This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/08/17/ai-can-translate-but-who-translates-the-risk/.

L'IA peut traduire. Mais qui traduit le risque ?

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.

Sommaire

Chaque déploiement d'IA déplace le risque vers quelqu'un. Dans la plupart des cas, il retombe sur la personne la moins capable de détecter une erreur et la moins autorisée à la contester.

Heather M. Grizzle Novara Consulting Group 17 août 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 res-

ponse 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?

Cette question ne concerne pas uniquement la langue des signes. C'est le problème central non résolu de la gouvernance de l'IA à l'heure actuelle, et il porte un nom qu'il vaut la peine d'employer sans détour. Appelons-le le fossé de responsabilité.

Qu'est-ce que le fossé de responsabilité de l'IA ?

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.

Ce fossé n'est pas d'abord un problème technique. Les modèles échouent de manière prévisible, et les ingénieurs sont généralement francs sur ces modes de défaillance. Le fossé est un problème institutionnel. Les organisations acquièrent des systèmes dont elles ne peuvent pas évaluer les résultats, les déploient dans des contextes où les erreurs ont des conséquences, et ne conservent aucun mécanisme pratique pour savoir si le système fonctionne.

Capacité, fiabilité et responsabilité sont trois types de preuves différents

La plupart des discussions d'achat traitent ces trois éléments comme un seul. Ce n'est pas le cas, et les distinguer est la démarche la plus utile qu'un acheteur institutionnel puisse adopter.

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 deployment, 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.

Une organisation qui accepte la preuve de capacité comme si elle réglait les deux autres questions n'a pas pris de décision d'achat. Elle a fait une hypothèse et l'a payée.

Pourquoi les lois sur la divulgation ne comblent pas le fossé

La réponse réglementaire la plus visible face au risque de l'IA aujourd'hui est la transparence. En vertu de l'article 50 du règlement européen sur l'intelligence artificielle (EU AI Act), des obligations devenues applicables le 2 août 2026 exigent des fournisseurs qu'ils marquent le contenu synthétique sous forme lisible par machine et exigent des déployeurs qu'ils divulguent les hypertrucages (deepfakes) et certains textes générés par IA sur des sujets d'intérêt public, sous peine d'amendes pouvant atteindre 15 millions d'euros ou trois pour cent du chiffre d'affaires annuel mondial total. La Commission européenne a accordé une période de grâce jusqu'au 2 décembre 2026 pour les systèmes génératifs déjà mis sur le marché.

Il s'agit là d'un travail significatif que je soutiens. Il vaut aussi la peine d'être précis sur ce qu'il accomplit réellement. L'article 50 est un régime de provenance. Il indique à une personne qu'une machine est intervenue. Il ne lui dit pas si la machine avait raison.

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.

Le schéma se répète aux États-Unis au niveau des États. Le Colorado a adopté la première loi étatique complète sur l′IA en 2024, avant de l′abroger et de la remplacer par le SB 26-189, signé le 14 mai 2026, dont les obligations substantielles pour les développeurs et les déployeurs s′appliquent à partir du 1er janvier 2027. Le texte de remplacement a supprimé les programmes obligatoires de gestion des risques, les évaluations d′impact algorithmique, et le devoir autonome de diligence raisonnable contre la discrimination algorithmique. Ce qui a survécu, ce sont un préavis avant utilisation, une divulgation en langage clair dans les trente jours suivant une décision défavorable, une conservation des dossiers pendant trois ans, et un droit de demander un examen humain significatif.

Je ne prétends pas que ces obligations survivantes soient sans valeur. Le préavis et la conservation des dossiers sont réels. Mais le préavis vous indique qu′un système a été utilisé. Les dossiers vous indiquent ce qu′il a fait. Ni l′un ni l′autre n′établit que le résultat était exact, et ni l′un ni l′autre n′attribue la responsabilité des conséquences s′il ne l′était pas. La direction prise par la réglementation ces deux dernières années a consisté à informer les gens qu′une IA était intervenue, tout en s′éloignant de l′idée de tenir quiconque responsable de ce qu′elle a produit.

L′asymétrie de vérification : le problème structurel sous-jacent

Voici la partie que je considère sous-examinée, et c′est là que le cas de la langue des signes cesse d′être un exemple de niche pour devenir l′illustration la plus claire disponible d′une condition générale.

Dans la plupart des déploiements d′IA, la personne la mieux placée pour détecter une erreur n′a aucune autorité pour agir en conséquence, et la partie ayant l′autorité d′agir n′a aucun moyen de la détecter. J′appellerais cela l′asymétrie de vérification, et c′est le mécanisme qui produit le fossé de responsabilité.

Considérez ce qu'une institution entendante voit lorsqu'elle déploie un système automatisé de langue des signes. Elle voit que le système a produit un résultat. Le personnel observe un avatar qui signe ou des sous-titres qui apparaissent. De ce point de vue, la transaction paraît complète.

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

Ce n'est pas une plainte technologique. C'est une répartition du pouvoir. L'institution détient l'autorité de déterminer si l'accès a eu lieu et manque de la capacité linguistique pour l'évaluer. La personne Sourde détient la capacité linguistique et manque de l'autorité. L'erreur, s'il y en a une, n'a nulle part où émerger.

Une fois qu'on en perçoit la forme, on la retrouve presque partout. Un patient ne peut pas évaluer une recommandation de triage assistée par l'IA. Un candidat ne peut pas inspecter le modèle qui l'a écarté. Un demandeur de prestations ne peut pas vérifier la détermination d'admissibilité. Une personne recevant des instructions juridiques traduites par machine, dans n'importe quelle langue, ne peut pas vérifier la traduction sans la maîtrise linguistique que la traduction est censée lui fournir. Dans chaque cas, la partie concernée absorbe un risque qu'elle ne peut pas mesurer, et l'institution qui déploie le système enregistre une transaction achevée.

