NOVARA
CONSULTING GROUP

The Procurement Theater Problem: Procedural Assurance, Evidentiary Assurance, and the Governance of Accessibility AI

Heather Grizzle · July 22, 2026

Checklists, demos and VPATs create confidence without evidence. Procurement theater, and the evidentiary assurance accessibility AI buyers need.

Contents

1. Executive Summary
2. The Policy Problem
3. Institutional Analysis
4. Governance Implications
5. NCG Policy Position
6. Implementation Considerations
7. Closing Position
8. References
9. Author's Note

Executive Summary

Institutions acquiring artificial intelligence systems for communication access are increasingly confident that they have managed risk, and that confidence is increasingly unearned. The confidence rests on artifacts: completed checklists, vendor documentation, scripted demonstrations, negotiated warranty language, filed accessibility statements. Each artifact is genuine. None of them, individually or in aggregate, is designed to answer the question the institution actually needs answered, which is whether the system will perform safely, reliably, and appropriately in the deployment context for which it is being acquired.

This issue names that condition procurement theater and argues that it is structural rather than malicious. Procurement professionals are rewarded for procedural completeness. Vendors are rewarded for persuasive demonstration. Both parties can discharge their obligations fully while the central performance question goes unexamined. The remedy is not more documentation. It is a shift in the governing standard from procedural assurance, which asks whether the required steps were taken, to evidentiary assurance, which asks whether the evidence collected was capable of discovering the truth.

Novara Consulting Group's position is that this distinction will become the defining competency of public sector AI governance, and that institutions should begin building the evidentiary capacity to make it now, before deployment rather than after failure.

The Policy Problem

Governance is not measured by activity. It is measured by whether the activities performed are capable of discovering the truth. Most current accessibility AI procurement fails this test not because the process is careless but because the process was designed for a different class of product.

Traditional procurement instruments assume that a product's capabilities are stable, observable, and largely disclosed by the manufacturer. Under that assumption, collecting artifacts is a reasonable proxy for verifying performance. Artificial intelligence systems mediating linguistic access violate every part of the assumption. Performance varies by signer, register, regional variation, lighting, camera geometry, latency, subject matter, and the presence or absence of the non-manual grammatical features from which this publication takes its name. A system that performs adequately in a controlled demonstration may degrade substantially in a clinical intake, a disciplinary hearing, or an emergency communication. Nothing in a standard artifact package is constructed to reveal that degradation.

Three conflations sustain the problem. Documentation is treated as validation, though documentation records what a vendor asserts rather than what a system does. Certification is treated as performance, though certification typically attests to process conformity rather than output quality. Conformance is treated as effectiveness, though conformance measures alignment to a technical standard rather than whether a deaf or hard of hearing user achieved functional communicative access. A procurement

process that cannot distinguish among these three concepts generates institutional risk in direct proportion to the number of forms it completes, because each completed form increases confidence without increasing knowledge.

Institutional Analysis

The incentive structure explains the persistence of the pattern more convincingly than any account of bad faith. Procurement offices operate under audit regimes that examine whether required steps occurred, not whether the steps were epistemically adequate. A file containing a vendor accessibility conformance report, a demonstration record, and negotiated contract language survives audit. A file documenting that the institution independently tested the system against representative deaf users in representative conditions, and found it wanting, creates schedule risk and political exposure. The system rewards the first file.

Vendors face a symmetrical incentive. Markets for accessibility AI reward demonstration quality, narrative coherence, and speed to contract. Vendors are rarely asked to disclose failure conditions, and disclosing them unilaterally imposes competitive cost. Public evidence does not yet support the claim that current commercial sign language AI systems have been evaluated against independently constructed, deployment representative test conditions with published results. In the absence of such evidence, marketing claims and self generated documentation become the institutional record by default.

The affected population absorbs the residual risk. Deaf and hard of hearing users are typically not parties to the procurement, are rarely represented in acceptance testing, and encounter the system only after institutional confidence has been established and the contract executed. By that point the burden of proof has inverted. The user is required to demonstrate that the system failed, against an institutional record asserting that due diligence was performed and the system was found compliant. This inversion is the operative harm of procurement theater, and it is produced entirely by legitimate actors following legitimate procedures.

Understood in the tools of government tradition, procurement is not an administrative formality but an instrument of public action, and the choice of instrument determines who is empowered to speak, what counts as proof,

> The decision not to look is itself a gover-

and where accountability comes to rest. Read alongside the causal stories literature, the prevailing account of accessibility AI failure is a story of accident, in which unpredictable technology occasionally underperforms and no actor bears responsibility. The evidentiary framing advanced here supports a different account, one of inadvertence at minimum: the failures are foreseeable, the conditions under which they occur are discoverable in advance, and the decision not to look is itself a governance decision.

nance decision.

