edge case; it is a documented, ordinary parenting practice performed by exactly the engaged guardians the policy assumes. This is not a story about a bad platform or a bad parent. It is a story about a binary. The parent had a nuanced judgment to express and the system had one bit to express it with, so the judgment was rounded to the nearest available lie.

Note precisely what was lost, because it is not abstract. The child in the hypothetical has an intellectual disability. He is a fourteen year old whose vulnerability profile is not a fourteen year old's. Whatever the platform would have done differently for a minor, it will not do. Whatever it might someday be required to do for a minor with a disclosed disability, it cannot do, because it has been informed by the only person who could have told it that there is an adult on the other end. The most protective act available to the most engaged parent in the population was to make his son invisible to the system he was letting his son into. That is not a parent's failure. It is a design with exactly one setting, and the setting is off.

D. What Delegation Returns

Under a two credential architecture the same parent presents a guardian credential alongside the child's own credential, and the platform receives an authorized minor. It knows the age. It applies the rules for that age. It maintains the escalation paths. It logs nothing about the parent's identity, because the delegation proves authorization rather than identity, and it is revocable by the same guardian at any time.

Every party gains. The platform gains a known minor rather than a phantom adult, which is better for it legally, since its exposure under child safety regimes turns on knowledge, and a platform that has been affirmatively told a minor is present under a documented guardian authorization is in a defensible position that a platform relying on a checkbox is not. The parent gains the ability to say yes without lying. The child gains the protections his parent wanted him to have.

And the regulator gains the thing it has never had, which is a legible record of the difference between a child whose guardian consented and a child who defeated a checkbox. Those are currently indistinguishable in every platform's data, which is one reason the empirical literature on minor use is as bad as it is. The reviews say so themselves: prevalence estimates for cybergrooming victimization vary widely, with the heterogeneity largely attributable to methodology, and the samples systematically omit the populations most at risk, including young people with special educational needs (Schittenhelm et al., 2024).

VI. The Ceremony

A. Physical Presence at Signup

The credential is presented once per account, at signup, by physical insertion of the token into the device, unlocked locally.

This is what a FIDO2 registration already is, and it is worth stating explicitly what the ceremony controls, because the requirement is often described loosely as verification in a way that overclaims. The ceremony controls how many accounts exist and who created them. It requires the credential holder, or a person the credential holder physically hands the token to, to be present at the moment of creation. It rate limits account creation to the speed of a human with a token in hand. It makes account creation a deliberate act rather than a click.

It does not control what the credential asserts. This distinction matters because a ceremony can be performed perfectly around a false assertion. If the token in the slot says the holder is over eighteen, the platform creates an adult account, and the ceremony has authenticated the lie rather than prevented it. The ceremony is necessary and it is not

sufficient, and the sufficiency comes from Part V rather than from here.

B. Unlock Modalities and Optionality

The token unlocks by fingerprint or by a PIN the holder sets. Both are local. Neither transmits anything. The plurality is not a convenience. It is Part III.B applied to this design's own front door: any single mandatory modality excludes someone, a fingerprint reader excludes the users the UK trial's disabled cohort stands for, and the mechanism is physiological rather than exotic, since ageing, dermatological conditions, finger wrinkling, and occupational wear each degrade fingerprint features to the point of recognition failure (Riaz et al., 2024); a PIN excludes a person who cannot form or retain one; so the design mandates neither and requires only that some local unlock succeed. An architecture that indicts biometric systems for exclusion and then requires biometric unlock would refute itself. This one does not.

A guardian who wants stronger control over the token may optionally link a time based one time password generator, so that unlocking requires a rotating code the guardian holds elsewhere. The word optionally is doing necessary work. This is an additional path for a guardian who wants it, not a requirement, and the token functions completely without it. A design that offers more paths to more people is accessibility. A design that mandates one path is the exclusion failure of Part III wearing new clothes.

The author notes one caution about this option without withdrawing it. A time based generator normally lives on a phone, and a phone is a device from the duopoly named in Part III.C. A guardian who chooses this path has voluntarily accepted a dependency the architecture otherwise avoids, which is their prerogative. The architecture must never require it, because the moment it does, the custody failure returns through a side door.

C. Unlinkability

The verifier never contacts the issuer. This is the property that separates the proposal from the surveillance system it will be accused of being, and it follows directly from the blind signature construction of Part II.A.

A naive implementation would have the platform query a government endpoint to confirm the credential. That implementation would give the state a log of every service every citizen ever joined, which is a comprehensive record of association, belief, health condition, and political activity, assembled by a system that was sold as a way to keep children out of pornography. The construction that avoids this is not exotic and it is not optional. Issuance happens once. Presentation is offline. The verifier checks a signature against a public key. Nobody queries anybody. Two parties know the drop and never meet at it.

