This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/08/04/the-evidence-that-wasnt-and-still-isnt-there/.

Las pruebas que no existían y que aún no existen

Heather M. Grizzle · August 4, 2026

Una investigación sobre la ausencia de evidencia independiente detrás de las mayores afirmaciones sobre la IA aplicada a la lengua de signos en 2026, desde SignGemma hasta las herramientas comerciales de traducción.

EL QUINTO PARÁMETRO

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

The Fifth Parameter, Vol. 1 · Número 13

AUTORA Heather M. Grizzle

ORCID 0009-0005-2605-8256

FECHA DE LA EVIDENCIA Vigente a agosto de 2026

PALABRAS CLAVE IA de lengua de señas; evaluación independiente; clases de evidencia; SignGemma; transparencia; confianza; acceso para personas Deaf

---

El hecho más importante sobre la IA y la lengua de signos en 2026 no es lo que ocurrió.

Es lo que no ocurrió.

Entre enero y agosto, las empresas anunciaron productos, hicieron demostraciones de sistemas, recaudaron fondos y hablaron sobre el futuro de la comunicación. Las nuevas herramientas se presentaron como más rápidas, más capaces y más cercanas al uso cotidiano. Algunas atrajeron titulares entusiastas. Otras llegaron con afirmaciones impresionantes sobre datos de entrenamiento, tiempos de respuesta o la cantidad de signos que podían reconocer.

Lo que no apareció fue evidencia independiente de que estos sistemas funcionaran como se anunciaba. Ningún tercero publicó una medición pública de la precisión, las

tasas de error o la comprensión por parte de los usuarios de ninguno de los principales sistemas examinados por el Índice. Las búsquedas en las fuentes requeridas no encontraron ningún punto de referencia independiente para Signapse, Sorenson AST, Rylo Sign, Sign-Speak, ChatSign, SignVrse, Talksign o SignGemma de Google.

La redacción formal del hallazgo es cuidadosa: no se "localizó" evidencia de Clase A "a partir de agosto de 2026". Esto deja abierta la posibilidad de que exista alguna evidencia en algún lugar fuera del registro público. Puede estar dentro de una empresa, en manos de un cliente o contenida en investigaciones que aún no se han publicado.

Pero la evidencia que no puede examinarse no puede sustentar una afirmación pública. Esta distinción importa porque el sector no ha estado callado. Ha producido un flujo constante de anuncios, alianzas, demostraciones e historias de financiación. La impresión es de un progreso rápido. Sin embargo, la evidencia pública necesaria para juzgar ese progreso no ha seguido el mismo ritmo.

Un campo puede estar en plena actividad sin volverse responsable. Puede generar noticias sin generar conocimiento.

Tres tipos de evidencia

El Índice separa la evidencia en tres clases amplias.

- La Clase A es evidencia independiente. Proviene de un tercero e incluye mediciones que pueden examinarse: resultados de precisión, tasas de error, estudios de comprensión, comparaciones con otros sistemas o rendimiento en un punto de referencia publicado.
- La Clase B es evidencia producida por la empresa u organización responsable del sistema. Puede contener detalles técnicos útiles, pero la organización que hace la afirmación es también la organización que está siendo evaluada.
- La Clase C consiste en afirmaciones: declaraciones en anuncios, material de marketing, entrevistas, presentaciones o cobertura de prensa que no pueden verificarse de forma independiente.

Esto no significa que toda afirmación de Clase B o Clase C sea falsa. Una empresa puede describir su producto con honestidad. Un equipo de ingeniería puede informar de un resultado real. Una demostración puede mostrar un progreso técnico genuino. El problema es que ninguna de estas cosas responde a la pregunta central: ¿cómo funciona el sistema cuando alguien sin interés en su éxito lo mide?

En 2026, casi todo juicio público sobre la IA y la lengua de signos seguía dependiendo de material de Clase B o Clase C. Las empresas proporcionaban las afirmaciones, el lenguaje y, a menudo, las medidas mediante las cuales esas afirmaciones se entendían. La evaluación independiente estaba, en gran medida, ausente.

