Este documento foi traduzido do inglês por máquina e pode conter erros. O original em inglês está disponível em https://www.novaracg.com/2026/08/04/the-evidence-that-wasnt-and-still-isnt-there/.

A Evidência Que Não Existia E Ainda Não Existe

Heather M. Grizzle · 4 de agosto de 2026

Uma investigação sobre a evidência independente em falta por trás das maiores alegações de IA de língua gestual de 2026, do SignGemma às ferramentas de tradução comerciais.

O QUINTO PARÂMETRO

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

The Fifth Parameter, Vol. 1 · Edição nº 13

AUTORA Heather M. Grizzle

ORCID 0009-0005-2605-8256

DATA DA EVIDÊNCIA Atualizada até agosto de 2026

PALAVRAS-CHAVE IA de língua gestual; avaliação independente; classes de evidência; SignGemma; transparência; confiança; acesso Deaf

---

O facto mais importante sobre IA e língua gestual em 2026 não é o que aconteceu.

É o que não aconteceu.

Entre janeiro e agosto, empresas anunciaram produtos, demonstraram sistemas, angariaram fundos e falaram sobre o futuro da comunicação. Novas ferramentas foram apresentadas como mais rápidas, mais capazes e mais próximas do uso quotidiano. Algumas atraíram manchetes entusiásticas. Outras chegaram com alegações impressionantes sobre dados de treino, tempos de resposta ou o número de sinais que conseguiam reconhecer.

O que não apareceu foi evidência independente de que estes sistemas funcionavam conforme anunciado. Nenhuma terceira parte publicou uma medição pública de precisão, taxas de erro ou compreensão do utilizador para qualquer um dos principais sistemas examinados pelo Índice. Pesquisas nas fontes exigidas não encontraram ne-

nhum teste de desempenho independente para Signapse, Sorenson AST, Rylo Sign, Sign-Speak, ChatSign, SignVrse, Talksign ou o SignGemma da Google.

A formulação formal da conclusão é cuidadosa: nenhuma evidência de Classe A foi "localizada até agosto de 2026". Isto deixa em aberto a possibilidade de que alguma evidência exista algures fora do registo público. Pode estar dentro de uma empresa, na posse de um cliente ou contida em investigação ainda não publicada.

Mas evidência que não pode ser examinada não pode sustentar uma alegação pública. Esta distinção importa porque a indústria não tem estado silenciosa. Tem produzido um fluxo constante de anúncios, parcerias, demonstrações e histórias de financiamento. A impressão é de progresso rápido. No entanto, a evidência pública necessária para avaliar esse progresso não tem acompanhado o ritmo.

Um campo pode estar agitado sem se tornar responsável. Pode gerar notícias sem gerar conhecimento.

Três tipos de evidência

O Índice separa a evidência em três classes amplas.

- A Classe A é evidência independente. Provém de uma terceira parte e inclui medições que podem ser examinadas: resultados de precisão, taxas de erro, estudos de compreensão, comparações com outros sistemas ou desempenho num teste de referência publicado.
- A Classe B é evidência produzida pela empresa ou organização responsável pelo sistema. Pode conter detalhes técnicos úteis, mas a organização que faz a alegação é também a organização que está a ser avaliada.
- A Classe C consiste em afirmações: declarações em anúncios, materiais de marketing, entrevistas, apresentações ou cobertura de imprensa que não podem ser verificadas independentemente.

Isto não significa que toda alegação de Classe B ou Classe C seja falsa. Uma empresa pode descrever o seu produto com honestidade. Uma equipa de engenharia pode reportar um resultado real. Uma demonstração pode mostrar progresso técnico genuíno. O problema é que nenhuma destas coisas responde à questão central: como é que o sistema se comporta quando alguém sem interesse no seu sucesso o mede?

Em 2026, quase todo o julgamento público sobre IA e língua gestual ainda dependia de material de Classe B ou Classe C. As empresas forneceram as alegações, a lingua-

gem e, muitas vezes, as medidas pelas quais essas alegações eram compreendidas. A avaliação independente estava largamente ausente.

