NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/07/07/the-policy-problem-with-sign-language-ai-when-public-procurement-outruns-scientific-evidence/.

Das politische Problem der Gebärdensprach-KI: Wenn die öffentliche Beschaffung der wissenschaftlichen Evidenz voraus ist

Heather Grizzle · July 7, 2026

Governments are buying sign language AI before independent evidence exists. Why procurement is the checkpoint and what responsible buying looks like.

Inhalt

1. Der Moment der Beschaffung
2. Die Anziehungskraft des technologischen Versprechens
3. Das Problem der Evidenzknappheit
4. Warum Gebärdensprach-KI nicht einfach ein weiteres Barrierefreiheits-Tool ist
5. Wissenschaftliche Untersuchung vs. überzeugende Texte
6. Beschaffung als Kontrollpunkt der Governance
7. Was eine verantwortungsvolle Bewertung fragen sollte
8. Das Risiko, den Hype-Zyklus zu wiederholen
9. Die Rolle unabhängiger Bewertung und Governance
10. Evidenz vor Einführung

Regierungen beschaffen Sprachzugangssysteme, bevor die Evidenzbasis ausgereift ist. Dies ist eine öffentliche Rechenschaftskrise in Zeitlupe.

Der Moment der Beschaffung

Regierungsbehörden werden aufgefordert, bedeutende Beschaffungsentscheidungen über Gebärdensprach-KI-Systeme zu treffen, bevor die wissenschaftliche Evidenzbasis ausgereift ist. Dies ist keine Frage, ob die Technologie irgendwann ausreifen wird. Es ist eine Frage, ob gegenwärtige Herstellerbehauptungen als validierte Tatsache behandelt werden, obwohl sie unverifiziertes Versprechen bleiben. Diese Unterscheidung ist von Bedeutung, besonders wenn öffentliche Gelder und der Zugang der Deaf-Community auf dem Spiel stehen.

Over the past eighteen months, accessibility officers, procurement teams, and IT departments across federal, state, and municipal governments have begun receiving vendor pitches for sign language AI systems. These pitches emphasize innovation, efficiency, and inclusion. They promise to expand access at scale, reduce costs, and modernize language services. The messaging is compelling, and it arrives at an institutional moment when public sector budgets are constrained and accessibility mandates are expanding. The question being asked by procurement officers is simple: "Why wouldn't we adopt this?"

The answer is equally simple, but less frequently voiced: "Because we do not yet have independent evidence that these systems work as claimed, and we have no agreed standards for what 'working' means in the context of language access."

Dies ist kein technischer Einwand. Es ist ein Einwand der Governance. Es ist die Frage, ob öffentliche Institutionen Beschaffungsentscheidungen normalisieren sollten, denen die Rechenschaftsstrukturen fehlen, die notwendig sind, um die Menschen zu schützen, die von diesen Entscheidungen am stärksten betroffen sind.

Die Anziehungskraft des technologischen Versprechens

Gebärdensprach-KI ist für öffentliche Institutionen aus Gründen attraktiv, die innerhalb der Logik institutioneller Beschaffung vollkommen rational sind. Die Technologie verspricht Innovation in einem Bereich, in dem Innovation bislang nur sporadisch stattgefunden hat. Sie suggeriert Skalierbarkeit in Zusammenhängen, in denen Sprachdienste historisch gesehen menschliche Dolmetscher im Eins-zu-eins-Verhältnis erfordert haben. Sie bietet Kosteneffizienz in Systemen, die unter Budgetdruck stehen. Sie präsentiert sich als Inklusion und Empowerment, im Einklang mit den Rechten von Menschen mit Behinderungen und Vorgaben zur Barrierefreiheit.

Dies sind starke Anreize für die Beschaffung. Wenn eine Organisation gleichzeitig unter Druck steht, den Zugang zu erweitern, Kosten zu senken und neue Technologien einzuführen, wird eine Lösung, die alle drei Dinge verspricht, im Moment des Angebots schwer abzulehnen.

Beschaffungsbequemlichkeit ist jedoch nicht dasselbe wie die Einhaltung von Bürgerrechten. Institutionelle Effizienz ist nicht dasselbe wie linguistische Validität. Kosteneinsparungen sind nicht dasselbe wie die Sicherheit der Community. Die Anziehungskraft des technologischen Versprechens hat wiederholt dazu geführt, dass der öffentliche Sektor Systeme übernommen hat, bevor deren reale Auswirkungen vollständig verstanden wurden. Im Bereich der Barrierefreiheit ist dieses Muster besonders gefährlich, da die Menschen, die am ehesten Schaden erleiden, selten diejenigen sind, die über die Einführung der Technologie entscheiden.

Beschaffungsbequemlichkeit ist nicht dasselbe wie die Einhaltung von Bürgerrechten.