One layer remains that presentation unlinkability does not touch, and it should be named rather than waved at. Issuance leaves a record. A government that issues at birth holds a registry of who received which token, and whoever holds that registry holds, in principle, the seed of a cradle to grave corpus. The design constraint is therefore a requirement rather than an assumption: key material on the token must be generated so that issuance records cannot resolve presentations, meaning the issuer learns that a token exists and to whom it was mailed, and learns nothing afterward. Credential constructions with this property are established art, and the anonymous credential literature this paper already relies on was built for it (Camenisch & Lysyanskaya, 2001; Chaum, 1985). Where a deployment cannot meet the requirement, the registry must be treated as what it is, a surveillance asset, and placed under audit and access constraints commensurate with that description. The paper prefers the cryptographic answer and states the custodial one so no implementer can claim the question went unasked.

VII. Verified Peer Space

A. The Deliverable Is Composition, Not Exclusion

Everything above is defensive and the defensive framing is why the age assurance debate is losing. Keeping people out is what the architecture does. It is not what the architecture is for.

What a bound, delegable, universally distributed credential makes possible is a space whose composition is known. Not filtered. Not moderated after the fact. Not estimated by a model reading faces. Known, as an architectural property, before anyone speaks: every occupant of this space holds a credential attesting to an age within this band, bound to their own body, and no other credential opens the door.

A thirteen year old is talking to a thirteen year old. The twenty four year old is not in the room, and he is not in the room because he cannot hold a credential bound to a fourteen year old's finger, not because a classifier flagged his messages or a moderator read his profile.

This is the thing every child safety organization in the world says it wants. None of them can deliver it, because every scheme they have proposed dies on the binding problem before it ships. Binding is the whole reason this is worth building.

B. Dual Token Consent

Admission to a space designated for children requires two presentations. The child's credential, attesting the band. The guardian's credential, attesting authorization for this service.

This is the wristband at the indoor play zone, and the analogy is exact rather than decorative. The parent does not merely permit the child to be in the room. The parent hands the child to a specific operator, at a specific place, for a specific period, on the understanding that trained staff are present and watching. The consent and the supervision are one transaction. No parent would perform the first half without the second, and no operator would accept the child without the first.

Contrast the outdoor playground of Part I. It has a fence and a gate and no staff, and no parent leaves. Not because the fence fails, but because the fence was never the thing they were relying on. The difference between the two buildings is not the quality of the barrier. It is the presence of somebody whose job is watching, and the wristband is what makes that somebody's job legitimate rather than intrusive.

C. The Moderation Inversion

Here the architecture produces something the debate has not seen, and it is the strongest structural result in this paper.

The objection to intensive monitoring of communication platforms is an adult privacy objection, and it is a good one. When a platform combs voice and text chat, it combs the communications of adults who have done nothing wrong, who did not consent to being read, and whose privacy interest is real and legally cognizable. That objection has stalled child safety proposals for a decade and it deserves to have stalled them, because the proposals were asking to surveil everyone in order to find someone.

Now remove the adults.

In a space where every occupant holds a credential attesting to an age band below majority, and every occupant's guardian has affirmatively authorized this specific service, the adult privacy objection has no one left to make it. There is no adult inside whose communications are being read. There is no participant whose guardian did not agree to exactly

this. The intensive supervision that is unacceptable in a mixed space is the supervision the parents specifically consented to when they handed the child over, and it is the supervision they assumed was happening.

The gate does not merely keep the predator out. It manufactures the legitimacy of the lifeguard. That is the inversion, and it is available in no other configuration: verification is what makes supervision defensible, because verification is what guarantees there is nobody in the room whose rights the supervision would violate.

And the relationship runs in both directions, which is the part that makes this architectural rather than merely convenient. Consider the failure case. A guardian intercepts the token issued to his own child during the unbound window of Part IV.C, enrolls his own fingerprint, and now holds a credential attesting that he is eight years old. The cryptography is intact. The signature is valid. The binding is genuine, to the wrong body. The gate has not been circumvented; it has been captured, and it will attest to his age correctly and forever.

Nothing in the credential layer detects this. Nothing in the credential layer can, because from the credential's point of view nothing went wrong.

What detects it is the watching. A grown man cannot behave like an eight year old in a monitored room for six months. The tell is not in the credential, it is in the conduct, and conduct is exactly what supervision observes and what cryptography cannot. This is the same reason the indoor play zone tolerates adults walking in. Parents come and go through that gate constantly. The operator's safety model was never that no adult is present. It is that no adult is unobserved.

The division of labor is not decorative, because the threat model demands it. The best national prevalence data finds that perpetrators of online sexual abuse against minors are predominantly dating partners, friends, and acquaintances rather than adult strangers (Finkelhor et al., 2022), which means a gate that excludes unverified adults addresses only part of the risk surface, and the part it cannot address, harm arising between known peers inside the space, is exactly what the supervision layer exists to catch. So the gate establishes the population and the supervision catches the gate's failures. Neither layer is sufficient and neither is optional. That is not a policy preference about how much moderation is nice to have. It is the observation that a compromised credential is invisible to every mechanism except the one that watches behavior, and that credentials will be compromised, because the unbound window in Part IV.C guarantees a supply of them.

