Dowody, których nie było i wciąż nie ma

Heather M. Grizzle · August 4, 2026

Dochodzenie w sprawie brakujących niezależnych dowodów stojących za największymi twierdzeniami dotyczącymi AI języka migowego w 2026 roku, od SignGemma po komercyjne narzędzia tłumaczeniowe.

PIĄTY PARAMETR

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

The Fifth Parameter, tom 1 · numer 13

AUTOR Heather M. Grizzle

ORCID 0009-0005-2605-8256

DATA DOWODÓW Aktualne na sierpień 2026

SŁOWA KLUCZOWE AI języka migowego; niezależna ewaluacja; klasy dowodów; SignGemma; przejrzystość; zaufanie; dostęp dla osób Deaf

---

Najważniejszym faktem dotyczącym sztucznej inteligencji i języka migowego w 2026 roku nie jest to, co się wydarzyło.

Chodzi o to, co się nie wydarzyło.

Między styczniem a sierpniem firmy ogłaszały produkty, prezentowały systemy, pozyskiwały fundusze i wypowiadały się na temat przyszłości komunikacji. Nowe narzędzia przedstawiano jako szybsze, sprawniejsze i bliższe codziennego użytku. Niektóre przyciągnęły entuzjastyczne nagłówki. Inne pojawiły się z imponującymi twierdzeniami dotyczącymi danych treningowych, czasu reakcji lub liczby rozpoznawanych znaków.

To, co się nie pojawiło, to niezależny dowód na to, że systemy te działały tak, jak reklamowano. Żadna strona trzecia nie opublikowała publicznego pomiaru dokładności, wskaźników błędów ani zrozumiałości dla użytkownika w odniesieniu do żadnego z głównych systemów zbadanych przez Indeks. Przeszukanie wymaganych źródeł nie

ujawniło niezależnego testu porównawczego dla Signapse, Sorenson AST, Rylo Sign, Sign-Speak, ChatSign, SignVrse, Talksign ani SignGemma od Google.

Formalne sformułowanie ustalenia jest ostrożne: żaden dowód klasy A nie został „odnaleziony na sierpień 2026". Pozostawia to otwartą możliwość, że gdzieś poza rejestrem publicznym istnieją jakieś dowody. Mogą znajdować się wewnątrz firmy, być w posiadaniu klienta lub zawarte w badaniach, które nie zostały jeszcze opublikowane.

Jednak dowód, którego nie można zbadać, nie może stanowić podstawy publicznego twierdzenia. To rozróżnienie ma znaczenie, ponieważ branża nie milczała. Wytworzyła stały strumień ogłoszeń, partnerstw, demonstracji i historii o finansowaniu. Wrażenie jest takie, jakoby postęp był szybki. Jednak publiczne dowody potrzebne do oceny tego postępu nie nadążały.

Dziedzina może być pełna aktywności, nie stając się przy tym odpowiedzialną. Może generować wiadomości, nie generując wiedzy.

Trzy rodzaje dowodów

Indeks dzieli dowody na trzy szerokie klasy.

- Klasa A to dowód niezależny. Pochodzi od strony trzeciej i obejmuje pomiary, które można zbadać: wyniki dokładności, wskaźniki błędów, badania zrozumiałości, porównania z innymi systemami lub wyniki na opublikowanym benchmarku.
- Klasa B to dowód wytworzony przez firmę lub organizację odpowiedzialną za system. Może zawierać przydatne szczegóły techniczne, ale organizacja formułująca twierdzenie jest jednocześnie organizacją poddawaną ocenie.
- Klasa C składa się z twierdzeń: stwierdzeń w ogłoszeniach, materiałach marketingowych, wywiadach, prezentacjach lub relacjach prasowych, których nie można niezależnie zweryfikować.

Nie oznacza to, że każde twierdzenie klasy B lub C jest fałszywe. Firma może uczciwie opisywać swój produkt. Zespół inżynierski może zgłosić rzeczywisty wynik. Demonstracja może pokazywać prawdziwy postęp techniczny. Problem polega na tym, że żadna z tych rzeczy nie odpowiada na kluczowe pytanie: jak system działa, gdy mierzy go ktoś niezainteresowany jego sukcesem?

W 2026 roku niemal każda publiczna ocena dotycząca AI i języka migowego wciąż opierała się na materiałach klasy B lub C. Firmy dostarczały twierdzenia, język, a czę-

sto także miary, za pomocą których te twierdzenia były rozumiane. Niezależna ocena była w dużej mierze nieobecna.

Problem SignGemma

SignGemma od Google stanowi najwyraźniejszy przykład. Kiedy model został ogłoszony, przyciągnął uwagę, ponieważ wydawał się stawiać zasoby dużej firmy technologicznej za sztuczną inteligencją dla języka migowego. Powiązane z nim twierdzenia obejmowały trening na 10 000 godzin danych oraz opóźnienie poniżej 200 milisekund. Te liczby rozeszły się szeroko, ponieważ sugerowały zarówno skalę, jak i szybkość.

