Knight–Georgetown
Institute

Age Assurance Online: A Technical Assessment of Current Systems and Their Limitations

February 2026

Agenda

1. Use cases, roles, and assessment criteria
2. Age assurance architectures
3. Age signals
4. Key findings

Use Cases

Safer Defaults

Designed to provide users with an experience deemed more age-appropriate, such as restricting the use of personalized feeds or of notifications during certain hours.

Typically involve a long term relationship.

May be less of a perceived need for anonymous or pseudonymous access.

May be less incentive to circumvent if experience not adversely affected.

Blocking

Content and experience are blocked entirely for minors.

May support access by unidentified users without accounts and the expectation is that service providers block underage users with no previous history of interaction.

Even in cases where accounts are required, users may wish to remain pseudonymous or anonymous.

Minors may be more motivated to circumvent age assurance in these cases if it prevents them from accessing content or experiences that they want.

Age Assurance Roles

The **user** wants to access a specific type of content

The **service provider** (e.g., website, app provider) the user is trying to access

The **evaluator** determines whether the user is within the eligible age range

The **enforcer** controls the user's access to the content

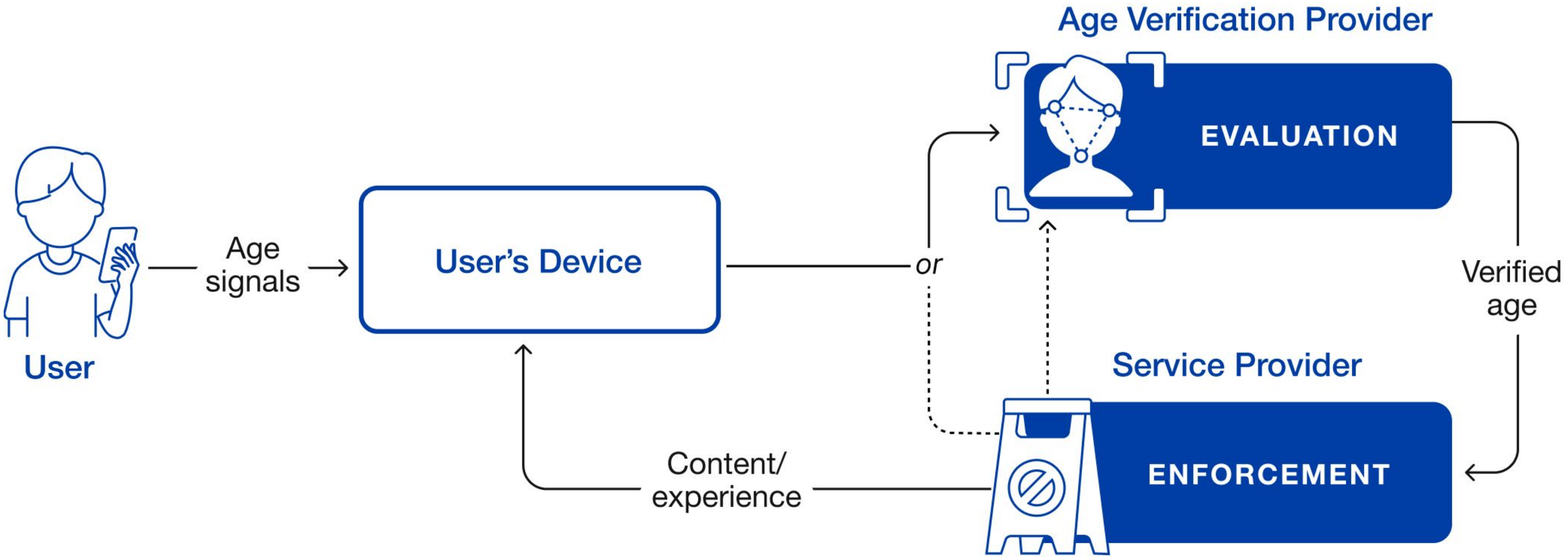
Assessment Criteria

- **Baseline accuracy**: the accuracy of the system in the absence of any attempts by the user to circumvent it.
- **Circumvention resistance:** the degree to which the system resists attempts by users to establish an age different from their true age.
- **Availability:** the degree to which the system will be usable by the eligible population.
- **Privacy:** the degree to which use of age assurance by a user reveals information that would not be accessible without the use of age assurance.

