Ce document a été traduit de l'anglais par une machine et peut contenir des erreurs. L'original en anglais se trouve à l'adresse https://www.novaracg.com/2026/08/04/the-evidence-that-wasnt-and-still-isnt-there/.

La preuve qui n'était pas là et qui n'y est toujours pas

Heather M. Grizzle · 4 août 2026

Une enquête sur les preuves indépendantes manquantes derrière les plus grandes affirmations sur l'IA en langue des signes de 2026, de SignGemma aux outils de traduction commerciaux.

LE CINQUIÈME PARAMÈTRE

Novara Consulting Group · Novara Press · Journal of Governance and Evidence in Accessibility AI

The Fifth Parameter, Vol. 1 · Numéro 13

AUTEURE Heather M. Grizzle

ORCID 0009-0005-2605-8256

DATE DES PREUVES Actualisé en août 2026

MOTS-CLÉS IA en langue des signes ; évaluation indépendante ; classes de preuves ; SignGemma ; transparence ; confiance ; accès des personnes Deaf

---

Le fait le plus important concernant l'IA et la langue des signes en 2026 n'est pas ce qui s'est passé.

C'est ce qui ne s'est pas produit.

Entre janvier et août, des entreprises ont annoncé des produits, présenté des systèmes, levé des fonds et évoqué l'avenir de la communication. De nouveaux outils ont été présentés comme plus rapides, plus performants et plus proches d'un usage quotidien. Certains ont suscité des titres enthousiastes. D'autres sont arrivés avec des affirmations impressionnantes concernant les données d'entraînement, les temps de réponse ou le nombre de signes qu'ils pouvaient reconnaître.

Ce qui n'est pas apparu, c'est une preuve indépendante que ces systèmes fonctionnaient comme annoncé. Aucun tiers n'a publié de mesure publique de la précision, des taux d'erreur ou de la compréhension des utilisateurs pour l'un des principaux

systèmes examinés par l'Index. Les recherches menées dans les sources requises n'ont trouvé aucun référentiel indépendant pour Signapse, Sorenson AST, Rylo Sign, SignSpeak, ChatSign, SignVrse, Talksign ou SignGemma de Google.

La formulation officielle de la conclusion est prudente : aucune preuve de Classe A n'a été « localisée en date d'août 2026 ». Cela laisse ouverte la possibilité que certaines preuves existent quelque part en dehors des archives publiques. Elles pourraient se trouver au sein d'une entreprise, être détenues par un client ou figurer dans des recherches qui n'ont pas encore été publiées.

Mais une preuve qui ne peut être examinée ne peut étayer une affirmation publique. Cette distinction importe car l'industrie n'est pas restée silencieuse. Elle a produit un flux constant d'annonces, de partenariats, de démonstrations et de récits de financement. L'impression donnée est celle d'un progrès rapide. Pourtant, les preuves publiques nécessaires pour juger de ce progrès n'ont pas suivi le rythme.

Un domaine peut être actif sans devenir responsable. Il peut générer de l'actualité sans générer de savoir.

Trois types de preuves

L'Index répartit les preuves en trois grandes classes.

- La Classe A est une preuve indépendante. Elle provient d'un tiers et comprend des mesures qui peuvent être examinées : résultats de précision, taux d'erreur, études de compréhension, comparaisons avec d'autres systèmes ou performance sur un référentiel publié.
- La Classe B est une preuve produite par l'entreprise ou l'organisation responsable du système. Elle peut contenir des détails techniques utiles, mais l'organisation qui formule l'affirmation est aussi celle qui est évaluée.
- La Classe C se compose d'assertions : déclarations dans des annonces, des supports marketing, des entretiens, des présentations ou une couverture médiatique qui ne peuvent être vérifiées de manière indépendante.

Cela ne signifie pas que chaque affirmation de Classe B ou de Classe C soit fausse. Une entreprise peut décrire son produit honnêtement. Une équipe d'ingénierie peut rapporter un résultat réel. Une démonstration peut montrer un véritable progrès technique. Le problème est qu'aucune de ces choses ne répond à la question centrale :

comment le système se comporte-t-il lorsqu'une personne n'ayant aucun intérêt à sa réussite le mesure ?

En 2026, presque tous les jugements publics sur l'IA et la langue des signes reposaient encore sur des documents de Classe B ou de Classe C. Les entreprises fournissaient les affirmations, le langage et, souvent, les mesures grâce auxquelles ces affirmations étaient comprises. L'évaluation indépendante était largement absente.