Piętnaście miesięcy później model wciąż nie został oficjalnie wydany. Pracownicy Google potwierdzili to na firmowym Forum Deweloperów AI. Dyskusja trwała aż do czerwca 2026 roku. Badacze z Pakistanu, Egiptu i innych krajów prosili o dostęp. Nie otrzymali go.

Stworzyło to nietypowy obiekt publiczny: model, który można było omawiać, ale nie testować, cytować, ale nie sprawdzać, oczekiwać, ale nie używać. Bez dostępu do modelu osoby postronne nie mogły odtworzyć zgłaszanych wyników. Nie mogły zbadać, w jaki sposób zestawiono jego dane treningowe, przetestować jego działania w różnych językach migowych ani odkryć, jak zachowywał się poza kontrolowanymi przykładami. Twierdzenia dotyczące skali treningu i opóźnienia pozostały więc klasą C.

Reputacja Google nie zmienia ich statusu. Autorytet może uczynić twierdzenie bardziej przekonującym, ale nie czyni go niezależnym. Powtarzanie nie zamienia twierdzenia w pomiar. Rzecz nie w tym, że SignGemma musiało ponieść porażkę. Rzecz w tym, że opinii publicznej nie dano żadnego wiarygodnego sposobu, by wiedzieć, czy odniosło sukces.

Dokładność to nie jedna liczba

Niezależne testowanie jest szczególnie istotne w technologii języka migowego, ponieważ słowo „dokładność” może ukrywać tyle samo, ile ujawnia. System może działać dobrze przy ograniczonym słownictwie zarejestrowanym w stabilnym oświetleniu, ale mieć trudności z naturalną rozmową. Może rozpoznawać znaki osób, których wygląd przypomina te z danych treningowych, działając słabo w przypadku innych. Może poprawnie identyfikować pojedyncze znaki, ale gubić sens w obrębie zdania. Może działać dla jednego języka migowego, będąc promowanym tak, jakby działał dla języka migowego w ogóle.

Języki migowe nie są zakodowanymi wersjami języków mówionych. Mają własną gramatykę, zróżnicowanie regionalne i kontekst kulturowy. Znaczenie może zależeć od ruchu, lokalizacji, czasu trwania, mimiki twarzy oraz relacji między znakami. System, który rozpoznaje dłonie, ale pomija twarz, może produkować słowa, tracąc jednocześnie zdanie. Nawet wysoki, nagłówkowy wskaźnik dokładności niewiele by nam powiedział, gdybyśmy nie wiedzieli, co zostało przetestowane, kto brał w tym udział i co uznano za wynik poprawny.

Czy system testowano na izolowanych znakach, czy na ciągłej rozmowie? Czy uczestnicy byli znani modelowi? Czy Głusi sygnatariusze oceniali, czy wynik miał sens? Czy błędy jedynie liczono, czy badano je pod kątem możliwej szkody? Czy test obejmował różny wiek, odcienie skóry, style migania i odmiany regionalne? To nie są szczegóły drugorzędne. Determinują one, co oznacza dana liczba.

Dlatego badania zrozumiałości mają znaczenie obok benchmarków technicznych. System może osiągnąć imponujący wynik, wciąż produkując wyjście, które użytkownicy uznają za mylące, nienaturalne lub wprowadzające w błąd. Ostatecznym testem nie jest to, czy model wykrywa ruch. Jest nim to, czy ludzie potrafią zrozumieć i polegać na tym, co jest im komunikowane.

Koszt antycypacji

Powtarzana obietnica, że skuteczny system lada moment się pojawi, niesie ze sobą konsekwencje. Nieopublikowany model może zajmować miejsce, które mogłyby wypełnić działające usługi. Może kształtować decyzje dotyczące finansowania, wpływać na dyskusje polityczne i zachęcać instytucje do odkładania inwestycji w pomoc świadczoną przez ludzi. Może sprawiać, że obecne niedociągnięcia wydają się tymczasowe, nawet gdy żaden harmonogram ani publiczne dowody nie potwierdzają tego przekonania.

To jest polityka wyczekiwania. Pomoc zawsze nadchodzi, więc brak pomocy w teraźniejszości traktowany jest jako mniej pilny. Dla użytkowników Głuchych koszt jest praktyczny. Szpital, uniwersytet, pracodawca czy urząd publiczny może wskazywać na nadchodzącą technologię jako dowód poprawy dostępności. Jednak obiecująca zapowiedź nie może przetłumaczyć konsultacji lekarskiej. Pokaz demonstracyjny nie może zagwarantować dostępu do sali lekcyjnej. Model niedostępny dla badaczy jest też niedostępny dla ludzi, których życie wykorzystuje się do uzasadniania jego znaczenia.

