AI Can Translate. But Who Translates the Risk?

Heather M. Grizzle · August 17, 2026

AI can translate, but who translates the risk? Explore the AI accountability gap, responsible AI governance, procurement standards, and sign language AI oversight.

Inhalt

Every AI deployment moves risk onto someone. In most cases it lands on the person with the least ability to detect an error and the least authority to challenge it.

Heather M. Grizzle Novara Consulting Group August 17, 2026

Earlier this month I wrote about Google DeepMind's sign language translation research and argued that a demonstration is not the same thing as evidence. The response told me something useful. Almost nobody disagreed that the distinction matters. Many people asked the obvious follow-up question instead: if a demonstration is not

enough, who is supposed to decide when a system is good enough to deploy, and who answers for it when it is not?

That question is not specific to sign language. It is the central unresolved problem in AI governance right now, and it has a name worth using plainly. Call it the accountability gap.

What is the AI accountability gap?

The AI accountability gap is the distance between what a system is capable of doing and what any identifiable party is answerable for when it does that thing badly. It opens whenever an organization deploys an AI system faster than it builds the means to verify the system's output, hear from the people the output affects, and correct or stop the deployment when the output is wrong.

The gap is not primarily a technical problem. Models fail in predictable ways, and engineers are usually candid about those failure modes. The gap is an institutional problem. Organizations are acquiring systems whose outputs they cannot evaluate, deploying them into settings where errors carry consequences, and retaining no practical mechanism to find out whether the system is working.

Capability, reliability, and accountability are three different kinds of evidence

Most procurement conversations treat these as one thing. They are not, and separating them is the most useful move an institutional buyer can make.

Capability evidence shows that a system can perform a task under conditions the vendor selected. Demonstrations, benchmark scores, and pilot footage are capability evidence. They answer the question "is this possible."

Reliability evidence shows how the system performs across the range of people, settings, and conditions the buyer will actually encounter, including the ones the vendor did not choose. Disaggregated performance data, independent testing, and error rates from live deployment are reliability evidence. They answer the question "how often is this wrong, and for whom."

Accountability evidence shows what happens when the system fails. It names who monitors performance, who receives complaints, who has authority to suspend the de-

ployment, what the affected person is owed, and how any of that is enforced. It answers the question "who is responsible."

Vendors compete on capability evidence because it is the cheapest to produce and the most persuasive to watch. Reliability evidence is expensive and unflattering. Accountability evidence is not really the vendor's to produce at all, because it describes the buyer's own governance, and most buyers have not built it.

An organization that accepts capability evidence as though it settled the other two questions has not made a purchasing decision. It has made an assumption and paid for it.

Why disclosure laws do not close the gap

The most visible regulatory response to AI risk right now is transparency. Under Article 50 of the EU AI Act (European Union Artificial Intelligence Act), obligations that became applicable on 2 August 2026 require providers to mark synthetic content in machine-readable form and require deployers to disclose deepfakes and certain AI-generated text on matters of public interest, with penalties reaching 15 million euros or three percent of total worldwide annual turnover. The European Commission allowed a grace period until 2 December 2026 for generative systems already placed on the market.

This is meaningful work and I support it. It is also worth being precise about what it does. Article 50 is a provenance regime. It tells a person that a machine was involved. It does not tell them whether the machine was right.

For a great deal of synthetic media, provenance is the whole question. If the concern is a fabricated video of a public figure, knowing the video is synthetic resolves it. But for AI systems that mediate communication, provide information, or influence a decision about a person, provenance is the easy half. A label reading "this translation was generated by AI" leaves the recipient exactly where they started, which is unable to tell whether the content is accurate.

The liability half of the European approach has moved in the opposite direction. The Commission's proposed AI Liability Directive, which would have eased the burden of proof for people harmed by AI systems, was formally withdrawn in October 2025. Transparency advanced. Redress did not.

The pattern repeats in the United States at the state level. Colorado passed the first comprehensive state AI law in 2024, then repealed and replaced it with SB 26-189, signed on 14 May 2026, whose substantive developer and deployer obligations apply from 1 January 2027. The replacement dropped the mandatory risk management programs, the algorithmic impact assessments, and the freestanding duty of reasonable care against algorithmic discrimination. What survived were pre-use notice, a plain-language disclosure within thirty days of an adverse decision, three-year record retention, and a right to request meaningful human review.