Le problème SignGemma

SignGemma de Google offre l'exemple le plus clair. Lorsqu'il a été annoncé, le modèle a attiré l'attention car il semblait placer les ressources d'une grande entreprise technologique derrière l'IA de la langue des signes. Les affirmations associées incluaient un entraînement sur 10 000 heures de données et une latence inférieure à 200 millisecondes. Ces chiffres ont largement circulé car ils suggéraient à la fois échelle et rapidité.

Quinze mois plus tard, le modèle n'avait toujours pas été officiellement publié. Le personnel de Google l'a confirmé sur le forum des développeurs IA de l'entreprise. La discussion s'est poursuivie jusqu'en juin 2026. Des chercheurs du Pakistan, d'Égypte et d'ailleurs ont demandé un accès. Ils ne l'ont pas obtenu.

Cela a créé un objet public inhabituel : un modèle qui pouvait être discuté mais non testé, cité mais non inspecté, anticipé mais non utilisé. Sans accès au modèle, les personnes extérieures ne pouvaient pas reproduire ses résultats rapportés. Elles ne pouvaient pas examiner comment ses données d'entraînement avaient été assemblées, tester sa performance selon différentes langues des signes ou découvrir comment il se comportait en dehors d'exemples contrôlés. Les affirmations concernant l'échelle d'entraînement et la latence sont donc restées de Classe C.

La réputation de Google ne change pas leur statut. L'autorité peut rendre une affirmation plus persuasive, mais elle ne la rend pas indépendante. La répétition ne transforme pas une assertion en mesure. Il ne s'agit pas de dire que SignGemma a nécessairement échoué. Il s'agit de dire que le public n'a reçu aucun moyen fiable de savoir s'il avait réussi.

L'exactitude n'est pas un chiffre unique

Les tests indépendants sont particulièrement importants dans la technologie de la langue des signes car le mot « exactitude » peut dissimuler autant qu'il révèle. Un sys-

tème peut bien performer sur un vocabulaire limité enregistré sous un éclairage stable mais avoir du mal avec une conversation naturelle. Il peut reconnaître les signes de personnes dont l'apparence ressemble à celles de ses données d'entraînement tout en performant mal pour d'autres. Il peut identifier correctement des signes individuels tout en perdant le sens à travers une phrase. Il peut fonctionner pour une langue des signes et être promu comme s'il fonctionnait pour la langue des signes en général.

Les langues des signes ne sont pas des versions codées des langues parlées. Elles ont leur propre grammaire, leurs variations régionales et leur contexte culturel. Le sens peut dépendre du mouvement, de l'emplacement, du minutage, de l'expression faciale et de la relation entre les signes. Un système qui reconnaît les mains mais manque le visage peut produire des mots tout en perdant la phrase. Même un taux d'exactitude global élevé nous en dirait peu sans savoir ce qui a été testé, qui y a participé et ce qui comptait comme un résultat correct.

Le système a-t-il été testé sur des signes isolés ou une conversation continue ? Les participants étaient-ils familiers du modèle ? Des signeurs Sourds ont-ils jugé si la sortie avait du sens ? Les erreurs ont-elles été simplement comptées, ou examinées pour un préjudice possible ? Le test incluait-il différents âges, teints de peau, styles de signes et variétés régionales ? Ce ne sont pas des détails secondaires. Ils déterminent ce que signifie le chiffre.

C'est pourquoi les études de compréhension importent autant que les référentiels techniques. Un système peut obtenir un score impressionnant tout en produisant une sortie que les utilisateurs trouvent confuse, non naturelle ou trompeuse. Le test ultime n'est pas de savoir si un modèle détecte le mouvement. C'est de savoir si les gens peuvent comprendre et se fier à ce qu'il communique.

Le coût de l'anticipation

La promesse répétée qu’un système efficace est sur le point d’arriver n’est pas sans conséquences. Un modèle non publié peut occuper un espace que des services fonctionnels pourraient autrement combler. Il peut influencer les décisions de financement, orienter les discussions politiques et encourager les institutions à reporter leurs investissements dans l’offre humaine. Il peut faire apparaître les insuffisances actuelles comme temporaires, même en l’absence de calendrier ou de preuve publique venant étayer cette croyance.

C’est la politique de l’anticipation. L’aide arrive toujours, si bien que son absence dans le présent est traitée comme moins urgente. Pour les usagers Deaf, le coût est concret. Un hôpital, une université, un employeur ou une autorité publique peuvent invoquer une technologie émergente comme preuve d’une accessibilité en progrès. Mais une annonce prometteuse ne peut pas interpréter une consultation médicale. Une démonstration ne peut pas garantir l’accès à une salle de classe. Un modèle inaccessible aux chercheurs est aussi inaccessible aux personnes dont la vie sert à justifier son importance.