O problema do SignGemma

O SignGemma da Google oferece o exemplo mais claro. Quando foi anunciado, o modelo atraiu atenção porque parecia colocar os recursos de uma grande empresa tecnológica por trás da IA para língua gestual. As alegações associadas a ele incluíam treino em 10.000 horas de dados e latência abaixo de 200 milissegundos. Estes números viajaram amplamente porque sugeriam tanto escala como velocidade.

Quinze meses depois, o modelo ainda não tinha sido oficialmente lançado. Funcionários da Google confirmaram isto no Fórum de Programadores de IA da empresa. A discussão continuou ao longo de junho de 2026. Investigadores do Paquistão, Egito e outros locais pediram acesso. Não o receberam.

Isto criou um objeto público invulgar: um modelo que podia ser discutido mas não testado, citado mas não inspecionado, antecipado mas não utilizado. Sem acesso ao modelo, pessoas de fora não conseguiam reproduzir os seus resultados reportados. Não conseguiam examinar como os seus dados de treino tinham sido reunidos, testar o seu desempenho em diferentes línguas gestuais ou descobrir como se comportava fora de exemplos controlados. As alegações sobre escala de treino e latência permaneceram, portanto, Classe C.

A reputação da Google não altera o seu estatuto. A autoridade pode tornar uma alegação mais persuasiva, mas não a torna independente. A repetição não transforma uma afirmação numa medição. A questão não é que o SignGemma deva ter falhado. A questão é que ao público não foi dada nenhuma forma fiável de saber se tinha sido bem-sucedido.

A precisão não é um único número

O teste independente é particularmente importante na tecnologia de língua gestual porque a palavra "precisão" pode ocultar tanto quanto revela. Um sistema pode ter um bom desempenho num vocabulário limitado gravado em iluminação estável, mas ter dificuldades numa conversa natural. Pode reconhecer sinais de pessoas cuja aparência se assemelha à dos dados de treino, enquanto tem um desempenho fraco para outras. Pode identificar sinais individuais corretamente mas perder o significado ao

longo de uma frase. Pode funcionar para uma língua gestual e ser promovido como se funcionasse para a língua gestual em geral.

As línguas gestuais não são versões codificadas de línguas faladas. Têm a sua própria gramática, variação regional e contexto cultural. O significado pode depender do movimento, localização, tempo, expressão facial e a relação entre sinais. Um sistema que reconhece as mãos mas perde o rosto pode produzir palavras enquanto perde a frase. Mesmo uma taxa de precisão de destaque elevada diria pouco a menos que soubéssemos o que tinha sido testado, quem tinha participado e o que contava como resultado correto.

O sistema foi testado em sinais isolados ou em conversação contínua? Os participantes eram familiares ao modelo? Os signatários Surdos julgaram se o resultado fazia sentido? Os erros foram apenas contados, ou foram examinados quanto a possíveis danos? O teste incluiu diferentes idades, tons de pele, estilos de sinalização e variedades regionais? Estes não são detalhes secundários. Determinam o que o número significa.

É por isso que os estudos de compreensão importam a par dos testes de referência técnicos. Um sistema pode alcançar uma pontuação impressionante enquanto ainda produz resultados que os utilizadores consideram confusos, pouco naturais ou enganadores. O teste final não é se um modelo deteta movimento. É se as pessoas conseguem compreender e confiar no que ele comunica.

O custo da antecipação

A promessa repetida de que um sistema eficaz está prestes a chegar tem consequências. Um modelo não lançado pode ocupar espaço que serviços funcionais poderiam de outra forma preencher. Pode moldar decisões de financiamento, influenciar discussões políticas e encorajar instituições a adiar o investimento na prestação humana. Pode fazer com que as deficiências atuais pareçam temporárias, mesmo quando nenhum calendário ou evidência pública sustenta essa crença.

Esta é a política da antecipação. A ajuda está sempre a chegar, pelo que a ausência de ajuda no presente é tratada como menos urgente. Para os utilizadores Surdos, o custo é prático. Um hospital, universidade, empregador ou autoridade pública pode apontar para tecnologia emergente como evidência de que a acessibilidade está a melhorar. Mas um anúncio promissor não pode interpretar uma consulta médica. Uma demonstração não pode garantir acesso a uma sala de aula. Um modelo indisponível para in-

vestigadores é também indisponível para as pessoas cujas vidas são usadas para justificar a sua importância.

