NOVARA
CONSULTING GROUP

Ten dokument został przetłumaczony z języka angielskiego maszynowo i może zawierać błędy. Oryginał w języku angielskim znajduje się pod adresem https://www.novaracg.com/2026/07/27/the-consent-gap-undisclosed-data-collection-in-sign-language-ai-interpreting-systems/.

Luka w zgodzie: nieujawnione gromadzenie danych w systemach tłumaczenia AI języka migowego

Heather Grizzle · 27 lipca 2026

Sign language AI depends on video of Deaf users. What vendors disclose about collection, retention, training use and sharing is inconsistent or absent.

Spis treści

1. Streszczenie
2. Problem polityki
3. Analiza instytucjonalna
4. Implikacje dla zarządzania
5. Stanowisko polityczne NCG
6. Kwestie wdrożeniowe
7. Bibliografia

Streszczenie

Sztuczna inteligencja języka migowego, niezależnie od tego, czy jest budowana do rozpoznawania, tłumaczenia czy generowania w oparciu o awatary, strukturalnie zależy od rejestracji wideo lub biometrycznej użytkowników Deaf i niedosłyszących. Ta zależność tworzy problem zarządzania danymi, który spotkał się ze znacznie mniejszą kontrolą instytucjonalną niż udokumentowane we wcześniejszych częściach tej serii błędy językowe i związane ze zrozumieniem. W całym krajobrazie dostawców, który NCG przeanalizowało w ramach struktury SLAT (Sign Language Access Trust), ujawnianie informacji o tym, co jest gromadzone, jak długo jest przechowywane, czy jest wykorzystywane do trenowania przyszłych modeli oraz czy jest udostępniane stronom trzecim,

jest w najlepszym razie niespójne, a w najgorszym razie nieobecne. Nie jest to twierdzenie, że jakikolwiek konkretny dostawca działał w złej wierze. Dostępne publicznie dowody nie pozwalają jeszcze na przypisanie intencji żadnej wymienionej z nazwy firmie, i niniejszy artykuł takiego przypisania nie dokonuje. Jest to twierdzenie, że rynek nie wypracował jeszcze standardu ujawniania informacji odpowiedniego do wrażliwości danych, o które chodzi, oraz że ta luka stanowi niedociągnięcie w zarządzaniu niezależne od postępowania jakiegokolwiek pojedynczego dostawcy. Przedstawione poniżej stanowisko polityczne argumentuje, że adekwatność zgody i ujawniania informacji powinna być traktowana jako kryterium dopuszczające w procesie zamówień publicznych, a nie jako kwestia drugorzędna, rozpatrywana dopiero po ocenie funkcjonalnej.

Problem polityki

Rejestracja języka migowego jest rejestracją biometryczną. Wideo przedstawiające ręce osoby migającej, gramatykę mimiczną i pozycję ciała nie stanowi danych wejściowych ubocznych, porównywalnych do zdania wpisanego w system oparty na tekście. Jest to informacja jednoznacznie identyfikująca, z definicji kodująca informacje o statusie niepełnosprawności, i w wielu jurysdykcjach spełnia ustawową definicję danych biometrycznych lub szczególnej kategorii danych podlegających wzmocnionej ochronie. Na przykład ustawa Illinois Biometric Information Privacy Act (BIPA) wymaga pisemnego powiadomienia i zgody przed zebraniem identyfikatorów biometrycznych oraz nakłada odpowiedzialność niezależną od udowodnienia rzeczywistej szkody. Ogólne rozporządzenie o ochronie danych Unii Europejskiej, GDPR (General Data Protection Regulation), klasyfikuje dane ujawniające informacje związane z niepełnosprawnością oraz określone dane biometryczne wykorzystywane do identyfikacji jako szczególną kategorię danych na mocy art. 9, wymagając wyraźnej podstawy prawnej wykraczającej poza zwykłą zgodę. Kalifornijskie ustawy CCPA (California Consumer Privacy Act) i CPRA (California Privacy Rights Act) nakładają własne obowiązki dotyczące ujawniania informacji i możliwości rezygnacji w odniesieniu do danych biometrycznych.

