NOVARA
CONSULTING GROUP

The Consent Gap: Undisclosed Data Collection in Sign Language AI Interpreting Systems

Heather Grizzle · July 27, 2026

Sign language AI depends on video of Deaf users. What vendors disclose about collection, retention, training use and sharing is inconsistent or absent.

Contents

1. Executive Summary
2. Policy Problem
3. Institutional Analysis
4. Governance Implications
5. NCG Policy Position
6. Implementation Considerations
7. References

Executive Summary

Sign language artificial intelligence, whether built for recognition, translation, or avatar-based generation, depends structurally on video or biometric capture of Deaf and hard of hearing users. This dependency creates a data governance problem that has received far less institutional scrutiny than the linguistic and comprehension failures documented in prior issues of this series. Across the vendor landscape NCG has reviewed under the SLAT (Sign Language Access Trust) framework, disclosure of what is collected, how long it is retained, whether it is used to train future models, and whether it is shared with third parties is inconsistent at best and absent at worst. This is not a claim that any specific vendor has acted in bad faith. Public evidence does not yet support attribution of intent to any named company, and this issue does not make that attribution. It is a claim that the market has not yet developed a disclosure standard adequate to the sensitivity of the data involved, and that this gap is a governance

failure independent of any individual vendor's conduct. The policy position below argues that consent and disclosure adequacy should be treated as a gating criterion in procurement, not a secondary consideration addressed after functional evaluation.

Policy Problem

Sign language capture is biometric capture. Video of a signer's hands, facial grammar, and body position is not incidental input, comparable to a typed sentence submitted to a text-based system. It is uniquely identifying, it encodes information about disability status by definition, and in many jurisdictions it meets the statutory definition of biometric or special category data subject to heightened protection. Illinois's Biometric Information Privacy Act (BIPA), for instance, requires written notice and consent before collection of biometric identifiers, and imposes liability independent of proof of actual harm. The European Union's GDPR (General Data Protection Regulation) classifies data revealing disability-adjacent information, and certain biometric data used for identification, as special category data under Article 9, requiring an explicit legal basis beyond ordinary consent. California's CCPA (California Consumer Privacy Act) and CPRA (California Privacy Rights Act) impose their own disclosure and opt-out obligations for biometric information.

Despite this regulatory landscape, consumer-facing sign language AI products frequently rely on general-purpose terms of service that were not drafted with Deaf users or biometric capture in mind. Three recurring deficiencies are visible across the vendor snapshots NCG has conducted: first, no separate or conspicuous disclosure that video input constitutes biometric data subject to specific statutory protections; second, no clear statement of whether captured video or derived features are retained for model training, and if so, for how long and under what deletion rights; third, no accessible-format disclosure, meaning the consent language itself is often not translated into ASL or presented in a format usable by the population whose data is being collected. This last point is not a technicality. A disclosure regime that requires literacy in a written second language to understand what is being done with one's own biometric data is not functioning as informed consent for a population for whom English literacy varies widely and for whom ASL is the primary language of comprehension.

is biometric capture.

Institutional Analysis

The institutional dynamics that produce this gap are structural rather than conspiratorial. Sign language AI is a young product category, and most vendors are small companies operating under the same generic privacy policy templates used across the software industry, adapted minimally if at all. Procurement processes at the institutional level, meanwhile, have historically evaluated these products on functional criteria: does the avatar render intelligibly, does the recognition system achieve acceptable accuracy, does the vendor have Deaf staff involved in development. Data governance has been treated as a legal or IT compliance question addressed after the substantive procurement decision, if it is addressed at all. This sequencing is the actual governance failure. By the time a data governance deficiency surfaces, either through a breach, a regulatory inquiry, or public criticism from the Deaf community whose data was involved, the institutional relationship and financial commitment are already established, and the leverage to demand remediation is correspondingly weaker.

There is also a training data provenance question that sits adjacent to the consent problem but is analytically distinct from it. Several sign language AI systems have been built in part on video corpora drawn from publicly posted content, including content created by Deaf ASL users and educators on social platforms. Whether that use falls within the scope of the platform′s own terms of service, whether it constitutes fair use under copyright law, and whether the individuals whose signing appears in that footage were meaningfully informed that their content would be used to train a commercial AI system, are open questions that NCG has not independently verified for any specific vendor and does not assert as established fact here. The point for governance purposes is that procurement officers evaluating these systems have almost no visibility into training data provenance, and current disclosure norms in the industry do not require vendors to provide it.

Governance Implications

An institution procuring sign language AI is not merely a purchaser of a translation tool. Where the system operates on live video of students, patients, employees, or members of the public, the institution is also a data controller or processor under most applicable privacy frameworks, and it inherits governance exposure from its ven-

dor's data practices whether or not that exposure was disclosed at the time of purchase. An institution that has not verified a vendor's data collection, retention, and training-use practices before deployment cannot represent to its own stakeholders, regulators, or the Deaf individuals whose data is being captured, that it has met its own data protection obligations. This is true regardless of the vendor's actual conduct, because the absence of verified disclosure is itself the governance defect, independent of whether any given vendor turns out to be handling data responsibly.

NCG Policy Position

NCG's position is that data governance and privacy disclosure adequacy should function as a gating criterion in sign language AI procurement, evaluated before functional performance rather than after it. A vendor's inability to produce clear, ASL-accessible, jurisdiction-appropriate disclosure of what biometric data is collected, how long it is retained, whether and how it is used in model training, and what deletion rights exist, should be treated as disqualifying for institutional deployment regardless of linguistic or technical performance elsewhere in the evaluation. This position is already operationalized in the SLAT framework's Data Governance and Privacy and Transparency and Disclosure domains, and this issue argues that those domains warrant elevated weighting relative to the framework's current calibration, given the statutory exposure institutions face and the comprehension barrier that non-accessible consent language creates for the affected population.

Implementation Considerations

Procurement officers evaluating sign language AI should require, as a condition of vendor response, a written data governance disclosure covering the four elements above, request that this disclosure be available in ASL video format in addition to written English, and treat vendor refusal or inability to produce this disclosure as an adverse finding independent of the system's functional performance. Institutions with existing deployments should conduct a retrospective review rather than waiting for contract renewal, given that governance exposure accrues from the date of deployment rather than the date of discovery. Where a vendor's training data provenance cannot be verified, institutions should request contractual warranties regarding lawful data sourcing rather than treating the absence of an answer as a neutral unknown.



THE FOUR DISCLOSURE ELEMENTS

What biometric data is collected, how long it is retained, whether and how it is used in model training, and what deletion rights exist.

References

- Biometric Information Privacy Act, 740 ILCS 14 (Illinois).
- California Consumer Privacy Act of 2018, as amended by the California Privacy Rights Act, Cal. Civ. Code § 1798.100 et seq.
- Regulation (EU) 2016/679 (General Data Protection Regulation), Art. 9.
- Federal Trade Commission Act, 15 U.S.C. § 45 (unfair or deceptive acts or practices).
- Novara Consulting Group, *Sign Language Access Trust (SLAT) Framework*, 2026.

TIPS

Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

How to cite

Grizzle, H. M. (2026, July 27). The Consent Gap: Undisclosed Data Collection in Sign Language AI Interpreting Systems. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003818