I am not arguing those surviving obligations are worthless. Notice and record retention are real. But notice tells you a system was used. Records tell you what it did. Neither establishes that the output was accurate, and neither assigns responsibility for the consequences if it was not. The direction of regulatory travel over the past two years has been toward telling people that AI was involved and away from making anyone answer for what it produced.

Verification asymmetry: the structural problem underneath

Here is the part I think is underexamined, and it is where the sign language case stops being a niche example and starts being the clearest available illustration of a general condition.

In most AI deployments, the person best positioned to detect an error has no authority to act on it, and the party with authority to act has no means of detecting it. I would call this verification asymmetry, and it is the mechanism that produces the accountability gap.

Consider what a hearing institution sees when it deploys an automated sign language system. It sees that the system produced output. Staff observe an avatar signing or captions appearing. From that vantage point the transaction looks complete.

Now consider what the Deaf person experiences. They receive output they cannot check against the source, delivered by a system they did not select, in a setting where the institution has already concluded that access was provided. If the rendering drops a negation, flattens a conditional, misplaces spatial reference, or omits the non-manual markers that carry grammatical meaning, the person may not know. If they do noti-

ce something is wrong, they are in the position of contesting the institution's own conclusion that communication succeeded.

That is not a technology complaint. It is a distribution of power. The institution holds the authority to determine whether access occurred and lacks the linguistic capacity to evaluate it. The Deaf person holds the linguistic capacity and lacks the authority. The error, if there is one, has nowhere to surface.

Once you see the shape of it, you find it nearly everywhere. A patient cannot evaluate an AI-assisted triage recommendation. An applicant cannot inspect the model that screened them out. A benefits claimant cannot audit the eligibility determination. A person receiving machine-translated legal instructions in any language cannot verify the translation without the fluency the translation exists to supply. In each case the affected party absorbs a risk they cannot measure, and the deploying institution records a completed transaction.

What makes signed language a particularly sharp case is the stakes attached to the content. Language in institutional settings carries consent, testimony, diagnosis, instruction, and legal effect. A mistranslation is not a degraded user experience. It is a defective consent form, an inaccurate medical history, or a misrepresented statement in a proceeding.

Accessibility law already answered part of this, and we are unlearning it

There is a workable answer already sitting in accessibility practice, which makes its current neglect harder to excuse.

Under Department of Justice regulation at 28 CFR 35.160, a public entity must give primary consideration to the auxiliary aid or service requested by the person with a disability. The rule for private public accommodations at 28 CFR 36.303 is weaker, requiring consultation while leaving the final choice with the provider, and that difference is itself instructive about where verification authority tends to erode. The National Association of the Deaf's minimum standards for video remote interpreting in medical settings state the underlying principle directly: a deaf or hard of hearing individual knows best which auxiliary aid or service will achieve effective communication. Those same standards go further and require the video interpreter to tell the provider

to terminate the remote session and obtain an on-site interpreter when the interpreter determines the remote arrangement is not producing effective communication.

Read that as a governance design rather than an accessibility rule and it is unusually well built. It locates the verification judgment with the person who can actually make it. It gives a professional in the room the authority to stop a deployment mid-use. It treats effectiveness as something to be determined in context rather than assumed from the presence of equipment.

Automated systems tend to strip both features out. There is no qualified interpreter positioned to call a halt, because the interpreter is what the system replaced. And the affected person's judgment about whether communication worked is quietly demoted from a legal standard to customer feedback.

The World Federation of the Deaf and the World Association of Sign Language Interpreters set a boundary on signing avatars back in 2018. Their position was that avatars are inappropriate for live, complex, or significant information, and that pre-recorded static content may be acceptable where Deaf people advised on the material and no live interaction is required. That statement is eight years old. The technology has advanced considerably since. The deployment boundary it drew has not been replaced by anything more rigorous, and it is still the clearest line available.