Il existe un autre coût. Lorsque l’attention publique se concentre sur des systèmes n’ayant fait l’objet d’aucun test indépendant, des organisations plus modestes, offrant des formes de soutien moins spectaculaires mais plus fiables, peuvent disparaître du champ de vision. Les interprètes humains, les services menés par la communauté et les pratiques d’accessibilité établies suscitent rarement le même enthousiasme qu’un nouveau modèle d’IA. On les juge comme des dépenses présentes, tandis que la technologie est valorisée comme une possibilité future. La comparaison est inégale. L’offre existante doit répondre de ses limites dès maintenant. Le système promis, lui, a le droit de rester parfait puisqu’il n’est pas encore arrivé.

Ce qu'exigerait la transparence

L’absence de preuve indépendante n’est pas difficile à corriger en principe. Les développeurs pourraient donner accès à des chercheurs qualifiés avant de formuler des affirmations générales sur les performances. Les évaluations pourraient être conçues avec des signeurs Deaf plutôt que simplement menées sur eux. Les jeux de test, les méthodes et les définitions pourraient être publiés. Les résultats pourraient être ventilés par langue, par contexte et par groupe d’utilisateurs plutôt que réduits à un seul chiffre phare.

Les chercheurs devraient également pouvoir signaler des échecs sans dépendre de l’autorisation de l’entreprise évaluée. Rien de tout cela n’éliminerait l’incertitude. Des études indépendantes peuvent être mal conçues. Des référentiels d’évaluation peuvent récompenser des formes étroites de performance. Les résultats publiés peuvent devenir obsolètes à mesure que les systèmes évoluent, et pourtant un examen imparfait vaut mieux qu’une visibilité maîtrisée. Un référentiel public donne aux autres matière à interroger, reproduire ou améliorer. Une affirmation non étayée ne leur donne qu’une matière à faire circuler.

La transparence signifierait aussi dire clairement ce qu'un système ne peut pas faire. Un produit conçu pour un cadre restreint ne devrait pas être présenté comme une solution générale. Un modèle testé sur une seule langue des signes ne devrait pas être autorisé à emprunter l'universalité apparente du mot « signe ». Un prototype devrait être identifié comme tel. Un modèle non publié ne devrait pas être traité comme une infrastructure publique.

Le cinquième paramètre

Les systèmes techniques sont souvent jugés selon des mesures familières : vitesse, échelle, coût et précision. La mesure manquante est la confiance. La confiance n'est pas une simple affirmation de performance supplémentaire. Elle résulte du fait de rendre les affirmations vérifiables. Elle grandit lorsque des chercheurs extérieurs peuvent inspecter un système, lorsque les utilisateurs peuvent décrire ses défaillances et lorsque les preuves restent disponibles une fois le cycle des annonces passé. C'est ***le cinquième paramètre*.**

Selon cette mesure, le domaine de l'IA appliquée à la langue des signes est entré en août 2026 avec un déficit sérieux. Il disposait de produits, de promesses et de publicité. Ce qui lui manquait, c'était des preuves indépendantes accessibles au public concernant les systèmes qui suscitaient le plus d'attention.

Cette absence ne doit pas être confondue avec un vide administratif temporaire. Elle fait partie de l'état actuel de la technologie. Tant que des évaluateurs indépendants ne pourront pas mesurer ces systèmes, la réponse honnête à de nombreuses questions sur leur performance n'est pas qu'ils fonctionnent, ni qu'ils échouent.

C'est que le public n'a pas reçu suffisamment de preuves pour savoir.

Déclaration d'intérêts L'auteure est fondatrice de Novara Consulting Group LLC, qui conçoit et concède sous licence le Sign Language Access Trust Index mentionné dans cet article, et dirige The Fifth Parameter.

Financement Aucun. Produit en interne par Novara Consulting Group.

Données et matériaux L'article s'appuie sur les archives publiques concernant les systèmes cités et sur le fil du Google AI Developers Forum consacré à SignGemma.

Divulgation de l'usage de l'IA Le texte de cet article est celui de l'auteure. L'assistance de l'IA n'a été utilisée que pour la mise en page dans le modèle de marque ; aucun argument, figure, citation ou référence n'a été généré par le modèle.

Citation suggérée Grizzle, H. M. (2026). La preuve qui n'était pas là. The Fifth Parameter. Vol 1 Numéro 13. Novara Press.

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.

Comment citer

Grizzle, H. M. (2026, August 4). La preuve qui n'était pas là et qui n'y est toujours pas. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003816