One assumption above must be stress tested, because an adversary will test it first. The verified space is safer than the open internet only while its verification holds, and a compromised credential inside a verified space is more dangerous than a stranger on an open platform, because the space's own guarantee is the predator's camouflage. The design's answer has three parts. First, the compromise classes are narrow and named: a borrowed sibling token, a coerced unlock, an enrollment performed by the guardian who is the threat, and each leaves the signed delegation trail of Part V rather than the anonymous void a checkbox leaves. Second, the supervision layer this section defends exists for exactly this residue; the inversion argument is that verified spaces are where intensive moderation is legitimate, and the compromised credential is the case the moderation is for. Third, the honest quantitative statement: the space remains safer than baseline only while the product of compromise rate and detection failure stays below the open internet's ambient rate, and that baseline is not hypothetical: the global meta-analytic estimate for past year online sexual solicitation of minors is 12.5 percent (Fry et al., 2025). That inequality is a monitorable operating condition rather than a fact the architecture may assume. A deployment that stops measuring it has stopped being the thing this paper proposes.

D. The Camera in the Corner

The instinct at this point is to reach for a zero retention rule: process for safety, store nothing, keep only what gets flagged. That instinct is wrong, and the author held it until the case against it became clear.

Consider the child who is groomed at thirteen and discloses at sixteen. She is not an edge case. Children typically delay disclosing experiences of abuse, and large scale national probability studies establish the prevalence of both delayed disclosure and non disclosure as the norm rather than the exception (McElvaney, 2013). The population level numbers have since been measured directly: in a nationally representative Australian sample of 8,503 people, 45.2 percent of those who had experienced child sexual abuse had never disclosed it before the survey itself, and among those who did disclose, the mean delay was 7.1 years, with one in ten people over 25 having waited more than twenty years (Mathews et al., 2025). The hypothetical child's three year delay sits well inside the measured mean. The delay is not incidental to the abuse; it is produced by it, and by the child's need to retain control over the disclosure process, which is why the reluctance takes years to wear off; the reviews describe disclosure as an iterative process unfolding over time rather than a discrete event, with the barriers consistently outweighing the facilitators (Alaggia et al., 2017). Much disclosure is not even initiated by the child: in a study of 1,737 cases, 43 percent were accidental detections and only 30 percent were purposeful disclosures (Collings, Griffiths, & Kumalo, 2005).

And when she finally speaks, the systematic evidence on disclosure pathways says the first person she tells will most likely be a peer rather than a parent or an authority (Manay & Collin-Vézina, 2019), which means the first disclosure plausibly happens inside the very space that holds the record. When she finally speaks, the record is three years old. Under a zero retention rule it does not exist. Her account is her word against a company that deleted the evidence pursuant to a policy that was described, at the time it was adopted, as protecting her.

Zero retention does not protect the child. It protects the operator, and it does so under a privacy banner, which is the most effective way to protect an operator that has ever been devised. A rule that destroys the only contemporaneous record of what happened to a child, and calls the destruction a safeguard, should be recognized for what it is.

Return to the play zone. There is a camera in the corner. It records everything, all day, and it keeps what it records for a good while. No parent in that building has ever called it surveillance, and no parent has ever asked for it to be turned off. Why not?

Not because of duration. The footage lasts longer than the visit by an order of magnitude. The reasons are three and none of them is about how long.

Nobody watches it. The footage is not a feed somebody enjoys. It sits, unwatched, until something happens.

The operator does not sell what is on it. Nobody is mining the footage to learn how eight year olds move so they can sell it to a toy company. The recording exists for one purpose and is used for that purpose.

And it gets pulled when a child says something happened. That is the whole reason it exists. A parent who is told, three weeks later, that something occurred, wants the footage to exist. The camera is not aimed at the children. It is aimed at the room, on the children's behalf.

The objection to retention was never duration. It was custody and purpose. Which gives the rule, and it has two halves that must not be collapsed.

Supervision is live and its output is a flag, not a corpus. The scanning that catches the compromised credential of Part VII.C operates in something close to real time, and what it emits is an alert. It does not emit a profile, a research dataset, an advertising segment, or a model training corpus. What it learns dies with the session unless it fires.

Retention is mandatory and escrowed. The communications are preserved, for a period long enough to survive delayed disclosure, and the operator holds them under a key the operator does not have. Preserve and do not read.

Access runs on legal process or on a verified report by a child or guardian, and every access is logged.

The two halves answer different objections and neither answers the other's. Escrow without live supervision means nobody catches the man behaving like an eight year old until someone reports him, which is years. Live supervision without escrow means the report, when it comes, arrives at an empty archive.

One argument must be given up here, and it is one the author found attractive. The claim that retention is safe because the platform does not know who the child is does not survive contact with the literature. The credential withholds a legal name. It does not withhold the child.

Replacing real names with pseudonyms before releasing a dataset is usually not sufficient to preserve anonymity, because the behavioral traces themselves contain information that links back to an individual, and sanitisation by adding noise does not eliminate the risk (Gambs, Killijian, & Núñez-del-Prado, 2014). The magnitudes are not marginal. Four spatio temporal points are sufficient to uniquely identify 95 percent of individuals in a mobility dataset (de Montjoye, Hidalgo, Verleysen, & Blondel, 2013), the same four point result holds for credit card metadata in a shopping context (de Montjoye, Radaelli, Singh, & Pentland, 2015), and up to 99 percent of users in the Netflix Prize dataset were matched to their public profiles using only movie ratings, exposing political, religious, and social views the users had never stated (Narayanan & Shmatikov, 2008).