Mimo tego krajobrazu regulacyjnego, produkty AI języka migowego skierowane do konsumentów często opierają się na ogólnych regulaminach usług, które nie zostały opracowane z myślą o użytkownikach Deaf ani o rejestracji biometrycznej. W dokonanych przez NCG przeglądach dostawców widoczne są trzy powtarzające się niedociągnięcia: po pierwsze, brak odrębnego lub wyraźnego ujawnienia, że wideo wejściowe

stanowi dane biometryczne podlegające konkretnym ochronom ustawowym; po drugie, brak jasnego oświadczenia, czy zarejestrowane wideo lub wyprowadzone z niego cechy są przechowywane do celów trenowania modelu, a jeśli tak, to jak długo i z jakimi prawami do usunięcia danych; po trzecie, brak ujawnienia w dostępnym formacie, co oznacza, że sam język zgody często nie jest przetłumaczony na ASL ani przedstawiony w formacie użytecznym dla populacji, której dane są gromadzone. Ta ostatnia kwestia nie jest formalnością. System zgody, który wymaga umiejętności czytania w pisanym drugim języku, aby zrozumieć, co dzieje się z własnymi danymi biometrycznymi, nie funkcjonuje jako świadoma zgoda dla populacji, w której znajomość języka angielskiego jest bardzo zróżnicowana, a ASL jest podstawowym językiem rozumienia.

Analiza instytucjonalna

Rejestracja języka migowego jest rejestracją biometryczną.

Dynamika instytucjonalna, która wytwarza tę lukę, ma charakter strukturalny, a nie spiskowy. AI języka migowego to młoda kategoria produktów, a większość dostawców to małe firmy działające na tych samych ogólnych szablonach polityki prywatności stosowanych w całej branży oprogramowania, dostosowywanych minimalnie, jeśli w ogóle. Tymczasem procesy zamówień na poziomie instytucjonalnym historycznie oceniały te produkty na podstawie kryteriów funkcjonalnych: czy awatar renderuje się w sposób zrozumiały, czy system rozpoznawania osiąga akceptowalną dokładność, czy dostawca zatrudnia osoby Deaf zaangażowane w rozwój produktu. Zarządzanie danymi traktowane było jako kwestia zgodności prawnej lub IT, rozpatrywana po podjęciu merytorycznej decyzji zakupowej, jeśli w ogóle była rozpatrywana. Ta kolejność stanowi rzeczywiste niedociągnięcie w zarządzaniu. Zanim niedociągnięcie w zarządzaniu danymi wyjdzie na jaw — poprzez naruszenie danych, dochodzenie regulacyjne lub publiczną krytykę ze strony społeczności Deaf, której dane były wykorzystywane — relacja instytucjonalna i zobowiązanie finansowe są już ustanowione, a możliwość wymuszenia naprawy jest odpowiednio słabsza.

Istnieje również kwestia pochodzenia danych treningowych, która sąsiaduje z problemem zgody, lecz jest analitycznie od niego odrębna. Kilka systemów AI języka migowego zostało zbudowanych częściowo na korpusach wideo pochodzących z publicznie udostępnionych treści, w tym treści tworzonych przez użytkowników i edukatorów ASL społeczności Deaf na platformach społecznościowych. To, czy takie wykorzystanie

mieści się w zakresie regulaminu danej platformy, czy stanowi dozwolony użytek na gruncie prawa autorskiego, oraz czy osoby, których migowanie widoczne jest w tym materiale, zostały w sposób znaczący poinformowane, że ich treści zostaną wykorzystane do trenowania komercyjnego systemu AI, to otwarte pytania, których NCG nie zweryfikowało niezależnie dla żadnego konkretnego dostawcy i których nie przedstawia tutaj jako ustalonych faktów. Z punktu widzenia zarządzania istotne jest to, że urzędnicy zajmujący się zamówieniami oceniający te systemy niemal nie mają wglądu w pochodzenie danych treningowych, a obecne normy ujawniania informacji w branży nie zobowiązują dostawców do jego dostarczenia.

