NOVARA
CONSULTING GROUP

Questo documento è stato tradotto dall'inglese tramite traduzione automatica e potrebbe contenere errori. L'originale in inglese si trova all'indirizzo https://www.novaracg.com/2026/08/04/the-evidence-that-wasnt-and-still-isnt-there/.

Le prove che non c'erano e che ancora non ci sono

Heather M. Grizzle · 4 agosto 2026

Un'indagine sulle prove indipendenti mancanti dietro le più grandi affermazioni sull'IA per la lingua dei segni del 2026, da SignGemma agli strumenti commerciali di traduzione.

IL QUINTO PARAMETRO

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

The Fifth Parameter, Vol. 1 · Numero 13

AUTRICE Heather M. Grizzle

ORCID 0009-0005-2605-8256

DATA DELLE PROVE Aggiornato ad agosto 2026

PAROLE CHIAVE intelligenza artificiale per la lingua dei segni; valutazione indipendente; classi di evidenza; SignGemma; trasparenza; fiducia; accesso per i Deaf

---

Il fatto più importante riguardo all'IA e alla lingua dei segni nel 2026 non è ciò che è accaduto.

È ciò che non è accaduto.

Tra gennaio e agosto, le aziende hanno annunciato prodotti, dimostrato sistemi, raccolto capitali e parlato del futuro della comunicazione. Nuovi strumenti sono stati presentati come più veloci, più capaci e più vicini all'uso quotidiano. Alcuni hanno attirato titoli entusiastici. Altri sono arrivati con affermazioni impressionanti su dati di addestramento, tempi di risposta o numero di segni riconoscibili.

Ciò che non è apparso è stata una prova indipendente che questi sistemi funzionassero come pubblicizzato. Nessuna terza parte ha pubblicato una misurazione pubblica dell'accuratezza, dei tassi di errore o della comprensione da parte degli utenti per nes-

suno dei principali sistemi esaminati dall'Index. Le ricerche nelle fonti richieste non hanno trovato alcun parametro di riferimento indipendente per Signapse, Sorenson AST, Rylo Sign, Sign-Speak, ChatSign, SignVrse, Talksign o SignGemma di Google.

La formulazione ufficiale del risultato è cauta: nessuna prova di Classe A è stata «reperita al mese di agosto 2026». Ciò lascia aperta la possibilità che alcune prove esistano da qualche parte al di fuori dei registri pubblici. Potrebbero trovarsi all'interno di un'azienda, in possesso di un cliente, o contenute in ricerche non ancora pubblicate.

Ma una prova che non può essere esaminata non può sostenere un'affermazione pubblica. Questa distinzione è importante perché il settore non è rimasto in silenzio. Ha prodotto un flusso costante di annunci, partnership, dimostrazioni e storie di finanziamento. L'impressione è quella di un progresso rapido. Eppure le prove pubbliche necessarie per valutare quel progresso non hanno tenuto il passo.

Un settore può essere affaccendato senza diventare responsabile. Può generare notizie senza generare conoscenza.

Tre tipi di evidenza

L'Index suddivide le prove in tre grandi classi.

- La Classe A è la prova indipendente. Proviene da una terza parte e include misurazioni che possono essere esaminate: risultati di accuratezza, tassi di errore, studi sulla comprensione, confronti con altri sistemi o prestazioni su un parametro di riferimento pubblicato.
- La Classe B è la prova prodotta dall'azienda o dall'organizzazione responsabile del sistema. Può contenere dettagli tecnici utili, ma l'organizzazione che avanza l'affermazione è anche l'organizzazione oggetto della valutazione.
- La Classe C consiste in asserzioni: dichiarazioni contenute in annunci, materiale di marketing, interviste, presentazioni o copertura stampa che non possono essere verificate in modo indipendente.

Questo non significa che ogni affermazione di Classe B o Classe C sia falsa. Un'azienda può descrivere onestamente il proprio prodotto. Un team di ingegneri può riportare un risultato reale. Una dimostrazione può mostrare un genuino progresso tecnico. Il problema è che nessuna di queste cose risponde alla domanda centrale: come si comporta il sistema quando qualcuno privo di interesse nel suo successo lo misura?

Nel 2026, quasi ogni giudizio pubblico sull'IA e la lingua dei segni dipendeva ancora da materiale di Classe B o Classe C. Le aziende fornivano le affermazioni, il linguaggio e, spesso, i criteri con cui quelle affermazioni venivano interpretate. La valutazione indipendente era in gran parte assente.