Jest jeszcze inny koszt. Gdy uwaga publiczna skupia się na systemach, które nie zostały niezależnie przetestowane, mniejsze organizacje oferujące mniej spektakularne, lecz bardziej niezawodne formy wsparcia mogą zniknąć z pola widzenia. Tłumacze języka migowego, usługi prowadzone przez społeczności i ugruntowane praktyki w zakresie dostępności rzadko wzbudzają takie samo podekscytowanie jak nowy model AI. Są oceniane jako bieżące wydatki, podczas gdy technologia jest ceniona jako przyszła możliwość. Porównanie jest nierówne. Istniejąca pomoc musi już teraz odpowiadać za swoje ograniczenia. Obiecany system może pozostać doskonały, ponieważ jeszcze się nie pojawił.

Czego wymagałaby przejrzystość

Brak niezależnych dowodów w zasadzie nie jest trudny do naprawienia. Twórcy mogliby udostępniać wykwalifikowanym badaczom dostęp przed wygłaszaniem szeroko zakrojonych twierdzeń o skuteczności. Ewaluacje mogłyby być projektowane wspólnie z osobami Głuchymi posługującymi się językiem migowym, a nie jedynie przeprowadzane na nich. Zbiory testowe, metody i definicje mogłyby zostać opublikowane. Wyniki mogłyby być przedstawiane w podziale na język, kontekst i grupę użytkowników, zamiast być sprowadzane do jednej nagłówkowej liczby.

Badacze powinni też mieć możliwość zgłaszania niepowodzeń bez konieczności uzyskiwania zgody firmy poddawanej ocenie. Nic z tego nie usunęłoby niepewności. Niezależne badania mogą być źle zaprojektowane. Benchmarki mogą nagradzać wąskie formy skuteczności. Opublikowane wyniki mogą się dezaktualizować wraz ze zmianami systemów, jednak niedoskonała kontrola jest lepsza niż zarządzana widoczność. Publiczny benchmark daje innym coś, co można zakwestionować, powtórzyć lub udoskonalić. Niepoparte twierdzenie daje im tylko coś, co można rozpowszechniać.

Przejrzystość oznaczałaby również jasne określenie, czego system nie potrafi. Produkt zaprojektowany dla ograniczonego zastosowania nie powinien być opisywany jako rozwiązanie uniwersalne. Model przetestowany na jednym języku migowym nie powinien móc zapożyczać pozornej uniwersalności słowa „migowy”. Prototyp powinien być określany jako prototyp. Nieopublikowany model nie powinien być traktowany jak infrastruktura publiczna.

Piąty parametr

Systemy techniczne są często oceniane według znanych miar: szybkości, skali, kosztu i dokładności. Brakującą miarą jest zaufanie. Zaufanie nie jest kolejnym twierdzeniem

o wydajności. Jest wynikiem uczynienia twierdzeń weryfikowalnymi. Rośnie ono, gdy zewnętrzni badacze mogą zbadać system, gdy użytkownicy mogą opisać jego usterki oraz gdy dowody pozostają dostępne po tym, jak cykl ogłoszeń przeminął. To jest ***piąty parametr*.**

Według tej miary dziedzina AI w dziedzinie języka migowego wkroczyła w sierpień 2026 roku z poważnym deficytem. Miała produkty, obietnice i rozgłos. Nie miała publicznie dostępnych, niezależnych dowodów dla systemów przyciągających najwięcej uwagi.

Ten brak nie powinien być mylony z tymczasową luką w dokumentacji. Jest częścią obecnego stanu technologii. Dopóki niezależni ewaluatorzy nie będą mogli zmierzyć tych systemów, uczciwa odpowiedź na wiele pytań o ich skuteczność nie brzmi, że działają, ani że zawodzą.

Chodzi o to, że opinii publicznej nie przedstawiono wystarczających dowodów, aby mogła to wiedzieć.

Deklaracja konfliktu interesów Autorka jest założycielką Novara Consulting Group LLC, która tworzy i licencjonuje Sign Language Access Trust Index przywoływany w tym artykule, oraz redaguje The Fifth Parameter.

Finansowanie Brak. Opracowano wewnętrznie przez Novara Consulting Group.

Dane i materiały Artykuł opiera się na publicznie dostępnych danych dotyczących wymienionych systemów oraz na wątku forum Google AI Developers Forum poświęconym SignGemma.

Ujawnienie użycia AI Tekst tego artykułu jest autorstwa autorki. Wsparcie AI wykorzystano wyłącznie do składu w markowym szablonie; żaden argument, wykres, cytat ani przypis nie został wygenerowany przez model.

Sugerowany cytat Grizzle, H. M. (2026). The evidence that wasn't there. The Fifth Parameter. Tom 1, numer 13. Novara Press.

Jak cytować

Grizzle, H. M. (2026, August 4). Dowody, których nie było i wciąż nie ma. Novara Consulting Group. https://www.novaracg.com/pl/2026/08/04/the-evidence-that-wasnt-and-still-isnt-there/