Implikacje dla zarządzania

Instytucja nabywająca AI języka migowego nie jest wyłącznie nabywcą narzędzia tłumaczeniowego. Tam, gdzie system działa na żywym materiale wideo studentów, pacjentów, pracowników lub przedstawicieli społeczeństwa, instytucja jest również administratorem lub podmiotem przetwarzającym dane w rozumieniu większości obowiązujących ram prawnych dotyczących prywatności, i dziedziczy ekspozycję zarządczą wynikającą z praktyk danych swojego dostawcy, niezależnie od tego, czy ta ekspozycja została ujawniona w momencie zakupu. Instytucja, która nie zweryfikowała praktyk dostawcy dotyczących gromadzenia, przechowywania i wykorzystania danych do trenowania modeli przed wdrożeniem, nie może zapewniać swoich interesariuszy, organów regulacyjnych ani osób Deaf, których dane są rejestrowane, że wypełniła własne obowiązki w zakresie ochrony danych. Jest to prawdziwe niezależnie od faktycznego postępowania dostawcy, ponieważ brak zweryfikowanego ujawnienia informacji sam w sobie stanowi wadę zarządzania, niezależnie od tego, czy dany dostawca ostatecznie okaże się odpowiedzialnie obchodzić z danymi.

Stanowisko polityczne NCG

Stanowisko NCG jest takie, że adekwatność zarządzania danymi i ujawniania informacji o prywatności powinna funkcjonować jako kryterium dopuszczające w zamówieniach na AI języka migowego, oceniane przed wydajnością funkcjonalną, a nie po niej. Niezdolność dostawcy do przedstawienia jasnego, dostępnego w ASL, odpowiedniego do jurysdykcji ujawnienia informacji o tym, jakie dane biometryczne są gromadzone, jak długo są przechowywane, czy i w jaki sposób są wykorzystywane w trenowaniu modelu oraz jakie istnieją prawa do usunięcia danych, powinna być traktowana jako dys-

kwalifikująca dla wdrożenia instytucjonalnego, niezależnie od wyników językowych lub technicznych w pozostałych częściach oceny. Stanowisko to jest już zoperacjonalizowane w domenach Zarządzania Danymi i Prywatności oraz Przejrzystości i Ujawniania Informacji struktury SLAT, a niniejszy artykuł argumentuje, że domeny te zasługują na podwyższoną wagę w stosunku do obecnej kalibracji struktury, biorąc pod uwagę ekspozycję ustawową, na jaką narażone są instytucje, oraz barierę zrozumienia, jaką niedostępny język zgody tworzy dla dotkniętej nią populacji.

Kwestie wdrożeniowe

Urzędnicy zajmujący się zamówieniami oceniający AI języka migowego powinni wymagać, jako warunku odpowiedzi dostawcy, pisemnego ujawnienia zarządzania danymi obejmującego cztery powyższe elementy, żądać, aby to ujawnienie było dostępne w formacie wideo ASL oprócz pisemnego języka angielskiego, oraz traktować odmowę lub niezdolność dostawcy do przedstawienia takiego ujawnienia jako niekorzystne ustalenie, niezależnie od wydajności funkcjonalnej systemu. Instytucje z istniejącymi wdrożeniami powinny przeprowadzić przegląd retrospektywny, zamiast czekać na odnowienie umowy, biorąc pod uwagę, że ekspozycja zarządcza narasta od daty wdrożenia, a nie od daty odkrycia. Tam, gdzie pochodzenie danych treningowych dostawcy nie może zostać zweryfikowane, instytucje powinny żądać gwarancji umownych dotyczących zgodnego z prawem pozyskiwania danych, zamiast traktować brak odpowiedzi jako neutralną niewiadomą.

> **CZTERY ELEMENTY UJAWNIENIA INFORMACJI**
>
> Jakie dane biometryczne są gromadzone, jak długo są przechowywane, czy i w jaki sposób są wykorzystywane w trenowaniu modelu oraz jakie istnieją prawa do usunięcia danych.

Bibliografia

- Biometric Information Privacy Act, 740 ILCS 14 (Illinois).
- California Consumer Privacy Act z 2018 r., ze zmianami wprowadzonymi przez California Privacy Rights Act, Cal. Civ. Code § 1798.100 i nast.
- Rozporządzenie (UE) 2016/679 (Ogólne rozporządzenie o ochronie danych), art. 9.

- Federal Trade Commission Act, 15 U.S.C. § 45 (nieuczciwe lub wprowadzające w błąd praktyki).
- Novara Consulting Group, *Struktura Sign Language Access Trust (SLAT)*, 2026.

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.

Jak cytować

Grizzle, H. M. (2026, July 27). Luka w zgodzie: nieujawnione gromadzenie danych w systemach tłumaczenia AI języka migowego. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003818