Il problema di SignGemma

SignGemma di Google offre l'esempio più chiaro. Quando fu annunciato, il modello attirò attenzione perché sembrava mettere le risorse di una grande azienda tecnologica al servizio dell'IA per la lingua dei segni. Le affermazioni ad esso associate includevano un addestramento su 10.000 ore di dati e una latenza inferiore a 200 millisecondi. Queste cifre si diffusero ampiamente perché suggerivano sia scala che velocità.

Quindici mesi dopo, il modello non era ancora stato rilasciato ufficialmente. Il personale di Google lo ha confermato sul forum aziendale AI Developers Forum. La discussione è proseguita fino a giugno 2026. Ricercatori provenienti da Pakistan, Egitto e altrove hanno chiesto di potervi accedere. Non lo hanno ottenuto.

Questo ha creato un oggetto pubblico insolito: un modello che poteva essere discusso ma non testato, citato ma non ispezionato, atteso ma non utilizzato. Senza accesso al modello, gli osservatori esterni non potevano riprodurre i risultati riportati. Non potevano esaminare come fossero stati raccolti i dati di addestramento, testare le prestazioni su diverse lingue dei segni o scoprire come si comportasse al di fuori di esempi controllati. Le affermazioni sulla scala dell'addestramento e sulla latenza rimasero quindi di Classe C.

La reputazione di Google non ne modifica lo status. L'autorevolezza può rendere un'affermazione più persuasiva, ma non la rende indipendente. La ripetizione non trasforma un'asserzione in una misurazione. Il punto non è che SignGemma debba per forza aver fallito. Il punto è che al pubblico non è stato dato alcun modo affidabile di sapere se avesse avuto successo.

L'accuratezza non è un numero unico

Il collaudo indipendente è particolarmente importante nella tecnologia per la lingua dei segni perché la parola «accuratezza» può nascondere tanto quanto rivela. Un sistema può funzionare bene su un vocabolario limitato registrato con illuminazione stabile ma avere difficoltà con la conversazione naturale. Può riconoscere i segni di persone il cui aspetto assomiglia a quello presente nei dati di addestramento mentre ha presta-

zioni scarse per altre. Può identificare correttamente singoli segni ma perdere il significato lungo una frase. Può funzionare per una lingua dei segni ed essere promosso come se funzionasse per la lingua dei segni in generale.

Le lingue dei segni non sono versioni codificate delle lingue parlate. Hanno una propria grammatica, variazione regionale e contesto culturale. Il significato può dipendere dal movimento, dalla posizione, dai tempi, dall'espressione facciale e dalla relazione tra i segni. Un sistema che riconosce le mani ma non coglie il viso può produrre parole perdendo la frase. Anche un tasso di accuratezza complessivo elevato ci direbbe poco a meno di sapere cosa è stato testato, chi ha partecipato e cosa è stato considerato un risultato corretto.

Il sistema è stato testato su segni isolati o su conversazione continua? I partecipanti erano familiari al modello? I segnanti Deaf hanno giudicato se l'output avesse senso? Gli errori sono stati semplicemente contati, o sono stati esaminati per un possibile danno? Il test ha incluso età diverse, tonalità di pelle diverse, stili di segnazione e varianti regionali diverse? Questi non sono dettagli secondari. Determinano cosa significhi il numero.

Ecco perché gli studi sulla comprensione contano tanto quanto i parametri tecnici di riferimento. Un sistema può ottenere un punteggio impressionante producendo comunque un output che gli utenti trovano confuso, innaturale o fuorviante. La prova finale non è se un modello rileva il movimento. È se le persone riescono a capire e a fare affidamento su ciò che comunica.

Il costo dell'anticipazione

La promessa ripetuta che un sistema efficace stia per arrivare ha delle conseguenze. Un modello non rilasciato può occupare uno spazio che servizi funzionanti potrebbero altrimenti riempire. Può orientare le decisioni di finanziamento, influenzare le discussioni politiche e incoraggiare le istituzioni a rimandare investimenti nella fornitura umana. Può far apparire temporanee le carenze attuali, anche quando nessun calendario o prova pubblica sostiene tale convinzione.

Questa è la politica dell'attesa. L'aiuto sta sempre per arrivare, quindi l'assenza di aiuto nel presente viene trattata come meno urgente. Per gli utenti Deaf, il costo è pratico. Un ospedale, un'università, un datore di lavoro o un'autorità pubblica possono indicare una tecnologia emergente come prova che l'accessibilità sta migliorando. Ma un annuncio promettente non può interpretare una consultazione medica. Una dimo-

strazione non può garantire l'accesso a un'aula. Un modello non disponibile ai ricercatori è anche non disponibile alle persone la cui vita viene usata per giustificarne l'importanza.