Há outro custo. Quando a atenção pública se concentra em sistemas que não foram testados independentemente, organizações mais pequenas que oferecem formas de apoio menos dramáticas mas mais fiáveis podem desaparecer de vista. Intérpretes humanos, serviços liderados pela comunidade e práticas de acessibilidade estabelecidas raramente recebem o mesmo entusiasmo que um novo modelo de IA. São julgados como despesas presentes, enquanto a tecnologia é valorizada como possibilidade futura. A comparação é desigual. A prestação existente tem de responder pelas suas limitações agora. Ao sistema prometido é permitido permanecer perfeito porque ainda não chegou.

O que a transparência exigiria

A ausência de evidência independente não é difícil de corrigir em princípio. Os programadores poderiam fornecer a investigadores qualificados acesso antes de fazer alegações de desempenho amplas. As avaliações poderiam ser concebidas com signatários Surdos em vez de meramente conduzidas sobre eles. Conjuntos de teste, métodos e definições poderiam ser publicados. Os resultados poderiam ser desagregados por língua, contexto e grupo de utilizadores em vez de comprimidos num único número de destaque.

Os investigadores também deveriam poder relatar falhas sem depender da permissão da empresa avaliada. Nada disto eliminaria a incerteza. Estudos independentes podem estar mal concebidos. Os testes de referência podem recompensar formas restritas de desempenho. Os resultados publicados podem tornar-se desatualizados à medida que os sistemas mudam, ainda assim um escrutínio imperfeito é melhor do que uma visibilidade gerida. Um teste de referência público dá aos outros algo para questionar, repetir ou melhorar. Uma alegação sem sustentação dá-lhes apenas algo para divulgar.

A transparência também significaria ser claro sobre o que um sistema não consegue fazer. Um produto concebido para um contexto restrito não deveria ser descrito como uma solução geral. Um modelo testado numa língua gestual não deveria ter permissão para apropriar-se da aparente universalidade da palavra “gestual”. Um protótipo deveria ser identificado como protótipo. Um modelo não divulgado não deveria ser tratado como infraestrutura pública.

O Quinto Parâmetro

Os sistemas técnicos são frequentemente avaliados por medidas familiares: velocidade, escala, custo e precisão. A medida em falta é a confiança. A confiança não é mais uma alegação de desempenho. É o resultado de tornar as alegações testáveis. Cresce quando investigadores externos podem inspecionar um sistema, quando os utilizadores podem descrever as suas falhas e quando a evidência permanece disponível depois de o ciclo de anúncios ter avançado. Este é ***o quinto parâmetro*.**

Segundo essa medida, o campo da IA de língua gestual entrou em agosto de 2026 com um sério défice. Tinha produtos, promessas e publicidade. O que não tinha era evidência independente disponível publicamente para os sistemas que atraem mais atenção.

Essa ausência não deve ser confundida com uma lacuna temporária de papelada. Faz parte da condição atual da tecnologia. Até que avaliadores independentes possam medir estes sistemas, a resposta honesta a muitas perguntas sobre o seu desempenho não é que funcionam, nem que falham.

É que o público não recebeu evidência suficiente para saber.

Declaração de interesses A autora é fundadora da Novara Consulting Group LLC, que constrói e licencia o Sign Language Access Trust Index referenciado neste artigo, e edita The Fifth Parameter.

Financiamento Nenhum. Produzido internamente pela Novara Consulting Group.

Dados e materiais O artigo baseia-se no registo público dos sistemas nomeados e no tópico do Google AI Developers Forum relativo ao SignGemma.

Divulgação do uso de IA O texto deste artigo é da própria autoria da autora. A assistência de IA foi utilizada apenas para a composição gráfica no modelo de marca; nenhum argumento, figura, citação ou referência foi gerado pelo modelo.

Citação sugerida Grizzle, H. M. (2026). The evidence that wasn't there. The Fifth Parameter. Vol 1 Edição 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.

Como citar

Grizzle, H. M. (2026, August 4). A Evidência Que Não Existia E Ainda Não Existe. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003816