Why the vendor demonstration cannot be the governance record

The demonstration is designed to be persuasive. That is its function, and there is nothing improper about it. The problem is that in the absence of an independent evaluation, the demonstration becomes the evidentiary basis for a procurement decision by default, because it is the only document in the room.

A demonstration establishes possibility under selected conditions. It cannot establish performance across the population the buyer serves, behavior outside controlled conditions, safety in high-consequence settings, or accountability when errors occur. Those are properties of a deployment, not properties of a model.

Federal policy already recognizes this in principle. OMB (Office of Management and Budget) guidance issued in April 2025 defines high-impact AI as systems whose output serves as a principal basis for decisions or actions with a legal, material, binding,

or significant effect on rights or safety, with the memo enumerating domains including civil rights, access to education, housing, employment, government benefits and services, and health. For those uses it requires pre-deployment testing, AI impact assessments before deployment and periodically thereafter, adequate human oversight, and timely human review with an opportunity to appeal for affected individuals. Agencies that cannot meet those minimums must safely discontinue the use. The companion acquisition memo requires contract terms giving the agency the ability to regularly monitor and evaluate the performance, risks, and effectiveness of the system.

Whatever one thinks of the details, the structure of that guidance is correct. It treats accountability as something specified in the contract, tested before launch, monitored after launch, and enforceable by suspension. That is a verification layer. Most institutions buying AI outside the federal acquisition system have nothing equivalent.

I would add one caution about assuming capability will arrive on schedule and close the gap on its own. In April 2026 the Department of Justice pushed the compliance deadlines for the ADA (Americans with Disabilities Act) Title II web accessibility rule back by roughly a year for each category of entity, moving large public entities to 26 April 2027 and smaller ones to 26 April 2028. Among the reasons the Department gave was that it had overestimated the capabilities, in staffing and in technology, of covered entities, and it stated plainly that advanced technology such as generative AI does not yet reliably automate the remediation of inaccessible content at scale. That is a federal regulator recording that AI capability claims outran delivery in an accessibility context. It is worth sitting with before treating any vendor timeline as a governance plan.

What institutions should require before deployment

The questions below are not exotic. They are the questions a competent buyer would ask about any system that carries consequence, and they are answerable.

Who evaluated this system, and were they independent of the vendor and of the purchaser?

Which populations, dialects, regional variants, and settings were included in testing, and what were the results for each rather than in aggregate?

What data was the system trained on, under what licence, and with what consent from the people whose language it captured?

What is the measured error rate in conditions resembling ours, and what happens to that rate as conditions degrade?

Who in our organization has the authority to suspend this deployment, and what triggers that authority?

How does an affected person report that the output was wrong, who receives that report, and what changes as a result?

What are we contractually entitled to know about performance after go-live, and what is our remedy if performance falls?

An organization that cannot answer these questions has not evaluated a system. It has accepted one. Novara Consulting Group built the Sign Language Access Trust Index to make these questions answerable in a form procurement can actually use, with defined evidence standards, indicator-level scoring, and a distinction between what a vendor asserts and what an independent party can confirm. The instrument is specific to sign language AI. The underlying discipline is not.

The next stage of AI is verification

The public conversation about AI has been organized around capability for three years. Can it write, can it reason, can it see, can it act, can it interpret human communication. Those questions are being answered quickly, and often in the affirmative.

The question that determines whether any of it is trustworthy is different. It is whether the institutions deploying these systems can tell when the systems are wrong, and whether the people affected by them have any standing to say so.

Right now, in most deployments, the answer to both is no. The risk has not been translated. It has been transferred to whoever is least able to see it.

Closing that gap is not a matter of waiting for better models. It requires evaluation independent of the vendor, evidence standards that survive contact with procurement, authority to stop a deployment located with someone in the room, and a channel through which the affected person's judgment counts for something. None of that is technically hard. It is institutionally inconvenient, which is a different problem and a more tractable one.

Frequently asked questions

What is the AI accountability gap? The AI accountability gap is the distance between what an AI system is capable of doing and what any identifiable party is answerable for when it does that thing badly. It appears when organizations deploy AI faster than they build the ability to verify its output, receive complaints from affected people, and suspend the deployment when it fails.