Das Problem der Evidenzknappheit

Many sign language AI systems are promoted with claims about accessibility, linguistic competence, inclusion, language access, and Deaf empowerment. These are claims that matter. They address real needs. But these claims are often not accompanied by the evidence structures that would normally be required before a public institution normalizes a system affecting fundamental aspects of people's lives.

Konkret werden diese Systeme häufig beworben, ohne dass Folgendes vorliegt:

- eine unabhängige Bewertung durch Forscher außerhalb der Herstellerorganisation;
- transparente und öffentlich zugängliche Datensätze;
- veröffentlichte Fehleranalysen, die dokumentieren, was das System nicht leistet und unter welchen Bedingungen;
- Belege aus dem realen Einsatz, die aus dauerhafter Nutzung in operativen Kontexten stammen; oder

- eine gemeinschaftlich informierte Risikobewertung, die mit gehörlosen Muttersprachlern der Gebärdensprache und den Bevölkerungsgruppen durchgeführt wird, die am stärksten auf Sprachzugang angewiesen sind.

Dies soll nicht heißen, dass alle Anbieter von Gebärdensprach-KI unbegründete Behauptungen aufstellen. Es soll heißen, dass selbst wenn Behauptungen in gutem Glauben gemacht werden, das Fehlen einer unabhängigen Überprüfung bedeutet, dass institutionelle Entscheidungen auf der Grundlage überzeugender Texte statt wissenschaftlicher Belege getroffen werden. Das ist nicht dasselbe. Vorführungen von Herstellern, Erfahrungsberichte von Testnutzern, Fallstudien und Marketingmaterialien können allesamt überzeugend sein und dennoch allesamt den Evidenzstandard verfehlen, der für eine verantwortungsvolle öffentliche Beschaffung notwendig ist.

The gap is not trivial. It is the space between "this technology shows promise" and "this technology is ready for public deployment." That space is where public policy should insist on accountability.

Warum Gebärdensprach-KI nicht einfach ein weiteres Barrierefreiheits-Tool ist

Gebärdensprache ist eine vollwertige natürliche Sprache, kein Feature, keine Untertitelebene und keine Verbesserung des Kundenservices. Sie hat eine komplexe phonologische Struktur, syntaktische Regeln, morphologische Prozesse und pragmatische Normen. Sie ist die Muttersprache der Deaf-Communities und die Grundlage für Bildung, berufliche und bürgerschaftliche Teilhabe von Millionen Menschen weltweit.

Dies ist von Bedeutung, weil Fehler in der Gebärdensprach-KI mehr als nur Unannehmlichkeiten verursachen können. Sie können das Verständnis von Informationen beeinträchtigen, die Gesundheit, rechtlichen Status, Bildungszugang, Beschäftigungsfähigkeit, Notfallreaktionen und bürgerliche Rechte betreffen. Ein Untertitelsystem, das gelegentlich Dialoge verpasst, ist ein Barrierefreiheits-Feature mit einer bekannten Einschränkung. Ein Gebärdensprachsystem, das gelegentlich Bedeutung falsch wiedergibt, ist ein potenzieller Vektor für Schaden.

Die Fehlertoleranz für Gebärdensprach-KI unterscheidet sich folglich von der Fehlertoleranz weniger sprachlich grundlegender Technologien. Ein System, das in 95 Prozent der Fälle korrekt funktioniert, mag für eine ergänzende Funktion akzeptabel sein. Es ist möglicherweise nicht akzeptabel als primärer Sprachzugangsmechanismus in

einem Gerichtsverfahren, einer medizinischen Konsultation oder einem Bildungsumfeld. Öffentliche Behörden müssen nicht nur verstehen, was ein System gut macht, sondern auch, was es schlecht macht, und ob diese Fehler in dem spezifischen Kontext, in dem das System eingesetzt wird, toleriert werden können.

Wissenschaftliche Untersuchung vs. überzeugende Texte

Die Unterscheidung zwischen wissenschaftlicher Untersuchung und überzeugenden Texten ist nicht akademisch. Es ist eine Frage der öffentlichen Rechenschaftspflicht.

Herstellerpräsentationen, Produktvorführungen, Erfahrungsberichte, Fallstudien und Marketingsprache erfüllen wichtige kommunikative Funktionen. Sie erzählen kohärente Geschichten. Sie betonen Vorteile. Sie schaffen Schwung. Aber sie sind überzeugende Werkzeuge, keine wissenschaftlichen Belege. Wissenschaftliche Belege werden durch Prozesse erzeugt, die darauf ausgelegt sind, geprüft, hinterfragt, repliziert und angefochten zu werden. Sie beinhalten eine unabhängige Bewertung. Sie dokumentieren Unsicherheit und Fehlermodi, nicht nur Erfolge. Sie unterliegen Peer-Review und öffentlicher Prüfung.