El problema de SignGemma

SignGemma, de Google, ofrece el ejemplo más claro. Cuando se anunció, el modelo atrajo la atención porque parecía poner los recursos de una gran empresa tecnológica al servicio de la IA para la lengua de signos. Las afirmaciones asociadas incluían entrenamiento con 10.000 horas de datos y una latencia inferior a 200 milisegundos. Estas cifras se difundieron ampliamente porque sugerían tanto escala como velocidad.

Quince meses después, el modelo aún no se había publicado oficialmente. El personal de Google lo confirmó en el Foro de Desarrolladores de IA de la empresa. El debate continuó hasta junio de 2026. Investigadores de Pakistán, Egipto y otros lugares solicitaron acceso. No lo obtuvieron.

Esto creó un objeto público inusual: un modelo del que se podía hablar pero no probar, que se podía citar pero no examinar, que se podía anticipar pero no usar. Sin acceso al modelo, los observadores externos no podían reproducir sus resultados declarados. No podían examinar cómo se había recopilado sus datos de entrenamiento, probar su rendimiento en distintas lenguas de signos ni descubrir cómo se comportaba fuera de ejemplos controlados. Las afirmaciones sobre la escala de entrenamiento y la latencia permanecieron, por tanto, en Categoría C.

La reputación de Google no altera su estatus. La autoridad puede hacer que una afirmación sea más persuasiva, pero no la hace independiente. La repetición no convierte una afirmación en una medición. La cuestión no es que SignGemma tuviera que haber fracasado. La cuestión es que al público no se le dio ninguna manera fiable de saber si había tenido éxito.

La precisión no es un solo número

Las pruebas independientes son especialmente importantes en la tecnología de lengua de signos porque la palabra «precisión» puede ocultar tanto como revela. Un sistema puede funcionar bien con un vocabulario limitado grabado en condiciones de iluminación estables, pero tener dificultades con una conversación natural. Puede reconocer signos de personas cuya apariencia se asemeja a la de los datos de entrenamiento

mientras funciona mal con otras. Puede identificar signos individuales correctamente pero perder el significado a lo largo de una frase. Puede funcionar para una lengua de signos y promocionarse como si funcionara para la lengua de signos en general.

Las lenguas de signos no son versiones codificadas de las lenguas habladas. Tienen su propia gramática, variación regional y contexto cultural. El significado puede depender del movimiento, la ubicación, el tiempo, la expresión facial y la relación entre los signos. Un sistema que reconoce las manos pero pasa por alto el rostro puede producir palabras mientras pierde la frase. Incluso una alta tasa de precisión general nos diría poco a menos que supiéramos qué se había probado, quién había participado y qué se consideraba un resultado correcto.

¿Se probó el sistema con signos aislados o con conversación continua? ¿Eran los participantes familiares para el modelo? ¿Juzgaron personas Sordas si el resultado tenía sentido? ¿Se limitaron a contar los errores, o se examinaron por su posible daño? ¿Incluyó la prueba distintas edades, tonos de piel, estilos de signado y variedades regionales? Estos no son detalles secundarios. Determinan qué significa el número.

Por eso los estudios de comprensión importan junto a los parámetros técnicos. Un sistema puede lograr una puntuación impresionante y, aun así, producir un resultado que los usuarios encuentran confuso, poco natural o engañoso. La prueba definitiva no es si un modelo detecta el movimiento. Es si las personas pueden entender y confiar en lo que comunica.

El costo de la anticipación

La promesa repetida de que un sistema eficaz está a punto de llegar tiene consecuencias. Un modelo no publicado puede ocupar el espacio que los servicios en funcionamiento podrían llenar. Puede condicionar las decisiones de financiación, influir en los debates políticos y animar a las instituciones a posponer la inversión en provisión humana. Puede hacer que las carencias actuales parezcan temporales, incluso cuando ningún calendario ni evidencia pública respalda esa creencia.

Esta es la política de la anticipación. La ayuda siempre está por llegar, así que la ausencia de ayuda en el presente se trata como algo menos urgente. Para los usuarios Sordos, el coste es práctico. Un hospital, una universidad, un empleador o una autoridad pública pueden señalar una tecnología emergente como prueba de que la accesibilidad está mejorando. Pero un anuncio prometedor no puede interpretar una consulta médica. Una demostración no puede garantizar el acceso a un aula. Un modelo no dis-

ponible para los investigadores tampoco está disponible para las personas cuyas vidas se utilizan para justificar su importancia.