A platform holding years of a pseudonymous minor's messages holds her typing cadence, session times, device, network, social graph, vocabulary, and interests. If four points identify a person from where they walked, and movie ratings identify them well enough to infer their politics, then three years of a child's conversation is not close to anonymous, and the credential is not the protection. Custody and purpose limitation are the protection, and they have to be built rather than inferred from the absence of a name.

Which is also why the operator's identity matters more than any rule about duration. A platform that scans children's messages for safety will scan them for everything else it can monetize, because that is what the business is, and a promise not to is worth what promises are worth. Escrow is the difference between a company that has agreed not to read the archive and a company that cannot.

VIII. The Discharge Failure

The most likely way this architecture fails in practice is not technical. It is that operators treat the gate as a discharge.

The pattern is familiar from every compliance regime that has ever specified a mechanism, and it is one of the most heavily documented findings in the sociology of law: organizations respond to legal requirements by elaborating formal structures that function as visible symbols of compliance, constructed to serve managerial interests as much as regulatory goals (Edelman, 1992). Give a platform a verification requirement and the platform will build the verification, document the verification, and then point to the verification whenever anything goes wrong inside. And the pattern has a second act that makes it worse than operator laziness: courts come to recognize and legitimate the symbolic structures, conferring legal benefit on what began as a gesture of compliance, so the discharge is eventually ratified by the very institutions that were supposed to police it (Edelman et al., 1999). We verified. Everyone in the space is thirteen. The certificate is on the wall.

The predictable consequence is that moderation budgets fall. The space is verified, so what is there to moderate. The fence was expensive and the fence is the safety program now. And the verified space becomes less supervised than the unverified space it replaced, which is a worse outcome than building nothing, because the parents who would have supervised are now relying on a certificate.

The fence gets built and the lifeguard gets fired.

The rule that prevents this must be stated in the architecture rather than left to enforcement discretion. Attestation establishes who is in the space. It discharges no duty owed to them once they are in it. A verified peer space is a space where supervision is legitimate and therefore mandatory, not a space where supervision has been rendered unnecessary. The credential is a premise of the duty and not a substitute for it.

Part VII.C supplies the reason this rule is not merely desirable. A gate treated as a discharge is not simply an operator being cheap. It is an operator relying on the one layer that fails silently. The captured credential attests correctly, forever, to the wrong body, and no amount of cryptographic rigor detects it, because from the gate's perspective nothing is wrong. The only detector is the watching. An operator who fires the lifeguard on the strength of the fence has removed the sole mechanism capable of finding the failure the fence cannot see.

The outdoor playground of Part I is the picture of this. Its fence works. Its gate latches. And the reason no parent leaves is that a working fence with nobody behind it protects against the wrong thing: it keeps out what wanders in, and it does nothing whatsoever about what walks in on purpose.

This failure has a name. Trust laundering, in its general form, is the substitution ritual by which a mechanism that gestures at a protection is permitted to stand in for the protection, and the audiences who should demand the protection accept the gesture. The variant here carries a difference worth marking: the mechanism works. The gate does what it says. The failure is not deception about the mechanism's function. It is the substitution of a working mechanism for a different obligation it was never addressed to. That variant is harder to litigate, because the operator can truthfully say the gate performs exactly as specified, which is why the non discharge rule has to be written down rather than inferred.

IX. What This Does Not Solve

A paper proposing an architecture is obliged to state the architecture's limits with the same specificity as its claims. Six limits are named here. None is fatal. A proposal that concealed them would deserve to fail.

A. The Guardian Who Is the Threat

An adult with physical control over a child can compel the child to unlock a token. No credential architecture stops this. None has ever claimed to, and any vendor who claims otherwise is selling something.

This is not a gap in the design so much as a category the design was never addressed to. Credential systems govern the relationship between a person and a service. They do not govern the relationship between a person and the adult in the room with them. The harm in question is a physical control harm, and physical control harms are addressed by mandatory reporting, by the supervision Part VII makes legitimate, and by the criminal law, none of which this architecture displaces and all of which it should be understood to presuppose.

The same analysis applies to the guardian who authorizes carelessly. A system cannot distinguish a permission granted after deliberation from a permission granted to end an argument, and it should not try, because building a system that sorts parents into competent and incompetent categories requires an authority to do the sorting and that authority is more dangerous than the problem. The statutory consent regimes already in force do not attempt this distinction and are right not to.

What follows holds wherever a gate protects children, and it holds here: a gate's effectiveness depends on an engaged guardian, and guardian engagement correlates inversely with the risk the child faces. The mediation literature documents both halves. Lower parental mediation is associated with higher online risk, while the engaged, active mediation styles concentrate among parents with more education and digital skill (Cabello-Hutt et al., 2018), and the large multi-country surveys find enabling mediation deployed where parent or child is already digitally skilled, with the

restrictive styles that suppress opportunity falling on the families with less (Livingstone et al., 2017). The gate therefore does the most for the children who were already safest and the least for the children in the most danger, and it will report success, because the children it failed were never in the denominator.