Wenn öffentliche Behörden Beschaffungsentscheidungen treffen, entscheiden sie implizit, welcher Form von Wissen sie vertrauen. Wenn sie der Überzeugungskraft von Herstellern mehr vertrauen als unabhängigen Belegen, treffen sie eine Governance-Entscheidung mit Konsequenzen. Sie sagen damit, dass institutionelle Effizienz und Innovationsschwung wichtiger sind als die unabhängige Überprüfung von Behauptungen, die die öffentliche Rechenschaftspflicht betreffen.

Dies ist kein Argument gegen Innovation. Es ist ein Argument für die Disziplin, die Innovation in öffentlichen Institutionen begleiten sollte. Die öffentliche Beschaffung existiert, um der Öffentlichkeit zu dienen. Sie erfordert Belege, nicht nur Versprechen.

Beschaffung als Kontrollpunkt der Governance

Beschaffung ist einer der wichtigsten Eingriffspunkte in der Technologie-Governance, und sie wird häufig als Transaktion statt als Entscheidung behandelt.

Once a system is purchased, integrated into institutional workflows, trained into staff processes, and normalized as standard operating procedure, changing course becomes difficult. The installed base creates inertia. The sunk costs create resistance to alternatives. The staff who have learned to work around the system's limitations develop wor-

karounds rather than escalate problems. The system becomes infrastructure, and infrastructure is hard to change.

Deshalb muss Beschaffung als Kontrollpunkt der Governance behandelt werden. Es ist der Moment, in dem eine öffentliche Behörde Hebelwirkung hat, um Rechenschaftspflicht einzufordern. Es ist der Moment, in dem Hersteller am empfänglichsten für Anforderungen sind und am ehesten bereit sind, Belege vorzulegen. Es ist der Moment, bevor das System normalisiert und schwer kritisch zu bewerten wird.

Responsible procurement should therefore require independent evidence before adoption, not require apologies after deployment. The question should not be "How do we make this system work?" but rather "Do we have sufficient evidence that this system is appropriate for this use case?" The burden of proof should rest with the vendor and the evidence, not with the institution and hope.

Was eine verantwortungsvolle Bewertung fragen sollte

Institutionen, die sich der Beschaffung von Gebärdensprach-KI nähern, sollten klare, spezifische Fragen stellen, bevor sie Kaufentscheidungen treffen. Diese Fragen sollten sich auf Belege, Transparenz, Einbindung der Community, Risikobewertung und Governance konzentrieren.

- Wer hat die Bewertung dieses Systems durchgeführt, und war diese Person unabhängig vom Hersteller?
- Welche Gebärdensprachen, Dialekte, Register und Nutzerpopulationen wurden getestet, und welche Populationen wurden von den Tests ausgeschlossen?
- Waren gehörlose Muttersprachler der Gebärdensprache an der Bewertung beteiligt, und in welcher Funktion?
- Wurden DeafBlind-Nutzer, Menschen mit Behinderungen und mehrfach marginalisierte Nutzer gezielt getestet?
- Welche dokumentierten Fehlermodi gibt es, und unter welchen Bedingungen funktioniert das System schlecht?
- Was geschieht in Situationen mit hohem Risiko, etwa bei medizinischer Kommunikation, Gerichtsverfahren oder im Bildungsbereich?
- Kann der Anbieter auf Grundlage von Belegen klar benennen, wofür das System nicht genutzt werden sollte?

- Legt das System seine eigene Unsicherheit offen, oder präsentiert es Ergebnisse mit falscher Sicherheit?
- Gibt es einen in den Einsatz integrierten Prozess für menschliches Eingreifen oder Eskalation?
- Kann die öffentliche Einrichtung die Systemleistung nach dem Kauf unabhängig prüfen, oder beschränkt sich die Leistungsbewertung auf vom Anbieter bereitgestellte Kennzahlen?

Dies sind keine feindseligen Fragen. Es sind die Fragen, die verantwortungsbewusste Institutionen stellen sollten, bevor sie Entscheidungen treffen, die den grundlegenden Sprachzugang für Deaf-Gemeinschaften betreffen.

Das Risiko, den Hype-Zyklus zu wiederholen

Öffentliche Institutionen haben wiederholt Technologien übernommen, bevor die langfristigen Auswirkungen vollständig verstanden waren. Spracherkennungssysteme wurden normalisiert, bevor die Fehlerraten bei akzentgefärbter Sprache vollständig dokumentiert waren. Prädiktive Algorithmen wurden in der Strafjustiz eingesetzt, bevor Verzerrungen in den Trainingsdaten systematisch untersucht wurden. Automatisierte Untertitelungssysteme wurden eingeführt, bevor ihre Grenzen bei komplexen Dialogen und spezialisierten Fachgebieten kartiert waren.