Hay otro coste. Cuando la atención pública se concentra en sistemas que no han sido evaluados de forma independiente, las organizaciones más pequeñas que ofrecen formas de apoyo menos espectaculares pero más fiables pueden desaparecer de la vista. Los intérpretes humanos, los servicios dirigidos por la comunidad y las prácticas de accesibilidad consolidadas rara vez reciben el mismo entusiasmo que un nuevo modelo de IA. Se los juzga como gastos presentes, mientras que la tecnología se valora como posibilidad futura. La comparación es desigual. La provisión existente debe responder por sus limitaciones ahora. Al sistema prometido se le permite seguir siendo perfecto porque aún no ha llegado.

Lo que exigiría la transparencia

La ausencia de evidencia independiente no es, en principio, difícil de corregir. Los desarrolladores podrían dar acceso a investigadores cualificados antes de hacer amplias afirmaciones sobre el rendimiento. Las evaluaciones podrían diseñarse junto con personas Sordas signantes en lugar de simplemente realizarse sobre ellas. Los conjuntos de prueba, los métodos y las definiciones podrían publicarse. Los resultados podrían desglosarse por lengua, contexto y grupo de usuarios en lugar de comprimirse en una única cifra destacada.

Los investigadores también deberían poder informar de los fallos sin depender del permiso de la empresa evaluada. Nada de esto eliminaría la incertidumbre. Los estudios independientes pueden estar mal diseñados. Los parámetros de referencia pueden premiar formas estrechas de rendimiento. Los resultados publicados pueden quedar obsoletos a medida que los sistemas cambian; aun así, un escrutinio imperfecto es mejor que una visibilidad gestionada. Un parámetro de referencia público da a otros algo que cuestionar, repetir o mejorar. Una afirmación no respaldada solo les da algo que difundir.

La transparencia también significaría ser claro sobre lo que un sistema no puede hacer. Un producto diseñado para un entorno restringido no debería describirse como una solución general. Un modelo probado en una lengua de signos no debería poder apropiarse de la aparente universalidad de la palabra «signo». Un prototipo debería identificarse como prototipo. Un modelo no publicado no debería tratarse como infraestructura pública.

El quinto parámetro

Los sistemas técnicos suelen juzgarse por medidas familiares: velocidad, escala, costo y precisión. La medida que falta es la confianza. La confianza no es otra afirmación de rendimiento. Es el resultado de hacer que las afirmaciones sean verificables. Crece cuando investigadores externos pueden inspeccionar un sistema, cuando los usuarios pueden describir sus fallos y cuando la evidencia sigue disponible después de que el ciclo de anuncios ha quedado atrás. Este es ***el quinto parámetro*.**

Según esa medida, el campo de la IA para la lengua de signos entró en agosto de 2026 con un déficit grave. Tenía productos, promesas y publicidad. Lo que no tenía era evidencia independiente disponible públicamente para los sistemas que atraían la mayor atención.

Esa ausencia no debe confundirse con una laguna temporal en la documentación. Es parte de la condición actual de la tecnología. Hasta que los evaluadores independientes puedan medir estos sistemas, la respuesta honesta a muchas preguntas sobre su rendimiento no es que funcionen, ni que fallen.

Es que al público no se le ha dado suficiente evidencia para saber.

Declaración de intereses La autora es fundadora de Novara Consulting Group LLC, que desarrolla y licencia el Sign Language Access Trust Index referenciado en este artículo, y dirige The Fifth Parameter.

Financiación Ninguna. Producido internamente por Novara Consulting Group.

Datos y materiales El artículo se basa en el registro público de los sistemas mencionados y en el hilo del Google AI Developers Forum relativo a SignGemma.

Divulgación sobre el uso de IA El texto de este artículo es autoría propia de la autora. La asistencia de IA se utilizó únicamente para la maquetación dentro de la plantilla de marca; ningún argumento, cifra, cita o referencia fue generado por el modelo.

Cita sugerida Grizzle, H. M. (2026). La evidencia que no estaba. The Fifth Parameter. Vol. 1, número 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.

Cómo citar

Grizzle, H. M. (2026, August 4). The Evidence That Wasn't And Still Isn't There. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003816