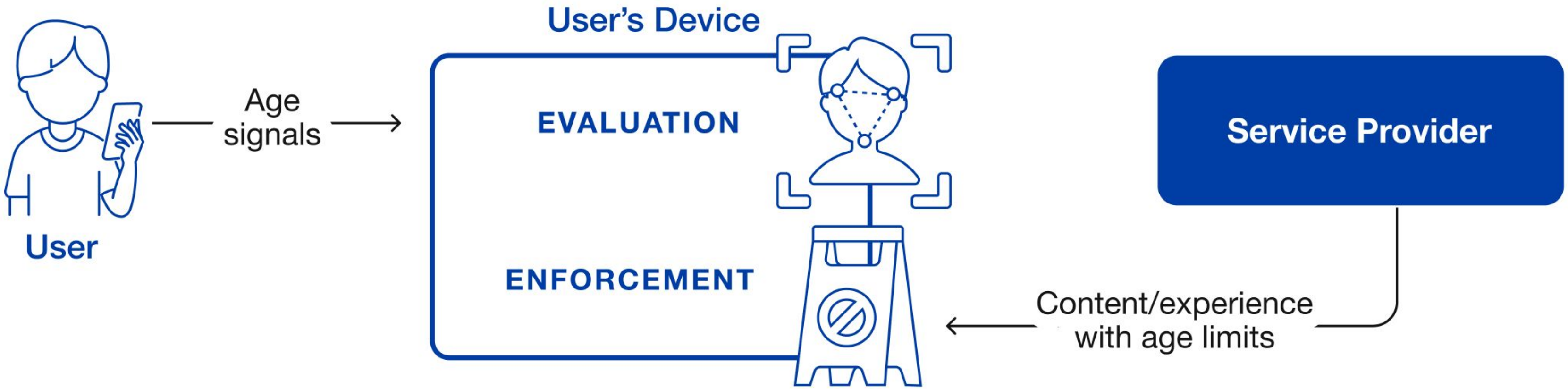
Age Assurance Architectures

1. Server-Based Evaluation and Enforcement
2. Device-Based Evaluation and Enforcement
3. Device-Based Evaluation, Server-Based Enforcement

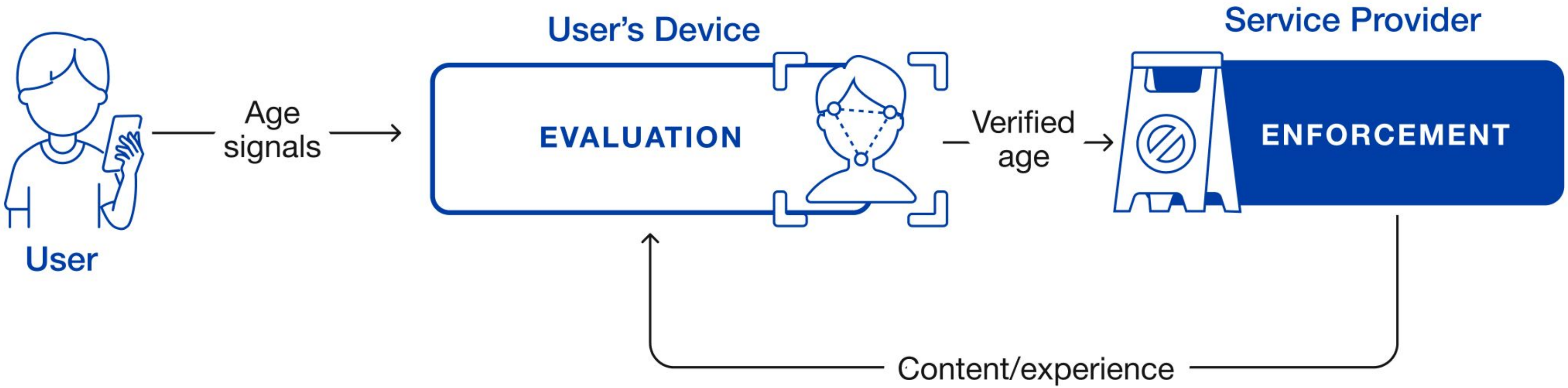
Server-Based Age Evaluation and Enforcement