Is AI transparency the same as AI accountability? No. Transparency rules, including Article 50 of the EU AI Act, generally establish provenance, meaning they tell people that content was generated or manipulated by AI. Accountability establishes who is responsible for the accuracy of that content and what the affected person is owed when it is wrong. A system can be fully labelled and still be unaccountable.

Who is legally responsible when an AI system causes harm? In most jurisdictions this remains unsettled. The European Commission withdrew its proposed AI Liability Directive in October 2025, and several United States state frameworks have narrowed toward notice and record-keeping obligations rather than substantive duties of care. In practice, responsibility is determined by contract terms, sector-specific law such as disability and civil rights statutes, and general liability doctrine rather than by AI-specific rules.

Why is sign language AI a useful test case for AI governance? Because it makes verification asymmetry visible. The institution deploying the system usually cannot evaluate the output, and the Deaf person who can evaluate it usually has no authority to challenge the institution's conclusion that access was provided. The same structure operates in medical, hiring, and benefits AI, but is harder to observe.

What should an organization require before deploying an AI system that affects people? Independent evaluation rather than vendor demonstration, performance data disaggregated by population and setting, disclosed training data provenance, a named person with authority to suspend the deployment, a functioning complaint channel for affected people, and contractual rights to monitor performance after launch.

What is the SLAT Index? The Sign Language Access Trust Index is Novara Consulting Group's evaluation instrument for sign language AI systems. It sets defined evidence standards across governance, linguistic, transparency, accessibility, and deployment domains, scores systems at the indicator level, and distinguishes vendor asserti-

on from independently confirmable fact so that procurement decisions rest on evidence rather than demonstration.

Heather M. Grizzle is the founder of Novara Consulting Group, which advises institutions on the governance and procurement of AI accessibility technology. She is the author of the SLAT Index Evaluation Standard.

Related reading: The Demo Is Not the Evidence: What Google DeepMind's Sign Language AI Must Prove Next

---

Quellen

- European Commission, Quick Facts: Transparency rules for AI systems (Article 50, applicable 2 August 2026). https://digital-strategy.ec.europa.eu/en/factpages/quick-facts-transparency-rules-ai-systems
- European Parliament Legislative Train Schedule, AI Liability Directive (withdrawal, October 2025). https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-ai-liability-directive
- U.S. Department of Justice, Extension of Compliance Dates for Nondiscrimination on the Basis of Disability: Accessibility of Web Information and Services of State and Local Government Entities, 20 April 2026. https://www.federalregister.gov/documents/2026/04/20/2026-07663/extension-of-compliance-dates-for-nondiscrimination-on-the-basis-of-disability-accessibility-of-web
- Office of Management and Budget, M-25-21, Accelerating Federal Use of AI through Innovation, Governance, and Public Trust (April 2025). https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf
- Office of Management and Budget, M-25-22, Driving Efficient Acquisition of Artificial Intelligence in Government (April 2025). https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf
- Colorado SB 26-189, repealing and reenacting the SB 24-205 framework, signed 14 May 2026, substantive obligations applying from 1 January 2027. https://leg.colorado.gov/bills/sb26-189

- 28 CFR 35.160, Communications (Title II, primary consideration). https://www.ecfr.gov/current/title-28/chapter-I/part-35/subpart-E/section-35.160
- 28 CFR 36.303, Auxiliary aids and services (Title III, consultation standard). https://www.ecfr.gov/current/title-28/chapter-I/part-36/subpart-C/section-36.303
- National Association of the Deaf, Minimum Standards for Video Remote Interpreting Services in Medical Settings. https://nad.org/minimum-standards-for-video-remote-interpreting-services-in-medical-settings/
- World Federation of the Deaf and World Association of Sign Language Interpreters, Statement on Use of Signing Avatars, 14 March 2018, updated 14 April 2018. https://wasli.org/wp-content/uploads/2023/07/WFD-and-WASLI-Statement-on-Avatar-FINAL-14032018-Updated-14042018-1.pdf

Zitierweise

Grizzle, H. M. (2026, August 17). AI Can Translate. But Who Translates the Risk?. Novara Consulting Group. https://www.novaracg.com/de/2026/08/17/ai-can-translate-but-who-translates-the-risk/