What the design does provide should be stated, because it is material. A hostile guardian who abuses delegation under this system signs his name to it. The authorization is dated, scoped, and legible, and when the abuse surfaces, an investigator holds a signed, dated authorization rather than an absence of records. The checkbox world leaves no difference between the guardian who consented and the child who lied; this one leaves a dated record bearing an adult's signature, and signed records are the raw material of prosecution. The architecture cannot prevent the hostile guardian. It can refuse him deniability.

That is not an argument against the gate. It is an argument against the gate being the program. The children whose guardians will never engage are reached by education and by an environment that is safe without a guardian's participation, which is a separate body of work and not this one.

B. The Lost Token and the Dead Algorithm

One reading must be closed off before the limit is stated. A lost or stolen token is not a security event. The token without its holder's local unlock is inert (Part II), so whoever finds or takes it holds a paperweight, and nothing the token protects travels with it. What loss costs is access, and access is exactly the thing this paper's own standard refuses to treat as small.

The second limit is operational and has two faces, replacement and rotation, and the paper's own standard makes both mandatory to address, because exclusion is the binding constraint and a child whose token is lost, broken, or confiscated is excluded for exactly as long as replacement takes. The replacement path is therefore part of the architecture, not an implementation detail: reissue rides the same distribution rail as issuance, initiated by the guardian or by the child through any institution that can witness identity, a school, a library, a clinic, and the latency budget is a published design requirement against which deployments are measured, in days rather than weeks. Confiscation deserves its own sentence, because a token taken by the abusive adult of the preceding subsection is not lost, and a replacement process that notifies the household defeats the child it exists for; the reissue path must be exercisable by the child through a mandated reporter without guardian notice, and that requirement is stated here so it cannot be dropped in silence.

Rotation is the same rail on a longer clock. A credential issued at birth must outlive algorithms, and nothing cryptographic issued in 1946 remained sound in 2026. The token is therefore the durable anchor and the key material the replaceable element: keys are rotated by reissue, the architecture assumes multiple reissues per holder per lifetime as its normal case rather than its failure case, and cryptographic agility is inherited from the distribution system instead of engineered into an eighty year artifact. A proposal that fixed key material for a lifetime would deserve the objection. This one converts the objection into a maintenance schedule.

C. The Unwritten Invoice

The third limit is that this paper prices nothing, and the omission is deliberate rather than evasive. Hardware cost per token, reader infrastructure, replacement logistics, and program administration are real numbers, and the honest statement is that no current public figures exist from which to derive them; the last comparable public trial ran in 2004 and its costs are two decades stale, which Part IV.C already establishes for its error rates. Pricing is downstream of the pilot this paper calls for, and any figure asserted here would be ornamental rather than derived. What can be said without a pilot is the shape of the appropriation: issuance at birth alongside the birth certificate and the Social Security

number, two documents the United States has issued universally, at public expense, through appropriations politics, for generations. The cost profile is a policy choice with two working precedents, not a novel fiscal category, and the paper's claim is only that the choice be made with the costs written down, which is more than the wallet based alternative has offered.

D. The Band Boundary

Age bands are hard cuts and the things they govern are continuous, and the paper has no clean answer to what happens where the two meet. The arbitrariness is not speculative; the age at which European jurisdictions let an adolescent consent for themselves in research settings ranges from twelve to eighteen depending on the country, while the capacity literature finds cognitive appraisal advancing sharply from age ten and plateauing around sixteen, a curve no single statutory line can honor (Samdal et al., 2023).

A sixteen year old and a thirteen year old meet inside a verified space and become friends. The friendship is unremarkable and nothing about it is a problem; the space was built to contain exactly this. Three years pass. One of them is nineteen and one of them is sixteen, and the architecture severs the friendship on a birthday, because the older one's credential no longer opens the door.

Note that the architecture does this without anything having gone wrong. It is not detecting a harm. It is enforcing a boundary that was drawn to keep a stranger out and is now falling on a friendship it watched grow. Every hard band does this at its edges, and the edges are populated, because friendships do not respect the year they were sorted into.

The obvious repairs each fail in a recognizable way. Grandfathering the relationship means the architecture maintains a permanent record of who is friends with whom, which is a social graph held by the state and is worse than the problem. Overlapping bands push the cut somewhere else without removing it. Guardian authorized exceptions return the question to the guardian, which is defensible and reintroduces every problem Part IX.A identifies. Sliding windows keyed to the age gap reduce the severity without addressing the mechanism.

The author's honest position is that this is unsolved, that the underlying difficulty is not technical, and that societies do not agree with themselves about it. The same jurisdictions that would find a nineteen year old's continued friendship with a sixteen year old alarming permit marriage at younger ages with parental consent. A design cannot resolve a boundary the society drawing it has not resolved. What a design can do is refuse to pretend the boundary is natural, and state that the severing is a cost the architecture imposes rather than a harm it prevents.

E. The Bystander

The fifth limit is the one this paper is most at risk of being credited with solving, and it is stated here so the two problems are never conflated.