In accessibility specifically, this pattern is especially consequential. The people most affected by accessibility technology decisions are rarely the people making those decisions. Procurement officers, IT directors, budget officers, and executives make the purchasing call. Users from the affected community often discover the technology's limitations only after deployment, when the purchasing decision has already been locked in and alternatives are no longer being considered.

Das Risiko, diesen Zyklus bei Gebärdensprach-KI zu wiederholen, ist real. Die Dynamik nimmt zu. Die Anbieterpräsentationen werden ausgefeilter. Der Druck zur Einführung wächst. Und die Evidenzbasis, obwohl sie wächst, ist noch nicht ausgereift genug, um das derzeit vorgeschlagene Ausmaß der Einführung zu stützen.

Die Rolle unabhängiger Bewertung und Governance

Novara Consulting Group wurde gegründet, um eine sichtbare Lücke in der Beschaffungspraxis zu schließen. Öffentlichen Institutionen fehlt ein strukturierter Ansatz zur

Bewertung von Gebärdensprach-KI-Systemen vor deren Einführung, und diese Lücke birgt Risiken sowohl für die Institution als auch für die Communitys, die am stärksten auf Sprachzugang angewiesen sind.

Der Sign Language Access Trust (SLAT) Index ist ein auf Governance ausgerichtetes Rahmenwerk, das Institutionen dabei helfen soll, vor dem Kauf von Gebärdensprach-KI-Systemen bessere Fragen zu stellen. Es handelt sich nicht um einen technischen Standard und auch nicht um ein Zertifizierungssystem. Es ist ein strukturierter Bewertungsansatz, der sich auf Transparenz der Evidenz, Governance-Mechanismen, Einbindung der Community, Bewertung sprachlicher Risiken, Einsatzkontext und Rechenschaftsstrukturen konzentriert.

SLAT bewertet Systeme anhand von zehn Governance-Bereichen:

> **DIE ZEHN SLAT-GOVERNANCE-BEREICHE**
>
> Produkttransparenz, Unabhängigkeit der Bewertung, Dokumentation der Datensätze, Fehleranalyse, Validierung in der Praxis, Nutzereinbindung, Einsatz-Schutzmaßnahmen, Eskalation und Eingriffsmöglichkeit, Leistungsüberwachung und Rechenschaftspflicht gegenüber der Community

Für jeden Bereich fragt SLAT, ob Anbieter Belege vorlegen können, die den Standardkriterien entsprechen, oder ob Lücken ungeprüft bleiben.

Das Ziel ist nicht, Innovation zu blockieren. Es geht darum, sicherzustellen, dass vor dem Kauf die richtigen Fragen gestellt werden und dass die Antworten einer unabhängigen Prüfung unterzogen werden.

Evidenz vor Einführung

Die Zukunft der Gebärdensprach-KI sollte nicht allein durch die Dynamik des Marketings entschieden werden. Wenn öffentliche Einrichtungen Systeme kaufen, die den Sprachzugang für Deaf-Gemeinschaften vermitteln, müssen sie vor der Einführung unabhängige Belege verlangen. Diese Belege sollten vor der Verpflichtung geprüft werden, nicht nach dem Einsatz diskutiert werden. Die Frage ist nicht, ob Gebärdensprach-KI vielversprechend ist. Die Frage ist, ob öffentliche Institutionen bereit sind, dieses Versprechen verantwortungsvoll zu steuern.

The distinction between innovation and accountability is not a constraint on progress. It is a prerequisite for progress that is safe. Public agencies that treat procurement as a governance checkpoint, not a transaction, will find vendors and systems that are willing to meet that standard. The ones that do will set a precedent for the field. The ones that don't will bear the consequences.

ÜBER DEN AUTOR

Heather M. Grizzle-Odland

Author Note: Heather M. Grizzle-Odland is Founder and CEO of Novara Consulting Group LLC, an AI governance and accessibility consulting firm. She is the developer of the Sign Language Access Trust (SLAT) Index and a doctoral student in Public Policy at Liberty University's Helms School of Government. She writes from inside the Deaf community and is a non-speaking ASL user based in Minnesota.

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.

SUPPORT NCG

NCG does this work without vendor money, so our findings stay ours. If it has been useful to you, you can support NCG's independent work here. NCG is a for-profit company, not a charity. Contributions are not tax-deductible and buy no service or influence.

Zitierweise

Grizzle, H. M. (2026, July 7). Das politische Problem der Gebärdensprach-KI: Wenn die öffentliche Beschaffung der wissenschaftlichen Evidenz voraus ist. Novara Consulting Group. https://www.novaracg.com/de/2026/07/07/the-policy-problem-with-sign-language-ai-when-public-procurement-outruns-scientific-evidence/