C'è un altro costo. Quando l'attenzione pubblica si concentra su sistemi che non sono stati testati in modo indipendente, le organizzazioni più piccole che offrono forme di sostegno meno spettacolari ma più affidabili possono scomparire dalla vista. Interpreti umani, servizi guidati dalla comunità e pratiche di accessibilità consolidate raramente ricevono lo stesso entusiasmo di un nuovo modello di IA. Vengono giudicati come spese attuali, mentre la tecnologia viene valutata come possibilità futura. Il confronto è impari. La fornitura esistente deve rispondere ora dei propri limiti. Al sistema promesso è permesso rimanere perfetto perché non è ancora arrivato.

Cosa richiederebbe la trasparenza

L'assenza di prove indipendenti non è difficile da correggere in linea di principio. Gli sviluppatori potrebbero fornire ai ricercatori qualificati l'accesso prima di avanzare ampie affermazioni sulle prestazioni. Le valutazioni potrebbero essere progettate insieme ai segnanti Deaf anziché semplicemente condotte su di loro. Set di test, metodi e definizioni potrebbero essere pubblicati. I risultati potrebbero essere disaggregati per lingua, contesto e gruppo di utenti anziché compressi in un'unica cifra riassuntiva.

I ricercatori dovrebbero inoltre poter segnalare i fallimenti senza dipendere dal permesso dell'azienda oggetto della valutazione. Nulla di tutto ciò eliminerebbe l'incertezza. Gli studi indipendenti possono essere progettati male. I parametri di riferimento possono premiare forme ristrette di prestazione. I risultati pubblicati possono diventare obsoleti man mano che i sistemi cambiano, eppure un controllo imperfetto è meglio di una visibilità gestita. Un parametro di riferimento pubblico dà agli altri qualcosa da mettere in discussione, ripetere o migliorare. Un'affermazione non supportata dà loro solo qualcosa da far circolare.

La trasparenza significherebbe anche essere chiari su ciò che un sistema non può fare. Un prodotto progettato per un contesto ristretto non dovrebbe essere descritto come una soluzione generale. Un modello testato su una sola lingua dei segni non dovrebbe poter prendere in prestito l'apparente universalità della parola «segno». Un prototipo dovrebbe essere identificato come tale. Un modello non rilasciato non dovrebbe essere trattato come infrastruttura pubblica.

Il quinto parametro

I sistemi tecnici vengono spesso giudicati in base a misure familiari: velocità, scala, costo e accuratezza. La misura mancante è la fiducia. La fiducia non è un'altra rivendicazione di prestazioni. È il risultato del rendere le affermazioni verificabili. Cresce quando ricercatori esterni possono esaminare un sistema, quando gli utenti possono descriverne i fallimenti e quando le prove restano disponibili dopo che il ciclo degli annunci è passato. Questo è ***il quinto parametro*.**

Secondo questo criterio, il settore dell'IA per la lingua dei segni è entrato nell'agosto 2026 con un grave deficit. Aveva prodotti, promesse e pubblicità. Ciò che non aveva erano prove indipendenti disponibili pubblicamente per i sistemi che attiravano maggiore attenzione.

Questa assenza non dovrebbe essere scambiata per una lacuna temporanea nella documentazione. È parte della condizione attuale della tecnologia. Finché i valutatori indipendenti non potranno misurare questi sistemi, la risposta onesta a molte domande sulle loro prestazioni non è che funzionano, né che falliscono.

È che al pubblico non sono state fornite prove sufficienti per sapere.

Dichiarazione di interessi L'autrice è fondatrice di Novara Consulting Group LLC, che sviluppa e concede in licenza il Sign Language Access Trust Index citato in questo articolo, e cura The Fifth Parameter.

Finanziamento Nessuno. Prodotto internamente da Novara Consulting Group.

Dati e materiali L'articolo si basa sulla documentazione pubblica relativa ai sistemi citati e sul thread del Google AI Developers Forum riguardante SignGemma.

Dichiarazione sull'uso dell'IA Il testo di questo articolo è opera dell'autrice. L'assistenza dell'IA è stata utilizzata solo per l'impaginazione nel modello grafico ufficiale; nessuna argomentazione, figura, citazione o riferimento è stato generato dal modello.

Citazione suggerita Grizzle, H. M. (2026). The evidence that wasn't there. The Fifth Parameter. Vol 1 Numero 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.

Come citare

Grizzle, H. M. (2026, August 4). Le prove che non c'erano e che ancora non ci sono. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003816