A gate operates on a person who approaches it. The entire architecture above, every part of it, is a description of what happens when a human being presents themselves to a service and asks to be let in. Presentation is the precondition. Without it, there is nothing for the credential to do.

Ambient sensing systems do not wait to be approached. A camera in a domestic space captures the child who walks past it. She did not open an application. She did not create an account. She did not authenticate, did not consent, did not request, and does not know the device is there. She is not a user of anything. She is a person in a room where a machine is running.

No credential reaches her. Not this one and not any other. Zero-knowledge proofs do not reach her. Hardware tokens do not reach her. Face estimation does not reach her, except by doing to her the exact thing the architecture exists

to prevent; the predictive biometrics literature is explicit that age estimation from facial data is a mature and expanding capability (Fairhurst et al., 2017), which is a reason to keep it away from bystanders rather than a way to protect them. The question of what she is owed cannot be answered by asking what she presented, because she presented nothing, and any regime that keys its duties to the requester will discover that she is outside all of them.

Protecting her requires a rule keyed to the data subject rather than the requester. Her face is hers whether or not she asked for anything. That is a different proposal requiring a different argument, and the only thing this paper claims about it is that the present architecture is not a substitute for it and must never be offered as one.

The gate is for the child who knocks. Something else is needed for the child who was already in the room.

F. Adoption and the Interests Behind the Objections

The sixth limit is that none of this happens unless a government does it, and the argument for why a government would is not a technical argument.

The custody analysis in Part III.C predicts the opposition with some precision. The parties best positioned to build a wallet based alternative are the parties that gain most from wallet based custody, and they will oppose a standalone token on grounds framed as technical: fragmentation, cost, user experience, hardware logistics, replacement burden. Each of those objections has a real component and each of them also happens to preserve an intermediary position worth a great deal of money. The author does not claim the objections are made in bad faith. The author claims only that their proponents' interests should be legible to anyone weighing them, and that the burden of showing why the identity layer of a regulatory scheme should be operated by its targets falls on the people proposing that arrangement rather than on the people questioning it.

One further objection in this family should be named plainly, because platform counsel will raise it before platform engineering does. Verification creates knowledge and knowledge creates duty. A platform that cannot know a user's age enjoys the safe ignorance the current regime rewards, and consuming this credential ends that ignorance forever. No general counsel volunteers for that trade without a price, so the price must be stated: adoption takes the shape of a safe harbor or it does not happen. Verification in exchange for a defined duty ceiling, a platform that consumes the credential and honors its bands receives a specified, bounded obligation set in place of open ended constructive knowledge liability. That concession is made deliberately rather than reluctantly. Bounded duties discharged in full are worth more to children than unbounded duties every platform structures itself to avoid.

X. CONCLUSION

The age assurance debate has spent a decade arguing about whether to trade privacy for child safety, and the trade has never been necessary. The cryptography that dissolves the tension was published in 1983. The hardware that closes the binding gap is a W3C Recommendation. Two European states have run national credential systems for a combined thirty years.

What has never been tried is giving the credential away.

Every proposal on the table makes the subject go and get it, and the going and the getting are where the exclusion lives, because a fee is a means test, an office is a mobility test, a document is a status test, a smartphone is a wealth test, and a fingerprint reader with one modality is a disability test. None of those tests appear in the policy debate as tests. They appear as implementation details, which is how exclusion has always entered infrastructure: not by decision, but by the absence of one.

The proposal is to invert the distribution. Issue a boolean predicate, blind signed, on free standalone hardware, at

birth, to everyone, with no attribute inside it and no query on presentation. Add a second credential so that a guardian can authorize a service without asserting that their child does not exist. Then use the composition the credentials make possible to build the one thing the field has wanted and never had: a space where a child is talking to a child, where every guardian has consented, and where supervision is therefore not surveillance because there is nobody inside whose rights it would violate.

And then say plainly what it does not do. It does not stop the adult holding the child's hand. It does not reach the child who never approached. It does not discharge a single duty a platform owes to the children it has now positively identified.

The contribution here is not cryptographic. Every primitive in this paper was invented by someone else and most of them are older than the web. The contribution is the recognition that age assurance failed on distribution while everyone was arguing about privacy, and that a credential nobody has to ask for is the only kind that reaches the people every other design leaves out.

And then the second recognition, which is the one the title carries. Nobody leaves a child at the outdoor playground. The fence is good, the gate latches, and every parent in the world knows without being told that a barrier with nobody behind it is not a safety program. Parents leave children at the indoor play zone, and the reason is not a better fence. It is that somebody is watching, and the wristband is what lets them.

The age assurance debate has spent ten years perfecting a latch. The latch was never the point. The point is the room you can leave a child in, and a decade spent arguing about the gate has produced a generation of fences with nobody behind them, which is worse than no fence at all, because a fence is what persuades the parent to walk away.

ACKNOWLEDGMENTS

The author discloses the use of speech-to-text software (Wispr Flow) and Anthropic's Claude as assistive technology for a diagnosed neurodevelopmental condition. All intellectual contributions, research questions, theoretical arguments, methodological decisions, and conclusions are solely the author's own.