Governance Implications

Shifting from procedural to evidentiary assurance requires institutions to ask a different set of questions, and to treat the answers as acceptance criteria rather than background information. What evidence supports the claim being made, and is that evidence performance evidence or process evidence? Who generated the evidence, under what commercial relationship to the outcome, and with what methodological transparency? Can the result be independently reproduced by a party with no interest in the procurement succeeding? Under what specific conditions does the system fail, at what rate, and with what severity distribution across user populations? And when those failures occur in deployment, which party bears responsibility, through what mechanism, on what timeline, and with what remedy available to the affected individual?

These questions do more than raise the evidentiary bar. They relocate accountability. A procurement file that answers them establishes what the institution knew, when it knew it, and what it accepted. A procurement file that does not answer them establishes only that the institution completed a process. Under emerging governance expectations, including risk based frameworks that require organizations to characterize context of use and document residual risk, the second file is becoming increasingly difficult to defend.

The competency at issue is the institutional capacity to distinguish confidence from evidence. That capacity is not resident in checklists. It is resident in evaluation instruments, in the ability to specify representative test conditions, and in the willingness to treat a vendor claim as a hypothesis rather than a finding.

NCG Policy Position

Novara Consulting Group holds that accessibility AI acquisition should be governed by evidentiary assurance, and that the following positions follow from that principle.

Vendor generated documentation should be treated as a claim requiring verification, never as verification itself. Its function in the procurement record is to establish what the vendor asserted, which is a matter of accountability rather than a matter of proof

Independent evaluation should be conducted under conditions representative of the actual deployment context, including the user populations, linguistic registers, subject matter, and environmental conditions the system will encounter in service. Evaluation conducted under any other conditions answers a question the institution did not ask.

Failure conditions should be affirmatively disclosed and contractually addressed before award. A system whose failure modes are undocumented has not been evaluated, regardless of how much documentation accompanies it.

Deaf and hard of hearing stakeholders should participate in acceptance criteria and evaluation design, not solely in post deployment feedback. Participation after institutional confidence has formed is consultation in name only.

Responsibility for failure should be assigned explicitly at the point of acquisition, with defined remedy and escalation available to the affected individual rather than solely to the contracting institution.

Implementation Considerations

Institutions adopting an evidentiary posture should anticipate three practical frictions. The first is timeline. Independent evaluation extends acquisition cycles, and the extension must be defended to leadership in advance rather than absorbed silently by the procurement office. The most durable defense is comparative: the cost of pre deployment evaluation against the cost of remediation, contract termination, or litigation arising from failures discovered in service.

The second is capability. Most institutions do not currently possess the internal expertise to design representative evaluation conditions for sign language AI, and the absence of that expertise is frequently misread as a reason to rely on vendor documentation. It is more accurately read as a specification for external evaluation capacity.



The third is market response. Vendors accustomed to artifact based procurement may resist evaluation requirements, and some will decline to participate. Institutions should recognize that a vendor's unwillingness to be evaluated under deployment representative conditions is itself evidence, and is among the more informative signals available during acquisition.

Implementation need not be total to be meaningful. Institutions may begin by adding evidentiary questions to existing solicitation instruments, by requiring failure condition disclosure as a responsiveness criterion, or by conditioning award on successful independent evaluation. Each step moves the institutional record from confidence toward evidence.

Closing Position

The future of AI governance will not be determined by which organization collects the largest number of compliance documents. It will be determined by which organizations build procurement systems capable of distinguishing confidence from evidence.

> **THE FIFTH PARAMETER**
>
> Effective governance begins where procedural confidence ends and independent evidence begins.

References

- Salamon, L. M. (Ed.). (2002). *The Tools of Government: A Guide to the New Governance*. Oxford University Press.
- Stone, D. (2012). *Policy Paradox: The Art of Political Decision Making* (3rd ed.). W. W. Norton.
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework* (AI RMF 1.0). NIST AI 100-1.
- International Organization for Standardization. (2023). ISO/IEC 42001:2023, Information technology - Artificial intelligence - Management system.

- Information Technology Industry Council. *Voluntary Product Accessibility Template* (VPAT). Accessibility conformance reporting instrument referenced under Section 508 of the Rehabilitation Act.
- Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence.
- Novara Consulting Group. *The Fifth Parameter*, Issue 002, "The Procurement Trap."
- Novara Consulting Group. *The Fifth Parameter*, Issue 011 / Policy No. 2, "The Accountability Gap."

Author's Note

Heather M. Grizzle-Odland is the Founder and CEO of Novara Consulting Group LLC, where she focuses on AI accessibility governance, sign language AI procurement, institutional accountability, and public-sector technology oversight. She is the developer of the Sign Language Access Trust Framework™ and the author of The Fifth Parameter, a publication examining the governance questions institutions often overlook when evaluating emerging technologies.

Her work centers on evidence, accountability, Deaf-led governance, and the gap between technical conformance and real-world accessibility.

> **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.

How to cite

Grizzle, H. M. (2026, July 22). The Procurement Theater Problem: Procedural Assurance, Evidentiary Assurance, and the Governance of Accessibility AI. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003831