Device-Based Evaluation and Enforcement



Device-Based Evaluation, Server-Based Enforcement



Assessment of Age Assurance Architectures

	Baseline Accuracy	Circumvention	Availability	Privacy
Server-Based Evaluation and Enforcement	Depends on underlying age signals. Applying the correct jurisdictional policy depends on the server being able to determine the user's location.	Vulnerable to location spoofing via VPNs and to injection attacks on untrusted devices (for apps) and on the web generally.	High if untrusted devices are acceptable. Much lower if trusted devices are required to prevent injection attack.	Evaluators frequently learn information about the user, which can be abused.
Device-Based Evaluation	Depends on how the device determines the user's age.	Depends on whether the user can obtain an unlocked device or get an adult to obtain one for them. Circumvention is easier on desktop.	Device-based enforcement only restricts behavior on devices which are configured to enforce restrictions. Mobile app users on non-upgraded devices may be excluded.	Service providers do not learn anything other than that the user is in the eligible age range. Any user who wants an unrestricted experience must undergo age assurance.

Age Signals

1. Commercial and Government Records
2. Government IDs
3. Facial Age Estimation
4. Behavioral Signals

Commercial and Government Records

Banking Records: Banks already identify customers. Evaluator asks user to login to their bank account and bank returns an answer.

Credit cards: Some jurisdictions restrict credit cards to 18+. Evaluator asks for credit card number and does a small charge, then reverses it.

Mobile Network Operator Verification: Some jurisdictions (UK) provide a restricted experience for under 18s. Evaluator queries operator for status.

Other records: Evaluator asks user for a long-term identifier such as phone number or email address and then looks for records that indicate the user's age.

Assessment of Commercial and Government Records (I)

	Baseline Accuracy	Circumvention	Availability	Privacy
Banking Records	High.	Easy with access to an adult's account. Difficult otherwise.	Depends on having a bank account. A significant fraction of adults do not. Low availability for below 18s. Requires the bank to provide an API.	Evaluator does not learn the user's identity, but bank learns about age assurance. Evaluator learns the user's banking institution.
Mobile Network Operator Verification	Depends on the MNO's procedures for verifying age.	Easy with cooperation of an adult or temporary access to an adult's phone.	Only available in jurisdictions that impose default restrictions on mobile phones. Not practical for under 18s. Requires carriers to provide an API	Evaluator learns the user's mobile number.

Assessment of Commercial and Government Records (II)

	Baseline Accuracy	Circumvention	Availability	Privacy
Credit Cards	Depends on issuer's procedures for verifying age.	Easy with cooperation of an adult or temporary access to an adult's credit card.	Only available in jurisdictions where credit cards are age-restricted. Depends on having a credit card, which a significant number of adults do not. Low availability for under 18s.	Evaluator learns the user's credit card number and usually postal code, and may learn the user's name and address if payment processor requires it.
Other Commercial and Government Records	Unknown. Reported false reject rates in excess of 10%.	Depends on the identifying information used. For birthdate, address, and SSN, fairly easy. Email address or mobile number verification is easy to circumvent with assistance of an adult, difficult otherwise.	Depends on quality of records. Reported false reject rates in excess of 10% suggests that this may be low.	Evaluator learns the user's identity or a proxy for their identity such as email address. Stored records are difficult to anonymize.