Correspondence concerning this article should be addressed to Travis Gilly, Real Safety AI Foundation, Illinois, United States. Email: t.gilly@ai-literacy-labs.org.

REFERENCES

Alaggia, R., Collin-Vézina, D., & Lateef, R. (2017). Facilitators and barriers to child sexual abuse (CSA) disclosures: A research update (2000–2016). *Trauma, Violence, & Abuse*, 20(2), 260–283.

Bundesrat Drucksache 139/21, *Entwurf eines Gesetzes zur Einführung eines elektronischen Identitätsnachweises mit einem mobilen Endgerät* [Draft Act introducing electronic proof of identity on a mobile device] (Ger. Feb. 12, 2021). https://dip.bundestag.de/drucksache/139-21/250501

Cabello-Hutt, T., Cabello, P., & Claro, M. (2018). Online opportunities and risks for children and adolescents: The role of digital skills, age, gender and parental mediation in Brazil. *New Media & Society*, 20(7), 2411–2431.

Camenisch, J., & Lysyanskaya, A. (2001). An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In B. Pfitzmann (Ed.), *Advances in Cryptology: EUROCRYPT 2001* (pp. 93–118). Springer.

Carah, N., Demaio, S., Holly, L., Kickbusch, I., & Williams, C. (2024). Beyond banning: Our digital world must be made safer for young people. *Health Promotion Journal of Australia*, 36(1). https://doi.org/10.1002/hpja.938

Cavallini, C., et al. (2025). Early adolescents and exposure to risks online: What is the role of parental mediation styles? *Social Sciences*.

Chaudhuri, B. (2022). Programmed welfare: An ethnographic account of algorithmic practices in the public distribution system in India. *New Media & Society*, 24(4), 887–902. https://doi.org/10.1177/14614448221079034

Chaum, D. (1983). Blind signatures for untraceable payments. In D. Chaum, R. L. Rivest, & A. T. Sherman (Eds.), *Advances in Cryptology: Proceedings of Crypto 82* (pp. 199–203). Springer.

Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§6501–6506.

Chudnovsky, M., & Peeters, R. (2021). A cascade of exclusion: Administrative burdens and access to citizenship in the case of Argentina's national identity document. *International Review of Administrative Sciences*.

Collings, S. J., Griffiths, S., & Kumalo, M. (2005). Patterns of disclosure in child sexual abuse. *South African Journal of Psychology*, 35(2), 270–285. https://doi.org/10.1177/008124630503500207

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3, 1376. https://doi.org/10.1038/srep01376

de Montjoye, Y.-A., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221), 536–539. https://doi.org/10.1126/science.1256297

Gesetz über eine Karte für Unionsbürger und Angehörige des Europäischen Wirtschaftsraums mit Funktion zum elektronischen Identitätsnachweis [eID-Karte-Gesetz] [Act on a card for Union citizens and EEA nationals with an electronic proof of identity function] (Ger. June 21, 2019). https://www.gesetze-im-internet.de/eidkg/

Common Sense Media. (2025, April 10). *AI risk assessment: Social AI companions*. Common Sense Media. (Rating: unacceptable; use case review; conducted with input from the Stanford Brainstorm Lab for Mental Health Innovation.)

Edelman, L. B. (1992). Legal ambiguity and symbolic structures: Organizational mediation of civil rights law. *American Journal of Sociology*, 97(6), 1531–1576.

Edelman, L. B., Uggen, C., & Erlanger, H. S. (1999). The endogeneity of legal regulation: Grievance procedures as rational myth. *American Journal of Sociology*, 105(2), 406–454.

European Parliament and Council. (2024). Regulation (EU) 2024/1183 of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework. *Official Journal of the European Union*, L, 2024/1183.

Fairhurst, M., Li, C., & Da Costa-Abreu, M. (2017). Predictive biometrics: A review and analysis of predicting personal characteristics from biometric data. *IET Biometrics*, 6(6), 369–378. https://doi.org/10.1049/iet-bmt.2016.0169

Finkelhor, D., Turner, H., & Colburn, D. (2022). Prevalence of online sexual offenses against children in the US. *JAMA Network Open*, 5(10), e2234471.

Fry, D., et al. (2025). Prevalence estimates and nature of online child sexual exploitation and abuse: A systematic review and meta-analysis. *The Lancet Child & Adolescent Health*.

Gambs, S., Killijian, M.-O., & Núñez-del-Prado Cortez, M. (2014). De-anonymization attack on geolocated data. *Journal of Computer and System Sciences*, 80(8), 1597–1614. https://doi.org/10.1016/j.jcss.2014.04.024

Gilly, T., & Mattheus, R. (2026). Harm blindness in digital child protection policy: A systematic stakeholder, sentinel and legal analysis of Australia's under-16 social media ban. In K. Ringmar Sylwander, G. Shekhawat, & S. Livingstone (Eds.), *Children's rights under pressure in a digital world: Proceedings of the International Communication Association pre-conference* (pp. 60–65). Digital Futures for Children centre, LSE and 5Rights Foundation. https://researchonline.lse.ac.uk/id/eprint/138632/

Isikut tõendavate dokumentide seadus [Identity Documents Act] (Est. Feb. 15, 1999), Riigi Teataja. https://www.riigiteataja.ee/akt/77253

Livingstone, S., Ólafsson, K., Helsper, E. J., Lupiáñez-Villanueva, F., Veltri, G. A., & Folkvord, F. (2017). Maximizing opportunities and minimizing risks for children online: The role of digital skills in emerging strategies of parental mediation. *Journal of Communication*, 67(1), 82–105.