Ce qui rend la langue des signes un cas particulièrement aigu, ce sont les enjeux attachés au contenu. Le langage dans les contextes institutionnels porte consentement, témoignage, diagnostic, instruction et effet juridique. Une erreur de traduction n'est pas une expérience utilisateur dégradée. C'est un formulaire de consentement défectueux, un antécédent médical inexact, ou une déclaration mal représentée dans une procédure.

Le droit de l'accessibilité a déjà répondu en partie à cette question, et nous sommes en train de l'oublier

Il existe déjà une réponse viable dans la pratique de l'accessibilité, ce qui rend son abandon actuel d'autant plus difficile à justifier.

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.

À lire cela comme une conception de gouvernance plutôt que comme une règle d'accessibilité, on constate qu'elle est exceptionnellement bien construite. Elle situe le jugement de vérification chez la personne qui peut réellement l'exercer. Elle donne à un professionnel présent sur place l'autorité d'arrêter un déploiement en cours d'utilisation. Elle traite l'efficacité comme quelque chose à déterminer en contexte plutôt qu'à supposer du seul fait de la présence de l'équipement.

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.

La Fédération mondiale des sourds et l'Association mondiale des interprètes en langue des signes ont fixé une limite concernant les avatars signants dès 2018. Leur position était que les avatars sont inappropriés pour les informations en direct, complexes ou importantes, et que le contenu statique préenregistré peut être acceptable lorsque des personnes Sourdes ont conseillé sur le matériel et qu'aucune interaction en direct n'est requise. Cette déclaration a huit ans. La technologie a considérablement progressé depuis. La limite de déploiement qu'elle a tracée n'a été remplacée par rien de plus rigoureux, et elle demeure la ligne la plus claire disponible.

Pourquoi la démonstration du fournisseur ne peut pas tenir lieu de dossier de gouvernance

La démonstration est conçue pour être persuasive. C'est sa fonction, et il n'y a rien d'inapproprié à cela. Le problème, c'est qu'en l'absence d'évaluation indépendante, la démonstration devient par défaut la base probante d'une décision d'achat, car c'est le seul document présent dans la pièce.

Une démonstration établit une possibilité dans des conditions choisies. Elle ne peut établir ni la performance sur l'ensemble de la population que l'acheteur sert, ni le comportement en dehors des conditions contrôlées, ni la sécurité dans des contextes à conséquences élevées, ni la responsabilité en cas d'erreurs. Ce sont là des propriétés d'un déploiement, pas des propriétés d'un modèle.

La politique fédérale reconnaît déjà cela en principe. Les directives de l'OMB (Office of Management and Budget) publiées en avril 2025 définissent l'IA à fort impact comme les systèmes dont le résultat sert de base principale à des décisions ou des actions ayant un effet juridique, matériel, contraignant ou significatif sur les droits ou la sécurité, la note énumérant des domaines tels que les droits civils, l'accès à l'éducation, le logement, l'emploi, les prestations et services gouvernementaux, et la santé. Pour ces usages, elle exige des tests avant déploiement, des évaluations d'impact de l'IA avant le déploiement et périodiquement par la suite, une supervision humaine adéquate, et un examen humain rapide avec une possibilité d'appel pour les personnes concernées. Les agences incapables de satisfaire à ces minimums doivent cesser l'utilisation en toute sécurité. La note d'acquisition qui l'accompagne exige des clauses contractuelles donnant à l'agence la capacité de surveiller et d'évaluer régulièrement la performance, les risques et l'efficacité du système.

Quoi que l'on pense des détails, la structure de cette directive est correcte. Elle traite la responsabilité comme quelque chose de spécifié dans le contrat, testé avant le lancement, surveillé après le lancement, et applicable par suspension. C'est une couche de vérification. La plupart des institutions qui achètent de l'IA en dehors du système fédéral d'acquisition n'ont rien d'équivalent.

J'ajouterais une mise en garde à propos de l'hypothèse selon laquelle la capacité arrivera dans les délais et comblera le fossé d'elle-même. En avril 2026, le Département de la Justice a repoussé les échéances de conformité de la règle d'accessibilité web du Titre II de l'ADA (Americans with Disabilities Act) d'environ un an pour chaque caté-

gorie d'entité, faisant passer les grandes entités publiques au 26 avril 2027 et les plus petites au 26 avril 2028. Parmi les raisons invoquées par le Département figurait le fait qu'il avait surestimé les capacités, tant en personnel qu'en technologie, des entités concernées, et il a déclaré clairement que les technologies avancées telles que l'IA générative n'automatisent pas encore de manière fiable la remédiation du contenu inaccessible à grande échelle. Voilà un régulateur fédéral qui constate que les affirmations sur les capacités de l'IA ont devancé leur mise en œuvre effective dans un contexte d'accessibilité. Cela vaut la peine d'y réfléchir avant de traiter le calendrier d'un fournisseur comme un plan de gouvernance.

Ce que les institutions devraient exiger avant le déploiement

Les questions ci-dessous n'ont rien d'exotique. Ce sont les questions qu'un acheteur compétent poserait à propos de tout système porteur de conséquences, et elles appellent des réponses.

Qui a évalué ce système, et cette évaluation était-elle indépendante du fournisseur et de l'acheteur ?

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 assertion 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

---

Sources

- 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

Comment citer

Grizzle, H. M. (2026, August 17). L'IA peut traduire. Mais qui traduit le risque ?. Novara Consulting Group. https://www.novaracg.com/fr/2026/08/17/ai-can-translate-but-who-translates-the-risk/