Government IDs

Physical ID: Leverages existing IDs (driver’s licenses, national identity card). User shows their card to the camera and maybe takes a selfie.

Digital ID: User loads their physical ID onto their mobile device. Can then use it to remotely demonstrate their age.

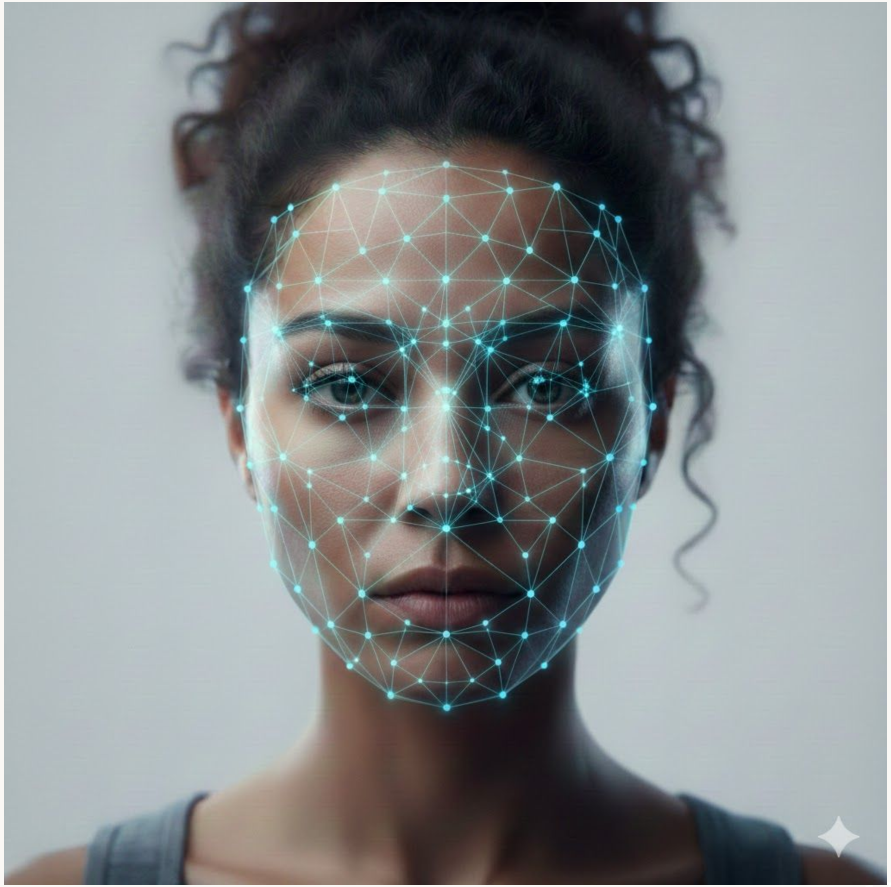
Digital ID with zero-knowledge proofs: Like regular digital IDs but uses cryptography to improve privacy.

Assessment of Government IDs

	Baseline Accuracy	Circumvention	Availability	Privacy
Physical IDs	High.	Users may acquire a fake ID or attempt to use a borrowed ID. Remote attack detection is difficult.	Depends on prevalence of the underlying credential. In jurisdictions where IDs are not mandatory, significant fractions of adults do not have them.	Evaluator learns the user's identity as well as other personal information such as address. Evaluators may be able to misuse face image if provided.
Digital IDs	High.	Depends on the security of the device. May be possible for an adult to enroll their ID in a minor's device or allow their device to be used for a one-time age assurance.	Depends on prevalence of the underlying credential. Also requires a device which can enroll that credential for age assurance, which is not currently available in most jurisdictions.	Only reveals the user's age eligibility and not identity. Allows for linkage with the assistance of the credential issuer. May allow for linkage between evaluators if credentials are reused.
Digital IDs with zero-knowledge proofs	High.	Same as for Digital IDs.	Same as for Digital IDs.	Only reveals the user's age eligibility and not identity.