Martin, A., et al. (2020). Exclusion and inclusion in identification: Regulation, displacement and data justice. *Information Technology for Development*.

Manay, N., & Collin-Vézina, D. (2019). Recipients of children's and adolescents' disclosures of childhood sexual abuse: A systematic review. *Child Abuse & Neglect*, 98, 104192.

Mathews, B., et al. (2025). Characteristics of first disclosure of child sexual abuse: Age, delay, recipient, and feeling of support. *Child Abuse & Neglect*.

McElvaney, R. (2013). Disclosure of child sexual abuse: Delays, non-disclosure and partial disclosure. What the research tells us and implications for practice. *Child Abuse Review*, 24(3), 159–169. https://doi.org/10.1002/car.2280

Mukherjee, A., & Sahay, S. (2019). Sinking under its own weight: Case of Aadhaar mediated entitlements in India. In P. Nielsen & H. C. Kimaro (Eds.), *Information and Communication Technologies for Development: Strengthening Southern-Driven Cooperation as a Catalyst for ICT4D* (pp. 472–485). Springer. https://doi.org/10.1007/978-3-030-18400-1_39

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy* (pp. 111–125). IEEE. https://doi.org/10.1109/SP.2008.33

Personalausweisgesetz [PAuswG] [Identity Card Act] (Ger. June 18, 2009), BGBl. I at 1346.

Personalausweisgebührenverordnung [PAuswGebV] [Regulation on fees for identity cards and eID cards for Union citizens and EEA nationals] (Ger. Nov. 1, 2010). https://www.gesetze-im-internet.de/pauswgebv/

Personalausweisverordnung [PAuswV] [Regulation on identity cards, eID cards for Union citizens and EEA nationals, and electronic proof of identity] (Ger. Nov. 1, 2010). https://www.gesetze-im-internet.de/pauswv/

Ólafsdóttir, B., et al. (2024). A case study on the inclusiveness of biometric technologies for individuals with congenital disabilities. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction*. ACM.

Poh, N., et al. (2025). Biometric bound credentials for age verification. In *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*.

Rao, U., & Nair, V. (2019). Aadhaar: Governing with biometrics. *South Asia: Journal of South Asian Studies*, 42(3), 469–481. https://doi.org/10.1080/00856401.2019.1595343

Riaz, I., et al. (2024). Loss of fingerprint features and recognition failure due to physiological factors: A literature survey. *Multimedia Tools and Applications*.

Rose, S. E., & Middling, L. R. (2025). Preteens social media use: Parents' and children's perceptions of what mediation approaches are used and why. *British Journal of Developmental Psychology*, 43(3), 771–786. https://doi.org/10.1111/bjdp.12552

Rosenberg, M., White, J., Garman, C., & Miers, I. (2023). zk-creds: Flexible anonymous credentials from zkSNARKs and existing identity infrastructure. In *2023 IEEE Symposium on Security and Privacy* (pp. 790–808). IEEE.

Samdal, O., Budin-Ljøsne, I., Haug, E., Helland, T., et al. (2023). Encouraging greater empowerment for adolescents in consent procedures in social science research and policy projects. *Obesity Reviews*, 24(S2). https://doi.org/10.1111/obr.13636

Schittenhelm, C., et al. (2024). Cybergrooming victimization among young people: A systematic review of prevalence rates, risk factors, and outcomes. *Adolescent Research Review*.

Thimm-Kaiser, M., & Keyes, K. M. (2025). US state policies regarding social media: Do policies match the evidence? *The Milbank Quarterly*, 103(S1), 337–365. https://doi.org/10.1111/1468-0009.70021

Wang, K., et al. (2023). Biometrics-based mobile user authentication for the elderly: Accessibility, performance, and method design. *International Journal of Human-Computer Interaction*.

UK Passport Service. (2005). *Biometrics enrolment trial: Report*. Atos Origin.

UKPS biometric enrolment trial. (2005). Biometric Technology Today. https://www.sciencedirect.com/

science/article/abs/pii/S0969476505703684

W3C. (2021). *Web Authentication: An API for publicly accessing public key credentials, Level 2* [W3C Recommendation]. World Wide Web Consortium. https://www.w3.org/TR/webauthn-2/

Zafeiropoulou, A., et al. (2025). Age verification in the context of the EUDI Wallet: Balancing privacy and security. In *Proceedings of the 16th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE.

Zhong, W., Luo, J., Luo, L., & Lyu, Y. (2025). The role of social robot in enhancing social–emotional skill development in children with autism: A three-level meta-analysis. *Journal of Computer Assisted Learning*, 41(6). https://doi.org/10.1111/jcal.70154