Facial Age Estimation

User uses their device camera to take a selfie or video. Evaluator uses an AI model to estimate the user’s age.

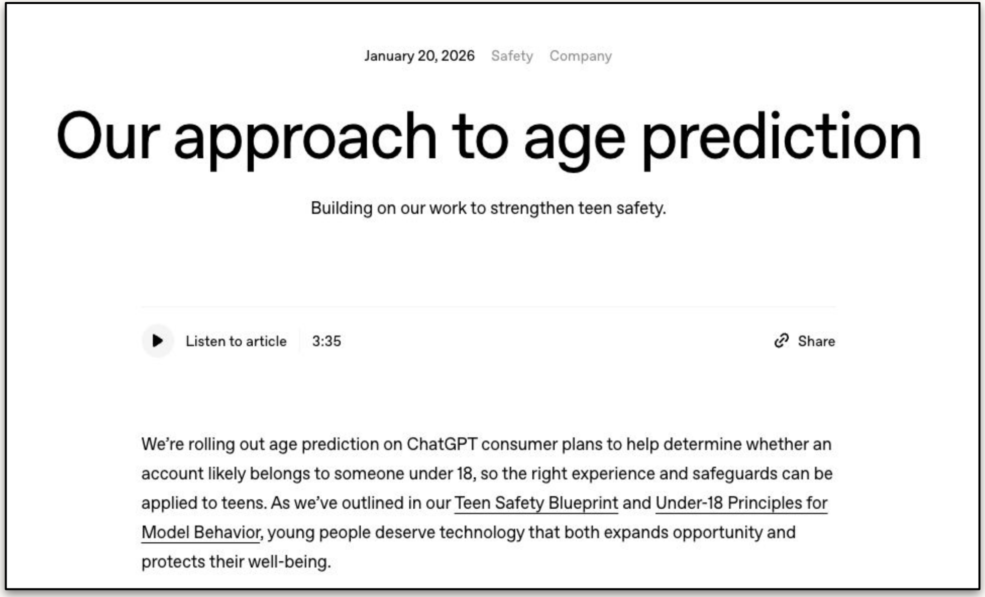


Assessment of Facial Age Estimation

	Baseline Accuracy	Circumvention	Availability	Privacy
Facial Age Estimation	Many users in the eligible age range are rejected.	Depends on the implementation. Vulnerable to presentation attacks and very vulnerable to injection attacks.	Requires a device with a camera. If trusted devices are required to prevent injection attacks, then cannot be used on the web.	Evaluator learns the user's face. May be able to use this to identify the user or misuse it in other ways.

Behavioral Signals

Many sites already measure user behavior for recommendation and advertising purposes. This information can be repurposed for age estimation.



January 20, 2026 Safety Company

Our approach to age prediction

Building on our work to strengthen teen safety.

We're rolling out age prediction on ChatGPT consumer plans to help determine whether an account likely belongs to someone under 18, so the right experience and safeguards can be applied to teens. As we've outlined in our Teen Safety Blueprint and Under-18 Principles for Model Behavior, young people deserve technology that both expands opportunity and protects their well-being.

Assessment of Behavioral Signals

	Baseline Accuracy	Circumvention	Availability	Privacy
Behavioral Signals	Unknown.	Unknown. Opening a new account or using privacy tools can prevent creation of a behavioral profile.	High. Challenging to use for primary age assurance because it cannot provide results for new users.	Requires storing and retaining a profile of user behavior, even if the provider does not already do so.

Key Findings

Multiple use cases

There are multiple use cases for age assurance, each with different requirements and challenges.

Privacy protection

The most commonly deployed age assurance approaches present privacy risks, even though more privacy-protective approaches are possible and becoming more widely available.

Multiple age signals

No single age signal is sufficient on its own.

Circumvention

All age assurance systems are vulnerable to circumvention.

Thank *You*